Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Scienze Biotecnologiche e Farmaceutiche

Ciclo XXXII

Settore Concorsuale: 03/D1

Settore Scientifico Disciplinare: CHIM/11

HOLOBIOMICS - Use of microbiomics for the exploration of microbial communities in holobionts.

Presentata da: Matteo Soverini

Coordinatore Dottorato **Supervisore**

Prof.ssa Maria Laura Bolognesi **Prof.ssa Patrizia Brigidi**

Esame finale anno 2020

SUMMARY

PREFACE

Microbiomics: *noun, (used as singular verb) | the scientific study of microbiomes.*

Unicellular organisms represent the first living organism appeared on Earth and can be undoubtedly considered one of the most successful forms of known life. Since the first establishment of life, dated at least ~4.3 billion years ago, they have populated every environment, pushing their limits in adaptability well beyond those possible for multicellular life. Bacteria, for example, have been successful in populating extreme environments such as abysses and clouds in the sky, finding growth and sustenance opportunities on unusual substrates and conditions. Being our species and all the others embedded within the microbial world, no multicellular organism can avoid contact with the microbial fraction. This interaction can have multiple natures and assume different shapes and significance, ranging from symbiosis to infection and disease. Due to the complex nature of these interactions, the study and the disentanglement of the underlined processes is a great challenge in contemporary biology, representing an endeavouring task for all the scientists involved. This does not mean that science has been steady, but rather many steps in this direction have been accomplished, allowing an even deeper knowledge of the microbiological world and its interaction with all the other ecosystems.

Although the hypothetical existence of microorganisms has been postulated since 300 years BCE, scientists started to empirically studying bacteria and other microscopic forms of life in late 19th Century, with the introduction of proper microscopes and instruments. Since this date we have progressively gained extensive knowledge of the microbial life and with the introduction, more than a decade ago, of the so-called next-generation sequencing (NGS) techniques we have exponentially increased our opportunities of shedding light on microbial communities. This revolution opened a 'golden era' in the new-born field of microbiomics, avoiding the culturing step that always represented a limiting factor in the characterization of particular and fastidious microorganisms. Furthermore, it is clear the advantage in retrieving all the taxonomic and functional information encoded within a microbiome directly by sequencing a sample deriving from an environment of interest. The huge amount of information produced in studies relying on NGS represents a challenging task, constituting the driver for the

creation of the computational microbiologist: a new figure alongside the molecular microbiologist and classic microbiologist. This researcher's work starts when the laboratory work ends, and all the sequencing process is completed: the aim of a computational microbiologist work is to deal with the vast amount of data generated by the sequencing process, producing biologically meaningful data. Current analysis tasks oversee the inclusion and merging of several -omics approaches (e.g. metagenomics, transcriptomics, metabolomics) whit the final objective of a better comprehension of the nature and the ways of integration and interaction of a microbial community in a specified environment. This knowledge set the foundations of the definition of 'holobionts': multipartite organisms in which the microbial part plays a fundamental role in the physiology and phenotype of the host organism.

During my PhD I have focused on these latter tasks, dealing with the characterization at different levels of various holobionts, ranging from wild animals to humans, giving attention at the bacterial, fungal and viral fractions in ecosystems. In the present work I report the main achievements of my research work, whose common denominator is the bioinformatic approach to microbiome data.

PART

Chapter 1 - Microbiomics pillars: sequencing and bioinformatics

- *Sequencing*
- *After the sequencing: bioinformatics applied to microbiology*
- *Marker-gene analysis bioinformatics*
- *Metagenomics bioinformatics*

I

BRIEF INTRODUCTION TO DNA SEQUENCING AND BIOINFORMATIC ANALYSIS

CHAPTER 1 – Microbiomics pillars: sequencing and bioinformatics

Sequencing

Once the nucleic acids are extracted and purified from the samples of interest, the sequencing process can take place. Sequencing consists in the determination of the exact sequence of nucleotides contained in a nucleic acid molecule (both DNA and RNA), retrieved using particular sequencing techniques. Due to the complex nature of nucleic acids, initial proceedings in this discipline were limited: it was only possible to determine the percentage of nucleotide composition, without knowing its exact sequence[1].

The first exact nucleotide chain sequence was produced in 1965, with the work of Robert Holley and colleagues: they produced the first whole nucleic acid sequence of alanine tRNA from *Saccharomyces cerevisiae*[2]. After this first achievement, scientists were able to produce the DNA sequence of other ribosomal and tRNA genes[3-4-5], despite the determination process of bases was restricted to short stretches of DNA, and still involved a considerable amount of analytical chemistry and fractionation procedures. A first breakthrough in the sequencing process was introduced in 1977 by Fred Sanger with the dideoxy terminators techniques[6]. The chain-termination technique foresaw the use of chemical analogues of the deoxyribonucleotides (dNTPs) that are the monomers of DNA strands: Dideoxynucleotides (ddNTPs) lack the 3′ hydroxyl group that is required for extension of DNA chains, and therefore cannot form a bond with the 5′ phosphate of the next dNTP. Mixing radio-labelled ddNTPs into a DNA extension reaction results in DNA strands of each possible length being produced, as the dideoxy nucleotides get randomly incorporated, halting further progression. By performing four parallel reactions containing each individual ddNTP base and running the results on four lanes of a polyacrylamide gel, a scientist is able to use autoradiography to infer what the nucleotide sequence in the

---

1 Holley R.W., Apgar J., Merrill S.H., Zubkoff P.L. Nucleotide and oligonucleotide compositions of the alanine-, valine-, and tyrosine-acceptor soluble ribonucleic acids of yeast. J. Am. Chem. Soc. 1961;83:4861–4862.

2 Holley R.W. Structure of a ribonucleic acid. Science. 1965;147:1462–1465.

3 Brownlee G., Sanger F. Nucleotide sequences from the low molecular weight ribosomal RNA of *Escherichia coli.* J. Mol. Biol. 1967;23:337–353

4 Cory S., Marcker K.A., Dube S.K., Clark B.F. Primary structure of a methionine transfer RNA from *Escherichia coli.* Nature. 1968;220:1039–1040.

5 Goodman H.M., Abelson J., Landy A., Brenner S., Smith J.D. Amber suppression: a nucleotide change in the anticodon of a tyrosine transfer RNA. Nature. 1968;217:1019–1024.

6 Sanger FS., Nicklen A.R.C. DNA sequencing with chain-terminating. Proc. Natl. Acad. Sci. 1977;74:5463–5467.

original template was, as there will be a radioactive band in the corresponding lane at that position of the gel. The simplicity, reliability and robustness of the process determined its success for years to come, being the election method for nucleic acid sequencing for decades. In 1991 the process allowed to produce the first automatic sequencer[7], paving the way in the direction of automatized sequencing. Years after this first-generation sequencing techniques a new method was introduced: the pyrosequencing. This approach based on luminescence consisted in a two-enzyme process in which ATP sulfurylase is used to convert pyrophosphate into ATP, which is then used as the substrate for luciferase, thus producing light in proportion to the amount of free pyrophosphate[8]. Libraries of DNA molecules are first attached to beads using specific adapter sequences, which then undergo a water-in-oil emulsion PCR[9] to coat each bead in a clonal DNA population, where, ideally, on average one DNA molecule ends up on one bead. These DNA-coated beads are then washed over a picoliter reaction plate that fits one bead per well; pyrosequencing then occurs as smaller bead-linked enzymes and dNTPs are washed over the plate, and pyrophosphate release is measured using a charged couple device (CCD) sensor beneath the wells. This setup is capable of producing reads around 400–500 base pairs (bp) long, for the million wells that would be expected to contain suitably clonally-coated beads[40]. Parallelization can be considered the distinctive element of second-generation sequencing, allowing researchers to completely sequence a single human's genome in a significant lower amount of time if compared to Sanger's first-generation sequencing. This method was commercialized by 454 (later purchased by Roche) and the most representative sequencer belonging to this series is the 454 GS FLX.

Other parallel sequencing techniques were introduced following the success of 454. The most important among them is arguably the Solexa method of sequencing, which was later acquired by Illumina[10]. In Solexa products adapter-bracketed DNA molecules are passed over a layer of complementary oligos bound to a flow-cell; a subsequent solid phase PCR produces neighboring clusters of clonal sequences from

---

7 Hunkapiller T., Kaiser R., Koop B., Hood L. Large-scale and automated DNA sequence determination. Science. 1991;254:59–67.

8 Nyrén P.I., Lundin A. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. Anal. Biochem. 1985;509:504–509.

9 Tawfik D.S., Griffiths A.D. Man-made cell-like compartments for molecular evolution. Nat. Biotechnol. 1998;16:652–656.

10 Voelkerding K.V., Dames S.a., Durtschi J.D. Next-generation sequencing: from basic research to diagnostics. Clin. Chem. 2009;55:641–658.

each of the individual original bonded DNA strands[11]. This process is called 'bridge amplification', due to replicating DNA strands having to arch over to prime the next round of polymerisation off neighbouring surface-bound oligonucleotides. The sequence is achieved in a sequencing-by-synthesis (sequencing requires the action of DNA polymerase) manner using fluorescent 'reversible-terminator' dNTPs. After the binding no further nucleotides can bind and elongate the chain as the fluorophore occupies the 3' hydroxyl position; this must be cleaved away before polymerization can continue, which allows the sequencing to occur in a synchronous way[12]. These modified dNTPs and DNA polymerase are washed over the primed, single-stranded flow-cell bound clusters in cycles. At each cycle, the nucleotide identity is monitored with a CCD by exciting the fluorophores with appropriate lasers, before enzymatic removal of the blocking fluorescent marker and continuation to the next position. First Solexa machines were initially only capable of producing very short reads (~ 35 bp long), but they had an advantage in that they could produce paired-end (PE) data, in which the sequence at both ends of each DNA cluster is obtained. This is achieved by first obtaining one read from the single-stranded flow-cell bound DNA, before performing a single round of solid-phase DNA extension from remaining flow-cell bound oligonucleotides and removing the already-sequenced strand. As the input DNA molecules are of an approximate known length, having PE data provides a greater amount of information. This can improve the accuracy of mapping reads to reference sequences, and aids in detection of spliced exons and rearranged DNA or fused genes. The standard Genome Analyzer version (GAIIx) was later followed by the Illumina HiSeq, a machine capable of even greater read length and depth, and then the Illumina MiSeq, which was a lower-throughput (but lower cost) machine with faster turnaround and longer read lengths.

Despite the actual widespread and usage of the second-generation techniques, technology moved forward and introduced the so-called third-generation techniques, based on the single-molecule-sequencing (SMS) approach. The first SMS technology was

---

11 Fedurco M., Romieu A., Williams S., Lawrence I., Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Res. 2006;34

12 Turcatti G., Romieu A., Fedurco M., Tairi A.-P. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. Nucleic Acids Res. 2008;36:e25.

developed in the lab of Stephen Quake[13], and worked broadly in the same manner that Illumina does, but without any bridge amplification; DNA templates become attached to the surface, and then propriety fluorescent reversible terminator dNTPs[14] are washed over one base a time and imaged, before cleavage and cycling the next base over. While relatively slow and expensive, this was the first technology to allow sequencing of long DNA fragments without any amplification, avoiding all potentially associated biases and errors. The most widely used third-generation technology is probably the single molecule real time (SMRT) platform from Pacific Biosciences, available on the PacBio range of machines[15]. During SMRT DNA polymerisation occurs in arrays of microfabricated nanostructures, which are essentially tiny holes in a metallic film covering a chip. These holes exploit the properties of light passing through apertures of a diameter smaller than its wavelength, which causes it to decay exponentially, exclusively illuminating the very bottom of the wells. This allows visualization of single fluorophore molecules close to the bottom of the hole, due to the zone of laser excitation being so small. A single DNA polymerase molecules inside the hole places them inside the laser-illuminated region: using the DNA library of interest and fluorescent dNTPs, the extension of DNA by single nucleotides can be monitored in real time, as fluorescent nucleotide being incorporated will produce a fluorescent blast, after which the dye is cleaved away, ending the signal for that position[16]. This process can sequence single molecules in a very short amount of time. Sequencing process occurs at the rate of the polymerase producing kinetic data: this allow the detection of modified bases as well as the production of long reads, up to 10 kb in length, which are useful for de novo genome assemblies[17].

After the sequencing: bioinformatics applied to microbiology

The sequencing process returns a huge amount of information to the scientist, needing further processing and elaboration through bioinformatic approaches. The primary sequencing outputs consist in .fastq files, a plain text format containing all the sequences generated for every sample. And the relative quality information. Generally

13 Braslavsky I., Hebert B., Kartalov E., Quake S.R. Sequence information can be obtained from single DNA molecules. Proc. Natl. Acad. Sci. U. S. A. 2003;100:3960–3964

14 Bowers J. Virtual terminator nucleotides for next-generation DNA sequencing. Nat. Methods. 2009;6:593–595.

15 van Dijk E.L., Auger H., Jaszczyszyn Y., Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30

16 Eid J. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323:133–138.

17 Schadt E.E., Turner S., Kasarskis A. A window into third-generation sequencing. Hum. Mol. Genet. 2010;19:R227–R240.

the sequences are machine-demultiplexed, meaning that the instrument automatically splits the reads in different files, one for each analyzed sample. Once the sequences produced by the sequencing run have been downloaded, the subsequent bioinformatics pipeline used vary according to the type of survey to be performed. Generally, genomics-based microbiomics studies relies on two different strategies: marker-gene analysis or whole-genome metagenomics. The methods underlying both types of surveys are shown below, with particular attention on different types of information that can be extrapolated from the microbial community.

Marker-gene analysis bioinformatics

Marker-gene surveys relies on the amplification of specific genomic regions subjected to slow rates of evolution and thus highly conserved inside a group of related organisms. The following discussion will focus on the phylogenetic characterization of the bacterial communities, representing the most widespread type of analysis today[18]. However the principles can also be applied to the study of fungal communities, in which a combination of sequencing of the 18S rRNA gene[19] and of the ITS regions (Internal Transcribed Spacers) [20] is used.

For bacteria, the elective marker region consists in the 16S rRNA gene[21], a component of the 30S small subunit of a prokaryotic ribosome that binds to the Shine-Dalgarno sequence. The structure of this gene and its crucial function in the bacterial metabolism decreed its successful use in genomics. Indeed, in addition to highly inter-specific conserved primer binding sites, 16S rRNA gene sequence contains hypervariable regions that provide species-specific signature sequences useful for taxonomic assignment to bacteria lineage[22], through the design of universal primers that can reliably produce the same sections of the 16S sequence across different taxa[23]. The bacterial 16S rRNA gene contains nine hypervariable regions (V1–V9), ranging from ~30 to 100 base

---

18 52,216 results using the 16S rRNA query vs 9,245 using the 18S and ITS query on NCBI PubMed

19 Meyer A., Todt C., Mikkelsen N. T. & Lieb B. (2010). "Fast evolving 18S rRNA sequences from Solenogastres (Mollusca) resist standard PCR amplification and give new insights into mollusk substitution rate heterogeneity". BMC Evolutionary Biology 10: 70.

20 Baldwin, Bruce G, Sanderson, Michael J, Porter, J. Mark; Wojciechowski, Martin F, Campbell, Christopher S, Donoghue, Michael J. (1995-01-01). "The ITS Region of Nuclear Ribosomal DNA: A Valuable Source of Evidence on Angiosperm Phylogeny". Annals of the Missouri Botanical Garden. 82 (2): 247–277.

21 Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A. 1977 Nov;74(11):5088-90.

22 Kolbert CP, Persing DH. Ribosomal DNA sequencing as a tool for identification of bacterial pathogens. Curr Opin Microbiol. 1999 Jun;2(3):299-305.

23 Větrovský T, Baldrian P (2013-02-27). "The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses". PLoS One. 8 (2): e57923.

pairs of length, that are important for the establishment of the secondary structure in the small ribosomal subunit[24]. The conservation levels vary broadly between hypervariable regions, with more conserved sections correlating to higher-level taxonomy (such as phyla or class) and less conserved sections to lower levels (such as genus and species)[25]. While the entire 16S rRNA gene sequence allows for comparison of all hypervariable regions, at approximately 1.500 base pairs represents a prohibitively expensive for studies seeking to identify or characterize diverse bacterial communities[52]. Modern genetic barcoding surveys are commonly conducted using Illumina machines[26], generally producing reads of 75–250 base pairs long. Although no hypervariable region can accurately and specifically classify all bacteria from domain to species, some are more reliable than others in the discrimination of specific taxonomic levels[53]. Gut community studies perform amplifications on a combination of semi-conserved hypervariable regions like the V4 (the most species-specific region of 16S rRNA gene[53]) and the V3 (the most reliable in identifying the potential pathogenic genera[27]).

Several software packages are used in the analysis of sequenced amplicons. The most commonly used are QIIME[28], RDP[29] and mothur[30]. RDP is a web-based tool, whereas QIIME and mothur are built as command-line interfaces. I will focus the discussion of computational resources related to QIIME and QIIME 2 software, representing the selected software used in the analysis here reported. QIIME is built as an ensemble of command-line scripts designed to assist users from raw sequence data and sample metadata to final results. After the first filtering processes the QIIME pipeline relies on the creation of OTUs (Operational Taxonomic Units)[31]. These taxonomic units

---

24 Gray MW, Sankoff D, Cedergren RJ (1984). "On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA". Nucleic Acids Research. 12 (14): 5837–52.

25 Yang B, Wang Y, Qian PY (March 2016). "Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis". BMC Bioinformatics. 17 (1): 135

26 Bartram AK, Lynch MD, Stearns JC, Moreno-Hagelsieb G, Neufeld JD (June 2011). "Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads". Applied and Environmental Microbiology. 77 (11): 3846–52.

27 Chakravorty S, Helb D, Burday M, Connell N, Alland D (May 2007). "A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria". Journal of Microbiological Methods. 69 (2): 330–9.

28 Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336.

29 Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 2009;37:D141–D145.

30 Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75:7537–7541.

31 Sokal & Sneath: Principles of Numerical Taxonomy, San Francisco: W.H. Freeman, 1963

consist in clusters of DNA sequences that shares a minimum threshold homology; in other words, OTUs are proxies for different microbial "species" at various taxonomic levels, in the absence of traditional systems of biological classification as are available for populations of macroscopic organisms. The specific OTU-clustering algorithm used can have a major influence in downstream analysis. OTU clustering algorithms can be divided in three categories: de novo, closed reference, and open reference. - In de novo OTU picking method, sequences are clustered into OTUs, without basing on any external reference sequences database as a template[32]. In contrast, closed-OTU picking uses a reference sequence database, and sample sequences that do not match the reference sequence database are discarded and not considered in further analysis. Finally, open-reference OTU picking is a two-step process consisting of first closed-reference OTU picking then followed by de novo clustering of sequences not previously aligned to the database. Is generally recommended to use open-reference OTU picking because this method retains all sequencing data, despite there are circumstances for which this picking method is not applicable.(i.e. when combining sequence data from different regions of the 16S rRNA gene). In reference-based OTU picking, sequences are compared against a reference database such as Greengenes[33], Ribosomal Database Project (RDP)[29], or SILVA[34]. The gut microbial community (especially for Human species) is well represented in the databases compared to other sample types.

Recently, a newer and refined version of QIIME has been released[35], representing a completely reengineered and rewritten system that is expected to facilitate reproducible and modular analysis of microbiome. In this second iteration of the tool the reproducibility of the results is a central focus, as well as the implementation of a new

---

32 Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol. 2005;71:1501–1506.

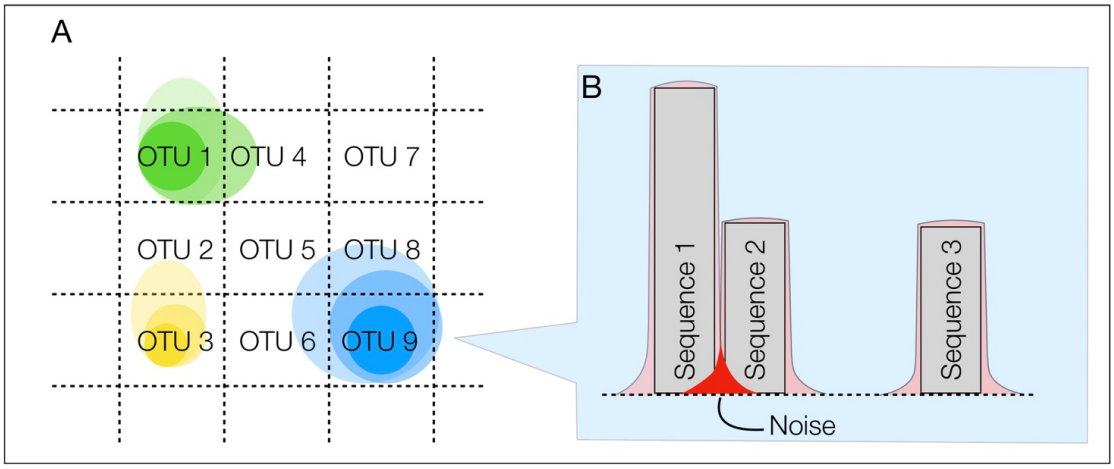
33 McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 2012;6:610–618.

34 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–D596.

35 Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, PruesseE, Rasmussen LB, Rivers A, Robeson MS 2nd, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019 Aug;37(8):852-857.

step of OTU denoising. This process is a key factor for the production of high-confidence OTUs from which retrieve a more precise taxonomic assignation, especially for closely related taxa. The process consists in applying an error model on each sequence of the pool, in order to discern the intrinsic sequence variability (due to its taxonomic origin) from the variability determined by the errors occurred during sequencing (mainly indels). This allows to correct the sequences 'noise', generated by errors arisen during the sequencing process. This filtering can be performed using pre-established error models tailored on Illumina sequencing machines (Deblur algorithm[36]) or can be dynamically computed and then applied on the sequences set (like DADA2 tool[37]).

The concepts underlying this correction are exposed in Box 1.



Box 1 | (**A**) Every sequences assigned to a specific OTU have an intrinsic noise generated by the sequencing machine. This noise can lead some sequences to trespass the 'boundaries' of a close OTU, sharing homology with a near taxonomic unit. Denoising process consist in the detection, modelling and superimposition of the error model, indicated as 'noise' in panel **B**.

Moreover, QIIME 2 introduces a new OTU picking method: insertion tree[38]. This method aligns the query sequences to a full-length sequences, and then that alignment is used to find the optimal location in the phylogenetic tree for the query sequence. This latter method differs from the other because it can also give information about the evolutionary relationships between the query sequences and known species.

Once the OTU profile is determined, the microbial community analysis continues through the determination of diversity profiles. A Microbiome diversity is typically described in

---

36 Amir A, McDonald D, Navas-Molina JA, et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems. 2017;2(2):e00191-16. Published 2017 Mar 7. doi:10.1128/mSystems.00191-16

37 Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583.

38 Mirarab S, Nguyen N, Warnow T. SEPP: SATé-enabled phylogenetic placement. Pac Symp Biocomput. 2012:247-58.

terms of within (i.e., alpha) and between the samples (i.e., beta) diversities. Both concepts were introduced in the mid 1900's by the American ecologist R. H. Whittaker[39], in his attempt to model and describe ecologically describe the vegetation present on Oregon's mountains. Alpha diversity describes the diversity intrinsically associated to the sample. In other words, alpha diversity indices answer at the question: 'how many species are there in the sample?'. Being computed using several metrics (e.g. Shannon, Chao1, Faith's Phylogenetic Diversity) its iteration returns an index representing the biodiversity of each sample.

Beta diversity analysis provide the measure of the degree to which samples differ from one another and can reveal aspects of microbial ecology that are not clear from looking at the alpha diversity. This analysis answers to the question: 'how similar are the samples among them?'. Beta diversity metrics can be clustered in different ways. First, they can be based on sequence abundance, (e.g., Bray-Curtis or weighted UniFrac) or qualitative (considering only presence-absence of sequences, e.g., binary Jaccard or unweighted UniFrac). Second, they can rely on phylogeny (e.g. both the UniFrac metrics) or not (Bray-Curtis, etc.). Both types of metrics results can be used for further analysis, like ordination techniques. The most used in microbial ecology is Principal Components analysis (PCoA). Principal coordinates (PCos) from a PCA are plotted against each other in bidimensional or multidimensional plots, to summarize the microbial community compositional differences between samples. In these graphical representation each point is a single sample, and the distance between points represents how compositionally different the samples are from one another.

Metagenomics bioinformatics

As already discussed, the marker-gene surveys are based on the amplification of phylogenetic markers with the ultimate goal of producing a taxonomic characterization of a microbial community. Metagenomic approaches do not relies on this process, since this kind of surveys relies on the sequencing of all the genetic material retrieved from a microbial environment, without any amplification of phylogenetic marker regions. This process allows scientists to obtain different kinds of information about the community: I) Characterization of the whole microbial ecosystem taxonomy and structure using different

[39] Whittaker, R. H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs, 30, 279–338.

tools respectively for the bacterial, viral and fungal fractions (e.g. MetaPhlan[40], RITA[41], Centrifuge[42], MGMapper[43], ViromeScan[44], HumanMycobiomeScan[45]). All the taxonomic results can then be elaborated using the analytical methods cited above for marker-gene surveys; II) metagenomic analysis makes possible the characterization of the functional microbial profile inside a community. This second aspect is of crucial importance and represents the distinguishing feature between the two methods. By the sequencing of the whole genomic material retrieved from a source of interest, metagenomic analysis allow to reconstruct the metabolic potential of a microbial community, inferring microbial-host and microbial-microbial interaction networks. A vast plethora of tools is available aimed at depicting the community's metabolic activity (e.g. MetaCV[46], SmashCommunity[47], HUMAnN[48], FANTOM[49]). The scientist can also create ad-hoc pipelines by combining methods of sequences alignment (e.g. bowtie2[50], bwa[51], blast[52]) on function-oriented databases such as KEGG[53] or EGG-NoG[54].

A quick overview of the two work-frames is reported in Box 2.

---

40 Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods*. 2012;9(8):811–814. Published 2012 Jun 10. doi:10.1038/nmeth.2066

41 Parks D, MacDonald N, Beiko R. BMC Bioinformatics. 2011;12:328.

42 Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res*. 2016;26(12):1721–1729. doi:10.1101/gr.210641.116

43 Petersen TN, Lukjancenko O, Thomsen MCF, Maddalena Sperotto M, Lund O, Møller Aarestrup F, Sicheritz-Pontén T. MGmapper: Reference based mapping and taxonomy annotation of metagenomics sequence reads. PLoS One. 2017 May 3;12(5):e0176469.

44 Treated later in the manuscript

45 Treated later in the manuscript

46 Liu J, Wang H, Yang H, et al. Composition-based classification of short metagenomic sequences elucidates the landscapes of taxonomic and functional enrichment of microorganisms. *Nucleic Acids Res*. 2013;41(1):e3. doi:10.1093/nar/gks828

47 M. Arumugam, E.D. Harrington, K.U.Foerstner, J. Raes, P. Bork SmashCommunity: a metagenomic annotation and analysis tool Bioinformatics, 26 (2010)

48 S. Abubucker, N. Segata, J. Goll, A.M.Schubert, J. Izard, B.L. Cantarel, *et al.*Metabolic reconstruction for metagenomic data and its application to the human microbiome PLoS Comput Biol, 8 (2012)

49 K. Sanli, F.H. Karlsson, I. Nookaew, J.Nielsen FANTOM: functional and taxonomic analysis of metagenomes BMC Bioinformatics, 14 (2013)

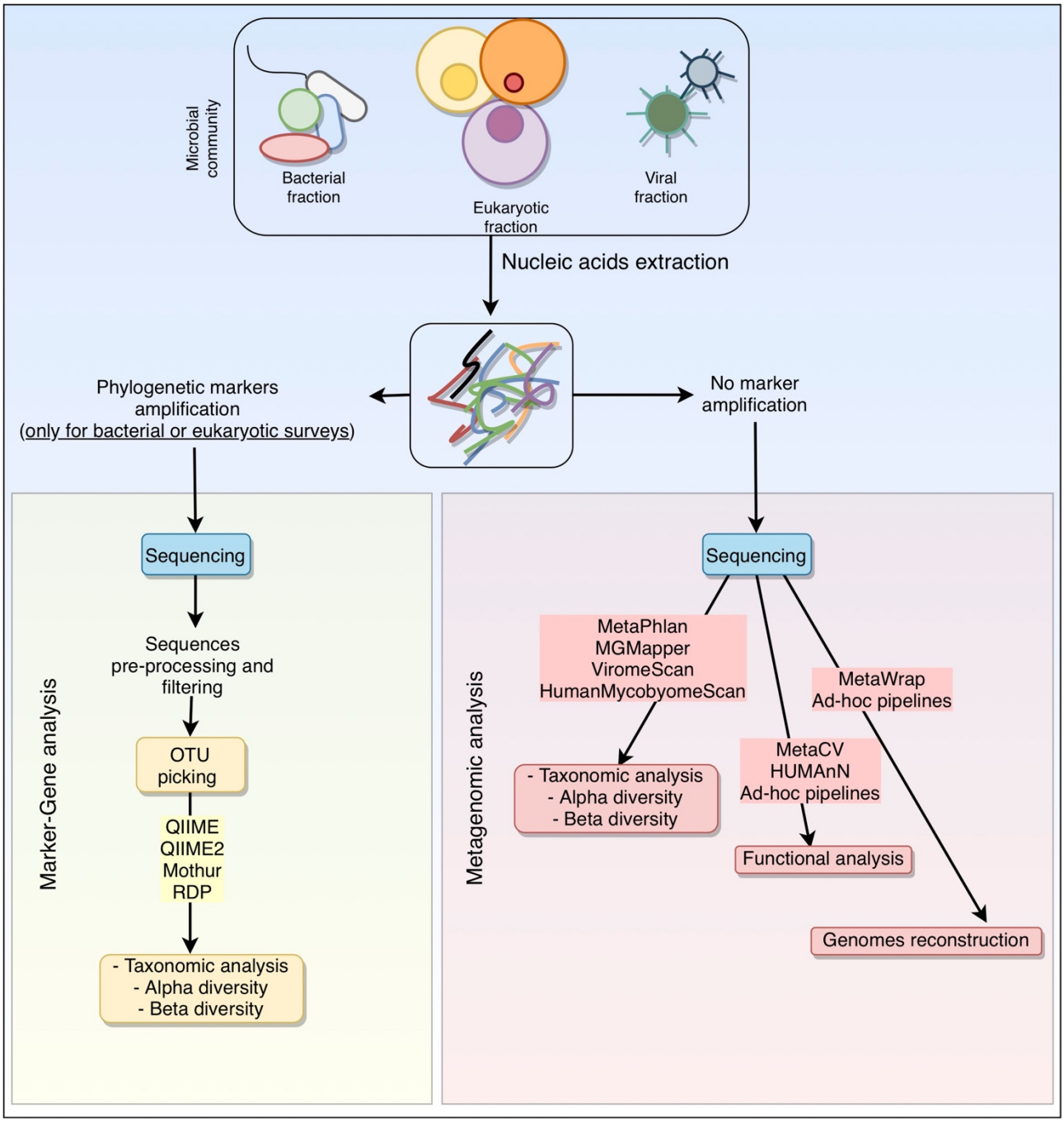
50 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

51 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760. doi:10.1093/bioinformatics/btp324

52 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403-10.

53 Tanabe M, Kanehisa M. Using the KEGG database resource. Curr Protoc Bioinformatics. 2012 Jun;Chapter 1:Unit1.12.

54 Huerta-Cepas J, Szklarczyk D, Forslund K, Cook et al., eggNOG hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2016 Jan 4;44(D1):D286-93.



Box 2 | Schematic representation of the analysis workflow in the context of marker-gene and metagenomic surveys.

PART

Chapter 2 - Bacteria

- Biology, ecology and study of bacterial communities

Chapter 3 - The gut bacterial community as a resource in adaptive processes of holobionts

- The gut bacterial community as an adaptive tool in human and other mammalians

Chapter 4 - Gut bacterial community plasticity in health and disease

- The gut bacterial community plasticity among populations in health and disease

II BACTERIA

CHAPTER 2 - Bacteria

Introduction

Microbiology consists in the study of all the organisms that cannot be seen by naked eye[55]. The most significant portion of this microcosmos is represented by bacteria, viruses and unicellular fungi[56]. Each one of these groups own its personal evolutionary history and has appeared on Earth during different periods. Bacteria are the oldest living form we have any trace of, with evidences dating their appearance around 4 billion years ago, roughly 500 million years after the planet formation.[57]

These organisms possess a long and complex evolutionary history, having succeeded over time in the colonization of nearly every context, populating also environments that are normally prohibitive for other life forms. Bacterial life, for example, flourishes in the Marianas' trench, the deepest point on Earth[58]. However their colonization ability pushes over, and bacteria are reported to inhabit deeply buried marine sediments at nearly 2.5 km under the seafloor[59] and hypothesized at depths of 19 km[60] , being able to survive and retrieve nourishment in extreme conditions and completely isolated from the terrestrial biosphere. As a last example of their colonization ability it is worth of mention the exploitation of aerial environments, such as clouds: bacteria play important roles in several atmospheric processes, being involved in precipitations[61] (properly called bio-precipitations) and in chemical reactions involving carbon molecules[62].

---

55 Brok, Madigan and Martinko (2006)

56 Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. A new view of the tree of life. Nat Microbiol. 2016 Apr 11;1:16048.

57 Dodd MS, Papineau D, Grenne T, Slack JF, Rittner M, Pirajno F, O'Neil J, Little CT. Evidence for early life in Earth's oldest hydrothermal vent precipitates. Nature. 2017 Mar 1;543(7643):60-64.

58 Kato C, Li L, Nogi Y, Nakamura Y, Tamaoka J, Horikoshi K. Extremely barophilic bacteria isolated from the Mariana Trench, Challenger Deep, at a depth of 11,000 meters. Appl Environ Microbiol. 1998 Apr;64(4):1510-3

59 Inagaki F, et al., DEEP BIOSPHERE. Exploring deep microbial life in coal-bearing sediment down to ~2.5 km below the ocean floor. Science. 2015 Jul 24;349(6246):420-4.

60 Stoddart P. Light carbon stable isotopes in aragonite veins, Lopez Island, WA: evidence for deep life? 2014 GSA Annual Meeting in Vancouver, British Columbia

61 Joly M, Attard E, Sancelme M, Deguillaume L, Guilbaud C, Morris CE, Amato P, Delort AM, Ice nucleation activity of bacteria isolated from cloud water, Atmospheric Environment, Volume 70, 2013, 392-400,

62 Husárová S, Vaïtilingom M, Deguillaume L, Traikia M, Vinatier V, Sancelme M, Amato P, Matulová M, Delort AM, Biotransformation of methanol and formaldehyde by bacteria isolated from clouds. Comparison with radical chemistry, Atmospheric Environment, Volume 45, Issue 33, 2011, 6093-6102

Relations with other organisms

Enclosed between these two extreme scenarios we can find the rest of the biosphere, in which bacteria live and flourish, surrounding all the multicellular organisms. Bacteria do not just limit to co-inhabit environments with animals: during evolution they have established complex relationships with multicellular counterparts, exploiting niches in which survive and flourish. It is a matter of fact that the vast number of known animals own microbial communities living outside and inside them, regulating crucial aspects in their host's phenotype. Ranging from sponges[63] to humans[64] all the multicellular branches of the tree of life coexists with bacterial communities that populate distinct areas of their organism. Humans, for example, harbor different microbial communities in different body districts: the skin[65], the oral cavity[66] and the gastrointestinal tract[67] are examples of niches in which bacterial communities thrive.

In most cases bacteria are commensals[68], coexisting without creating any harm, while others can establish a structured mutualistic relationship, conferring at the host's organism important and supportive functions. The mutualistic fraction of the gut microbiota plays an important role in a vast array of metabolic and adaptive processes, providing its host of otherwise neglected functions. The gut community is well-known to be involved in the energy metabolism regulation[69], in enabling the extraction of energy from host-indigestible sources[70-71], in the process of education and modulation of the

63 Schmitt S, Tsai P, Bell J, Fromont J, Ilan M, Lindquist N, Perez T, Rodrigo A, Schupp PJ, Vacelet J, Webster N, Hentschel U, Taylor MW. Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. ISME J. 2012 Mar;6(3):564-76.

64 Harmsen HJ, de Goffau MC. The Human Gut Microbiota. Adv Exp Med Biol. 2016;902:95-108. doi: 10.1007/978-3-319-31248-4_7.

65 Chen, Y. E., Fischbach, M. A., & Belkaid, Y. (2018). Skin microbiota-host interactions. *Nature*, *553*(7689), 427–436. doi:10.1038/nature25177

66 Lamont RJ, Koo H, Hajishengallis G. The oral microbiota: dynamic communities and host interactions. Nat Rev Microbiol. 2018 Dec;16(12):745-759.

67 Hooper LV, Gordon JI. Commensal host-bacterial relationships in the gut. Science. 2001 May 11;292(5519):1115-8.

68 Neville BA, Forster SC, Lawley TD. Commensal Koch's postulates: establishing causation in human microbiota research. Curr Opin Microbiol. 2018 Apr;42:47-52.

69 Bergman, E. N. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* 70, 567–590 (1990)

70 Martens, E. C. et al. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* 9, e1001221 (2011)

71 Amato KR, Leigh SR, Kent A, Mackie RI, Yeoman CJ, Stumpf RM, Wilson BA, Nelson KE, White BA, Garber PA. The gut microbiota appears to compensate for seasonal diet variation in the wild black howler monkey (Alouatta pigra). Microb Ecol. 2015 Feb;69(2):434-43.

immune system of vertebrates[72], while an involvement in cognitive and behavioural functions is growing in the number of evidences [73-74 -75].

Beyond the importance in host's physiology, an ensemble of aspects confined to the contemporary sphere, the bacterial communities have been depicted as one of the key drivers in the evolutionary process that concerns all multicellular life forms, generating the so-called '*hologenome theory of evolution*'[76]. In this context, the association between the bacterial fraction and the host is considered a single evolutionary unit, called 'holobiont'. This biological entity is directly and comprehensively subjected to the evolutionary forces, and the bacterial component is a key player in the process, conferring the potential of fast adaptability to the host[77]. The subject of the evolutionary process is, in fact, not only the genome of the host, but the resulting genome of the holobiont, the so-called 'hologenome'. This term refers to the whole set of genomes present in an organism and differs from the classical concept that the host genome alone is the only target of the selective process.

The study of the bacterial communities.

The study of bacterial life sets its roots in clinical microbiology, as the first interest of scientists was to shed light on the involvement of bacteria in disease. With this aim, Robert Koch identified the first bacteria as causative agents of tuberculosis, cholera and anthrax at the beginning of 19th century, formulating the well-known postulates[78]. This milestone set the foundations for the future development of bacteriology as a scientific discipline, and in the successive years the main effort of microbiologists was oriented in the development of methods aimed at studying organisms in a laboratory environment. In 1860 Pasteur fashioned a media of yeast, ash, candy sugar and ammonium salts,

---

72 Y.K. Lee, S.K. Mazmanian Has the microbiota played a critical role in the evolution of the adaptive immune system? Science, 330 (2010), pp. 1768-1773

73 Abdel-Haq R, Schlachetzki JCM, Glass CK, Mazmanian SK. Microbiome-microglia connections via the gut-brain axis. J Exp Med. 2019 Jan 7;216(1):41-59.

74 Ticinesi A, Tana C, Nouvenne A, Prati B, Lauretani F, Meschi T. Gut microbiota, cognitive frailty and dementia in older individuals: a systematic review. *Clin Interv Aging*. 2018;13:1497–1511.

75 Yoo BB, Mazmanian SK. The Enteric Network: Interactions between the Immune and Nervous Systems of the Gut. Immunity. 2017 Jun 20;46(6):910-926.

76 Zilber-Rosenberg I, Rosenberg E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiol Rev. 2008 Aug;32(5):723-35.

77 Alberdi A, Aizpurua O, Bohmann K, Zepeda-Mendoza ML, Gilbert MTP. Do Vertebrate Gut Metagenomes Confer Rapid Ecological Adaptation? Trends Ecol Evol. 2016 Sep;31(9):689-699

78 Koch R. Die Ätiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des Bacillus Anthracis, Cohns Beitrage zur Biologie der Pflanzen (1876)

producing the first reproducible terrain on which cultivate and study bacteria. This medium contained the basic requirements for microbial growth: nitrogen, a carbon source and vitamins[79]. In developing this media Pasteur noticed some points: that particular chemical features of the medium can promote or impede the development of any one microorganism and that competition occurs among different microorganisms for the nutrients contained in the media, which can lead to some species outgrowing and dominating a culture. During the subsequent decades the culturing methods have been progressively refined, making scientists capable to grow and insulate specific bacteria using selective media and conditions.

Despite the legacy of these pioneers in the field still exists and plays a pivotal role in modern microbiology, microbiology now relies on the support and integration of molecular biology, a complementary discipline created in late 1930s[80]. In particular, the knowledge of the DNA and its structure, promoted among others by the work of James Watson and Francis Crick[81], opened new opportunities and approaches in the biological disciplines. Molecular biology applied within bacteriology increased the knowledge of the bacterial world using different techniques: PCR[82], gel electrophoresis[83], macromolecule probing and DNA microarray[84]. These advances, even if of fundamental importance, do not allow to face one of the main problems with which microbiology has had to deal since the beginning: the unculturable bacteria. These microorganisms represent in certain cases more than 90% of the overall bacterial life on Earth[85] and, to date, cannot be cultured using any method. This is due to stringency in nourishment requirements and/or growth conditions, making it impossible to reproduce the natural habitat in a controlled environment. To fulfill this gap, nucleic acids sequencing techniques (extensively reported in chapter 1) have been applied to microbiology.

---

79 Tseng, C.K. (1946): Colloid Chemistry, Vol.6., in Alexander, J. (Ed.) Colloid Chemistry, New York: Reinhold Publishing Corp., p629.

80 Weaver W. Molecular biology: origin of the term. Science. 1970 Nov 6;170(3958):581-2.

81 Watson JD, Crick FH. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. J.D. Watson and F.H.C. Crick. Published in Nature, number 4356 April 25, 1953. Nature. 1974 Apr 26;248(5451):765.

82 Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol. 1986;51 Pt 1:263-73.

83 Tiselius A, A new apparatus for electrophoretic analysis of colloidal mixtures, transact of Chem Soc, 1937

84 Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. Nature. 1985 Mar 7-13;314(6006):67-73.

85 Wilson MJ, Weightman AJ, Wade WG. Applications of molecular ecology in the characterisation of uncultured microorganisms associated with human disease. Rev Med Microbiol 1997;8: 91-101

CHAPTER 3 - The gut bacterial community as a resource in adaptive processes of holobionts

The gut bacterial community represents a fast and effective tool in adaptive processes, overcoming the limit in the time required for the host genome to modify itself[77]. In this direction, here I present five studies I have conducted on this topic, investigating the eco-adaptive phenomena in human and other mammalian ecosystems (dolphin, naked mole-rat, sea turtle and horse). I had an active role in each of the presented studies, mainly focusing my activity in data preparation and analysis, as well in the hypothesis generation.

Section 3.1 - Variations in the post-weaning human gut metagenome profile as result of *Bifidobacterium* acquisition in the western microbiome

Introduction

Since 2010, several studies have been conducted with the specific aim to explore GM variation across human populations with different subsistence practices, lifestyles and geographical origin. These human GM surveys revealed the existence of robust bacterial compositional and functional subgroups, so far generally reflective of the variations in subsistence strategy: hunter-gatherer, rural agricultural, and urban industrial Western lifestyle[86-87-88-89].

The findings from this new and emerging field of research, which combines human microbiology ecology and anthropology, resulted in two main conclusions with important implications for both human evolutionary history and human health: first, humans co-evolved with symbiont microbial ecosystems, which have co-adapted along the trajectory of subsistence change across human evolutionary history, from hunter-gatherers to rural

86 Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., et al. (2012). Human gut microbiome viewed across age and geography. *Nature* 486, 222–227.

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

88 Obregon-Tito, A. J., Tito, R. Y., Metcalf, J., Sankaranarayanan, K., Clemente, J. C., Ursell, L. K., et al. (2015). Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat. Commun.* 6:6505.

89 Rampelli, S., Schnorr, S. L., Consolandi, C., Turroni, S., Severgnini, M., Peano, C., et al. (2015). Metagenome sequencing of the Hadza hunter-gatherer gut microbiota. *Curr. Biol.* 25, 1682–1693.

agricultural to the most recent development of completely industrialized societies[51]; second, despite the considerable variation in rural and traditional lifestyles, urban industrial populations stand apart as having a distinctly altered GM profile. Indeed, the GM of urban industrial populations seems to universally share certain compositional qualities, such as: (I) an overall compression of microbial diversity as measured by phylogeny and the number of unique taxa[90], (II) the loss of the so-called microorganisms "old friends", *Treponema* and *Succinivibrio*[91], and (III) the acquisition of *Bifidobacterium* as typical inhabitant of the adult gut[50-51].

Showing a relative abundance that ranges from 3 to 10% of the total ecosystem, bifidobacteria are an abundant bacterial component in the GM of urbanised populations adults, and also dominates the GM ecosystem of breast-fed infants, where this bacterial family accounts on average for 80% of the total community[92]. The characterization of the bifidobacterial pangenome – 18,181 *Bifidobacterium* specific Cluster of Orthologous Genes (BifCOGs) from 47 sequenced type strains – revealed the sugar-degrading functions of this microorganism and remarked the adaptation to the human gut environment[93]. Through comparisons of the GM of Hadza hunter-gatherers and urban industrial Italians, Schnorr et al[50] highlighted for the first time the substantial lack of *Bifidobacterium* from the GM of some traditional populations. The authors report the lack of bifidobacteria in adult Hadza hunter-gatherers as a consequence of the post-weaning GM composition drived by the absence of dairy foods, while the continued consumption of dairy into adulthood is one of the possible vectors by which many Westernized populations maintain a relatively large bifidobacterial presence. To date, comparative gut metagenome surveys do not specifically explore the impact of *Bifidobacterium* acquisition on the functional configuration of the GM of Western adults.

In order to examine changes to the GM as a result of these community shifts, here I investigate how the loss of *Treponema* and the acquisition of *Bifidobacterium* influenced the human gut metagenomic profile. To this aim, I compared gut metagenome functions

---

90 Segata, N. (2015). Gut microbiome: westernization and the disappearance of intestinal diversity. *Curr. Biol.* 25, R611–R613.

91 Blaser, M. J., and Falkow, S. (2009). What are the consequences of the disappearing human microbiota? *Nat. Rev. Microbiol.* 7, 887–894.

92 Turroni, F., Peano, C., Pass, D. A., Foroni, E., Severgnini, M., Claesson, M. J., et al. (2012). Diversity of bifidobacteria within the infant gut microbiota. *PLoS ONE7*:e36957.

93 Milani, C., Lugli, G., Duranti, S., Turroni, F., Mancabelli, L., Ferrario, C., et al. (2015a). Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut. *Sci. Rep.* 5, 15782.

assigned to *Treponema* and *Bifidobacterium* retrieved from downloadable GM metagenomic data for both Hadza hunter-gatherers and urban Italians. Findings reveal interesting functional gains in the urbanized microbiome corresponding to the post-weaning retention of *Bifidobacterium* as a symbiont microorganism, suggesting an opportunistic yet important role of this taxon in our adaptation to the urban environment.

Methods.

- Sample Collection and Shotgun Sequencing

The Illumina shotgun sequences used in this study were downloaded from the National Center for Biotechnology Information – Sequence Read Archive (NCBI SRA; SRP056480, Bioproject ID PRJNA278393).

- *Bifidobacterium* and *Treponema* Species Identification within Italian and Hadza Metagenomes

In order to identify the *Bifidobacterium* and *Treponema* species in Italian and Hadza populations, respectively, the 16S rDNA sequences within the assembled metagenomes were taxonomically selected using the assign_taxonomy.py script of the QIIME pipeline[28], against the Greengenes database[32]. The assignment at species level was performed by blastn[52] of the *Treponema* and *Bifidobacterium* 16S rDNA sequences against the entire NCBI nucleotide database and the top hit results for each sequence were retained for further analysis.

- Characterization of the CAZyme Repertoire Assigned to *Bifidobacterium* and *Treponema* in the Gut Metagenome.

Reads from a total of 38 individual GM metagenomes, 27 Hadza and 11 Italians[89] were downloaded and used for this study. Reads were assembled into contigs using MetaVelvet[94] with 350 bp as insert length. Predicted open reading frames (ORFs) were determined by FragGeneScan[95] on assembled contigs, using the –w 0 option for the fragmented genomic sequences and the parameter –t complete. From the translated ORFs I detected the CAZymes-coding sequences using hmmscan tool from the HMMER

94 Namiki T, Hachiya T, Tanaka H, Sakakibara Y. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res. 2012 Nov 1;40(20):e155.

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

software package[96] and the dbCAN database[97]. The outputs were processed using a custom script, selecting only the sequences that showed a minimum identity of 30% to the query sequences and an alignment length of at least 100 residues. In order to identify CAZymes derived from *Bifidobacterium* and *Treponema*, were retrieved the nucleotide sequences of the CAZymes detected with hmmscan from the FragGeneScan output, and then blasted them against the NCBI nucleotide database. Only the sequences that showed as best hit an assignment to *Bifidobacterium* for the Italian samples or *Treponema* for the Hadza samples were retained for further analysis. On the basis of the coverage of the contigs, was inferred information concerning the abundance of CAZymes. To compare the data among samples, I used a normalized CAZyme abundance by dividing the CAZyme coverages of every correspondent contig for the giga-bases of every correspondent sample.

- Read-Mapping Approach for the Detection of *Bifidobacterium* and *Treponema* Functions Involved in the Adaptation to the Gut Environment.

High quality reads for each sample were aligned to *Bifidobacterium*- or *Treponema*-assigned genes encoding bile acid adaptation, host interaction, and polysaccharide catabolism using bowtie2[50] and setting the alignment parameters to –sensitive-local. As reference for the alignment, two different databases containing orthologous genes from the NCBI genomes of the previously detected *Treponema* or *Bifidobacterium* species were created. Specifically, the databases contain genes for alpha-amylase, beta-galactosidase, mannanase, cellulase, pectinase, and xylanase. Furthermore, the databases were implemented with the sequences of the bile efflux pump, bile salt hydrolase, exopolysaccharide synthase, fimbrial subunit FimQ, sortase, galactosyl transferase, and undecaprenyl- phosphate phosphotransferase, since they were reported as genes that facilitate commensal-host cross-talk in *Bifidobacterium*[98]. In the event that the *Bifidobacterium* or *Treponema* NCBI genomes did not contain the above-mentioned genes, the databases were supplemented with genes belonging to the taxonomically

---

96 Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195.

97 Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40, W445–W451

98 Ferrario, C., Milani, C., Mancabelli, L., Lugli, G. A., Duranti, S., Mangifesta, M., et al. (2016). Modulation of the eps-ome transcription of bifidobacteria through simulation of human intestinal environment. *FEMS Microbiol. Ecol* 92:fiw056.

closest annotated microorganism. The reads that aligned with a reference using bowtie2, were extracted and their taxonomy was further verified by blastn[52] against the entire NCBI nucleotide database. Notably, in the case that the best hits of the blastn search were not assigned to *Bifidobacterium* or *Treponema*, I did not consider those reads for further analysis. The number of hits for each gene was normalized by the number of base pairs in the input file and in the correspondent reference in order to compare the results.

Results and discussion

Was first identified the diversity of the *Bifidobacterium* and *Treponema* species in urban Italian and Hadza GM by reconstructing the full 16S rDNA gene from assembled metagenomes. Italian samples contain sequences belonging to *Bifidobacterium fecale*, *Bifidobacterium pseudocatenulatum*, *Bifidobacterium adolescentis*, *Bifidobacterium coryneforme*, *Bifidobacterium bifidum*, *Bifidobacterium longum*, *Bifidobacterium angulatum*, and *Bifidobacterium dentium*. On the other hand 16S rDNA sequences assigned to *Treponema porcinum*, *Treponema bryantii*, *Treponema succinifaciens*, *Treponema parvum*, and *Treponema berlinense* were found in the GM of the Hadza hunter-gatherers. However, it must be acknowledged that these taxonomic assignments are limited by present whole genome for *Treponema* species, most of which have been characterized by work on human pathogens, rather than commensal members of the GM. In order to compare the specific carbohydrate-degrading functions conferred by *Bifidobacterium* and *Treponema* in the Italian and Hadza microbiomes, I identified a total of 5.4 million ORFs, of which 14,512 mapped to CAZymes for the Italian samples and 74,651 for the Hadza samples (See Figure 1 for details about the workflow).

Notably, the Hadza metagenomes contain significantly more CAZymes per-subject, in terms of ORFs assigned to CAZymes per million of reads, respect to the Italian metagenomes (mean ± SD, Hadza: 233 ± 86, Italians: 137 ± 78). Was then profiled the saccharolytic repertoire of *Bifidobacterium* and *Treponema* in the Italian and Hadza GM as relative abundance at the CAZyme category level based on the coverage of taxonomically assigned contigs (Figure 2A). *Bifidobacterium* showed a higher presence of glycosyl transferase (GT) and carbohydrate esterase (CE), with respect to *Treponema*. On the other hand, *Treponema* were more enriched in glycoside hydrolase (GH) and carbohydrate

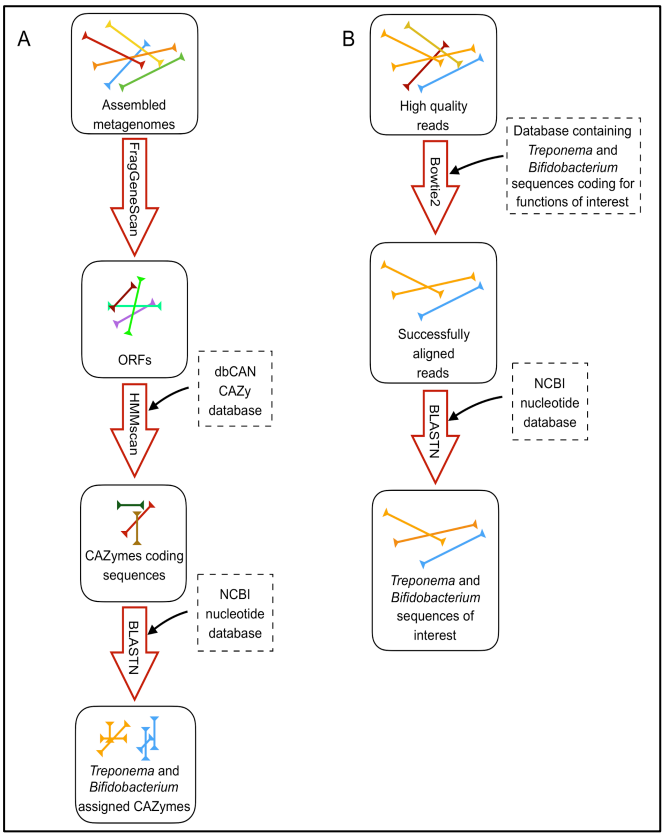


FIGURE 1 | Schematic representation of the analysis workflow. **(A)** Pipeline for the identification and assignment of *Treponema* and *Bifidobacterium* CAZymes on assembled metagenomes: (I) ORFs detection using FragGeneScan; (II) detection of the CAZyme-coding ORFs by using hmmscan against the dbCAN CAZy database; (III) taxonomy assignment to CAZyme-coding sequences by blastn against the NCBI nucleotide database. **(B)** Pipeline for the identification of *Treponema*and *Bifidobacterium* sequences coding for functions involved in the adaptation to the gut environment: (I) alignment of high quality reads to databases containing the selected *Treponema* or *Bifidobacterium* functions using bowtie2; (II) blasting of the successfully aligned reads against the NCBI nucleotide database to confirm the taxonomy.

binding module (CBM). At the CAZyme family level, I revealed the 4 families that constitute the core *Bifidobacterium* CAZyme repertoire: GH13 (GH family acting on substrates containing α-glycosidic linkages), GH3 (GH family that groups together exo-acting β-D-glucosidases, α-L-arabinofuranosidase, β-D-xylopyranosidase and *N*-acetyl-β-D-glucosaminidase), GT2 (GT family containing cellulose synthase, mannan synthase, and several monosaccharide-/oligosaccharide-transferases), and GT4 (GT family containing sucrose synthase, glucosyl transferase, and several phosphorylases). The sum of ORFs assigned to these four major families comprises 77% of the total detected CAZyme cohort. The relative abundance of the *Bifidobacterium* and *Treponema* CAZyme families detected in the Italian and Hadza samples reveals several differences in the potential carbohydrate-degrading functional contributions of these two microorganisms (Figure 2B). *Bifidobacterium* have a greater abundance of genes involved in the degradation of lactate, which is produced from pyruvate in the fermentation of simple sugar and commonly found in sour milk as well as in other lacto-fermented foods (family GH2). Emphasis on monosaccharide catabolism is evidenced by enrichment in gene families

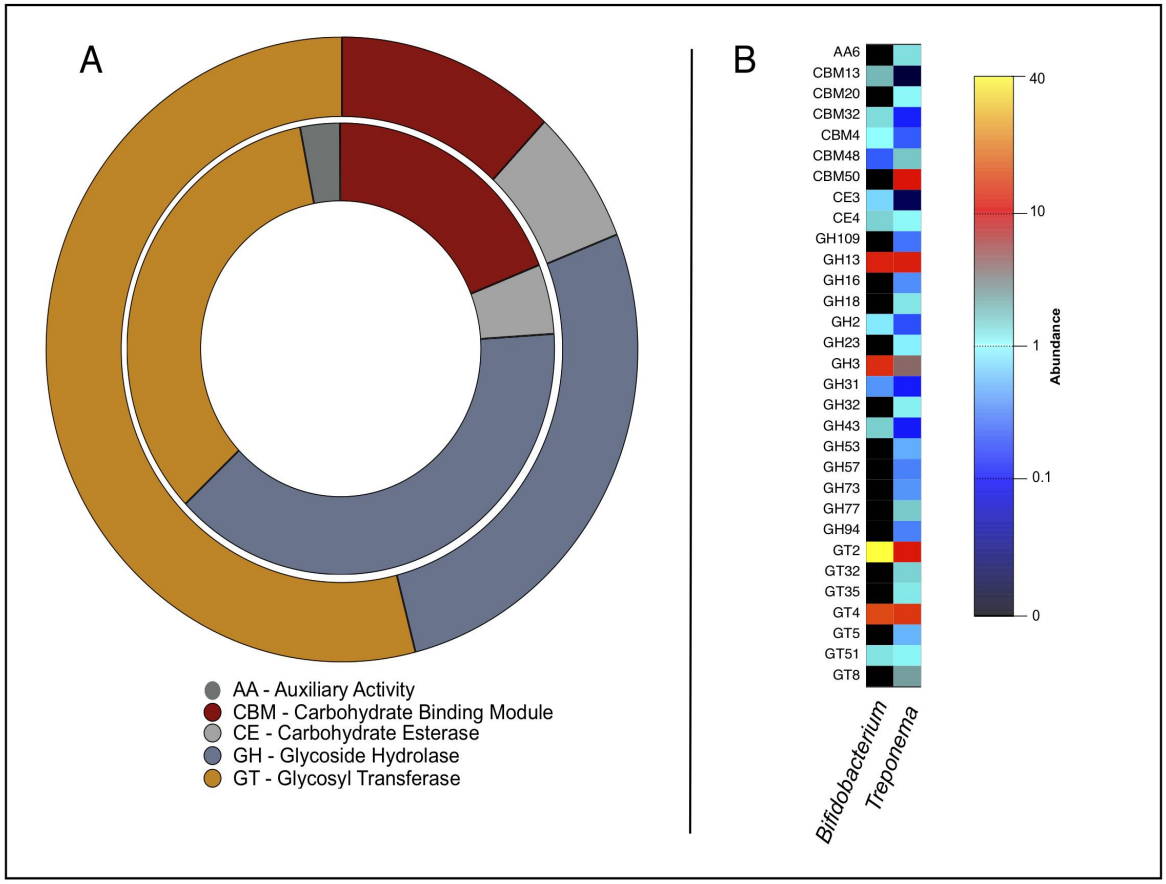


FIGURE 2 | Overview of *Bifidobacterium* and *Treponema* CAZyme repertoires in the Hadza and Italian samples. **(A)** Normalized relative abundance of the CAZyme category levels for *Bifidobacterium* and *Treponema*: auxiliary activity (AA), carbohydrate binding module (CBM), carbohydrate esterase (CE), glycoside hydrolase (GH), and glycosyl transferase (GT) categories. **(B)** Comparison between the *Bifidobacterium* and *Treponema* CAZyme family profiles. The relative abundance of each family is indicated by the color key.

also found that are involved in the degradation of α- and β-glucans (GH3 and GH31), this illustrates an ability of *Bifidobacterium* to retrieve energy also from more complex carbohydrates that are commonly present in the cellulosic biomass of plant foods in the Italian diet: salads, fruits, nuts, cereals and their product derivatives. In addition, *Bifidobacterium* are also enriched in genes involved in the catabolism of sucrose (GH31), which is widely distributed in nature, but robustly manifest in the industrial food products that are consumed daily by most urban populations. Further evidence of these functions comes from detection of a higher abundance of CBM families for lactose, galactose and β-glucans (CBM4, CBM13, and CBM32) in *Bifidobacterium*, with respect to *Treponema*. In contrast, the CAZyme profile of *Treponema* within the Hadza metagenome is mainly

devoted to degradation of glucans, galactans, and fructans (GH16, GH32, and GH53), which are sugar polymers that comprise hemicellulose (galactans) and inulin (fructans). The monosaccharide of galactans, galactose, is also expressed in mucilages and glycoproteins that derive from the human host, as well as a number of vegetable-derived carbohydrates. Both sugar polymers are largely implicit in difficult-to-digest plant polysaccharides that escape small intestine absorption and are instead fermented by the colonic microbiota. The Hadza diet is rich in such unrefined plant foods that contain indigestible polysaccharides such as berries, baobab fruit, and particularly tubers. *Treponema* are also enriched in two CAZyme α-amylase families (GH57 and GH77), which are unlike the typical α-amylase GH13 family because they have a conserved trans-glycosylating region. Finally, *Treponema* are better equipped to metabolize peptidoglycans due to a wide range of acetyl-glucosaminases and peptidoglycan-lyases (GH23, GH73, and GH109). These activities were confirmed by the detection of high levels of CBM families for peptidoglycans and α-glucans (CBM50 and CBM48).

I further investigated the presence of the genes involved in host interaction and immune modulation in the ORFs attributed to *Bifidobacterium* and *Treponema*. Analyses confirm that the enzymes involved in the production of EPS and pili, namely EPS synthase, undecaprenyl-phosphate phosphotransferase, galactosyl transferase, sortase, and fimbrial subunit FimQ, are typical of the *Bifidobacterium* ORFs detected in the Italian metagenome, while virtually absent in the *Treponema* ORFs retrieved from the Hadza GM ecosystem (Figure 3). Finally, was evaluated the presence of genes involved in bile tolerance as mechanisms of bacterial adaptation to the human host. Bile salts are detergent-like compounds with strong antimicrobial activity[99], and intestinal bacteria have had to evolve strategies to tolerate physiological concentrations of bile salts to colonize the intestine. Interestingly, two representative enzymes, which contribute to bile resistance and adaptation to gut environment, the bile-inducible efflux transporters and the bile salt hydrolase, are present in within the *Bifidobacterium* gut metagenome functions, but are not detected in *Treponema* ORFs (Figure 3).

[99] Begley, M., Gahan, C. G., and Hill, C. (2005). The interaction between bacteria and bile. *FEMS Microbiol. Rev.* 29, 625–651.

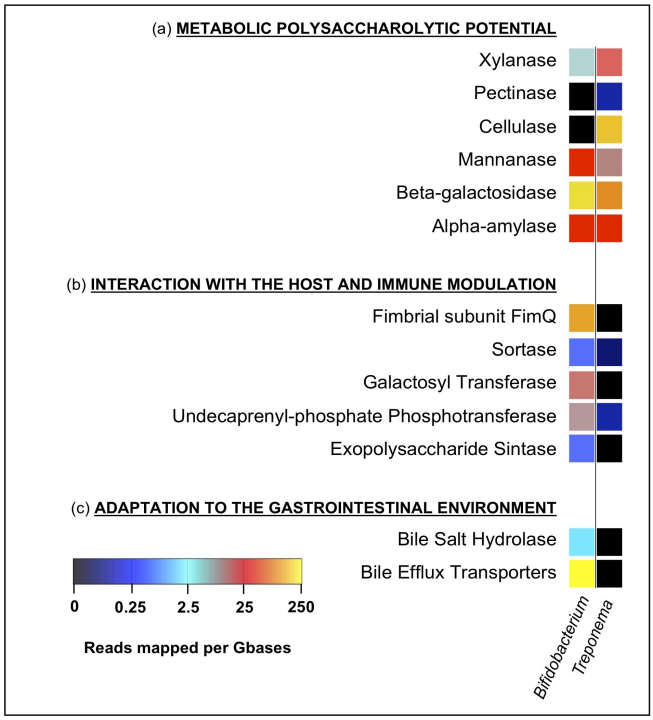


FIGURE 3 | Profile of *Bifidobacterium* and *Treponema* functions involved in the adaptation to the host environment. Polysaccharide metabolism **(A)**; interaction with the host and immune modulation **(B)**; adaptation to the human gastrointestinal environment **(C)**. Color key represents reads mapped per giga-bases of the sample of origin.

Conclusion

Findings suggest possible co-evolutionary implications for the loss of *Treponema* and the acquisition of *Bifidobacterium* as a stable component of the post-weaning GM ecosystem from post-industrial urban populations. Capable of heterogeneous carbohydrate metabolism, which ranges from complex plant polysaccharides to simpler sugars such as lactose and sucrose, *Bifidobacterium* are well suited to handle the degradative tasks imposed by a typical Western diet. Conversely, the progressive loss of more challenging microbiota accessible carbohydrates in the Western diet[100], such as hemicellulose and inulin, would help partially explain the extinction of a more specialized fiber degrader such as *Treponema* from the Western GM ecosystem. Furthermore, unlike *Treponema*, *Bifidobacterium* evolved the capacity to establish an intense microbe–host connection, which may help support a continuous and abundant Bifidobacterial presence in adults, allowing this commensal to outcompete other opportunistic, but functionally

100 Sonnenburg, E. D., and Sonnenburg, J. L. (2014). Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metab.* 20, 779–786.

diverse, microbiota. The acquisition of *Bifidobacterium* as a stable component of the GM ecosystem in small-scale rural agriculturalists, reminiscent of early human farmers, and modern Westernized populations, may therefore engage the functionalities of the host immune system, providing new adaptive solutions in response to changing selective pressures during the restructuring of human diet and society.

Section 3.2 - The bottlenose dolphin (Tursiops truncatus) gut microbiota

Introduction

Cetaceans have evolved from herbivorous terrestrial artiodactyls closely related to ruminants and hippopotamuses. *Delphinidae*, a family included in this order, represent an extreme and successful re-adaptation of mammalian physiology to the marine habitat and piscivorous diet. The anatomical aspects of *Delphinidae* success are well understood, whereas some physiological aspects of their environmental fitness are not yet defined, such as the gut microbiota composition and the adaptation to their dietary niche. Mammalians have evolved a beneficial relationship with symbiotic intestinal microorganisms collectively known as the gut microbiota. Providing the host with an additional source of essential nutrients and representing an endogenous defense against the colonization by opportunistic pathogens[101] , this mutualistic microbial counterpart has represented a strategic evolutionary advantage for the mammalian lineage, contributing, at least in part, to the evolutionary success of this class[102] . Expanding the host metabolic potential, the symbiont intestinal microorganisms have been a key factor for the mammalian radiation, allowing exploiting new dietary niches[103]. For instance, the adaptation to a plant-based diet during the Quaternary has depended on the acquisition of a new and characteristic symbiont microbial community and digestive tract anatomy. This provided essential nutrients to the host from otherwise inaccessible plant material[104].

In an attempt to understand the role of gut microbes in the mammalian evolution process, some studies comparing the gut microbiota structure among different mammalian species, including terrestrial carnivores, omnivores and herbivores, have been conducted[69; 105]. The results from these studies indicate that the composition of the mammalian gut microbiota is principally driven by diet, resulting in robust distinctive phylogenetic and functional ecosystem profiles for carnivores, omnivores and herbivores.

---

101 Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Im- munol* 2009;9:313–23.

102 Nelson TM, Rogers TL, Brown MV. The gut bacterial community of mammals from marine and terrestrial habitats. *PLoS One*2013;8:e83655.

103 McFall-Ngai M, Hadfield MG, Bosch TC *et al*. Animals in a bacterial world, a new imperative for the life sciences. *P Natl Acad Sci USA* 2013;110:3229–36.

104 Ley RE, Hamady M, Lozupone C *et al*. Evolution of mammals and their gut microbes. *Science* 2008a;320:1647–51.

105 Muegge BD, Kuczynski J, Knights D *et al*. Diet drives conver- gence in gut microbiome functions across mammalian phylogeny and within humans. *Science* 2011;332:970– 4.

However, a significant amount of gut microbiome variation was also connected to phylogeny and gut morphology, both recognized as possible additional drivers influencing the mammalian gut microbiota composition. For instance, in some mammalians with atypical diets for their clade, such as the herbivorous panda bear[106] and myrmecophagous mammalians[107], the hosted gut microbial communities were more similar to those of their close relatives than to those of other mammalians with comparable diet. These cases of phylogenetic inertia highlight the importance of the host phylogeny in constraining the range of variation of the gut microbiota composition in response to diet. Cetaceans, like whales and dolphins, represent a model of primary importance to understand how diet, phylogeny and gut morphology have combined to drive the mammalian-gut microbiota co-evolution process. Indeed, evolved from herbivorous terrestrial artiodactyls related to cows and hippopotamuses[108], cetaceans, as well as their ruminant relatives, still retain a multi-chambered foregut, in spite of a carnivorous diet[109]. In the present work, is explored the gut microbiota in bottlenose dolphins (*Tursiops truncatus*). Dolphins belong to odontocetes, possess processing teeth and are characterized by a piscivorous diet[110] despite having a multi-chambered stomach, which is unusual for carnivores[111]. Thus, the gut microbial ecosystem of dolphins could represent a new and peculiar gut microbiota-host configuration, specifically adapted to the carnivorous diet and evolved in aquatic mammalians. Here, is characterised the gut microbiota ecosystem from 9 adult bottlenose dolphins, each one sampled at least two times in a period of 6 months. The dolphin gut microbiome was compared with that of 33 mammalian species from Muegge et al.[70] and baleen whales from Sanders et al.[112]. It is also assessed the gut microbial ecosystem of one breast-fed calf at birth (meconium), 2 and 7 months of age, representing the first glimpse on the dynamics of the gut microbiome assembly in aquatic mammalians.

---

106 Zhu L, Wu Q, Dai J *et al*. Evidence of cellulose metabolism by the giant panda gut microbiome. *P Natl Acad Sci USA*2011;108:17714–9.

107 Delsuc F, Metcalf JL, Wegener Parfrey L *et al*. Convergence of gut microbiomes in myrmecophagous mammals. *Mol Ecol*2013;23:1301–17.

108 Gatesy J, Geisler JH, Chang J *et al*. A phylogenetic blueprint for a modern whale. *Mol Phylogenet Evol* 2013;66:479–506.

109 Langer P. Evidence from the digestive tract on phylogenetic relationships in ungulates and whales. *J Zool Syst* 2001;39: 77–90.

110 BlancoC,Salomo`nO,RagaJA.Dietofthebottlenosedolphin(*Tursiops truncatus*) in the western Mediterranean Sea. *J Mar Biol Assoc U.K.* 2001;81:1053–8

111 Reidenberg JS. Anatomical adaptations of aquatic mammals.*Anat Rec (Hoboken)* 2007;290:507–13.

112 Sanders JG, Beichman AC, Roman J *et al*. Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. *Nat Commun* 2015;6:8285.

Methods

- Animals and sample collection

Nine adult bottlenose dolphins (*T. truncatus*) (4 females and 5 males) and one calf housed at Oltremare (Riccione RN, Italy) were used for the present study. The dolphins were maintained in public display in outdoor pools. Captive bottlenose dolphins are constantly monitored under the ethical code enforced by European law, and the sample collection was scheduled to overlap veterinary medical health control programs. Diets consisted of frozen fish, including herring (*Clupea harengus*), capelin (*Mallotus villosus*), sprat (*Sprattus sprattus*), blue whiting (*Micromesistius poutassou*), mackerel (*Scomber scombrus*), and squid (*Loligo opalescens*) and were formulated to meet individual animal requirements.
From each adult animal, at least 2 fecal samples were collected during the 6-month study period. All samples were collected from unrestrained animals, using routine husbandry positive reinforcement training techniques and following EAAM standards for facilities housing bottlenose dolphins (www.eaam.org/housing_standards) and the EU directive for wild animals kept in zoos (http://eur-lex.europa.eu). The animals used in the present study were trained for fecal sample collection as part of a standard clinical examination. Fecal samples were placed in a plastic container and kept frozen at -20°C until analysis. Meconium was collected directly from the water with a net immediately after the newborn calf expelled it, and immediately frozen at -20°C. Milk was collected from a mother dolphin when her calf was two days old as part of normal medical procedures to perform cultural examination and to eliminate the presence of mastitis.

- Bioinformatics and statistics

DNA was extracted and sequenced, and amplicons were submitted on MG-RAST under project ID 15865. For further informations on DNA processing refer to Methods section of Soverini et al[113]
To analyse the raw sequences, a pipeline combining PANDAseq and QIIME[28] was used. High-quality reads were clustered into operational taxonomic units (OTUs) at 97% similarity threshold using UCLUST. Taxonomy was assigned through matching a representative sequence for each cluster against the Greengenes database[33] (May 2013

113 Soverini M, Quercia S, Biancani B, Furlati S, Turroni S, Biagi E, Consolandi C, Peano C, Severgnini M, Rampelli S, et al. 2016. The bottlenose dolphin (*Tursiops truncatus*) fecal microbiota. *FEMS Microbiol Ecol* 92:fiw055.

release). All singleton OTUs were discarded. Alpha rarefactions were analysed by using the Faith's phylogenetic diversity, Chao1, observed species, and Shannon index metrics. Beta diversity was estimated by computing Jaccard dissimilarity. Jaccard distances were used for Principal Coordinates Analysis (PCoA) and plotted by the rgl and vegan packages of R. Data separation in the PCoA was tested using a permutation test with pseudo F-ratios (function adonis in the vegan package). Heat map analysis was performed using the R ggplot2 and ape packages. To perform all statistical analysis, R software (version 3.1.3) was used. Significant differences were assessed by Wilcoxon signed rank sum test. When appropriate, a paired test was used. Where necessary, P values were corrected for multiple comparisons using the Benjamini-Hochberg method. $P < 0.05$ was considered as statistically significant.

Results and discussion

- The gut microbiota composition in bottlenose dolphins

A total of 29,005,176 high-quality reads were clustered into 11,465 operational taxonomic units (OTUs) at 97% identity (mean, 356 ± 51 OTUs per sample). At the phylum level, Firmicutes (mean relative abundance (r.a.) ± SEM, 56 ± 4.8%) and Proteobacteria (27 ± 8.8%) dominated the gut microbiota ecosystem of adult dolphins. Actinobacteria (5 ± 8.8%), Bacteroidetes (3 ± 0.8%), Fusobacteria (4 ± 2.2%) and Tenericutes (3 ± 2.3%) were subdominant phyla (Figure 4).

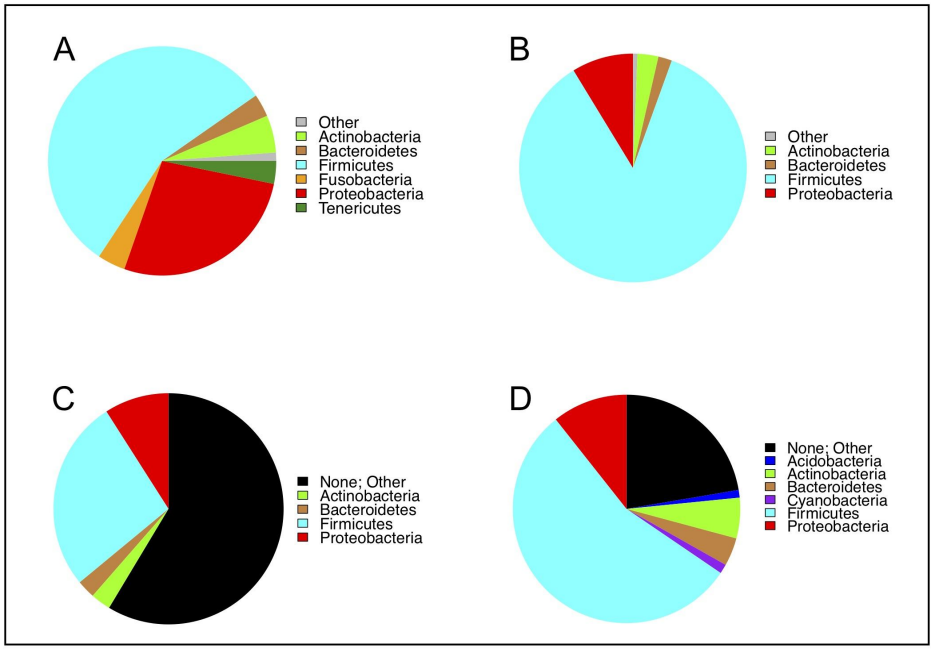


Figure 4 | The bottlenose dolphin microbiota at phylum level. Pie charts summarizing the phylum-level microbiota composition in the feces from nine adult bottlenose dolphins (**A**) and one calf (**B**), as well as in meconium (**C**) and maternal milk microbial ecosystem (**D**). Only phyla with a relative abundance ≥1% in at least 20% of subjects (A), or in at least 1 sample (B–D) are represented.

The most represented families were Clostridiaceae (16 ± 4.7%), *Vibrionaceae* (12 ± 4.3%), *Staphylococcaceae* (6 ± 1.9%), *Lactobacillaceae* (7 ± 2.9%), *Peptostreptococcaceae* (7 ± 2.7%), *Ruminococcaceae* (5 ± 1.4%), *Fusobacteriaceae* (4 ± 2.2%) and *Pasteurellaceae* (4 ± 1.9%) (Figure 5A). As already documented in humans[114], the breast-fed dolphin calf showed a peculiar compositional structure of the gut microbial ecosystem, with relevant differences when compared to the adult counterpart. In particular, in the calf gut microbiota we observed a clear predominance of Firmicutes (mean r.a. 86%), while Proteobacteria, the second dominant phylum in the adult ecosystem, was largely subdominant (8%), together with Actinobacteria (3%) and Bacteroidetes (2%) (Figure 1B). The dominant families in the calf microbial ecosystem were *Clostridiaceae* (34%) and *Peptostreptococcaceae* (31%). *Ruminococcaceae* (5%), *Enterobacteriaceae* (5%), *Enterococcaceae* (3%), *Staphylococcaceae* (3%), *Lachnospiraceae* (4%), *Streptococcaceae* (2%), *Prevotellaceae* (1%) and *Sphingomonadaceae* (1%) were present at a lower abundance (Figure 5B).

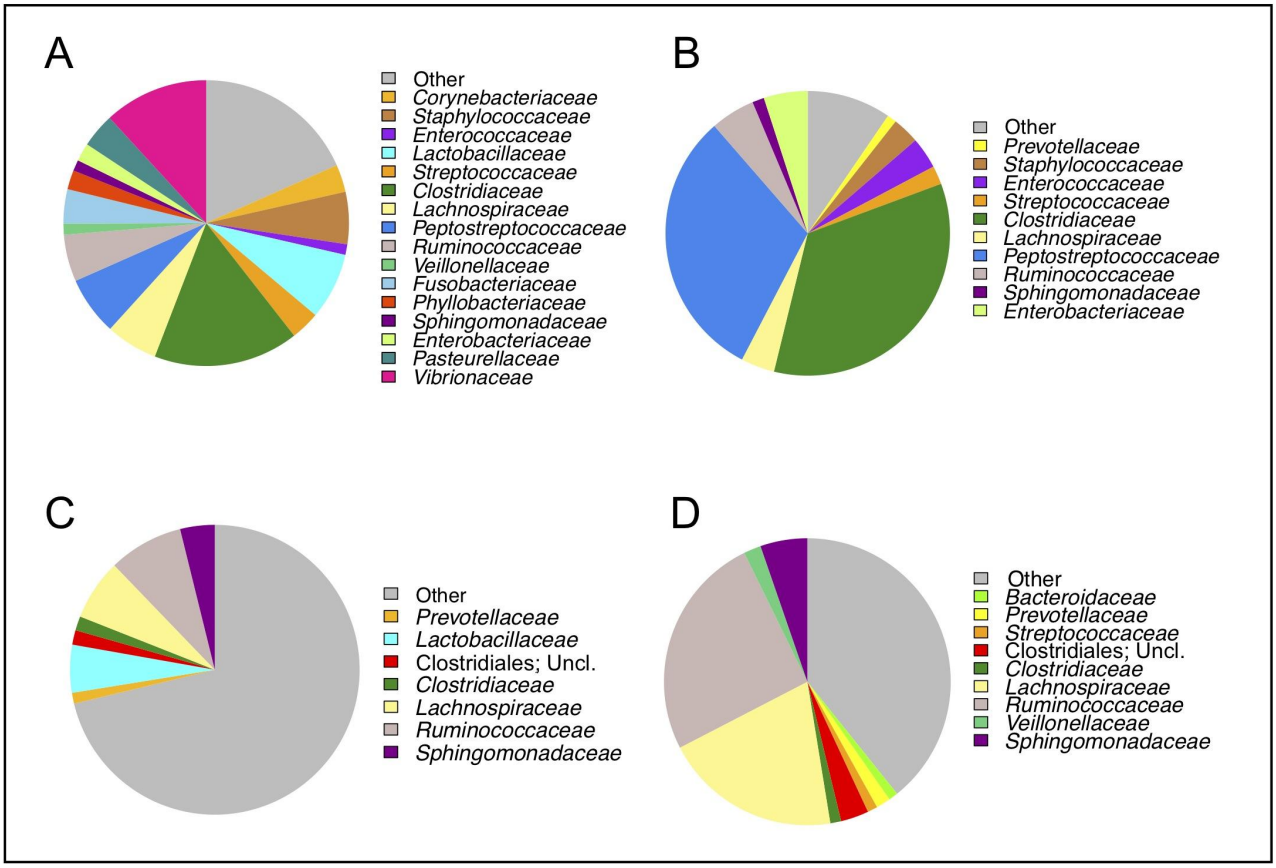


Figure 5 | The bottlenose dolphin microbiota at family level. Pie charts summarizing the family-level gut microbiota composition of adult dolphin (**A**), dolphin calf (**B**), meconium (**C**) and maternal milk (**D**).

[114] Candela M, Biagi E, Turroni S *et al.* Dynamic efficiency of the human intestinal microbiota. *Crit Rev Microbiol* 2013;41:165–71

In order to assess similarities among the 27 dolphin samples, including meconium and the mother's milk, a Ward-linkage hierarchical clustering of the weighted UniFrac distance matrix of the phylogenetic profiles was performed (Figure 6).

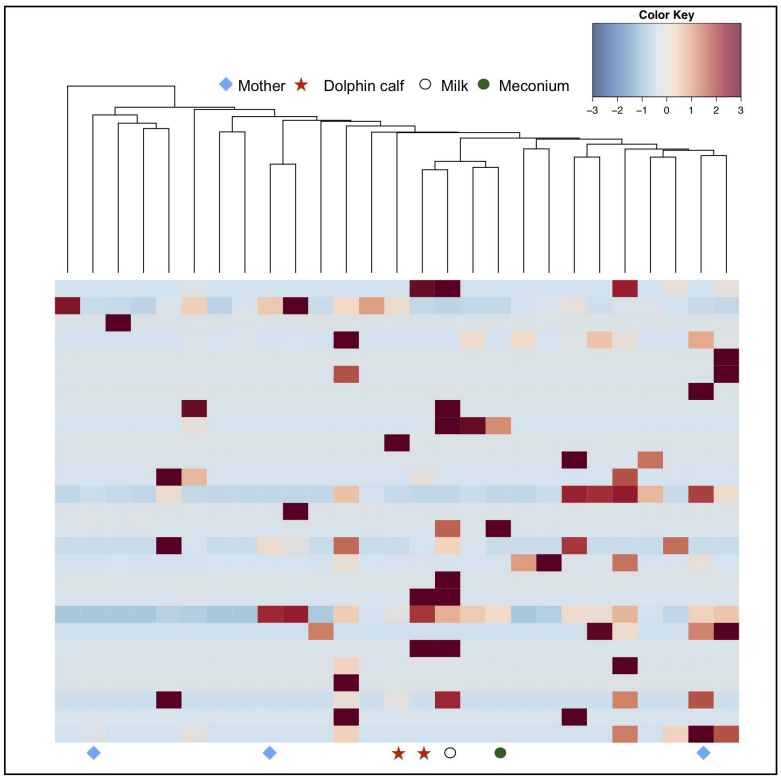


Figure 6 | Ward-linkage hierarchical clustering based on weighted UniFrac distance matrix of the microbiota profiles from fecal samples of nine adult bottlenose dolphins and one calf, maternal milk and meconium. Different time points from the same individual are labelled with a letter according to the temporal sequence of sampling (See Table S1, Supporting Information for more detail). Red star, fecal samples of the dolphin calf; blue twisted square, fecal samples of the calf mother; white circle, the maternal milk; green circle, meconium.

Demonstrating the high degree of temporal variability of the dolphin individual gut microbiome profile, fecal samples from the same adult individual collected at different time points did not cluster together. The only samples close to each other were those from the breast-fed calf collected at 2 and 7 months of age. Interestingly, both calf samples clustered in close proximity to the mother's milk, suggesting the importance of the milk microbial ecosystem for the process of gut microbiome assembly in breast-fed dolphins. Indeed, the compositional structure of the maternal milk microbiota well mirrored that of the gut microbiota observed in the calf.

For instance, at the phylum level the milk ecosystem was dominated by Firmicutes (r.a. 55%), while Proteobacteria (11%), Actinobacteria (6%), Bacteroidetes (4%), Acidobacteria (1%) and Cyanobacteria (1%) were subdominant components (Figure 4D). The most represented families in milk were Ruminococcaceae (25%) and Lachnospiraceae (20%) (Figure 5D). Finally, being dominated by Firmicutes with Proteobacteria as subdominant phylum, the microbial ecosystem from meconium showed a compositional structure that

generally approximated the calf gut microbiota at 2 and 7 months, as well as the milk microbial ecosystem (Figure 4C, Figure 5C). Interestingly, a considerable fraction (~60%) of OTUs detected in the meconium could not be assigned to any phylum.

- Phylogenetic differences in the gut microbiota compositional structure between dolphins, terrestrial mammalians and baleen whales

In order to explore distinctive features of the dolphin gut microbial ecosystem within the Mammalia class, the gut microbiota from the 9 adult bottlenose dolphins was compared with that previously published from 33 mammalian terrestrial species and 5 baleen whales. The PCoA ordination of the Jaccard distances of the genus-level gut microbiome compositional structures resulted in a significant segregation among terrestrial mammalian herbivores, omnivores, carnivores, baleen whales and dolphins (permutation test with pseudo F-ratios $P < 0.001$) (Figure 7). In the figure, the bacterial genera with the largest contribution to the ordination space are also shown. In particular, *Lactobacillus, Staphylococcus, Peptostreptococcus* and unclassified *Clostridiaceae* were the microbial genera characterizing the dolphin gut microbial ecosystem.

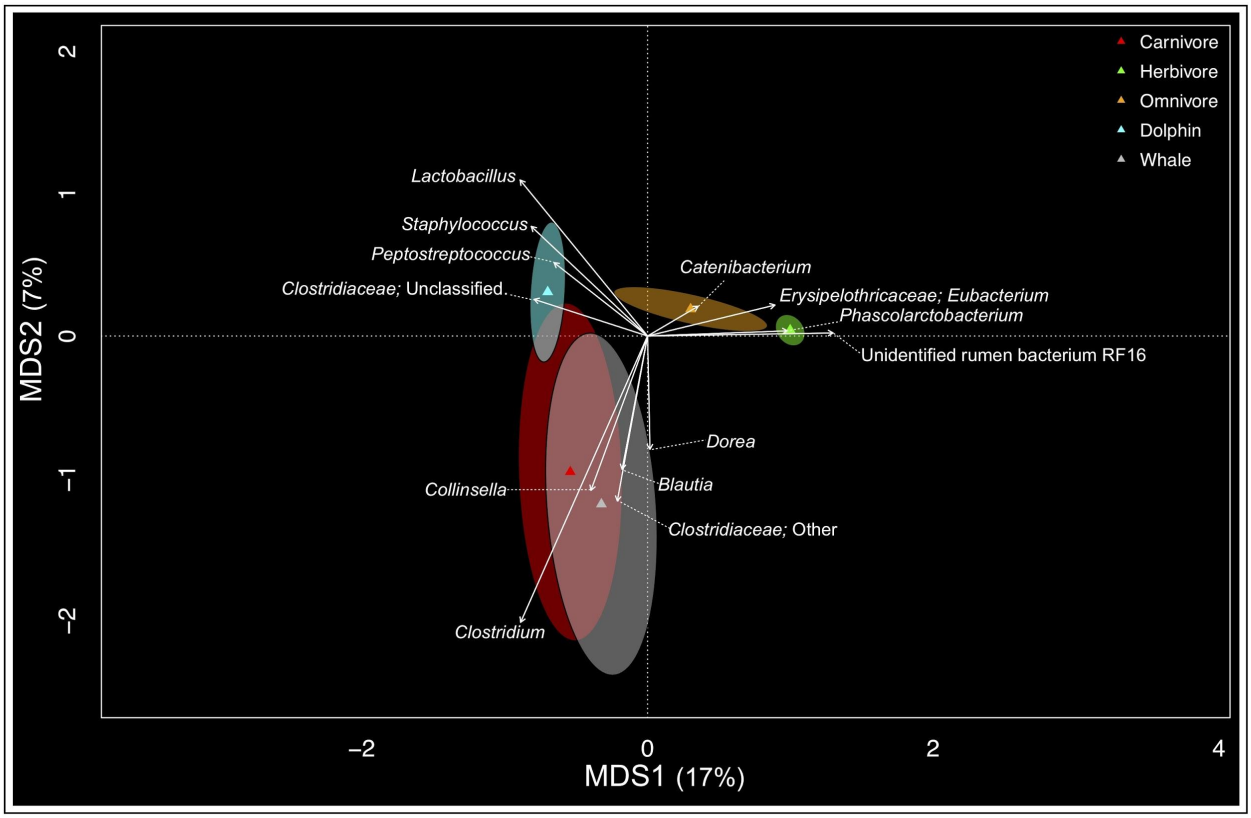


Figure 7. PCoA obtained by Jaccard distance matrix showing the separation between carnivores, herbivores, omnivores, dolphins and baleen whales based on their gut microbial composition.

According to the findings, when compared to other mammalians dolphins show a distinctive compositional structure of the gut microbial ecosystem, being dominated by *Clostridiaceae*, *Vibrionaceae*, *Staphylococcaceae*, *Lactobacillaceae* and *Peptostreptococcaceae*. Since this microbial pattern is overall different from the prey microbiota (herrings and mackerels)[115-116], for the dolphins of the present study a predominantly endogenous nature for the gut microbiota ecosystem can be hypothesized. The data are in agreement with what reported by Bik et al.[117]. The authors explored the microbiota composition from rectal specimens from 48 bottlenose dolphins, of which 38 were under the care of the US Navy Marine Mammal Program (MMP) in San Diego Bay, and 10 were free-ranging animals sampled during capture-release health assessments in Sarasota Bay, Florida. According to their findings, dolphins possess a characteristic rectal microbial community, which is dominated by Firmicutes, Proteobacteria and Fusobacteria, and shows a sharp boundary with the microbial communities from the surrounding water and fish diet.

Previous studies observed only subtle differences between the gut microbiota structures from animals raised in artificial and natural environment, showing that co-specific diversification is robust to captivity[69]. Thus, it is reasonable to expect that the peculiarities of the dolphin gut microbial ecosystem observed in this study remain robust without regarding the animal provenience. However, it should be noted that cetaceans are constantly immersed in the microbial assemblages present in the water, which can influence the gut microbiota composition. As the water microbial communities can be profoundly different when comparing artificial systems and native habitats, differences in the gut microbiota composition between dolphins from artificial environments and natural habitats cannot be excluded. A possible glimpse in this direction has been provided by Bik et al.[82]. Indeed, while the authors discovered significant differences in the oral microbial community from MMP and wild free-ranging dolphins, the corresponding differences between the rectal microbial communities did not reach the statistical significance. From a first sight, the overall structure of the dolphin gut microbiome, as

---

115 Svanevik CS Lunestad BT Characterisation of the microbiota of Atlantic mackerel (*Scomber scombrus*) *Int J Food Microbiol* 2011 151 164 70

116 Olsen MA Aagnes TH Mathiesen SD Digestion of herring by indigenous bacteria in the minke whale forestomach *Appl Environ Microb* 1994 60 4445 55

117 Bik EM Costello EK Switzer AD et al. Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea *Nat Commun* 2016 3 10516

determined in the present study, resembles that observed in terrestrial mammalian carnivores. However, at a closer look, specific adaptive declinations of the dolphin gut microbiota to the piscivorous diet in marine environment are evident. Indeed, the dolphin gut microbial ecosystem shares several compositional features with terrestrial mammalian carnivores, such as the high abundance of proteolytic *Fusobacteriaceae*, *Enterobacteriaceae, Enterococcaceae, Streptococcaceae, Peptostreptococcaceae* and *Clostridiaceae*, and the corresponding low amount of fibre-degrading microorganisms. On the other hand, in addition to these carnivore-like gut microorganisms, in the dolphin gut microbiome we also observed compositional features that specifically reflect the adaptation to the piscivorous diet and marine environment. Indeed, as already observed in marine carnivorous fishes[118-119-120] and mammalian piscivores living in marine environment[121] , the gut microbiota of dolphins is enriched in Alpha and Gammaproteobacteria, *Staphylococcaceae*, *Corynebacteriaceae* and *Lactobacillaceae*.

Taken together, the data suggest that, even if still retaining a multi-chambered foregut derived from their herbivorous terrestrial ancestor (artiodactyls), dolphins possess a gut microbiota ecosystem

similar to that of marine piscivores. This indicates the importance of the gut microbiota ecosystem in dolphins as an adaptive partner, strategic for the occupation of new dietary niches in the marine environment. However, it should be noted that free-ranging dolphins opportunistically eat non-fish prey as well, and their overall diet composition can vary with geographical location and season. Thus, further studies sampling dolphins in their natural habitat will be needed to explore connections between their actual diet and the gut microbiota layout.

Within cetaceans, dolphins and baleen whales show different gut microbiome profiles that mirror their respective dietary niches. While dolphins possess a gut microbiota structure with declinations that approximate what observed in marine piscivores, the baleen whale gut microbiome shares compositional and functional

---

118 Sullam EK Essinger SD Lozupone CA et al. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis *Mol Ecol* 2012 21 3363 78

119 Estruch GM Collado C Peñaranda DS et al. Impact of fishmeal replacement in diets for gilthead sea bream (*Sparus Aurata*) on the gastrointestinal microbiota determined by pyrosequencing the 16S rRNA gene *PLoS One* 2015 10 e0136389

120 Llewellyn MS McGinnity P Dionne M et al. The biogeography of the atlantic salmon (*Salmo salar*) gut microbiome *ISME J* 2015 DOI: 10.1038/ismej.2015.189

121 Nelson TM Rogers TL Brown MV The gut bacterial community of mammals from marine and terrestrial habitats *PLoS One* 2013 8 e83655

similarities with that of terrestrial herbivores[77]. In particular, when compared with dolphins, baleen whales were significantly depleted in microorganisms characteristic of marine piscivores, such as Alpha and Gammaproteobacteria, *Staphylococcaceae*, *Corynebacteriaceae* and *Lactobacillaceae*, while correspondingly enriched in fibre-fermenting *Spirochaetaceae*[122]. Similar to their herbivorous artiodactyl ancestors, baleen whales possess a blind-end caecum between the ileum and colon, which is absent in odontocetes. This would allow filtering feeding baleen whales—a relatively recent innovation within cetaceans—to accommodate in their gut fibrolytic microorganisms specialized in chitin fermentation, a strategic factor to extract energy from the most abundant biopolymer in the sea[123]. Thus, baleen whales have evolved the ability to capitalize the large amount of chitin present in the crustacean exoskeleton introduced with their diet through the microbiota fermentation of this complex polysaccharide. On the other hand, dolphins have occupied a different nutritional niche, becoming piscivores and sharing aspects of their microbiota with more diverged groups like carnivorous marine fishes. According to findings, the gut microbiome ecosystem of the breast-fed dolphin calf possesses a characteristic configuration different from that detected in adults. In particular, the calf gut microbiota was rich in proteolytic *Clostridiaceae* and *Peptostreptococaceae*, and depleted in microorganisms characteristic of marine piscivores. This latter feature is shared with the milk microbiota that, besides the low abundance of Proteobacteria¸ is characterized by the enrichment in the fibre-degrading families *Ruminococcaceae* and *Lachnospiraceae*. Even if limited, our data highlight the overall compositional similarity between the microbial ecosystem of the calf gut and the mother's milk, suggesting the importance of the milk microbiome in the process of mammalian gut microbiota assembly in marine environment.

Conclusion

In conclusion, the adaptation to the piscivorous diet and marine environment has been the major driver modeling the co-diversification between dolphins and their intestinal microbes. The high degree of compositional similarities between the gut microbiota of dolphins and carnivorous marine fishes suggests that in dolphins the

122 Obregon-Tito AJ Tito RY Metcalf J et al. Subsistence strategies in traditional societies distinguish gut microbiomes *Nat Commun* 2015 6 6505

123 Beier S Bertilsson S Bacterial chitin degradation—mechanisms and ecophysiological strategies *Front Microbiol* 2013 4 149

adaptation to the marine environment involved the compositional convergence of their gut microbiota with that of marine fishes, overcoming the phylogenetic inertia.

Section 3.3 - Unraveling the gut microbiome of the long-lived naked mole-rat

Introduction

By preserving the biological homeostasis of the holobiont, the gut microbiota has a role of primary importance in supporting human longevity[124]. However, only few hypotheses on the mechanisms involved have been advanced. Longevity is a tricky trait to be studied in humans, because it is a rare event, with an incredible amount of confounding genetic, lifestyle and clinical variables, both past and present. Still, the microbiota of human populations with extraordinary longevity rate is being investigated across geographical zones [125] - [126] and interesting hypotheses on the role of the microbiome in health-maintenance and adaptation during aging are being advanced.

In this scenario, the naked mole-rat (*Heterocephalus glaber*) might represent an extremely interesting model to study health and longevity, since, like for human beings, in naked mole rat the selection against aging is strongly reduced[127]. This eusocial, subterranean mouse-sized mammal, native to the arid and semi-arid regions of the Horn of Africa, occupies underground mazes of sealed tunnels and lives a very long life (30 years, approximately 8 times longer than common mice and rats) in large colonies[128]. Phylogenetically, this small rodent is classified within the newly-defined family *Heterocephalidae*, separated from the other African mole-rat species (*Bathyergidae*)[129]. The naked mole-rat shows few age-related degenerative changes[130], displays an elevated tolerance to oxidative stress[131], and its fibroblasts have shown resistance to heavy metals, DNA damaging agents, chemotherapeutics and other poisonous chemicals[132]. Moreover, this mammals show remarkably small susceptibility to both spontaneous cancer and

---

124 Biagi, E., Candela, M., Fairweather-Tait, S., Franceschi, C. & Brigidi, P. Aging of the human metaorganism: the microbial counterpart. *Age (Dordrecht, Netherlands)* 34, 247–267 (2012)

125 Biagi, E. *et al.* Gut Microbiota and Extreme Longevity. *Curr. Biol.* 26, 1480–1485 (2016).

126 Kong, F. *et al.* Gut microbiota signatures of longevity. *Curr. Biol.* 26, R832–833 (2016).

127 Skulachev, V. P. *et al.* Neoteny, Prolongation of Youth: From Naked Mole Rats to "Naked Apes" (Humans). *Physiol. Rev.* 97, 699–720 (2017).

128 Lewis, K. N. *et al.* Unraveling the message: insights into comparative genomics of the naked mole-rat. *Mamm. Genome* 27, 259–278 (2016).

129 Patterson, B. D. & Upham, N. S. A newly recognized family from the Horn of Africa, the Heterocephalidae (Rodentia: Ctenohystrica). *Zool. J. Linn. Soc.* 172, 942–963 (2014).

130 Grimes, K. M., Reddy, A. K., Lindsey, M. L. & Buffenstein, R. And the beat goes on: maintained cardiovascular function during aging in the longest-lived rodent, the naked mole-rat. *Am. J. Physiol. Heart Circ. Physiol.* 307, H284–291 (2014).

131 Perez, V. I. *et al.* Protein stability and resistance to oxidative stress are determinants of longevity in the longest-living rodent, the naked mole-rat. *Proc. Natl. Acad. Sci. USA* 106, 3059–3064 (2009).

132 Salmon, A. B., Sadighi Akha, A. A., Buffenstein, R. & Miller, R. A. Fibroblasts from naked mole-rats are resistant to multiple forms of cell injury, but sensitive to peroxide, ultraviolet light, and endoplasmic reticulum stress. *J. Gerontol. A Biol. Sci. Med. Sci.* 63, 232–241 (2008).

induced tumorigenesis[133-134-135]. These features of the naked mole-rat are maintained throughout their long lifespan, making this rodent a putative animal example of impressively prolonged "healthspan". Moreover, the within-colony low genetic diversity (possibly due to the high inbreeding rate)[136], the climatologically stable underground habitats, and the constant diet (mainly tubers and other underground plant storage organs), make the naked mole-rat a unique model for studying the microbiota-host interaction, focusing on the ability of the gut microbes to contribute to adaptation and health maintenance during aging.

Methods

- Sample collection and storage.

Study subjects were captured and detained from the Rift Valley ecosystem in the eastern part of Ethiopia. Briefly, the fecal samples from each animal were collected and immediately frozen in a liquid nitrogen tank and transported to Leipzig, Germany, and stored at −80 °C prior to further analysis. The study was approved and permitted by Ethiopian Wild Life and Agricultural Authorities (reference number 31/25/08 dated on 19th November, 2015). Subject collection and sampling were performed in accordance with the Ethiopian Wild Life Law guideline and regulation.

- Data analysis and bioinformatics

For information on DNA extraction sequencing and chemical analysis refer to Debebe et al.[137]. Raw sequences were processed using a pipeline combining PANDAseq[138] and QIIME[28]. Sequencing reads were deposited in the National Center for Biotechnology Information Sequence Read Archive. High-quality reads were binned into operational taxonomic units (OTUs) according the taxonomic threshold of 97% using UCLUST[139].

---

133 Miyawaki, S. *et al*. Tumour resistance in induced pluripotent stem cells derived from naked mole-rats. *Nat. Commun*. 7, 11471 (2016).

134 Seluanov, A. *et al*. Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. *Proc. Natl. Acad. Sci. USA* 106, 19352–19357 (2009)

135 Liang, S., Mele, J., Wu, Y., Buffenstein, R. & Hornsby, P. J. Resistance to experimental tumorigenesis in cells of a long-lived mammal, the naked mole-rat (Heterocephalus glaber). *Aging cell* 9, 626–635 (2010).

136 Ingram, C. M., Troendle, N. J., Gill, C. A., Braude, S. & Honeycutt, R. L. Challenging the inbreeding hypothesis in a eusocial mammal: population genetics of the naked mole-rat, Heterocephalus glaber. *Mol. Ecol*. 24, 4848–4865 (2015).

137 Debebe T, Biagi E, Soverini M, Holtze S, Hildebrandt TB, Birkemeyer C, Wyohannis D, Lemma A, Brigidi P, Savkovic V, König B, Candela M, Birkenmeier G. Unraveling the gut microbiome of the long-lived naked mole-rat. Sci Rep. 2017; 7:9590.

138 Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 13, 31 (2012).

139 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461 (2010).

Chimera filtering was performed by discarding all singleton OTUs. Taxonomy was assigned using the RDP classifier against Greengenes database[33] (May 2013 release) and relative abundances at different phylogenetic levels were calculated. Alpha rarefaction was analysed by using Chao1, PD whole tree, observed species, and Shannon index metrics in order to verify the saturation of the sequencing method. OTU assignment was performed as above and genus-level relative abundances were calculated. Bray-Curtis distances were computed based on genus-level profiles using R software (https://www.r-project.org/) and the libraries vegan and stats. Principal Components Analysis (PCoA) was performed and a 3D graphical representation was obtained by using the R package rgl. Biodiversity of samples was quantified by computing Simpson diversity index using the function "diversity" of the R package vegan and the genus-level relative abundances for each considered samples. Metagenome imputation of Greengenes-picked OTU was performed using PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States)[140] with default settings. The KEGG (Kyoto Encyclopedia of Genes and Genomes) Onthology (KO) database54 was used for functional annotation. Mann-Whitney U test was used to assess for significant differences between naked mole-rat, mouse and human imputed metagenome profiles. The p-values were corrected for multiple comparisons using the Bonferroni method. Corrected $p < 0.05$ was considered as statistically significant.

Results and discussion

In order to obtain an ecological perspective on the naked mole-rat microbiota composition, the obtained profiles at genus-level were compared to that of humans, wild mice (*Mus musculus*) and other different mammals, in a PCoA based on Bray-Curtis distances between samples (Figure 8). Naked mole-rat microbiota clustered separately from both mice and western humans, with the mixed mammals dispersed in between. The dispersion of the samples might have been influenced by the fact that most of the animals in the study of Muegge et al.[70] were kept captive in the same zoo environment, but it is still interesting to see that the naked mole-rat intestinal ecosystem emerged as a differently assembled microbiota. This could be linked to both the peculiar physiology

140 Langille, M. G. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821 (2013).

and genetics of this rodent, and to the fact that it is the first completely subterranean mammal of which the microbiota have been studied. Interestingly, the closest animal sample to the naked mole-rat cluster belonged to the capybara (*Hydrochoerus hydrochaeris*), with which the naked mole-rat shares the suborder *Hystricognathi*. This confirmed the dominant influence of the mammalian phylogeny in determining the gut microbiota structure[141].

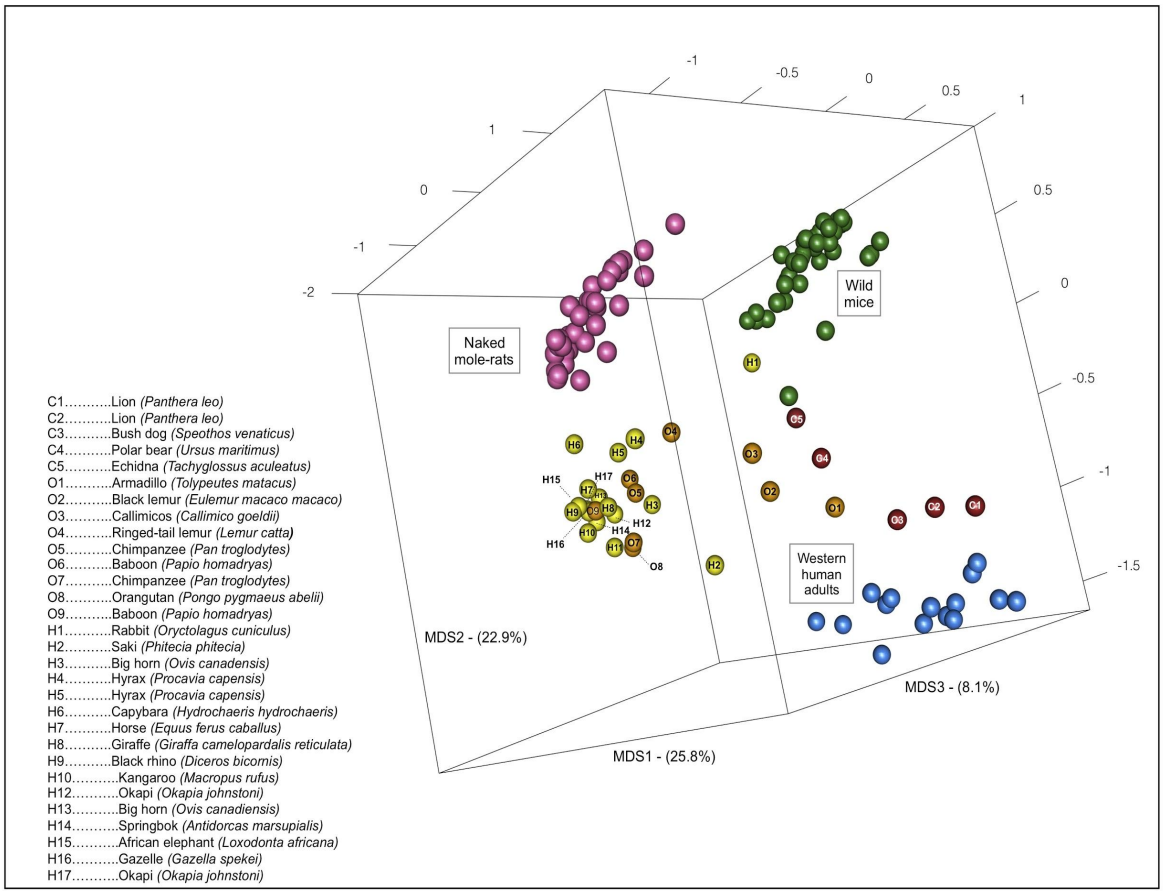


Figure 8 | 3D PCoA obtained by Bray-Curtis distance matrix showing the separation between naked mole-rats (pink), mice (green), western human adults (blue) and a group of different terrestrial mammalian species (carnivores in red (C1-C5), omnivores in orange (O1-O9), herbivores in yellow (H1-H17)) based on their gut microbial composition. Gut microbiota composition of terrestrial mammalian species was retrieved from Muegge *et al*[70]., identification of these mammals is provided in the legend (left). First, second and third principal component are showed, accounting for 25.8%, 22.9% and 8.1% of the total variance in the dataset.

Naked mole-rat microbiota profile was then compared with other gut microbial ecosystems: wild mice, western humans, Hadza subjects from Tanzania and centenarians.. While wild mice and western human adults were selected as representative of the most explored reference ecosystem for mammalians, microbiomes from supercentenarians and

141 Delsuc, F. *et al.* Convergence of gut microbiomes in myrmecophagous mammals. *Mol. Ecol.* 23, 1301–1317 (2014)

the Hadza hunter gatherers were chosen as representative of particularly successful holobiont adaptement, the first considered to support longevity[90] and the second the host homeostasis in a complex environment[88]. In the context of this family-level comparative analysis (Figure 9), with respect to wild mice and humans, the naked mole-rat microbiota showed an expanded relative contribution of families from the phylum Bacteroidetes, with a more pronounced inter-phylum diversity (6 families with rel.ab. > 0.8% vs 3 or 4 in wild mice and the three human populations). Interestingly, bacteria of the family *Bacteroidaceae*, i.e. the most abundant Bacteroidetes member both in western humans, was not represented in the naked mole-rat. On the other hand, their Bacteroidetes fraction was composed mostly by *Prevotellaceae*, *Paraprevotellaceae*, *Porphyromonadaceae* and the recently identified family *S24–7*. This peculiar configuration, with the exception of the *S24–7*, resembles the one observed in the human rural population Hadza[51] (Figure 9). Another trait that the naked mole-rat microbiota had in common with the Hadza one was the presence of *Spirochaetaceae* (10.9% and 2.8% in average, respectively), and in particular of the genus *Treponema*. This genus was represented in the naked mole-rat microbiota by a diversified population (763 OTUs, related to around 20 different *Treponema* species), with an average diversity of ten *Treponema* species per individual at > 0.01%, and three species at > 1%. Five species (*T. amylovorum*, *T. brennaborense*, *T. porcinum*, *T. succinifaciens*, *T. zuelzerae*) were found in all naked mole-rat samples at > 0.01% (never totally absent), with *T. porcinum* as the most frequently represented (>1%, average rel.ab. 5%, in all individuals). It is likely that *Treponema*, similar to the genus *Prevotella*, increases the ability of the naked mole-rat to digest and extract nutrition from fibrous naturally occurring plants, of which both the naked mole-rat and the Hadza hunter-gatherers diet are enriched, since this genus includes proficient cellulose and xylan hydrolyzers. *Treponema* is indeed considered as an "old friend" and it is assumed that this taxa has been lost from human gut flora due to industrialization and modern lifestyle.

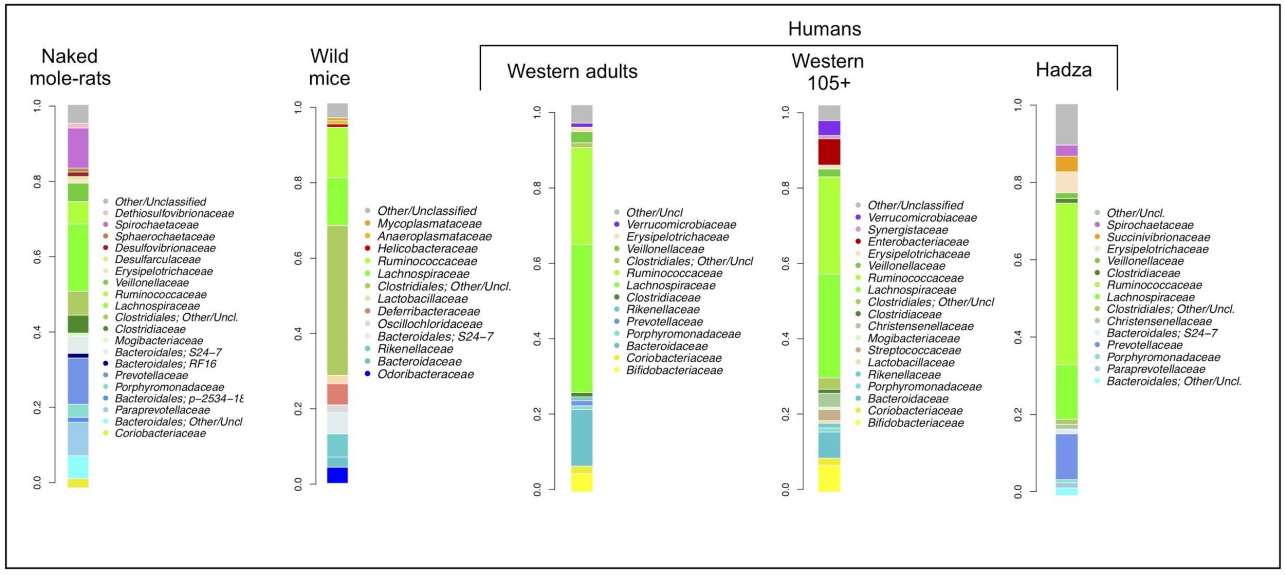


Figure 9 | Family level gut microbiota average profiles of naked mole-rats, wild mice, western human adults and supercentenarians and rural humans (Hadza). Families with average rel.ab > 0.8% are plotted. Color legends are reported for each profile to improve readability. Bacteroidetes and Firmicutes families are plotted in shades of blue and green, respectively.

Finally, the naked mole-rat microbiota showed appreciable abundance of bacterial families able to use sulfate, sulfite or other sulfur-containing molecules as terminal electron acceptor for fermentative and/or respiratory metabolism, such as *Desulfovibrionaceae* (average rel.ab. 1.2 ± 0.5%), *Desulfarculaceae* (0.9 ± 0.4%), and *Dethiosulfovibrionaceae* (1.2 ± 0.4%). In particular, bacteria of the family *Desulfarculaceae*, non-fermenting microbes that oxidize organic substrates completely to carbon dioxide[142], have never been observed in the gut ecosystem of any animal[143]. The ecological role of these bacteria in the naked mole-rat gut is difficult to foresee, however, it is interesting to point out that the subsoil of the Rift Valley, in which these animals dig their tunnel and which they rarely leave, is enriched in sulphate[144].

This peculiar configuration confers an adaptive advantage to the holobiont, providing the host with a terminal electron acceptor to support an alternative and peculiar oxidative metabolism in the gut. This could represent a new mutualistic configuration oriented to the desulphurisation of the gut environment, avoiding adverse consequences for the host .

142 Kuever, J. The family Desulfarculaceae. In "The Procaryotes". *Springer Berlin Heidelberg*, 41–44 (2014).

143 Rabus, R. *et al.* A Post-Genomic View of the Ecophysiology, Catabolism and Biotechnological Relevance of Sulphate-Reducing Prokaryotes. *Adv. Microb. Physiol.* 66, 55–321 (2015).

144 Itanna, F. Sulfur distribution in five Ethiopian Rift Valley soils under humid and semi-arid climate. *J. Arid Environ.* 62, 597–612 (2005).

Inferred metagenomics obtained by PiCRUST analysis and comparison between the KEGG pathways relative abundances in the naked mole-rat, western humans and mice, showed that the gut metagenome of the naked mole-rat was significantly enriched in pathways related to the *tryptophan metabolism* (naked mole-rat, 0.18% of the totality of KEGG pathways; wild mouse, 0.15% (Mann Whitney $P = 0.003$); western humans, 0.10% ($P = 0.001$), as well as *glycine*, *serine*, *and threonine metabolism*(naked mole-rat, 0.88%; wild mouse, 0.77% ($P = 0.03$); western humans, 0.82% ($P < 0.0001$)).

Conclusion

Eterocephalus glaber possesses a unique gut microbiome composition, which is the result of the host phylogeny and adaptation to its ecological niche. This microbiome layout has many compositional and functional peculiarities - such as the propensity for an oxidative metabolism, an enhanced capacity to produce SCFA and mono- and disaccharides, as well as the peculiar structure within Bacteroidetes, the high load and diversity of *Spirochetaceae* and the presence of *Mogibacteriaceae* - some of which are shared with gut microbial ecosystems considered as models of healthy aging, as well as metabolic and immune homeostasis. This suggest a possible role of the gut microbiota as a universal contributor to mammalian health and fitness, which goes beyond the host phylogeny, supporting health and longevity of the mammalian host. Moreover, even if confirmatory functional studies need to be carried out, findings seem to suggest a capacity of the naked mole-rat gut microbiota to utilize soil sulphate as a terminal electron acceptor to sustain an anaerobic oxidative metabolism in the gut. This could represent an unprecedented ecological equilibrium and an evidence of the importance of the gut microbiota in the adaptive process. Specific for subterranean animals, this sulfate-dependent metabolism may further highlight the importance of the gut microbial ecosystem as an adaptive partner for the mammalian biology, which exerted a strategic role in the eco-evolutionary processes.

Section 3.4 - Fecal bacterial communities from Mediterranean loggerhead sea turtles (*Caretta caretta*)

Introduction

Loggerhead sea turtle (*Caretta caretta*) is the most common and widespread sea turtle species in the Mediterranean basin, regarded as 'vulnerable' by the International Union for Conservation of Nature (IUCN) red list of threatened species[145]. The Adriatic Sea represents an important migratory route for loggerheads, even if nesting along the northern Adriatic coast is to be considered exceptional[146]. Loggerhead sea turtle is a generalist carnivorous species, feeding prevalently upon benthic animals in neritic areas, during juvenile and adult life[147].

Defined as important ecological indicators, sea turtles are considered pollution 'flagship species', whose health status draws attention to raise awareness about the conditions of the marine environment[148]. For this reason *C. caretta* represents an important model organism to search for molecular biomarkers of both organismal and ecosystem health. The bacteria harboured within the gastrointestinal tract of all vertebrates could be a source for such markers. Animal microbiomes have been linked to changes in various host's features, such as growth rate and size, metabolism, phylogeny, ecology, and evolutionary history[149]

Studies on gut microbiomes of different animals have provided a wealth of ecological and evolutionary information. However, for what concerns reptiles, the most interesting molecular studies have focused on species with peculiar feeding strategies, such as alligators, crocodiles and pythons, characterized by alternate periods of starving and active digestion[150 - 151]. As for sea turtles, the gut microbiota composition has been more

---

145 http://www.iucnredlist.org/details/3897/0

146 Lazar, B., Margaritoulis, D., and Tvrtkovic, N.(2004) Tag recoveries of the loggerhead sea turtle *Caretta caretta* in the eastern Adriatic Sea: implications for conservation. *J Mar Biol Assoc UK* 84: 475–480.

147 Bjorndal, K.A. (1997) Foraging ecology and nutrition of sea turtles. In *The Biology of Sea Turtles*. Lutz, P.L., and Musick, J.A. (eds). Boca Raton, FL: CRC Press, pp. 199–232.

148 Foti, M., Giacopello, C., Bottari, T., Fisichella, V., Rinaldo, D., and Mammina, C. (2009) Antibiotic resistance of gram negatives isolates from loggerhead sea turtles (*Caretta caretta*) in the Central Mediterranean Sea. *Mar Pollut Bull* 58: 1363–1366.

149 Colston, T.J., and Jackson, C.R. (2016) Microbiome evolution along divergent branches of the vertebrate tree of life: what is known and unknown. *Mol Ecol* 25: 3776–3800.

150 Costello, E.K., Gordon, J.I., Secor, S.M., and Knight, R. (2010) Postprandial remodeling of the gut microbiota in Burmese pythons. *ISME J*4: 1375–1385.

151 Keenan, S.W., Engel, A.S., and Elsey, R.M.(2013) The alligator gut microbiome and implications for archosaur symbioses. *Sci Rep*3: 2877.

thoroughly explored in *Chelonia mydas* (green turtle), than in other species, because of the interesting shift from omnivorous to herbivorous diet that these turtles undergo during juvenile age[152]. In spite of being the most common and studied sea turtle, *C. caretta* has been only sketchily explored for its gut microbiome, in a single preliminary study that included four feces and six intestinal mucosa samples from eight individuals[153], highlighting a gap of knowledge in the study of this relevant indicator of marine ecosystem health. Here, is provided an overview of the fecal microbiota composition in 29 loggerhead sea turtles, each one sampled twice during the stay in the Sea Turtles Rescue Centre of the 'Fondazione Cetacea' (Riccione, Italy), where medical attention and necessary therapies are provided to stranded, drifted or captured animals from the upper–west Adriatic Sea coast.

Methods

- Samples collection

Samples were taken from 29 loggerhead sea turtles (*C. caretta*) hosted at the Sea Turtles Rescue Centre of the 'Fondazione Cetacea', Riccione, Italy (43°59.133'N; 12°41.465'E). The study population included turtles found stranded or captured by fishery nets in the northern Adriatic Sea (Figure 10). Turtles were kept in the centre for cure and rehabilitation, hosted in single tanks or tanks separated by a septum, for a variable length of time, before being released.

Figure 10 | Geographic area of recovery of Mediterranean loggerhead sea turtles. Map of the Adriatic Sea; highlighted is the western Adriatic coast where turtles were found.

-DNA sequencing and bioinformatics

For details on DNA extraction from fecal samples and tank water and its processing refer to article from Biagi et al[154]. Raw sequences were

152 Ahasan, M.S., Waltzek, T.B., Huerlimann, R., and Ariel, E. (2017) Fecal bacterial communities of wild–captured and stranded green turtles (*Chelonia mydas*) on the great barrier reef. *FEMS Microbiol Ecol* 93: 12.

153 Abdelrhman, K.F., Bacci, G., Mancusi, C., Mengoni, A., Serena, F., and Ugolini, A. (2016) A first insight into the gut microbiota of the sea turtle *Caretta caretta*. *Front Microbiol* 7: 1060.

154 Biagi, E. , D'Amico, F. , Soverini, M. , Angelini, V. , Barone, M. , Turroni, S. , Rampelli, S. , Pari, S. , Brigidi, P. and Candela, M. (2019), Fecal bacterial communities from Mediterranean loggerhead sea turtles (*Caretta caretta*). Environmental Microbiology Reports, 11: 361-371.

processed using QIIME[28] pipeline. Sequencing reads were deposited in MG–Rast under project ID 84794. High–quality reads, as selected using the default values in QIIME, were binned into operational taxonomic units (OTUs) according to a 97% similarity threshold using UCLUST[104], through an open–reference strategy. Taxonomy was assigned using the RDP classifier against Greengenes database[33] (May 2013 release). Singleton OTUs were discarded to exclude chimeric sequences from downstream analysis. Alpha rarefactions were analysed by using PD whole tree, observed OTUs and Shannon index metrics. Beta diversity was estimated by computing weighted and unweighted UniFrac distances. For the descriptive analysis of the ecosystem, few interesting OTUs that were listed as unclassified after Greengenes taxonomy assignment underwent a subsequent BLAST analysis, with the 16S ribosomal RNA (Bacteria and Archaea) database as a reference in order to obtain species–level assignment. Statistical analysis was performed using R version 3.1.3 (https://www.r-project.org/) and the packages made4 and vegan. Correlations between variables were tested by using Kendall and Spearman tests. Kruskal–Wallis test was used for multiple comparisons, followed by Tukey post–hoc test when appropriate. The p values were corrected for multiple comparisons using the Benjamini–Hochberg method. Principal Coordinates Analysis (PCoA) was performed on weighted and unweighted UniFrac distances, as well as on Bray–Curtis distances calculated on genus–level relative abundances, to explore inter–sample variability, in relation to different covariates (e.g., antibiotics usage, days of hospitalization at the date of sampling, days of starving, closeness to the meal and CCL). A permutation test with pseudo F ratios (function 'adonis' in the vegan package) was used to determine the significance of separation on PCoA plots. The contribution of covariates to the ordination space was found by using the function 'envfit' of the R package vegan. The impact of different variables on the microbiota structure was also explored by the Random Forest machine learning algorithm[155], using the R packages RandomForest and rfPermute function. SourceTracker 2, a Phyton implementation of SourceTracker[156], was used to

155 Breiman, L. (2001) Random forests. *Mach Learn* 45: 5–32.

156 Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., *et al.* (2011) Bayesian community–wide culture–independent microbial source tracking. *Nat Methods* 8: 761–763.

evaluate the proportional contributions of seawater and tank water microbial communities (at OTU level) to the composition of the fecal microbiota in hospitalized sea turtles.

Results and discussion

At phylum level the fecal microbiota of C. caretta is averagely dominated by Firmicutes and Fusobacteria [average relative abundance (rel.ab.) ± SD, 46.5% ± 17.2% and 26.5% ± 18.5%, respectively] (Figure 12A). The high abundance of Fusobacteria constitutes a similarity trait with the gut microbiome of marine mammals, especially those with a fish–based diet, like dolphins[78] and seals[157], and other carnivorous reptiles such as alligators[158].

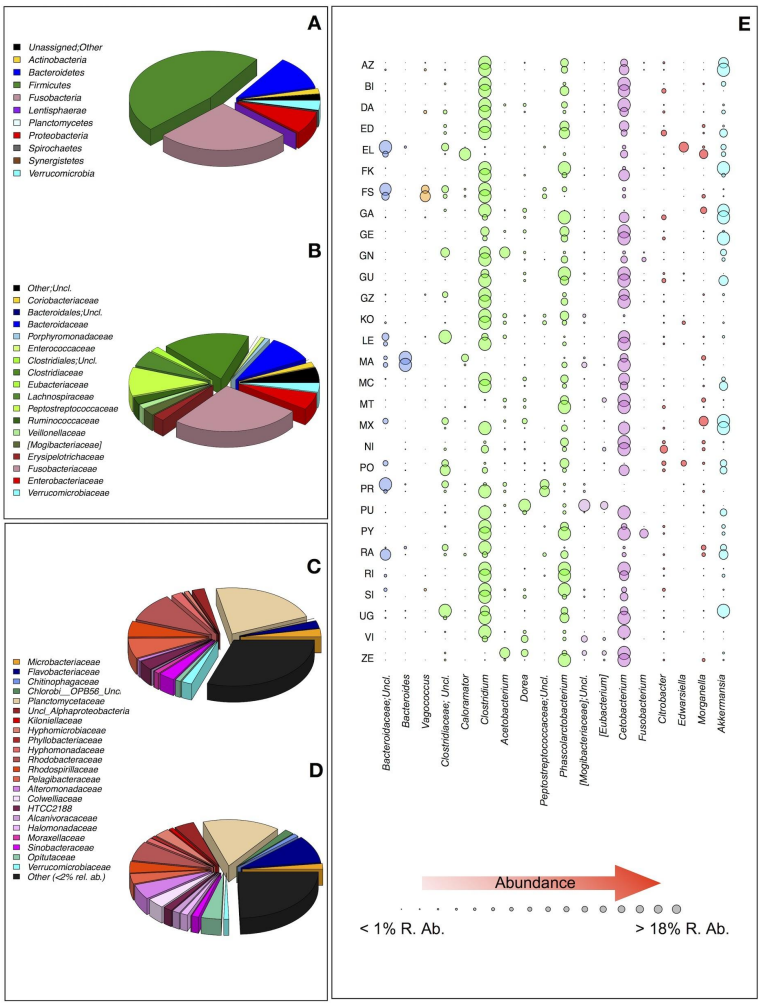


Figure 12 | Microbiota profiles of Mediterranean loggerhead sea turtles feces, sea and tank water. Average phylogenetic profiles of the gut microbiome of sea turtles provided as pie charts at phylum (**A**) and family (**B**) level; bacterial taxa were filtered for graphical representation as > 0.1% and > 0.5%, respectively. Relevant information on the gut microbiota profile at genus level is provided for each sample (**E**) using a representation in which the circle size is proportional to the genus relative abundance. Bacterial genera present at relative abundance > 10% in at least one sample are represented (horizontal) and the two samples taken from each turtle (vertical, listed using a 2–letter code) are plotted close to each other. Family–level phylogenetic profile of the microbial communities detected in seawater (**C**) and tank water (**D**) is also provided. Bacterial families contributing > 2% to the water microbial ecosystems are plotted.

157 Numberger, D., Herlemann, D.P., Jürgens, K., Dehnhardt, G., and Schulz–Vogt, H. (2016) Comparative analysis of the fecal bacterial community of five harbor seals (*Phoca vitulina*). *Microbiology* 5: 782–792.

158 Keenan, S.W., Engel, A.S., and Elsey, R.M.(2013) The alligator gut microbiome and implications for archosaur symbioses. *Sci Rep*3: 2877.

On the contrary, in the gut microbiota of herbivorous reptiles, such as lizards, iguanas, terrestrial tortoises and Galapagos giant turtles, Fusobacteria are largely subdominant or not present[159-160]. Interesting parallels in aquatic adaptation between sea turtles and marine mammals have been pointed out for other physiological functions, that is, the long and deep–diving ability and the tolerance to hypoxia[161]. In the case of the gut ecosystem composition, phenomena of convergent evolution, related to both the shared environment and the similar diet, might have driven resemblances between the gut microbiota profiles of these phylogenetically distant animals with similar adaptive needings. Within Firmicutes the most represented families are *Clostridiaceae* and *Peptostreptococcaceae* (17.8% ± 12.0% and 10.2% ± 7.5%, respectively) (Figure 12B), a trait that *C. caretta* shares with terrestrial carnivores (19% and 16%, respectively, in average in lions, cheetah and hyena[162]). Conversely, both herbivorous green turtles (*C. mydas*) and hindgut–fermenting terrestrial tortoises (*Gopherus polyphemus*) are known to host a fecal Firmicutes population distributed mainly between the families *Lachnospiraceae* and *Ruminococcaceae*[126], well–known metabolizers of complex carbohydrates of plant origin. At the genus level (Figure 12E), the fecal microbiota is prevalently dominated by the genus *Cetobacterium* (rel.ab. 25.6% ± 19.2%; the most abundant genus in 30 out of 58 samples, and 10 turtles out of 29 maintain this dominance across the two sampling times), belonging to the family *Fusobacteriaceae*. Indeed, most of the OTUs assigned to this family are classified as *Cetobacterium somerae* (88.3% of the *Fusobacteriaceae* diversity in terms of OTUs count), with a single OTU (OTU16285) constituting more than 90% of the Fusobacteriaceae diversity in 40 out of 58 samples. In all other cases, two *Fusobacterium* OTUs (assigned to the species *Fusobacterium varium* (OTU6166) and Fusobacterium perfoetens (OTU14010)] were the most abundant *Fusobacteriaceae* sequences. *C. somerae* has been reported as abundant in the gut microbiota of different freshwater fishes, both carnivorous (channel catfish, bluegill,

---

159 Hong, P.Y., Wheeler, E., Cann, I.K., and Mackie, R.I. (2011) Phylogenetic analysis of the fecal microbial community in herbivorous land and marine iguanas of the Galápagos Islands using 16S rRNA–based pyrosequencing. *ISME J*5: 1461–1470.

160 Yuan, M.L., Dean, S.H., Longo, A.V., Rothermel, B.B., Tuberville, T.D., and Zamudio, K.R. (2015) Kinship, inbreeding and fine–scale spatial structure influence gut microbiota in a hindgut–fermenting tortoise. *Mol Ecol* 24: 2521–2536.

161 Lutcavage, M.E., and Lutz, P.L. (2003) Diving physiology. In *The Biology of Sea Turtles.* Lutz, P.L., and Musick, J.A. (eds). Boca Raton, FL: CRC Press, pp. 277–296.

162 Nelson, T.M., Rogers, T.L., and Brown, M.V.(2013) The gut bacterial community of mammals from marine and terrestrial habitats. *PLoS One* 8: e83655.

largemouth bass) and herbivores (a few different kinds of carp)[163]. *Cetobacterium* was also detected in the gut microbiota of the green turtle *C. mydas*, and a hypothetical role in high–efficiency production of vitamin B12 has been suggested[164].Clostridium is the most abundant genus of the *Clostridiaceae* family, with a general contribution to the whole microbiota of 14.8% ± 10.7%; *Clostridium* is the dominant genus in 9 samples out of 58 (Figure 12E). OTUs assigned to this genus are dominated by a single one (OTU87965, accounting for 47.9% ± 26.9% of the *Clostridium* diversity on average) assigned to the species Clostridium perfringens in 26 out of 58 samples. More generally, OTUs assigned to *C. perfringens* account for 70% of the total *Clostridium* diversity on average. Other abundant *Clostridium* OTUs were identified by blast analysis and show the highest identity scores with the species *Clostridium swellfunianum* (OTU10531; accounting for 4.6% ± 10.3% of the *Clostridium* diversity in average), *Clostridium chromiireducens* (OTU26039 and OTU75627; 9.2% ± 17.0% and 1.5% ± 5.5%, respectively) and *Clostridium quinii* (OTU34970; 2.5% ± 5.6%). All these *Clostridia* species were firstly isolated from contaminated soils, mud and wastewaters[165-166]. If confirmed by further, targeted studies, the presence of bacteria likely to be adapted to polluted environments, or even able to decontaminate pollutant agents (as in the case of *C. chromiireducens*), within the gut of sea turtles might be an adaptive trait to one of the most polluted marine environments worldwide, the Adriatic Sea[167]. Indeed, since sea turtles are long–lived animals, the biomagnification of chemical and biological stressors is likely to be great, resulting in an adaptive response in the holobiont, favoring the selection of detoxifying strains. It might be possible that long–living marine holobionts, such as sea turtles, whales, dolphins or tuna fish, can act as 'scavengers' for bacterial strains able to confer resistance to pollutants, such as microplastics or heavy metals, whose load is

163 Larsen, A.M., Mohammed, H.H., and Arias, C.R. (2014) Characterization of the gut microbiota of three commercially valuable warmwater fish species. *J Appl Microbiol* 116: 1396–1404.

164 Ahasan, M.S., Waltzek, T.B., Huerlimann, R., and Ariel, E. (2018) Comparative analysis of gut bacterial communities of green turtles (*Chelonia mydas*) pre–hospitalization and post–rehabilitation by high–throughput sequencing of bacterial 16S rRNA gene. *Microbiol Res* 207: 91–99.

165 Svensson, B.H., Dubourguier, H.C., Prensier, G., and Zehnder, A.J.B. (1992) *Clostridium quinii* sp. nov., a new saccharolytic anaerobic bacterium isolated from granular sludge. *Arch Microbiol* 157: 97–103.

166 Inglett, K.S., Bae, H.S., Aldrich, H.C., Hatfield, K., and Ogram, A.V. (2011) *Clostridium chromiireducens* sp. nov., isolated from Cr(VI)–contaminated soil. *Int J Syst Evol Microbiol* 61: 2626–2631.

167 Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., *et al*. (2008) A global map of human impact on marine ecosystems. *Science* 319: 948–952.

unfortunately increasing in marine ecosystems worldwide[168], playing a role in the future marine discovery of bacteria useful for bioremediation, mitigation or adaptation to polluted environments. Principal Coordinates Analysis (PCoA) based on unweighted UniFrac distances among samples shows that the individuality of the fecal microbiota profile of captive turtles is often not maintained (Fig. 13A). By analysing the coordinates on the PCoA plot for each specimen, we quantified the distance between the two samples belonging to the same turtle; the obtained values are not significantly correlated to the number of days of hospitalization ($p > 0.05$, Kendall and Spearman correlation tests). Indeed, the length of captivity at the sampling date does not impact on gut microbiota profiles, as assessed by both PCoA analysis on unweighted UniFrac distances [$p > 0.05$, permutation multivariate ANOVA based on distance matrix (Adonis)] and correlation tests with the relative abundance of bacterial genera ($p > 0.05$, using both Kendall and Spearman methods). The number of starving days since the beginning of hospitalization before regular feeding started does not impact on the gut microbiota composition or biodiversity ($p > 0.05$, using both Kendall and Spearman methods). A time of starvation is common in sea turtles at the beginning of their stay in the Rescue Centre, and it can be related to physical injuries and/or pain, cold stunning, or captivity stress, that cause a lack of appetite. The lack of correlation between the duration of this starvation and the fecal–associated bacteria suggests a high level of resilience of the sea turtle gut microbiota. PCoA analysis, based on both weighted and unweighted UniFrac distances, and also Bray Curtis distances calculated on genus–level profiles, do not reveal any statistical separation between the two samples analyzed for each turtle, that is, far or close to the meal (Adonis for unweighted and weighted UniFrac, $p = 0.712$ and $p = 0.492$, respectively; for Bray Curtis, $p = 0.836$). Random Forest analysis confirms that closeness to the meal is not an impacting variable on the gut microbiota profile of sea turtles (error rate: 81.03%). In light of this, the other covariates possibly impacting on the sea turtle fecal community composition were explored by using a subset including only the first sample for each turtle. The second sample available of each turtle was included in a second subset on which the same statistical analysis was performed as a confirmation.

168 Deudero, S., and Alomar, C. (2015) Mediterranean marine biodiversity under threat: reviewing influence of marine litter on species. *Mar Pollut Bull* 98: 58–68.

According to PCoA and Adonis test based on both weighted and unweighted UniFrac distances, the antibiotics usage shows no impact on gut microbiota diversity in the studied sea turtles (Figure 13A). Further, the size (and consequently the age) of sea turtles was considered. The CCL is related to the age of loggerhead turtles through the von Bertalanffy function; according to Casale et al.[169], the turtles included in our study should not have reached the function asymptote, implying that they should be within the age range in which the bigger is the turtle the older it is. Here, we defined three groups of turtles: CCL1, composed of all the turtles with a CCL < 40 cm and so approximately within 0–5 years old; CCL2, composed of the turtles with CCL ranging between 41 and 60 cm with an approximate age between 6 and 14 years old; CCL3, including the biggest turtles that should be older than 15 years old with CCL > 60 cm. PCoA based on weighted and unweighted UniFrac distances reveals significant separation of the gut microbiota profiles by CCL group (Adonis: unweighted, $p = 0.005$; weighted, $p = 0.071$; Figure 13B), providing an interesting support to the relation between the gut microbiota and the size of the vertebrate host highlighted by Godon et al.[170]. Results were confirmed by repeating the same analysis on the second sample subset, that is, the farthest from the capture date (Adonis for both unweighted and weighted UniFrac, $p = 0.003$). Interestingly, the relative abundance of [*Mogibacteriaceae*] is correlated with the CCL value (Kendall tau = 0.39, $p < 0.001$; Spearman rho = 0.54, $p < 0.001$), highlighting the worthiness of exploring in future studies the functionality of this subdominant fraction of the gut microbiota, which seems to be curiously associated with long–living animals and humans[90-102].

169 Casale, P., Mazaris, A., and Freggi, D. (2011) Estimation of age at maturity of loggerhead sea turtle *Caretta caretta* in the Mediterranean using length–frequency data. *Endang Species Res* 13: 123–129.

170 Godon, J.J., Arulazhagan, P., Steyer, J.P., and Hamelin, J. (2016) Vertebrate bacterial gut diversity: size also matters. *BMC Ecol* 16: 12.

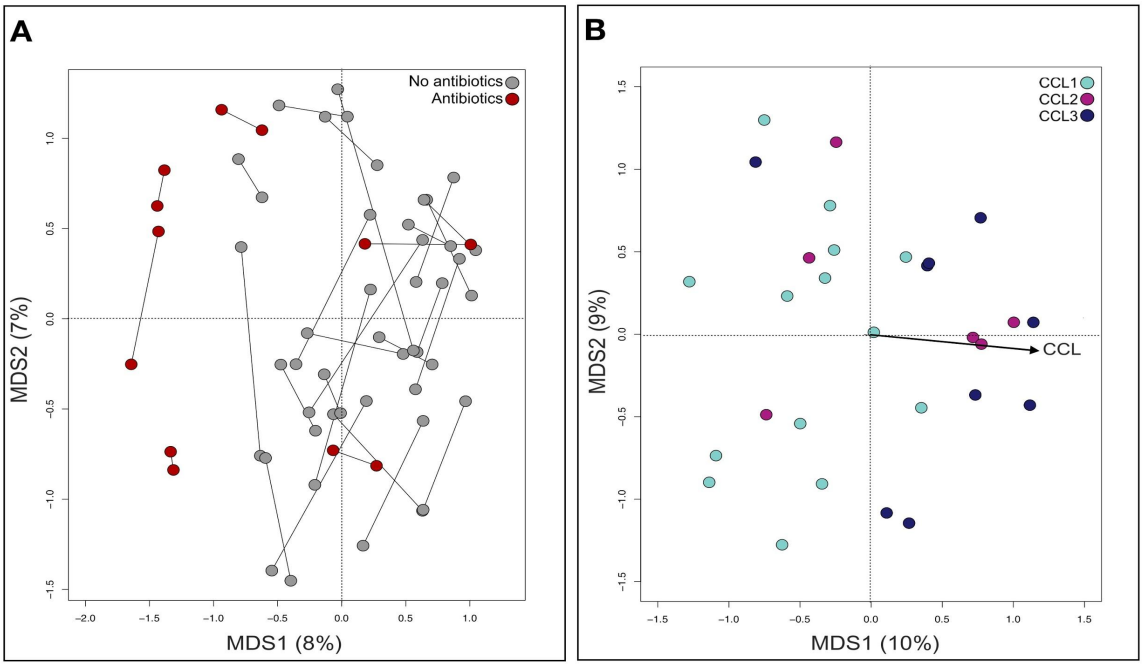


Figure 13 | Diversity of the sea turtle gut microbiota. Unweighted UniFrac distance PCoA showing (A) all samples, with the two samples belonging to the same turtle connected with a segment and samples taken from antibiotic–treated animals highlighted in dark red, (B) the first samples for each turtle by CCL group. CCL1 (CCL < 40 cm and within 0–5 years old – cyan), CCL2 (CCL comprised between 41 and 60 cm, and within 6–14 years old – magenta) and CCL3 (CCL > 60 cm and more than 15 years old – dark blue). Permutation test with pseudo *F* ratios based on distance matrix (Adonis), $p = 0.005$. Black arrow is obtained by fitting the CCL values for each sample within the ordination space (function envfit of the vegan R package).

The microbial community of the water recirculating in the sea turtle recovery tanks (two tanks were sampled) was compared with the microbial ecosystem of seawater (two samples taken off the coast) (Figure 12C and D). The tank water shows an overall lower diversity (tank water vs. seawater, average Shannon index: 5.95 vs. 6.75; average PD whole tree index: 38.7 vs. 60.4; average number of observed OTUs: 509.5 vs. 751.5). This is also highlighted by the higher percentage of the seawater microbial ecosystem composed of bacterial families with less than 2% abundance, compared with the tank water ecosystem (in black in Figure 12C and D). The tank water samples are comparatively depleted in *Planctomycetaceae*, bacteria abundant in microalgae biofilms and producers of bioactive compounds with antifungal and antibacterial activities[171], as well as in the alpha–proteobacterial families *Pelagibacteraceae*, one of the most abundant bacterial clades in the world's oceans[172], *Rhodospirillaceae* and

171 Graça, A.P., Calisto, R., and Lage, O.M. (2016) Planctomycetes as novel source of bioactive molecules. *Front Microbiol* 7: 1241.

172 Brown, M.V., Ostrowski, M., Grzymski, J.J., and Lauro, F.M. (2014) A trait-based perspective on the biogeography of common and abundant marine bacterioplankton clades. *Mar Genomics* 15: 17–28.

Rhodobacteraceae. Contamination by turtle feces is not evident in tank water, with the exception of the presence of *Peptostreptococcaceae* (mean relative abundance, 0.23% in tank water, < 0.05% in seawater), one of the most abundant bacterial families in the *C. caretta* gut ecosystem (Figure 12B). SourceTracker analysis, performed using seawater and tank water microbial communities as possible sources, highlights that only 0.02% on average (range 0%–0.3%) of the microbial OTUs found in captive sea turtles could be accounted for as deriving from the tank water microbial community (or emerging from the contamination of fecal specimens, which are collected from water). The percentage of OTUs possibly deriving from tank water does not show any correlation with the amount of time that each turtle spent in the Rescue Centre at the sampling date. Based on this observation, the gut microbiota of sea turtles emerges as a peculiarly separated 'nested ecosystem', with unexpectedly low exchange of microbial strains with the surrounding aquatic environment, a concept recently applied to sponge holobionts[173]. The ecosystem filtering performed by the sea turtle gut might thus be particularly efficient in not allowing the adhesion and/or permanence of water−originated microbes, also taking into account the long intestinal transit time shown by these animals[174].

Conclusion

Sea turtles have been around for hundreds of millions of years and represent an extraordinary model of co−evolution between holobionts. This work shows that loggerhead turtles share more gut microbiota features with marine mammals (i.e., dolphins and seals) or other carnivorous terrestrial vertebrates, than with the phylogenetically close, but herbivorous, green turtles or other terrestrial tortoises, as a demonstration of the adaptive function to diet and environment provided by the gut microbial component.

Preservation of the sea turtle population, which is being seriously endangered by human activities, is complicated by a lack of understanding of sea turtle ecology and by their long life cycle, spanning multiple habitats. Knowledge of how and to what extent the anthropogenic pressure is impacting on turtle ecology and biology, is needed for

---

173 Pita, L., Rix, L., Slaby, B.M., Franke, A., and Hentschel, U. (2018) The sponge holobiont in a changing ocean: from microbes to ecosystems. *Microbiome* 6: 46.

174 Di Bello, A., Valastro, C., Staffieri, F., and Crovace, A. (2016) Contrast radiography of the gastrointestinal tract in sea turtles. *Vet Radiol Ultrasound* 47: 351–354.

conservation programs to be effective[175]. The impact of temporary captivity on the gut microbiota profile of sea turtles is an important component of this scenario, especially in the co–evolutionary view of vertebrates, in which the bacterial counterpart is relevant in determining the health of the meta–organism.

Taken together, these findings support the reliability of the data as possibly representative of the loggerhead turtle gut microbiome. It is noteworthy that in the northern Adriatic Sea, loggerheads exhibit an early ontogenetic habitat shift, so that turtles with greater than 25 cm CCL already start to feed upon benthic animals[112]. The short–term temporary captivity state of the animals offers opportunity for detailed investigations, performed on a homogenous cohort in terms of diet, lifestyle (confined living space), and water chemical and biological features. This allowed to highlight a relation between the gut microbiota beta diversity and the size of the turtle (measured as CCL), that might be worth exploring in future studies focused on the long-life cycle of these animals, characterized by a long migratory route and the encounter of many different marine habitats.

175 Lutcavage, M.E., Plotkin, P., Witherington, B., and Lutz, P.L. (2003) Human impacts on sea turtle survival. In *The Biology of Sea Turtles*. Lutz, P.L., and Musick, J.A. (eds). Boca Raton, FL: CRC Press, pp. 387–409.

Section 3.5 - Early colonization and temporal dynamics of the gut microbial ecosystem in Standardbred foals.

Introduction

Horses strictly depend on the gut microbiota for their energetic homeostasis, representing a mammalian model of host-microbiota adaptation. Indeed, the gut microbial ecosystem exerts a very crucial role in horse nutritional biology, allowing the extraction of energy from their forage–based diet[176]. In particular, the major end–products of the gut microbiota catabolism of dietary fibre, i.e. the short–chain fatty acids acetate, propionate and butyrate, represent a key energy source for the horse, accounting for more than 50% of the total daily animal requirement[177-178]. The relevance of the gut microbial ecosystem in horse physiology is also highlighted by the deleterious impact of several gut microbiota–compromising factors on the horse health, such as antibiotics, dietary changes or gastrointestinal infections[179-180].

Despite the relevant role of the hindgut microbiota in horse health, only a few studies have focused on its characterization by next–generation sequencing (NGS)–based approaches and particularly on the early life ecosystem development in neonatal foals[142-181], but no molecular study has explored the perinatal gut microbial colonization process. In this scenario, by means of NGS of the 16S rRNA gene, is here explored the perinatal colonization process and temporal dynamics of the early gut microbiota establishment in Standardbred foals. This provide an extensive description of the microbial ecosystem of the foal gut, as well as of the mare amniotic fluid, milk and feces, highlighting the importance of the vertical transmission of microbiome components from the mare to the

---

176 Shepherd, M.L., Swecker, W.S., Jensen, R.V.and Ponder, M.A. (2012) Characterization of the fecal bacteria communities of forage–fed horses by pyrosequencing of 16S rRNA V4 gene amplicons. *FEMS Microbiol. Lett.* 326, 62–68.

177 Santos, A.S., Rodrigues, M.A., Bessa, R.J., Ferreira, L.M. and Martin–Rosset, W. (2011) Understanding the equine cecum–colon ecosystem: current knowledge and future perspectives. *Animal* 5, 48–56.

178 Brøkner, C., Austbø, D., Næsset, J.A., Blache, D., Bach Knudsen, K.E. and Tauson, A.H. (2016) Metabolic response to dietary fibre composition in horses. *Animal* 10, 1155–1163.

179 Båverud, V., Gustafsson, A., Franklin, A., Lindholm, A. and Gunnarsson, A. (1997) *Clostridium difficile* associated with acute colitis in mature horses treated with antibiotics. *Equine Vet. J.* 29, 279–284.

180 Chapman, A.M. (2001) Acute diarrhea in hospitalized horses. *Vet. Clin. N. Am.: Equine Pract.* 25, 363–380.

181 Costa, M.C., Arroyo, L.G., Allen–Vercoe, E., Stämpfli, H.R., Kim, P.T., Sturgeon, A. and Weese, J.S. (2012) Comparison of the fecal microbiota of healthy horses and horses with colitis by high throughput sequencing of the V3–V5 region of the 16S rRNA gene. *PLoS One*7, e41484.

foal. This process allows the transmission of acquired mutualistic traits among generations, maintaining adaptive advantages in the specie.

Methods

- Samples collection

Thirteen Standardbred mare–foal pairs were included in the study. Mares were hospitalised for attending parturition during foaling season. They were 4–20 years old with a parity of 1–10. Foals came from five different sires. Mares were admitted at about 310 days of gestation and received complete physical examination, blood count and biochemical exams at admission. Sample collection was planned for each pair as follows. During stage II labour, when the amniotic vesicle was clearly visible, at least 50 mL of amniotic fluid were collected by needle puncture of the amnion after swabbing with 70% ethanol, using sterile gloves and a 60−mL sterile syringe. At least 2 g of mare feces and meconium were directly harvested immediately after birth. At least 5 mL of colostrum (milk at T0) were collected immediately after parturition and before the first suckling, after udder swabbing with chlorhexidine and sterile water, and
using sterile gloves. Milk and mare and foal feces were concurrently collected at 24 h after birth (T1), and subsequently at 3, 5, 7 and 10 days of life (T3 to T10) when possible. All samples were immediately stored in sterile vials at −25°C, subsequently transferred at −80°C. A total of 164 samples was collected.

- DNA sequencing and bioinformatics

Information about DNA extraction and sequencing can be retrieved from Quercia et al.[182] and amplicon sequences were deposited in the MG−Rast database under accession 59129. Raw sequences were analysed using a QIIME[28] pipeline. Reads were filtered for quality using the split_library_fastq.py script of the QIIME pipeline with default values. High−quality reads were clustered into operational taxonomic units (OTUs) at 97% homology. In order to avoid chimeric sequences, all singleton OTUs were discarded. OTUs were taxonomically assigned using RDP classifier[183] against Greengenes database[32]

---

182 Quercia, S. , Freccero, F. , Castagnetti, C. , Soverini, M. , Turroni, S. , Biagi, E. , Rampelli, S. , Lanci, A. , Mariella, J. , Chinellato, E. , Brigidi, P. and Candela, M. (2019), Early colonization and temporal dynamics of the gut microbial ecosystem in Standardbred foals. Equine Vet J, 51: 231-237.

183 Qiong, W., Garrity, G.M., Tiedje, J.M. and Cole, J.R. (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261−5267.

(May 2013 release). Alpha diversity was computed using Chao1, Faith's phylogenetic diversity, Shannon index and observed species metrics. Beta diversity was computed using the Bray–Curtis distances, which were used as input for Principal Coordinates Analysis (PCoA) (packages vegan and rgl in R version 1.0.136). Permutation test with pseudo F−ratios was used to evaluate sample separation in the PCoA space ('adonis' function of vegan). Genera superimposition was performed using the 'envfit' function (vegan), and the reported genus vectors had significant correlation with the bidimensional space. All P values obtained from multiple comparisons were corrected using the Benjamini−Hochberg False Discovery Rate method. For the OTU sharing analysis, OTUs representing more than 0.01% of the total OTU number were considered.

Results and discussion

Considering the bacterial OTUs detected at a relative abundance greater than 0.5% in at least 33% of the samples, was obtained the core molecular structure for the microbiomes of meconium, amniotic fluid and the mare gut at delivery (Figure 14A–C). In agreement with previous studies[147-184], the core community of the mare gut showed the characteristic compositional layout of a carbohydrate-degrading and short−chain fatty acid−producing mutualistic microbiome, being enriched in Clostridiales, *Lachnospiraceae* and *Ruminococcaceae*. Conversely, the phylogenetic layout of the microbial communities from meconium and amniotic fluid shared a peculiar ecological structure, including OTUs from ubiquitous microorganisms and components from mare microbiomes at subdominant, but still relevant percentages. In particular, *Acinetobacter*, *Stenotrophomonas* and *Sanguibacter* were found to dominate the meconium ecosystem. These aerobic microorganisms are common inhabitants of soil or aquatic ecosystems, but occasionally they have been isolated as opportunistic bacteria from animal hosts[185]. On the other side, meconium also contained *Aerococcus* at relevant percentages (relative abundance 9%). Belonging to the Clostridiales order, this microorganism is a common

184 Ericsson, A.C., Johnson, P.J., Lopes, M.A., Perry, S.C. and Lanter, H.R. (2016) A microbiological map of the healthy equine gastrointestinal tract. *PLoS One* 11, e0166523.

185 Bello−Akinosho, M., Makofane, R., Adeleke, R., Thantsha, M., Pillay, M. and Chirima, G.J. (2016) Potential of polycyclic aromatic hydrocarbon−degrading bacterial isolates to contribute to soil fertility. *Biomed. Res. Int.* 2016, 5798593.

anaerobic fermenter of the mammalian gut ecosystem[186]. Finally, even if at subdominant levels, other common gut microbiota components, such as *Streptococcus*, *Enterococcus* and *Enterobacteriaceae*, were shown to be part of the core community of meconium.

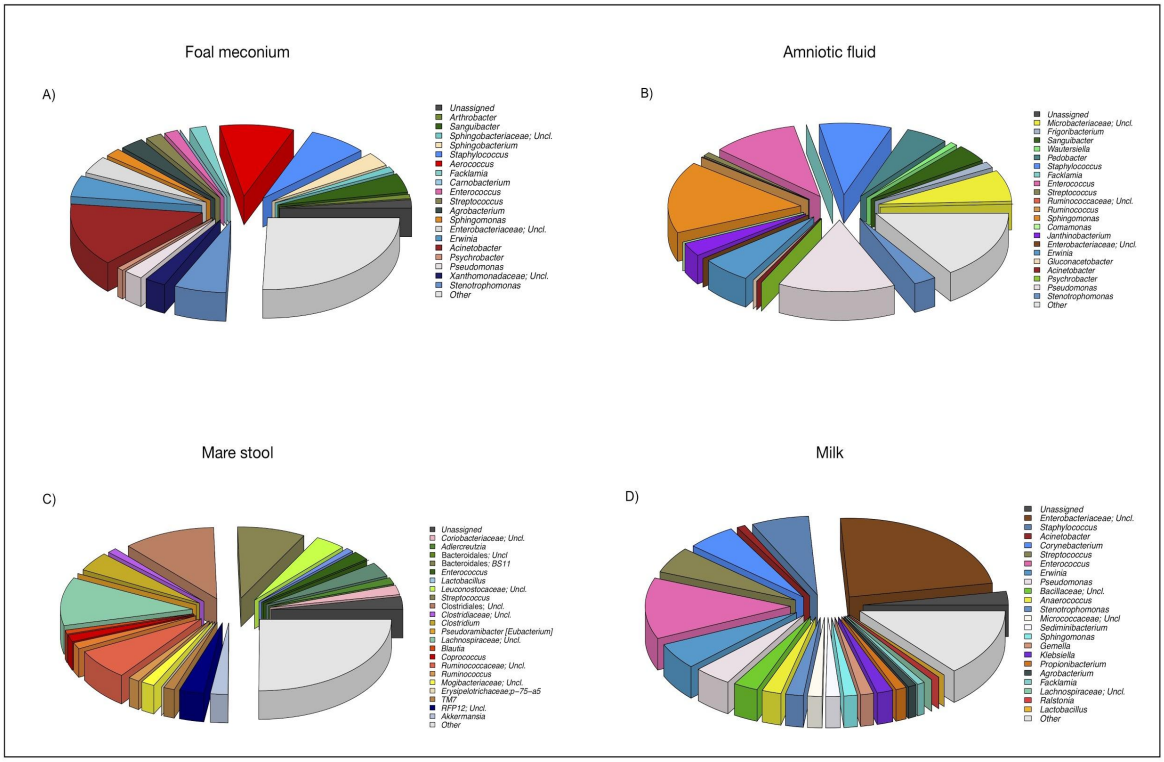


Figure 15 | Compositional structure of microbial communities of foal meconium, amniotic fluid, mare feces at delivery and milk. Pie charts representing the microbial ecosystem of meconium (**A**), amniotic fluid (**B**), mare feces at T0 (**C**) and milk (**D**). Only genera with a relative abundance ≥0.5% in at least 33% of subjects are shown; genera under the threshold were clustered in the 'Other' group.

In order to compare the α (intra−sample)−diversity of the three microbial ecosystems, the Shannon diversity curves were plotted (Figure 15). The mare gut microbiota proved to be the most diverse, with the highest OTU richness among the considered ecosystems. On the other hand, the amniotic fluid showed the lowest level of biodiversity, while meconium an intermediate level. Moreover, when the alpha diversity values associated with ecosystems were compared to each other, the difference was always significant (Wilcoxon rank−sum test, $P<0.001$).

186 Candela, M., Biagi, E., Maccaferri, S., Turroni, S. and Brigidi, P. (2012) Intestinal microbiota is a plastic factor responding to environmental changes. *Trends Microbiol.* 20, 385−391.

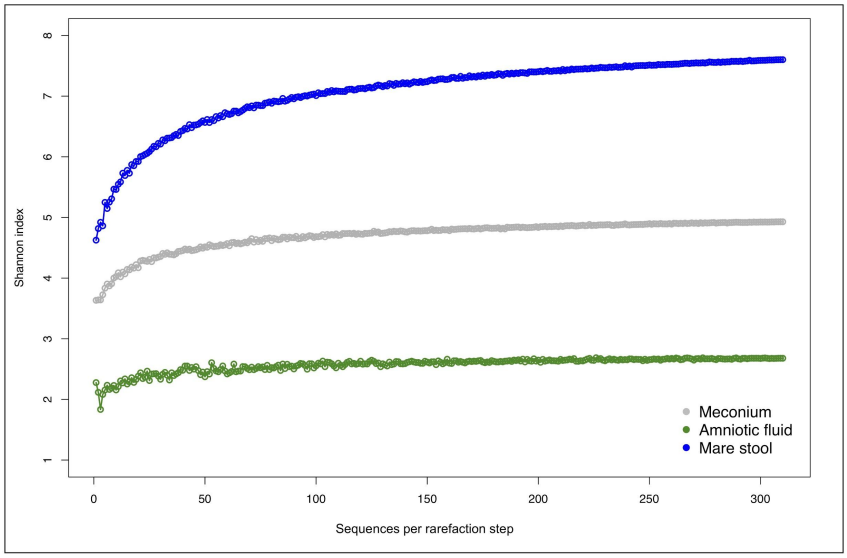


Figure 15 | Biodiversity of the mare gut microbiota, foal meconium and amniotic fluid at delivery. Alpha diversity was measured using the Shannon diversity. The mare microbial community showed the highest biodiversity index, followed by meconium and amniotic fluid (Wilcoxon rank–sum test, P<0.001).

When searching for bacterial OTUs shared between meconium and mare gut microbiota and amniotic fluid communities, I was successful in detecting 6 OTUs shared among the three microbial ecosystems, whereas 75 and 32 OTUs were shared between meconium and mare gut, or amniotic fluid, respectively (Figure 16). Surprisingly, no OTU was exclusively shared between the mare gut and the amniotic fluid. The existence of OTUs specifically shared between meconium and the two mare ecosystems suggests that each of them can provide a specific contribution to the meconium bacterial community in terms of microbial DNA components. In particular, while the mare gut microbiota seems to specifically contribute to the meconium community by providing microbial components from the gut ecosystem, the amniotic fluid could deliver a heterogeneous microbial subset, including microbial components from all mare microbiomes – such as the skin, oral and gut microbiome – and from cosmopolitan opportunistic bacteria. Even if the biological relevance of these microbiome components in the fetal developmental programming still remains to be determined, are here provided evidence in support of their presence in the fetal gut, as a result of a

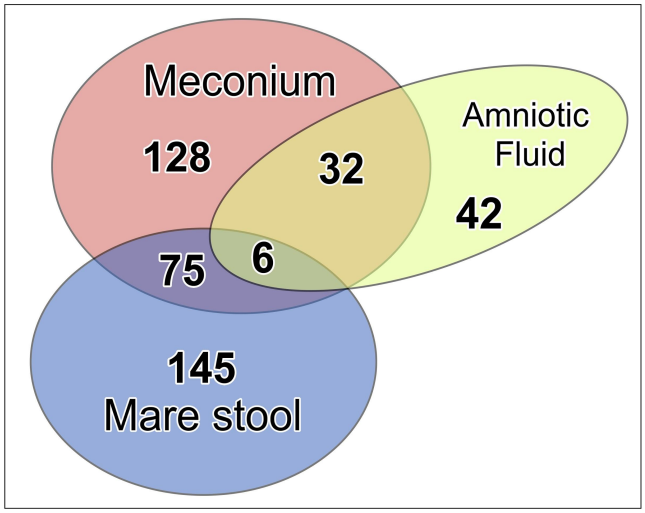


Figure 16 | OTU sharing between microbial ecosystems of mare stools, amniotic fluid and foal meconium at delivery.

scavenging process of microbial components from both amniotic fluid and the mare gut community. Data suggest the existence of possible internal transmission routes of microbial antigens, whereby mare microbiome factors are vertically transmitted to the fetus. Indeed, according to Perez et al.[187], dendritic cells from the mare penetrate the host epithelia, such as the intestinal one, sampling luminal bacteria or bacterial antigens that are then released into the placenta via the bloodstream[153-188]. Once the amniotic fluid is reached, these microbial factors may have access to the fetal gut, becoming part of the meconium ecosystem. Since the external route of mare microbiome transmission (i.e. coprophagy and suckling) is probably the sole route of transmission of live mare microbiomes to the foal[189], data suggest that the main biological function of the intrauterine transfer of microbial factors to the fetus may be the delivery of microbial antigens to the foal, priming its immune system to receive the subsequent vertical transmission of mare and environmental microbiomes at birth.

Finally, was explored the temporal dynamics of the early colonization process of the foal gut microbiota, from birth to day 10 (Figure 17). According to findings, the foal gut microbial ecosystem describes a peculiar developmental trajectory during the first days of life, progressively approaching the configuration typical of the adult gut. In particular, starting from meconium until the 3rd day of life, the foal gut microbiota undergoes gradual changes, with the progressive acquisition of microorganisms typical of the milk community, such as *Enterococcus* and *Enterobacteriaceae*. However, according to our data, this transitory state of the foal gut ecosystem rapidly changes with coprophagy. Indeed, from day 3 to 5 after birth, period corresponding to the first episode of coprophagy, the foal gut ecosystem acquires

microorganisms belonging to the core gut microbiota of the adult, such as *Prevotella*, *Blautia* and *Ruminococcus*. Being capable of providing the host with short–chain fatty acids from the degradation of dietary fibre, these microorganisms are strategic in horse biology and nutrition.

---

187 Perez, P.F., Dorè, J., Leclerc, M., Levenez, F., Benyacoub, J., Serrant, P., Segura–Roggero, I., Schiffrin, E.J. and Donnet–Hughes, A. (2007) Bacterial imprinting of the neonatal immune system: lessons from maternal cells?*Pediatrics* 119, e724–e732.

188 Jiménez, E., Marin, M.L., Martin, R., Odriozola, J.M., Olivares, M., Xaus, J., Fernandez, L. and Rodriguez, J.M. (2008) Is meconium from healthy newborns actually sterile? *Res. Microbiol*. 159, 187–193.

189 36Funkhouser, L.J. and Bordenstein, S.R.(2013) Mom knows best: the universality of maternal microbial transmission. *PLoS Biol.*11, e1001631.

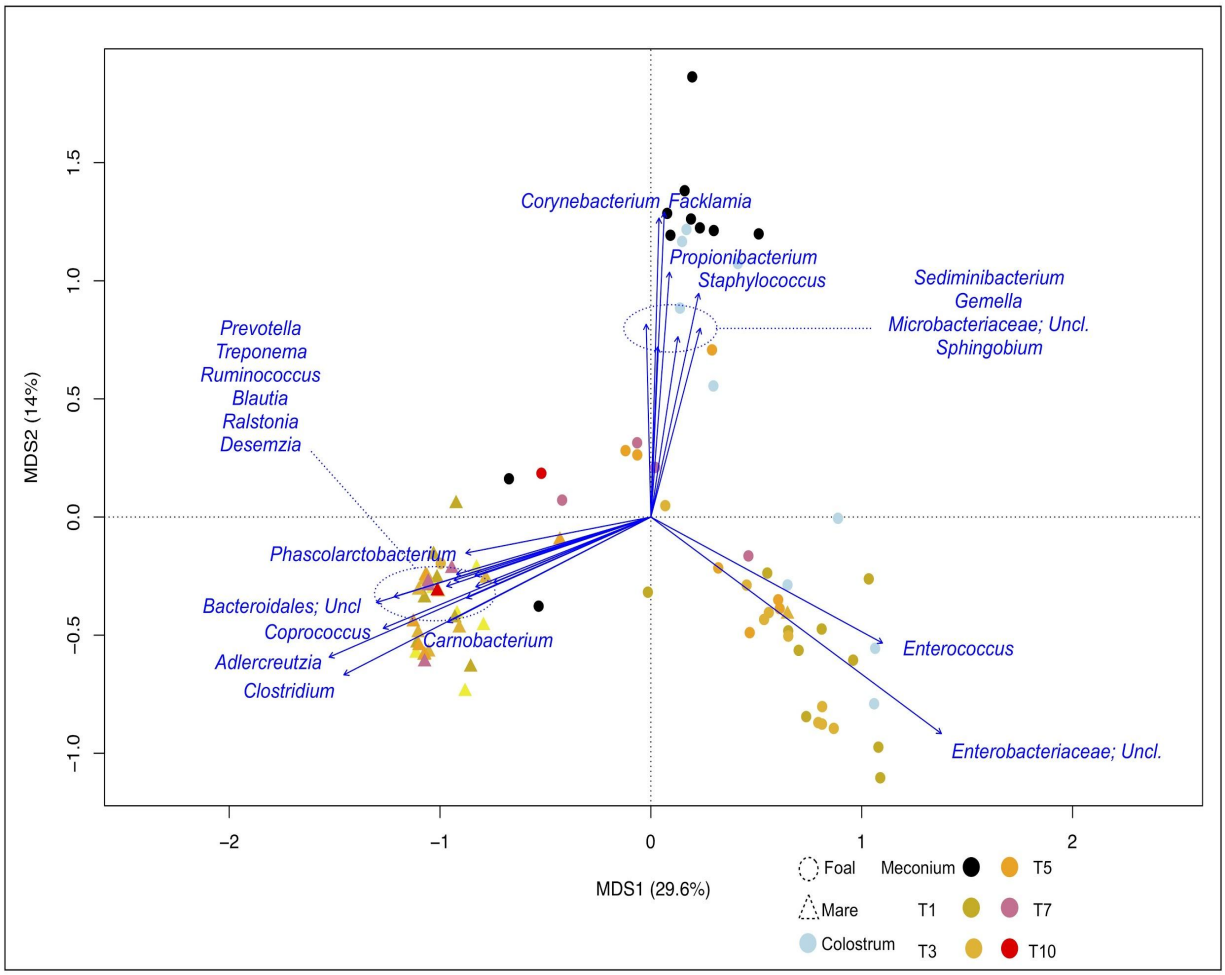


Figure 17 | Early temporal dynamics of the foal gut microbial ecosystem in relation to mare intestinal and milk microbial communities. Principal coordinates analysis (PCoA) based on the Bray–Curtis dissimilarity index of genus–level microbial communities from foal and mare feces, and milk samples taken on day 1, 3, 5, 7 and 10 after birth. Milk samples are represented as a plan of variation by an azure ellipse that comprises samples within 0.5 confidence interval. Bacterial genera showing significant correlation with the bidimensional space are represented. The arrows indicate the direction of increasing gradient, and their length is proportional to the correlation strength. MDS, multi–dimensional scale.

Costa et al.[181] already recognized the presence of a rich and rapidly changing bacterial community early in newborn foals. Although the comparison between the two studies is not straightforward – mainly due to the different timing in foal sampling – the data confirms the rapid convergence of the foal gut microbiota to the mare one early after birth, consistent with the beginning of coprophagy. As carefully reviewed by Perez–Muñoz et al.[153], a limitation of the study is that, despite efforts to avoid contamination during collection, the presence of contaminating environmental bacteria in the samples cannot be ruled out. Furthermore, the inability to discriminate between live and dead cells, a common limit of studies based on molecular approaches, does not make it possible to

determine whether meconium and amniotic fluid are actually populated by live bacteria or dead bacteria, or just by microbial debris, such as bacterial DNA

Conclusion

In conclusion, data suggest the possible presence of distinct microbial components in meconium and amniotic fluid, the former sharing microbial OTUs with both the mare gut microbiota and the amniotic fluid. This represents the first study in the equine species to provide evidence of the prenatal exposure of the fetus to microbial components (i.e. DNA) from the mare microbiomes. This finding paves the way for further studies aimed at comprehending the possible biological role of these processes in the foal developmental process, as suggested for human subjects. Furthermore, the description of the short–term temporal dynamics of the gut microbiome establishment in the foal has revealed the strategic importance of two steps in the process, milk feeding and coprophagy. The latter is particularly crucial to acquire from the mare key mutualistic gut microbiota components that, by providing the host with short–chain fatty acids, will support the nutritional and immunological health of the horse.

CHAPTER 4 - Gut bacterial community plasticity in health and disease

The adaptive potential of the gut microbiota sets its roots in the intrinsic dynamics and plasticity of the microbial ecosystem. This holobiont ability is a key factor in the eco-adaptive processes, allowing a fast reconfiguration in response to different environmental factors. In an eubiotic context, these fluctuations are of crucial importance, allowing a rapid adaptation to novel energy sources or different environmental inputs. However, this process can be also triggered by disease conditions, driving the gut microbiota to maladaptive configurations, corroborating the disease onset and perpetration.

Here I report a total of four studies in which I have investigated the plasticity potential of the bacterial community in humans, both in health and disease conditions, performing data analysis, as well as hypothesis testing.

Section 4.1 - Variation of Carbohydrate-Active Enzyme Patterns in the Gut Microbiota of Italian Healthy Subjects and Type 2 Diabetes Patients

Introduction

The human gut microbiota (GM) has been associated with various complex functions, essentials for the host health. Among these, it is certainly worth noting the degradation of the so-called microbiota-accessible carbohydrates (MACs), which the GM breaks down through specific enzymes, referred to as carbohydrate-active enzymes (CAZymes)[190]. This degradation constitutes the first step in the production of short-chain fatty acids (SCFAs), small microbial key molecules having multiple health-promoting effects for the host organism[191]. The decline in MAC dietary intake in urban Western populations forced the shrinkage of CAZyme repertoire in the GM[192], as shown by the literature comparing the microbiome layout between Western urban citizens and traditional rural populations. Even if this reduction in GM functional complexity has been

190 Gill S. R., Pop M., Deboy R. T., Eckburg P. B., Turnbaugh P. J., Samuel B. S., et al. (2006). Metagenomic analysis of the human distal gut microbiome. *Science* 312 1355–1359.

191 Koh, A., De Vadder, F., Kovatcheva-Datchary, P., and Bäckhed, F. (2016). From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell*165, 1332–1345. doi: 10.1016/j.cell.2016.05.041

192 Sonnenburg, E. D., and Sonnenburg, J. L. (2014). Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metab.* 20, 779–786.

associated with the onset of the so-called "diseases of civilization"[193], only few information regarding the CAZyme variation within Western populations has been provided to date, and its connections with diet and health are still unexplored. In this scenario, here is explored the GM-encoded CAZyme repertoire across two Italian adult cohorts, including healthy lean subjects consuming a Mediterranean diet and obese patients affected by type 2 diabetes, consuming a high-fat diet. In order to impute the CAZyme panel, a bioinformatic pipeline was specifically implemented. The study highlighted the existence of robust clusters of bacterial species sharing a common MAC degradation profile in the Italian GM, allowing the stratification of the individual GM into different steady states according to the carbohydrate degradation profile, with possible connections with diet and health.

Methods

- Determination of the pan-microbiome from Italian Healthy Subjects

The publicly available 16S rRNA sequencing data of the fecal samples of 16 Italian healthy subjects from Schnorr et al.[50] were downloaded from the MG-RAST website[194] and taxonomically characterized to the species level using the QIIME pipeline[28], with blastn[51] as an assignment method and the HMP gastrointestinal 16S rRNA dataset as reference sequences. The detected species were considered part of the so-called Italian "pan-microbiome," i.e., the virtual entity gathering the vast majority of bacterial species present in the GM of the Italian population. The assembled reference genomes of these bacterial species were downloaded from the NCBI genome section[195]. Then, to characterize the CAZyme repertoire of these microorganisms, the CAZyme identification pipeline developed by Soverini et al[196]. was applied. Briefly, ORFs were extracted from the assembled genomes using FragGeneScan 1.16[95]. From the translated ORFs, the CAZyme-coding sequences were detected using the hmmscan tool of the HMMER software package[96] and the dbCAN CAZyme database[97]. The outputs were further

193 Sonnenburg, E. D., Smits, S. A., Tikhonov, M., Higginbottom, S. K., Wingreen, N. S., and Sonnenburg, J. L. (2016). Diet-induced extinctions in the gut microbiota compound over generations. *Nature* 14, 212–215.

194 https://www.mg-rast.org

195 https://www.ncbi.nlm.nih.gov/genome

196 Soverini M, Rampelli S, Turroni S, Schnorr SL, Quercia S, Castagnetti A, Biagi E, Brigidi P, Candela M. Variations in the Post-weaning Human Gut Metagenome Profile As Result of Bifidobacterium Acquisition in the Western Microbiome. Front Microbiol. 2016 Jul 12;7:1058.

processed by a modified version of the script hmmscan-parser.sh, selecting only the ORFs that showed a minimum identity of 30% to the query sequences and an alignment length of at least 100 residues.

- Identification of CAZyme Co-abundance Groups within the Italian Pan-microbiome

The CAZyme profiles were used to generate CAZy co-abundance groups (CCGs), which were conceived as groups of bacterial species sharing a similar CAZyme profile. In brief, the CCGs were generated by applying hierarchical Ward-linkage clustering based on Spearman correlation coefficients to the abundances of glycosyl-hydrolase (GH) and auxiliary activity (AA) families detected in the bacterial genomes. Permutational multivariate analysis of variance (function "adonis" of the vegan package in R) was used to determine whether CCGs were significantly different from each other. CAZymes were also manually classified for their ability to degrade specific substrates by consulting the publicly available CAZy database[197]. Specifically, was evaluated the ability to degrade different types of MACs: resistant starch (RS), non-digestible carbohydrates (NDC), non-starch polysaccharides (NSP), and mucins/glycoproteins (M/G). When more than one activity was found, was selected the most relevant one, i.e., the one with the highest abundance of genes involved in the degradation of a given substrate.

- Assessment of Redundant Patterns of CAZymes in Italian Healthy Subjects and Type 2 Diabetes Patients

To explore CAZyme profiles in the Italian population in health and disease, we integrated the dataset used to determine the pan-microbiome with the 16S rRNA sequences of the GM from 40 patients affected by type 2 diabetes[198]. The sequences were downloaded from MG-RAST and analysed using QIIME[28] and the HMP database, as described above for healthy subjects. The CAZyme profile of each GM was obtained by quantifying the relative abundance of each CCG, as a sum of the relative contribution of component bacterial species. We then grouped the subjects using hierarchical Ward-linkage clustering based on Spearman correlation coefficients. Separation between clusters was

197 http://www.cazy.org

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

tested using the permutational multivariate analysis of variance. All statistical analyses were computed in R version 3.1.3 using R studio version 1.0.36.

Results and discussion

The pan-microbiome of the studied population included a total of 98 bacterial species and, in healthy subjects, it was dominated by *F. prausnitzii, E. rectale, R. bromii and B. adolescentis*, which also emerged as the most prevalent species in the GM from Italian healthy adults. When compared with previously characterized GM pan-genomes, such as that from the Chinese population[199] , the Italian one showed some peculiarities, i.e., the presence of *E. rectale* and *Bifidobacterium* and the absence of *Phascolarctobacterium* within the core community. Although both studies are based on relatively small cohorts, and a more extensive screening at the population level is needed, these data seem to suggest a certain level of country specificity in the gut microbiome structure, which may contribute to the immunological and metabolic peculiarities of the populations.

According to findings, the bacterial species belonging to the Italian pan-microbiome showed two different types of MAC-degrading profiles, essentially characterized by a high or low content of glycosyl-hydrolase-coding sequences, respectively. As expected, the CAZyme distribution in the various species of the Italian GM was heterogeneous, and the absolute number of CAZymes was independent from the genome size. However, it should be mentioned that, for each identified species, the analysis of the CAZyme content was performed by using the type strain reference genome deposited in the NCBI database and therefore our classification was blind with respect to the possible strain-level functional variability in the CAZyme profile.

199 Zhang J., Guo Z., Xue Z., Sun Z., Zhang M., Wang L., et al. (2015). A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography, and ethnicities. *ISME J.* 91979–1990.

The GM species of the Italian pan-microbiome were successfully clustered into four CCGs according to the similarity of the CAZyme pattern: *S. variabile* (CCG1), *E. rectale* (CCG2), *R. bromii [B. Obeum]* (CCG3), and *F. prausnitzii* (CCG4) (Figure 18). Interestingly, each of the identified CCGs was characterized by a peculiar structure in terms of CAZyme content (Figure 19).

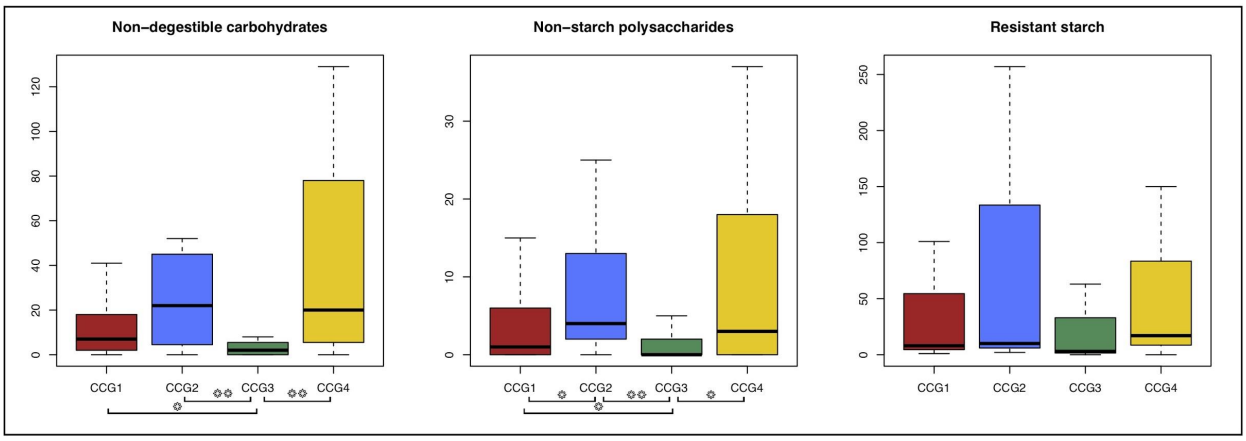


FIGURE 19 | Boxplots representing the distribution of the degradative potential of exogenous MACs in the different CAZyme Co-abundance Groups (CCGs). Square brackets at the bottom indicate a significant difference in raw abundances (single asterisk, *p*-values between 0.05 and 0.001; double asterisk, *p*-values below 0.001, Wilcoxon rank sum test).

In particular, *F. prausnitzii* and *B. obeum group*s were the most enriched CCGs in terms of represented CAZyme functions, whereas *F. prausnitzii* and *E. rectale* groups were the most equipped CCGs in terms of CAZymes specifically involved in the breakdown of non-digestible carbohydrates and non-starch polysaccharides (i.e., xylans, pectins, and mannans). These observations suggest that the Italian pan-microbiome is diversified in at least four patterns of carbohydrate degradation, raising several open questions related to: (I) the major determinants of the co-evolutionary processes underlying this differentiation; (II) the relative contribution of host genetics, lifestyle and diet as drivers of this functional convergence; (III) the ultimate connections between the observed CCGs and the host metabolic phenotype. When exploring the quantitative distribution of CCGs in the individual microbiota from all subjects analysed, we observed four robust clusters of subjects sharing a similar CCG profile, termed from CT1 to CT4. In particular, the CT1 and CT3 clusters included CCG4 (*F. prausnitzii* group) as the most prevalent CCG, being present in all individuals, and CCG3 (*B. obeum* group) and CCG1 (*S. variabile* group) as less prevalent, ancillary, and generally mutually exclusive groups. Conversely, CT4 was dominated by both CCG4 (*F. prausnitzii* group) and CCG2 (*E. rectale* group), which

equally shared the ecosystem. Finally, CT2 showed CCG2 (*E. rectale* group) as the most prevalent group and CCG4 (*F. prausnitzii* group) as ancillary and less prevalent group, except for three subjects that were dominated by CCG1 (*S. variabile* group). These

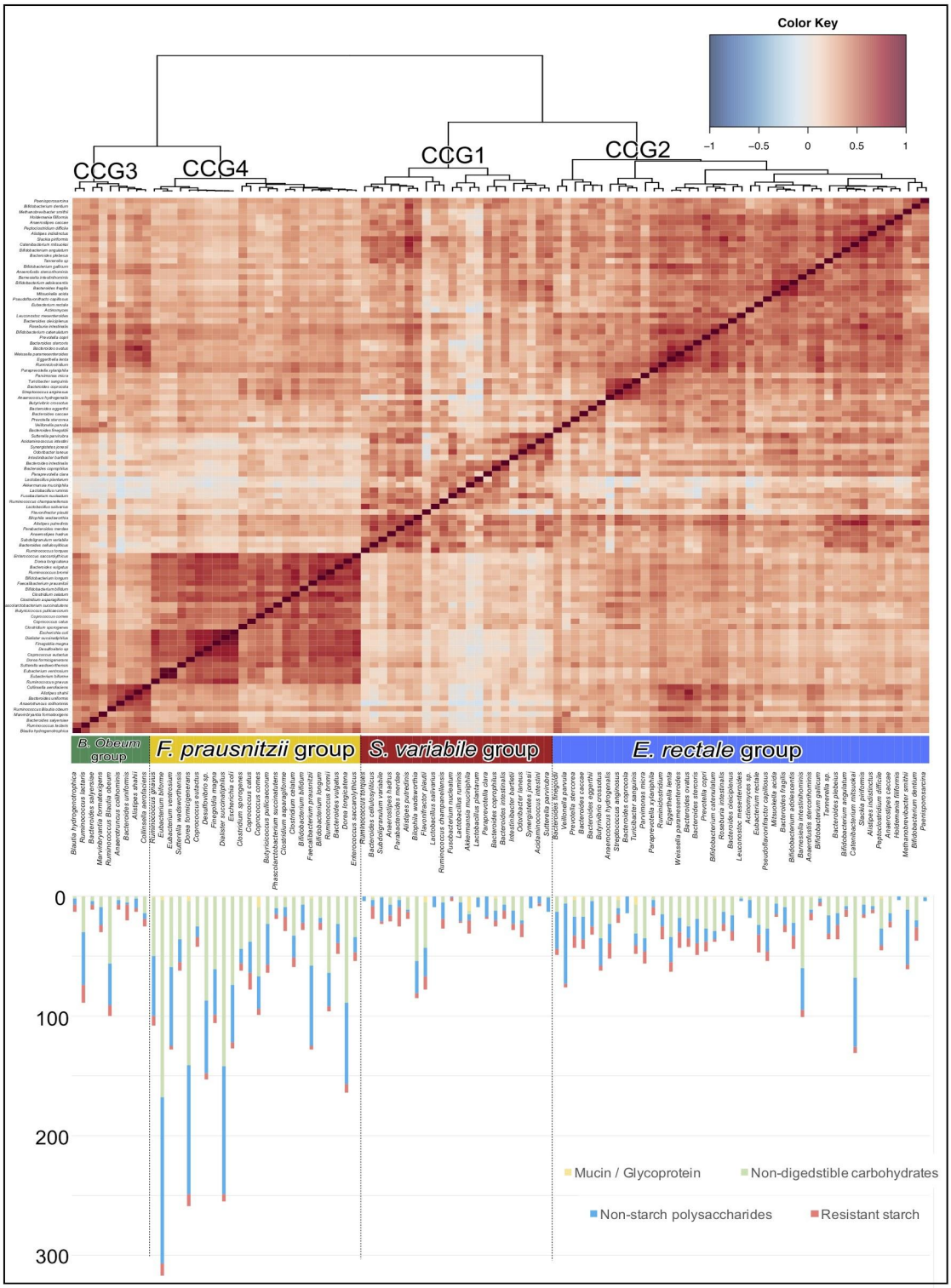
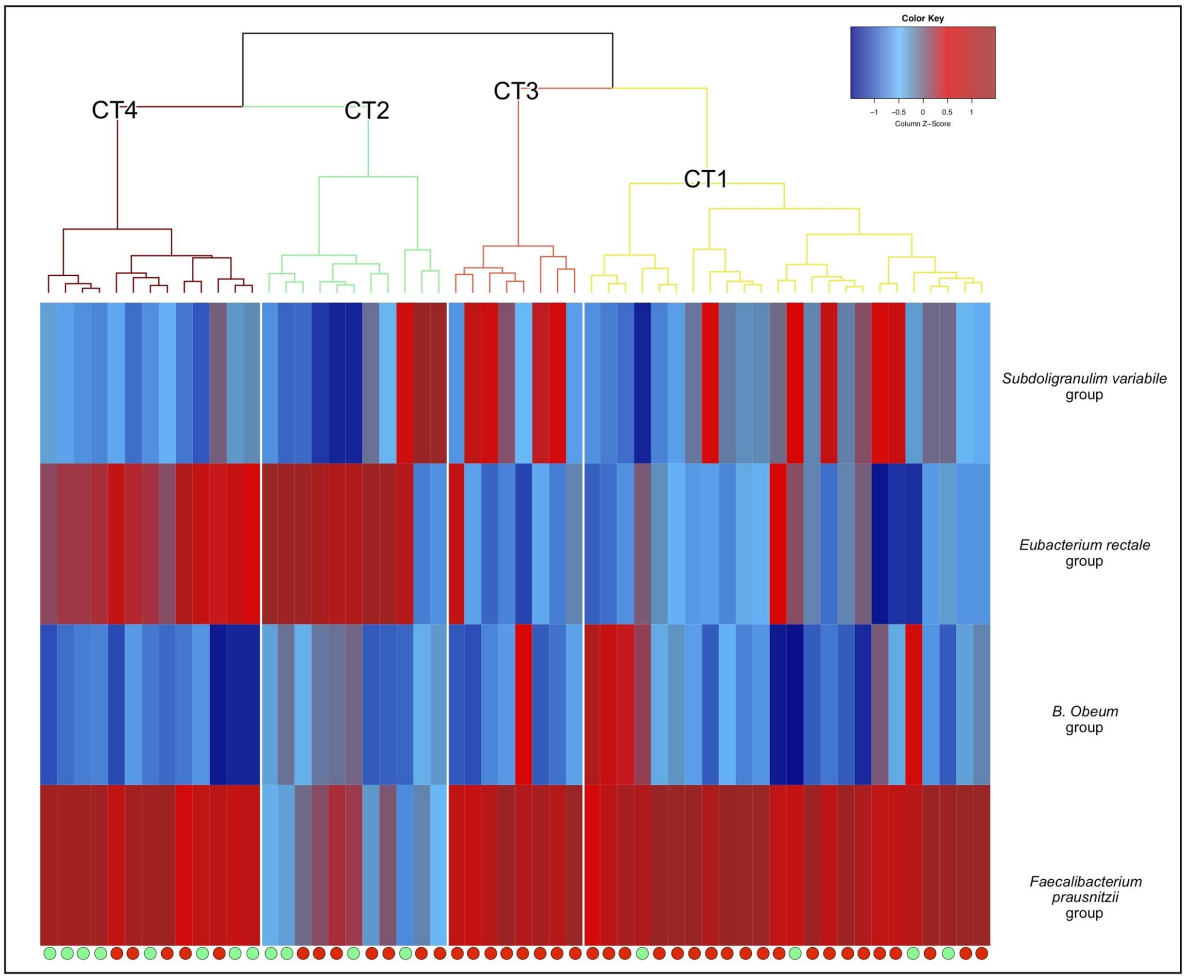


FIGURE 18 | Hierarchical clustering of the raw abundances of the CAZy glycosyl-hydrolase (GH) and auxiliary activity (AA) families in the bacterial species constituting the Italian pan-microbiome. Spearman distance and Ward's minimum variance method were used. Four CAZy Co-abundance Groups (CCGs) were identified and named according to the most abundant species in each group, as follows: CCG1 – *Subdoligranulum variabile* group (red), CCG2 – *Eubacterium rectale* group (blue), CCG3 – *Blautia obeum* group (green), and CCG4 – *Fecalibacterium prausnitzii* group (yellow). Below the heatplot are reported the counts of CAZyme-coding sequences belonging to the GH and AA families, detected in the type strain reference genomes for each bacterial species, divided by class of MACs: resistant starch (RS), non-digestible carbohydrates (NDC), non-starch polysaccharides (NSP), and mucins/glycoproteins (M/G).

observations are indicative of a different ecological behavior for the diverse CCGs. Indeed, while CCG4 (*F. prausnitzii* group) appears to co-exist with all other CCGs, CCG2 (*E. rectale* group) and the CCGs 1 (*S. variabile* group)/3 (*B. obeum* group) are mutually exclusive. Confirming this, none of the CTs showed the simultaneous presence of CCG2 and CCG1 and/or CCG3. Taken together, these data suggest that the GM-host co-evolution process has resulted in the establishment of four well-defined functional steady states, i.e., the four CTs, each determined by the CCG propensity to share the same gut environment, and each conferring to the host a specific pattern of CAZymes. In order to explore possible associations of these CTs with the host diet and health, was explored their variation in Italian healthy adults consuming a Mediterranean diet and obese T2D patients consuming a high-fat low-MACs diet (Figure 20). Interestingly, according to our data, most healthy subjects belonged to the CTs 2 and 4, which were characterized by the simultaneous presence of CCG4 (*F. prausnitzii* group) and CCG2 (*E. rectale* group). Conversely, the great majority of obese T2D patients belonged to CT1 and CT3, where CCG2 (*E. rectale* group) was substituted by CCG3 (*B. obeum group*) and/or CCG1 (*S. variabile* group). Although caution is needed in interpreting results, the analysis presented here suggests that a high-fat low-MACs diet, in the context of metabolic deregulation, such as obesity and T2D, could force changes in the GM CTs, supporting the presence of CCG1 (*S. variabile* group) and/or CCG3 (*B. obeum* group) to the detriment of CCG2 (*E. rectale* group). Interestingly, compared to CCG1 and CCG3, the CCG2 showed higher levels of enzymes involved in the degradation of non-digestible carbohydrates and non-starch polysaccharides, which are indeed abundant MACs in the Mediterranean dietary regimen. Though preliminary, the data highlight a possible adaptive or maladaptive nature for each of the four CT steady states that describe the Italian pan-microbiome. Indeed, the steady states CTs 2 and 4, that were generally found within healthy hosts, seem to be the result of an adaptive microbiome-host co-evolution process, in which the interplay between diet, gut microorganisms and the host can contribute to overall metabolic health. On the other hand, the CTs 1 and 3 that were associated with T2D and a high-fat low-MACs diet (permutation test with pseudo-F ratios $p<0.001$), may result from a maladaptive microbiome–host process, in which this type of diet has led to the



selection of CT steady states able to contribute to metabolic and/or immunological deregulation.

FIGURE 20 | Hierarchical clustering of the relative abundances of each CAZyme Co-abundance Group (CCG) in the gut microbiota (GM) of every subject. Bray–Curtis distance and Ward's minimum variance method were used. On the top: the four CAZyTypes (from CT1 to CT4) identified, i.e., clusters of different GM configurations with a similar carbohydrate-degrading profile. On the right: the four CCGs, named according to the most abundant species in each group (CCG1 – *Subdoligranulum variabile* group, CCG2 – *Eubacterium rectale*group, CCG3 – *Blautia obeum* group, and CCG4 – *Fecalibacterium prausnitzii* group). At the bottom: green dot, healthy lean subject; red dot, obese type 2 diabetic patient.

Conclusion

These findings highlighted the existence of specific and well-defined GM functional layouts (CAZyTypes, CTs) for what concerns the ecosystem capacity to metabolize MACs, and support the hypothesis that the human GM has the ability to reconfigure its own CAZyme functional layout in response to dietary changes, with possible implications for the host health and metabolic regulation.

Section 4.2 - Infant and Adult Gut Microbiome and Metabolome in Rural Bassa and Urban Settlers from Nigeria

Introduction

In recent years, science has witnessed a growing number of studies on the characterization of the human gut microbiome across the globe, in populations adhering to varying subsistence patterns, from more traditional to more urbanized[50-51-85-200]. In addition to providing valuable information on the specific adaptations of the gut microbiota to diet and other lifestyle factors, such studies have evolutionary relevance, as they recall ways of life that accompanied our history, from the hunting and gathering of our Paleolithic ancestors, to small-scale agriculture and permanent settlements of the Neolithic, to the post-industrial Westernized lifestyle. This body of literature has consistently illustrated distinctive signatures of the urbanization process in intestinal microbial communities, including reduced diversity, loss of bacterial taxa with carbohydrate-degradation specializations, and the appearance of microorganisms as a potential adaptive response to the changes in diet, environment, use of antibiotics, and hygiene practices, brought on by the modern lifestyle. However, most of the present studies have focused on the taxonomic variation of the gut microbiota in adult populations living distinct lifestyles, thus leaving a number of unanswered questions, especially about potential subsistence-driven alterations in metabolic networks and the broad-scale application of this body of data toward informing about the infant gut microbiome alterations. Moreover, most studies comparing microbiomes from hunter-gatherers, rural agriculturalists, and urbanized communities have so far dealt with geographically and culturally distant populations, with obvious confounding factors, such as relatedness and local environment. In an attempt to bridge these gaps, it is here characterized the fecal microbiota and metabolome of two Nigerian communities, the Bassa rural agriculturalists and urban individuals from four state capitals (Ilorin, Abeokuta, Ado Ekiti, and Ibadan) and the Nigerian capital city (Abuja), which also include infants aged <3 years. The Bassa are an agrarian community with limited contact with other

200 C. De Filippo, D. Cavalieri, M. Di Paola, M. Ramazzotti, J.B.Poullet, S. Massart, S. Collini, G. Pieraccini, P. Lionetti Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa Proc. Natl. Acad. Sci. USA, 107 (2010), pp. 14691-14696

populations, who live on a hill about 500 m away from the Chibiri village in Kuje Area Council (Abuja), where they moved from Kogi State (a distance of 160 km) about 100 years ago (Figure 21). Their community comprises about 70–80 people who primarily eat what they grow on their farm, such as tubers, grains, fruit, and other small crops. The Bassa can be considered an isolated group, but nonetheless maintain a self-sufficient rural horticultural subsistence. Microbial communities in the Usuma River, which is a daily feature in Bassa life, both for nourishment and physical exposure, were characterized as well. The urban dwellers recruited in our study were randomly selected from different ethnic groups, including *Hausas, Igbos, Yorubas* and *Ebira*, as representative of people geographically close to the Bassa but who are embracing a Western lifestyle. Compositional microbiome data and metabolome profiles from these populations were interpreted across subsistence strategies and age, and integrated with available data from worldwide populations, with varying degrees of traditional or urban lifeways. By exploring, at a finer geographic and age resolution than previous efforts, the variation of the human gut ecosystem along the transition from rural to urbanized communities, this study led to uncovering specific adaptive gradients, at both structural and functional scale.

Methods

- Subject Enrollment and Sample Collection

Figure 21 | Location of Bassa tribe.

Eighteen Bassa (nine adults and nine infants) participated in this study. The infants were younger than 3 years old, while the adults were of indeterminable age but presumably younger than 60 years. Sex information was not available. For urban volunteers, 12 infants (<3 years) and 18 adults (5–75 years) were recruited from state capitals of four states in South Western Nigeria (Ilorin, Kwara State; Abeokuta, Ogun State; Ado Ekiti, Ekiti State; Ibadan, Oyo State) and Nigeria capital city (Abuja, Northern Nigeria). Fecal samples were collected in mid-2015 (July–September) upon consent from the adults and assent from the youth

with consent granted by parents or guardians. Samples were processed in dry form with 97% ethanol, and then transported to Bologna (Italy) for analysis. The study was approved by the Institute of Advanced Medical Research and Training (IAMRAT), College of Medicine, University of Ibadan, Ibadan, Nigeria, with ethical approval number UI/EC/15/0050.

- DNA processing, metabolomics, bioinformatics and statistics

For DNA extraction procedures and metabolomic analyses refer to Ayeni et al.[201]. Amplicons generated in the context of this study are deposed in MG-RAST under accession number MGP83994.

Raw sequences were processed using a pipeline combining PANDAseq[103] and QIIME[28]. High-quality reads were binned into OTUs at 97% similarity using UCLUST[104]. Taxonomy was assigned using the RDP classifier against Greengenes database (May 2013 release). All singleton OTUs were discarded. Alpha diversity was computed after rarefaction to 8,480 sequences per sample (minimum sampling depth) using observed OTUs, Shannon and Faith's phylogenetic diversity (PD) indices. Beta diversity was estimated by computing weighted and unweighted UniFrac (16S rRNA data), Euclidean (metabolome), and Bray-Curtis (16S rRNA data and genus tables from worldwide populations) distances. Bacterial and metabolic CAGs were determined as previously described[202]; Wiggum plots were created using Cytoscape 3.2.1. For bacterial CAGs, genera with ≥0.1% relative abundance in at least 30% of subjects were considered. For metabolic CAGs, metabolites with ≥0.1% relative abundance in at least two subjects were included. Discriminatory metabolites between study populations were identified using Random Forests[121]. SourceTracker[122] was used to estimate the proportional contributions of traditional or urban sources to the microbiota of Bassa and urban Nigerians. All statistical analysis was performed in R 3.3.2 using R studio 1.0.136. Principal coordinate analysis (PCoA), PCA, Procrustes, adonis (permutation test with pseudo-F ratios), and ANOSIM tests were performed using the vegan package; the Random Forests analysis was carried out using the library package randomForest, SourceTracker using the corresponding package

201 Ayeni FA, Biagi E, Rampelli S, Fiori J, Soverini M, Audu HJ, Cristino S, Caporali L, Schnorr SL, Carelli V, Brigidi P, Candela M, Turroni S. Infant and Adult Gut Microbiome and Metabolome in Rural Bassa and Urban Settlers from Nigeria. Cell Rep. 2018 Jun 5;23(10):3056-3067

202 M.J. Claesson, I.B. Jeffery, S. Conde, S.E. Power, E.M.O'Connor, S. Cusack, H.M. Harris, M. Coakley, B.Lakshminarayanan, O. O'Sullivan, *et al.* Gut microbiota composition correlates with diet and health in the elderly Nature, 488 (2012), pp. 178-184

SourceTracker, and correlation (Spearman and Kendall tau) tests and non-parametric tests (Wilcoxon rank-sum test or Kruskal-Wallis test) were achieved using the stats package. p values were corrected for multiple comparisons using the Benjamini-Hochberg method when appropriate. A corrected p value <0.05 was considered statistically significant.

Results and discussion

This study demonstrates that two human communities living in a geographically proximate region in Nigeria follow a predictive pattern of dissimilarity in taxonomic and metabolic traits of the gut microbiome that mirror the traditional and/or rural versus urban and/or industrialized subsistence dichotomy. Importantly, these results allowed to witness specific traits that indicate a progressive adaption of the intestinal microbial ecosystem toward urbanization.

Consistent with prior findings, the data point to a reduced inter-individual variation in the microbiota of people adhering to a traditional lifestyle, with the well-known dominance of bacteria with high potential for fiber degradation (primarily *Prevotella*, *Treponema*, and *Succinivibrio*, but also *Ruminobacter*, *Phascolarctobacterium*, and *Butyrivibrio*), and the underrepresentation or even absence of common members of urban-industrial gut microbiomes (e.g., *Bacteroides*, *Bifidobacterium*, and a series of known SCFA producers, including *Blautia* and *Fecalibacterium*) (Figure 22).

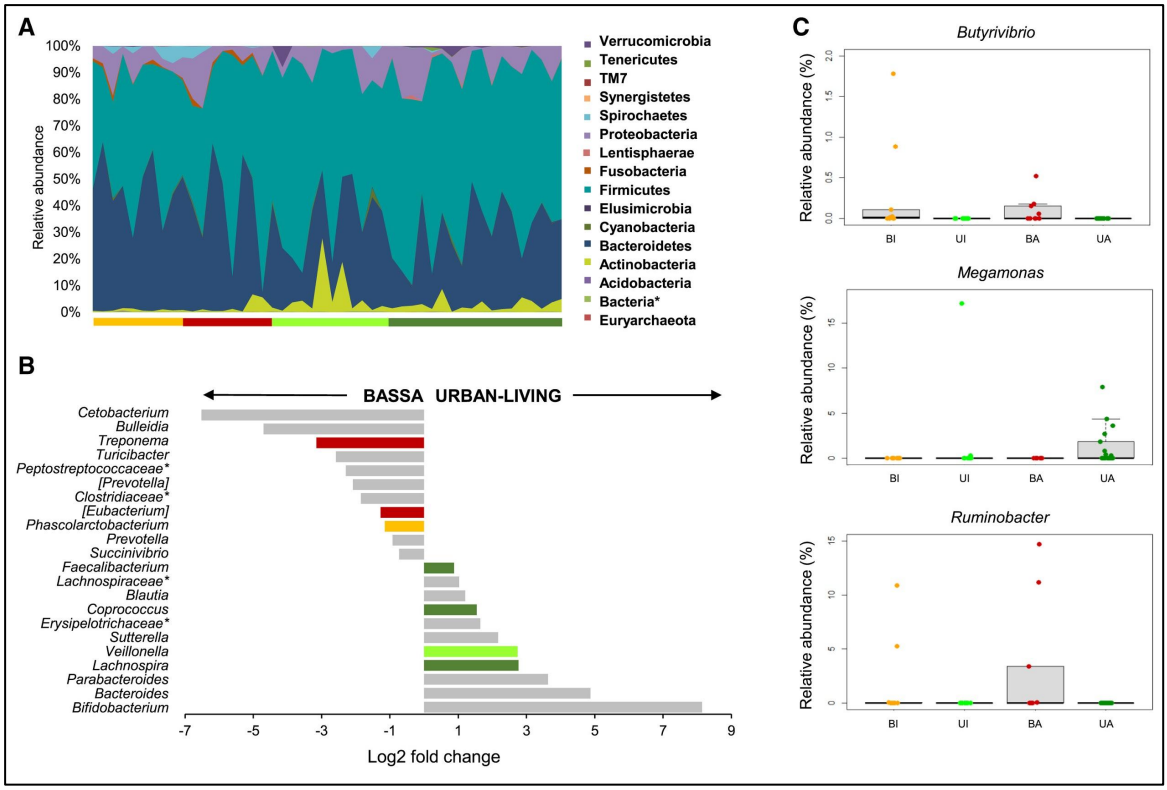


Figure 22 | Gut Microbiome Profile of Bassa and Urban Nigerians. (**A**) Relative abundances of phylum-level taxa. Bars below the area chart are colored by tribe and age (orange, Bassa infants; red, Bassa adults; green, urban infants; olive green, urban adults). (**B**) Log2 fold changes of the main discriminant genera between Bassa and urban Nigerians (**C**) Boxplots showing the relative abundance distribution of genera that were uniquely detected in the gut microbiota of Bassa (*Butyrivibrio* and *Ruminobacter*) or urban individuals (*Megamonas*).

The study also led to the identification of bacteria worthy of further investigation for their possible association with the lifestyle patterns of the study populations, i.e., *Cetobacterium* and *Bulleidia* for rural gut communities, and *Megamonas* and *Oscillospira* for urban microbiotas. *Cetobacterium* is a Fusobacteria genus indigenous to the digestive tract of freshwater fish[203], including *Tilapia*, which dominates the lower Usuma River reservoir as well as other West African water bodies[204]. The abundance of *Cetobacterium* in the gut microbial ecosystem of the Bassa individuals sampled in this study may be related to their regular consumption of fish and the close relationship they maintain with the Usuma River. Less information is available for *Bulleidia*, frequently associated with the human oral microbiome but recently identified as exclusive to the intestinal microbiota of Bangladeshi children living in an urban slum compared with upper-middle class suburban children from the United States[205]. On the other hand, the *Megamonas* species known so

203 C. Tsuchiya, T. Sakata, H. Sugita Novel ecological niche of *Cetobacterium somerae*, an anaerobic bacterium in the intestinal tracts of freshwater fish Lett. Appl. Microbiol., 46 (2008), pp. 43-48

204 A.S. Dan-kishiya A survey of the fishes of lower Usuma reservoir, Bwari, F.C.T. Abuja, Nigeria Rep. Opinion, 4 (2012), pp. 48-51

205 A. Lin, E.M. Bik, E.K. Costello, L. Dethlefsen, R. Haque, D.A.Relman, U. Singh Distinct distal gut microbiome diversity and composition in healthy children from Bangladesh and the United States PLoS ONE, 8 (2013), p. e53838

far are listed among Firmicutes members with more limited carbohydrate utilization capabilities, and to date they have been identified in urban contexts[206] and found to be differentially abundant according to ethnicity[207]. Similarly, *Oscillospira* is shown to increase in abundance with the switch to an animal-based diet, under high-bile conditions[208], thus likely reliant on fermentation products generated by other microbes or on host mucus glycans rather than primary fiber degradation. The high abundance of *Oscillospira* and the exclusive presence of *Megamonas* in the gut microbiota of urban Nigerians may be potential markers of the progressive urbanization and adoption of a Western lifestyle. Metabolomics results are reported in Figure 23.

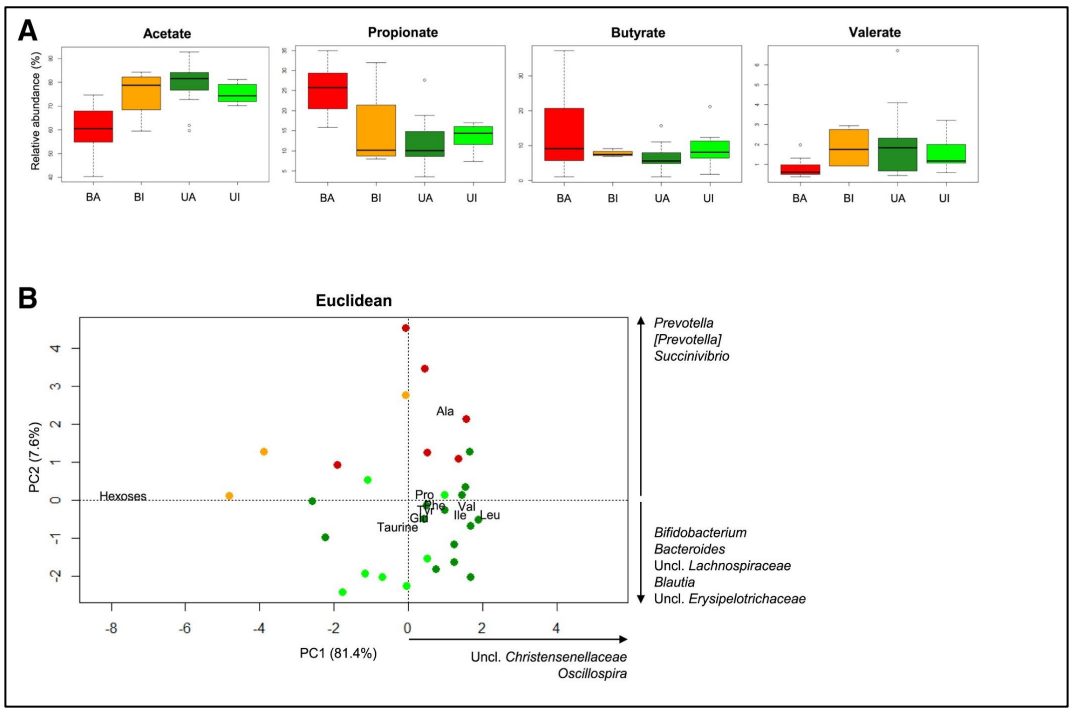


Figure 23 | Fecal Metabolome of Bassa and Urban Nigerians. (**A**) Boxplots showing the relative abundance distribution for short-chain fatty acids. Acetate and propionate levels are different between study groups ($p \leq 0.004$, Kruskal-Wallis test). Valerate is enriched in Bassa infants compared with adults ($p = 0.04$, Wilcoxon rank-sum test). (**B**) PCA of Euclidean distances between the metabolic profiles of the study populations, assessed using a semi-untargeted metabolomics approach- The main discriminant metabolites are mapped on the plot. Genera of the gut microbiota significantly correlated to PC1 and PC2 ($p < 0.05$, Kendall tau correlation test) are displayed at the bottom and on the right, respectively. BA, Bassa adults (red); BI, Bassa infants (orange); UA, urban adults (olive green); UI, urban infants (green).

206 S.H. Park, K.A. Kim, Y.T. Ahn, J.J. Jeong, C.S. Huh, D.H. Kim Comparative analysis of gut microbiota in elderly people of urbanized towns and longevity villages BMC Microbiol., 15 (2015), p. 49

207 J. Chen, E. Ryu, M. Hathcock, K. Ballman, N. Chia, J.E. Olson, H. Nelson Impact of demographics on human gut microbial diversity in a US Midwest population PeerJ, 4 (2016), p. e1514

208 L.A. David, C.F. Maurice, R.N. Carmody, D.B. Gootenberg, J.E. Button, B.E. Wolfe, A.V. Ling, A.S. Devlin, Y. Varma, M.A.Fischbach, *et al*. Diet rapidly and reproducibly alters the human gut microbiome Nature, 505 (2014), pp. 559-563

Bassa, especially infants, show an overall healthy profile with greater proportions of hexoses and fewer amounts of amino acids and biogenic amines compared with urban individuals[65]. The abundance of hexoses may be indicative of a diet high in microbiota-accessible carbohydrates, as that of the Bassa, heavily based on tubers, grains, and derived processed foods, as well as a variety of leafy soups, with the microbiota-dependent release of monosaccharides probably exceeding the enteric nutritional demands and thus excreted in the feces. With specific regard to Bassa infants, hexoses may also result from the digestion of the sugary and starchy liquid or semisolid foods they are fed with during weaning. On the other hand, the smaller amounts of amino acids and derivatives in Bassa feces may reflect less protein consumption compared with urban Nigerians and/or altered metabolisms or absorption.

When focusing on age, compositional data on urban infants corroborated what is well-known in developed countries, i.e., that the gut microbiota of infants aged < 3 years is unstable, with high inter-individual diversity and a taxonomic structure progressively approaching the more complex and stable adult-type microbiota. Conversely, different and at times opposite features were observed for the intestinal microbial ecosystem of Bassa infants that, compared with the adult counterparts, showed high biodiversity, lower inter-individual variability, and no difference at the various taxonomic levels. This in turn brought about a finding that lends clues to a lingering question from the work of Schnorr et al.[50], which is whether bifidobacteria are indeed absent in the kinetics of assembly and development of traditional microbiotas. The data on pre- and peri-weaned Bassa infants confirms that bifidobacteria, which are undoubtedly beneficial for Western human populations, are missing from certain traditional population infant guts. Collectively, these data are in contrast with the work of Yatsunenko et al.[85], which identified distinctive microbiome features in early childhood in rural populations, including greater inter-individual variation among children than adults, an increasing biodiversity with age, and the dominance of *Bifidobacterium*. From an ecological standpoint, it is possible to speculate that the extensive sharing of life within the Bassa community (in terms of lifestyle habits, contact with the environment, and usage of untreated river water, which indeed shows traces of microbiota components) results in a high degree of microbial

dispersal, thus allowing the human microbiome to behave as a meta-community[209]. In turn, the establishment of a meta-community, a feature probably common to traditional populations, has the potential to nullify the differences between infant and adult microbiomes, as mainly observed in Western populations, significantly shortening the microbiota assembly process. On the contrary, along with Westernization (involving sanitization, water treatment and other hygienic practices, and reduced life sharing with dispersal limitation), the human microbiome has lost its meta-community feature, resulting in increased individuality, and consequently driving the differentiation of the infant-type microbiome, as well as the trajectory of microbiome assembly, typical of Western populations. Indeed, the acquisition or, more likely, the extension of an infant-type microbiota in modern populations could be the result of neoteny in human evolution, favored by the establishment of profound differences in diet and lifestyle between infants and adults in modern societies.

The comparison of metabolite profiles from the Bassa with those from the Hadza hunter-gatherers reveals a shared pattern of enrichment in hexoses and reduction in amino acids and biogenic amines relative to urbanized counterparts. The differences between the populations of the present study are, however, less pronounced as those between Hadza and Italians, consistent with the less, and more recently, divergent lifestyle and environmental contexts of Bassa and urban Nigerians.

Conclusion

In summary, the microbial and metabolic characterization of the intestinal ecosystem of rural Bassa and urbanized individuals in Nigeria provided insights into the complex host-microbiome relationships across subsistence strategies, advancing our understanding of the changes in gut microbial communities and metabolic networks that probably accompanied human evolutionary history but, above all, stressing the relevance of the progressive adoption of a Western lifestyle as a major driver selecting for the loss of ancient signatures. Moreover, our findings support the existence of distinct trajectories

209 E.K. Costello, K. Stagaman, L. Dethlefsen, B.J. Bohannan, D.A. Relman The application of ecological theory toward an understanding of the human microbiome Science, 336 (2012), pp. 1255-1262

of development of the intestinal ecosystem in early life, depending on human ecological context.

Section 4.3 - Modulation of gut microbiota dysbioses in type 2 diabetic patients by macrobiotic diet

Introduction

Type 2 diabetes (T2D) is markedly increasing its prevalence in Westernised countries[210], and it represents a challenging problem for national healthcare systems[211]. Several insights provided evidence of an altered gut microbiota (GM) in T2D subjects, suggesting a possible role for gut micro-organisms in the disease onset[212-213].

Intestinal micro-organisms, and their metabolic products, have been shown to exert relevant functions in regulating host metabolic pathways. Although a mutualistic GM composition is crucial to support the host energy homeostasis, certain GM dysbioses can result in profound deregulations of the host metabolism, supporting the onset and consolidation of metabolic diseases, such as T2D[214]. Moreover, a pro-inflammatory layout of the gut microbial ecosystem has been suggested to be the basis of chronic inflammatory processes observed in T2D, and the concept of metabolic infection has been proposed[215]. As a result of an increased gut permeability, endotoxins from pro-inflammatory GM components can penetrate the epithelial barrier and aggravate metabolic inflammation and insulin resistance in T2D[216]. As diet has been recognised as a potent modulator of the composition and metabolism of the human GM[217], the possibility to improve metabolic control in T2D by developing selective diets that are able to correct the GM dysbioses has been considered[218].

Very recently, macrobiotic diet has been reported to be more effective than a control mediterranean diet (CTR), which is based on the dietary guidelines recommended by

---

210 Xu Y (2013) Prevalence and control of diabetes in Chinese adults. JAMA 310, 948.

211 Zhang P, Zhang X, Brown J, et al. (2010) Global healthcare expenditure on diabetes for 2010 and 2030. Diabetes Res Clin Pract 87, 293–301.

212 Xu Z, Malmer D, Langille MGI, et al. (2014) Which is more important for classifying microbial communities: who's there or what they can do? ISME J 8, 2357–2359.

213 Karlsson FH, Tremaroli V, Nookaew I, et al. (2013) Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature 498, 99–103.

214 Tilg H & Moschen AR (2014) Microbiota and diabetes: an evolving relationship. Gut 63, 1513–1521.

215 Burcelin R (2012) Regulation of metabolism: a cross talk between gut microbiota and its human host. Physiology (Bethesda) 27, 300–307

216 Cani PD, Amar J, Iglesias MA, et al. (2007) Metabolic endotoxemia initiates obesity and insulin resistance. Diabetes 56, 1761–1772.

217 David LA, Maurice CF, Carmody RN, et al. (2013) Diet rapidly and reproducibly alters the human gut microbiome. Nature 505, 559–563.

218 Everard A & Cani PD (2013) Diabetes, obesity and gut microbiota. Best Pract Res Clin Gastroenterol27, 73–83.

professional societies in Italy, for the improvement of metabolic control in T2D patients[219]. Specifically, the macrobiotic diet is enriched in complex carbohydrates, legumes, fermented products, sea salt and green tea, and it excludes fat and protein from animal source and added sugars. In a 21-d controlled open-label trial (MADIAB trial), fifty-six overweight T2D patients were randomised (1:1 ratio) to the macrobiotic diet or the CTR diet.

Methods

- Study design

The design of the MADIAB trial is described in Soare et al.[196]. Briefly, it was designed as a 21-d controlled open-label trial, in which the participants were assigned (1:1) to the macrobiotic diet or a CTR diet based on the dietary guidelines for T2D recommended by professional societies in Italy. The trial was conducted in accordance with the Declaration of Helsinki and the Good Clinical Practice guidelines, and the study was approved by the Institutional Review Board of University Campus Bio-Medico (trial registration number ISRCTN10467793; http://www.isrctn.com/ISRCTN10467793). Written informed consent was obtained from all subjects/patients. The Department of Endocrinology and Diabetes of the University Campus Bio-Medico in Rome (Italy) recruited overweight or obese (BMI 27–45 kg/m2) subjects, aged 40–77 years and affected by T2D. Associated metabolic syndrome was evaluated according to the National Cholesterol Education Program Adult Treatment Panel III criteria, although it was not an

inclusion criterion. Inclusion criteria were as follows: T2D diagnosed at least 1 year before the start of the trial, treated exclusively with dietary intervention, oral hypoglycemic drugs or both for 6 months before study entry. Exclusion criteria were as follows: the use of insulin either at present or at any time in the 2 year before the study, current use of corticosteroid therapy or any other drug that can interfere with carbohydrate metabolism, alcohol abuse and pregnancy. In addition, thirteen healthy controls, aged 21–40 years (mean age 32 years) and with 18·3–24·6 kg/m2 BMI, were enrolled for the study . All

219 Soare A, Khazrai YM, Del Toro R, et al. (2014) The effect of the macrobiotic Ma-Pi 2 diet vs. the recommended diet in the management of type 2 diabetes: the randomized controlled MADIAB trial. Nutr Metab 11, 39.

samples were immediately frozen at −20°C, and then transferred within 1 week to −80°C and stored there until processing.

- DNA processing, sequencing, bioinformatics and statistics

For further informations about 16S rRNA extraction and sequencing refer to Candela et al[220]. Amplicon sequences were deposited in the MG-RAST database under accession 17675.

Raw sequences were processed using a pipeline combining PANDAseq[103] and QIIME[28]. High-quality reads were binned into operational taxonomic units (OTU) at a 0.97 similarity threshold using UCLUST[104]. Taxonomy was assigned using the RDP (Ribosomal Database Project) classifier against Greengenes database[32] (May 2013 release). Chimera filtering was performed by discarding all singleton OTU. α Rarefaction was analyzed by using the Faith's phylogenetic diversity, Chao1, observed species and Shannon index metrics. β Diversity was estimated by computing weighted and unweighted UniFrac distances. Weighted UniFrac distances were used for principal coordinates analysis (PCoA) and plotted by the rgl and vegan packages of R. Data separation in the PCoA was tested using a permutation test with pseudo F ratios (function adonis in the vegan package). Heat-map analysis was performed using the R ggplot2 package.

Functional reconstruction of Greengenes-picked OTU was performed using PICRUSt[105] with default settings. The KEGG Orthology (KO) database[221] was used for functional annotation. Procrustes superimposition was conducted on the normalised KO gene data set and phylogenetic compositional data using vegan and rgl.

The correlation between age and GM diversity was computed by Kendall τ correlation test. All statistical analyses were performed in R, version 3.1.3. Significant differences were assessed by Wilcoxon's signed rank-sum test. When appropriate, a paired test was used. Where necessary, P values were corrected for multiple comparisons using the Benjamini–Hochberg method. P<0·05 was considered statistically significant.

220 Ref

221 https://www.genome.jp/kegg/pathway.html

Results and discussion

To characterise GM dysbioses in T2D at the enrolment, the compositional structure of the GM at T0 was compared with that of healthy controls. T2D patients were characterised by a significant reduction of the GM Shannon diversity index (P<0·05, Wilcoxon's signed rank-sum test). Even if it cannot be excluded that the age differences between T2D patients (mean age 66 years) and healthy controls (mean age 32 years) contribute, at least in part, to the observed differences in GM diversity, I failed to detect any significant correlation between age and microbiome diversity in the data set. The PCoA of the weighted UniFrac distances resulted in a significant segregation between the two groups (Figure 24 A (P<0·001, permutation test with pseudo F ratios), confirming the presence of compositional differences in the GM structure of T2D patients and healthy controls. To identify the microbial genera responsible for this separation, the bi-plot of the average bacterial coordinates weighted by the corresponding bacterial abundance per sample was superimposed on the PCoA plot (Figure 24B).

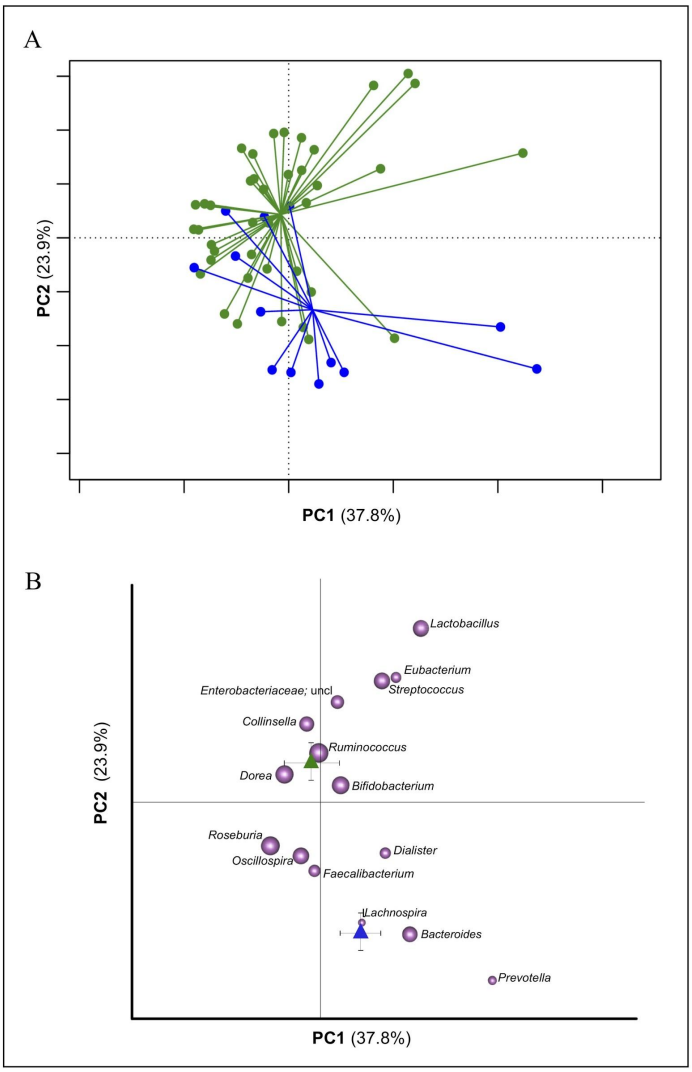


Figure 24 | Comparison of the gut microbiota compositional structure between overweight type 2 diabetes (T2D) patients at baseline and healthy controls. (**A**) Principal coordinates analysis (PCoA) based on weighted UniFrac distances shows separation between forty overweight T2D patients at T0 and thirteen normal-weight healthy controls. T2D patients. P<0·001; permutation test with pseudo F ratios. Green: T2D subjects; Blue: CTR subjects (**B**) Superimposition of microbial genera on the PCoA plot in order to identify the genera involved in this separation. Sphere width is proportional to the mean relative abundance of the genus across all samples. The two components explain 37·8 and 23·9 % of the variance, respectively.

The reduction of the GM compositional diversity in T2D corresponded to phylogenetic changes. T2D patients were indeed enriched in *Lactobacillus*, *Ruminococcus* and in several potential pro-inflammatory GM components, such as *Enterobacteriaceae*,

Collinsella and *Streptococcus*[222-223], whereas they were depleted in important health-promoting SCFA producers, such as members of *Lachnospiraceae*, *Fecalibacterium*, *Bacteroides* and *Prevotella*. Subsequently, were explored the changes in GM functions matching these compositional perturbations by inferred metagenomics (Figure 25). The data suggest deregulation in pathways involved in the metabolism of amino acids, lipids and secondary metabolites in the GM of T2D patients, including a reduced abundance of functions for the metabolism of d-arginine and d-ornithine, as well as of d-glutamine and d-glutamate, a corresponding increase in the metabolism of tyrosine, alanine, aspartate and glutamate, and a higher load of functions involved in arachidonic acid metabolism and polyketide sugar biosynthesis. The observed T2D-related dysbiotic microbial community could exert a multifactorial role in the disease onset, contributing to metabolic and immune deregulation. Indeed, the T2D GM is slightly depleted in fibrolytic health-promoting mutualists, fundamental for providing butyrate and propionate from the degradation of indigestible plant polysaccharides and starch, such as the butyrate-producing *Dorea*, *Lachnospira*, *Roseburia* and *Fecalibacterium*, and the propionate-producing *Bacteroides* and *Prevotella*.

Even if the biological relevance of this depletion of SCFA producers remains to be determined, it could result in the reduction of bioavailability of these crucial GM metabolites in the gut, with consequences on the host metabolic and immunological homeostasis. For instance, butyrate and propionate are important for host glucose control[181], insulin sensitivity regulation, insulin signalling and intestinal gluconeogenesis[224]. In parallel, the observed increase of potential pro-inflammatory micro-organisms in the gut of T2D patients, such as *Enterobacteriaceae*, *Collinsella* and Streptococcus, could further contribute to raise the host inflammatory level, supporting the evolution of insulin resistance[225].

In the subset of forty diabetic participants – twenty-one assigned to the macrobiotic diet and nineteen to the CTR diet – was then explored the efficacy of the nutritional interventions in supporting the recovery of a mutualistic GM configuration in

---

222 Faber F & Bäumler AJ (2014) The impact of intestinal inflammation on the nutritional environment of the gut microbiota. Immunol Lett 162, 48–53

223 Kamada N, Seo SU, Chen GY, et al. (2013) Role of the gut microbiota in immunity and inflammatory disease. Nat Rev Immunol 13, 321–335.

224 Russell WR, Hoyles L, Flint HJ, et al. (2013) Colonic bacterial metabolites and human health. Curr Opin Microbiol 16, 246–254

225 52. Johnson AMF & Olefsky JM (2013) The origins and drivers of insulin resistance. Cell 152, 673–684.

T2D patients (Figure 26). Primary and secondary outcomes were re-analysed for this patient subset included in the gut microbiome study, confirming that the macrobiotic diet was associated with a greater reduction in fasting blood glucose, total serum cholesterol,

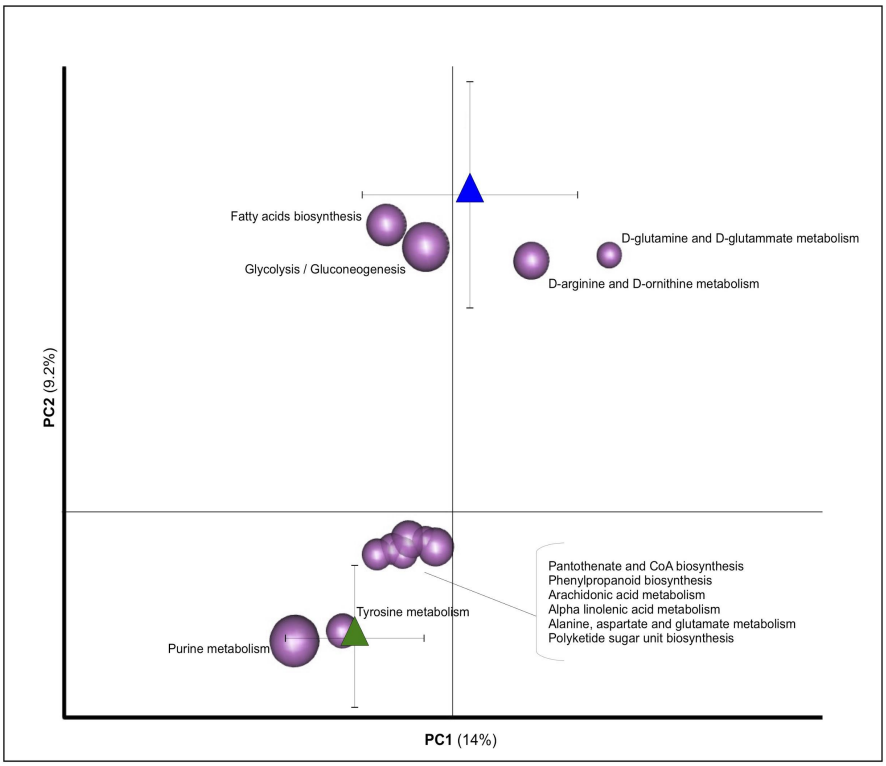


Figure 24 | Functional dysbioses of the gut microbiome in type 2 diabetes (T2D) patients. Metabolic pathways were superimposed on the principal component analysis plot based on Euclidean distances, and the pathways responsible for the separation are shown. An external file that holds a picture, illustration, etc. An external file that holds a picture, illustration, etc. Sphere width is proportional to the mean relative abundance of the function across all samples. (Green: T2D subjects. Blue: CTR subjects.

CRP and IL-6 in T2D patients. According to gut microbiome data, both macrobiotic and CTR diet were able to modulate the GM dysbioses in T2D patients, supporting the recovery of a healthy-like compositional structure and resulting in an increased ecosystem diversity, which represents a strategic feature for a healthy GM ecosystem[65]. According to imputed metagenomics, only the macrobiotic diet resulted in a significant modulation of the functional microbiome layout in T2D patients. In particular, the decrease of several markers of functional GM dysbioses in T2D patients, such as imbalances in alanine metabolism, arachidonic acid metabolism and polyketide sugar biosynthesis, was observed. Moreover, the macrobiotic diet favoured the reduction of GM functions related to oxidative phosphorylation and glycosphingolipids biosynthesis. Anaerobic respiration provides an ecological advantage for *Enterobacteriaceae* in an inflamed gut[226], whereas glycosphingolipids are powerful bacterial modulators of the host inflammatory response[227]. Thus, the reduction in abundance of these pathways further suggests the

226 Faber F & Bäumler AJ (2014) The impact of intestinal inflammation on the nutritional environment of the gut microbiota. Immunol Lett 162, 48–53

227 Wieland Brown LC, Penaranda C, Kashyap PC, et al. (2013) Production of α-galactosylceramide by a prominent member of the human gut microbiota. PLoS Biol 11, e1001610.

potential of macrobiotic diet to counteract the ongoing bloom of pro-inflammatory pathobionts in T2D.

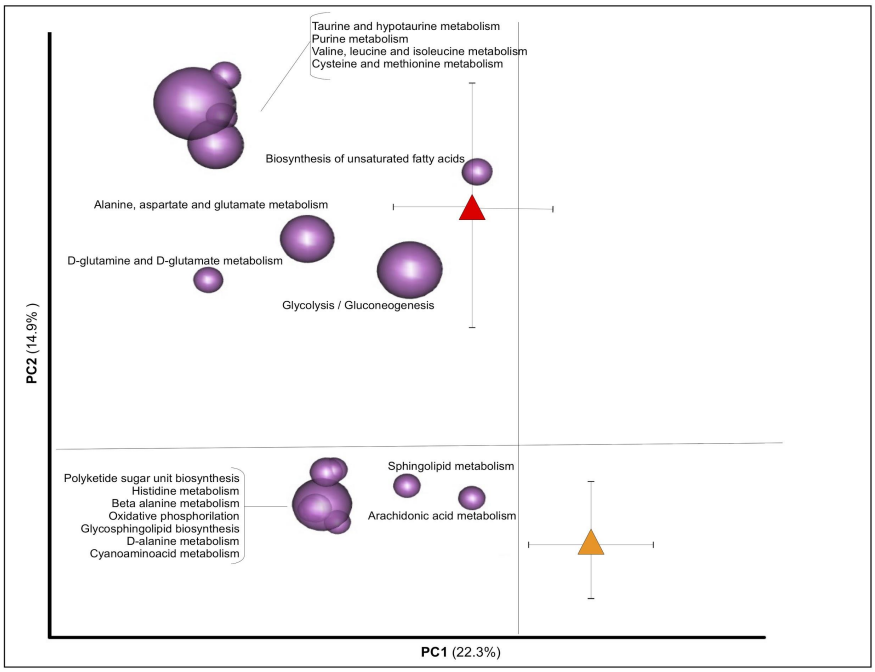


Figure 26 | Impact of macrobiotic dietary intervention on the functional configuration of the gut microbiome in T2D patients. Metabolic pathways were superimposed on the principal component analysis plot based on Euclidean distances in T2D patients before (T0 - Orange) and after (T1 - Red).

Conclusion

Both macrobiotic and CTR diets showed the potential to support the recovery of GM-host mutualism in T2D patients, favoring the restoration of carbohydrate-degrading SCFA-producing GM components, thus promoting metabolic control of T2D patients. Differently from the CTR diet, the macrobiotic diet was effective in counteracting the rise of possible pro-inflammatory micro-organisms in T2D patients. This suggests that the diet may have the potential to reduce GM-dependent pro-inflammatory stimuli in the gut that, increasing chronic inflammation, can lead to insulin resistance in T2D. Even if direct causation still needs to be proved, and this conclusion must be taken with adequate caution, this peculiar property shown by the macrobiotic diet could partly explain the greater improvements in metabolic control following that dietary intervention compared with the CTR diet.

Section 4.4 - Gut resistome plasticity in pediatric patients undergoing hematopoietic stem cell transplantation

Introduction

The rate of infection by antibiotic-resistant bacteria (ARB) is continuously raising worldwide, particularly because of the selective pressure resulting from the increasing usage of broad-spectrum antibiotics[228]. This burden of ARB is of particular relevance for hematological patients, who undergo frequent antimicrobial prophylaxis and treatments[229]. The prolonged exposure to health care settings may indeed favor the progressive accumulation of antimicrobial resistance (AMR) genes in the gut microbiome (GM) of patients[230]. Consequently, opportunistic ARB can accumulate in intestinal niches, where they can take advantage of the chemotherapy-induced damage to the gut epithelium and the overlapping neutropenia, spreading through the gut wall and causing life-threatening systemic infections[231]. In patients who have received an allogeneic hematopoietic stem cell transplantation (HSCT), systemic infections with ARB have indeed been associated with a non-relapse mortality rate from 36 to 95%[232-233]. Furthermore, gut colonization by ARB and associated systemic infections may strongly influence the process of immune system recovery following HSCT, thus affecting the incidence of acute Graft-versus-Host Disease (aGvHD)[234].

The gut resistome has recently been recognized as an important and dynamic reservoir of AMR genes, which can no longer be ignored when assessing antibiotic resistance[235-236]. In fact, it represents a basin of AMR genes that can be transferred to

---

228 Roca, I. *et al.* The global threat of antimicrobial resistance: science for intervention. *New Microbes New Infect.* 6, 22–29

229 Mikulska, M. *et al.* Aetiology and resistance in bacteraemias among adult and paediatric haematology and cancer patients. *J. Infect.* 68, 321–331,

230 Macesic, N., Morrissey, C. O., Cheng, A. C., Spencer, A. & Peleg, A. Y. Changing microbial epidemiology in hematopoietic stem cell transplant recipients: increasing resistance over a 9-year period. *Transpl. Infect. Dis.* 16, 887–896

231 Shono, Y. & van den Brink, M. R. M. Gut microbiota injury in allogeneic haematopoietic stem cell transplantation. *Nat. Rev. Cancer.* 18, 283–295

232 Kim, S. B. *et al.* Incidence and risk factors for carbapenem- and multidrug-resistant *Acinetobacter baumannii* bacteremia in hematopoietic stem cell transplantation recipients. *Scand. J. Infect. Dis.* 46, 81–88,

233 Girmenia, C. *et al.* Infections by carbapenem-resistant *Klebsiella pneumoniae* in SCT recipients: a nationwide retrospective survey from Italy. *Bone Marrow Transplant.* 50, 282–288

234 Sadowska-Klasa, A., Piekarska, A., Prejzner, W., Bieniaszewska, M. & Hellmann, A. Colonization with multidrug-resistant bacteria increases the risk of complications and a fatal outcome after allogeneic hematopoietic cell transplantation. *Ann. Hematol.* 97, 509–517

235 Holler, E. *et al.* Metagenomic analysis of the stool microbiome in patients receiving allogeneic stem cell transplantation: loss of diversity is associated with use of systemic antibiotics and more pronounced in gastrointestinal graft-versus-host disease. *Biol. Blood Marrow Transplant.* 20, 640–645

236 Gibson, M. K., Pesesky, M. W. & Dantas, G. The yin and yang of bacterial resilience in the human gut microbiota. *J Mol Biol.* 426, 3866–3876

passenger pathogens or opportunistic bacteria by horizontal gene transfer, with serious repercussions on human health[237].

In this scenario, the molecular assessment of the structure, ecology and evolution of the gut resistome in HSCT patients has become of strategic importance, allowing to understand the dynamics that govern the ARB establishment in such subjects. Particularly, the gut resistome characterization by shotgun metagenomics has been indicated as a unique and sensitive approach to understanding the genetic and biological effects of AMR in HSCT.

In the present study, was performed a whole-genome shotgun (WGS) metagenome sequencing of the fecal DNA from eight subjects (four developing aGvHD and four aGvHD-negative) from Biagi et al.[238]. This work provides glimpses on the gut resistome structure and its evolutionary trajectory in HSCT paediatric patients, before and after transplantation.

Methods

- Sample collection

This study used genomic DNA extracted from fecal samples of eight paediatric patients from Biagi et al.[205], who underwent allo-HSCT for high risk acute leukemia. Four patients out of the eight developed moderate (I-II grade) to severe (III-IV stage) aGvHD. Fecal samples were collected before HSCT and at different time points after the transplant, up to about 85 days post-HSCT, for a total of 32 samples (Figure 26). Because of episodes of febrile neutropenia occurred after the chemotherapy, patients received an empirical treatment based on a third-generation cephalosporin with activity against Pseudomonas before HSCT. Informed consent was obtained for all the subjects enrolled by parents and/or legal guardians. The study was approved by the Ethics Committee of the Sant'Orsola-Malpighi Hospital-University of Bologna (ref. number 19/2013/U/Tess). All methods were performed in accordance with the relevant guidelines and regulations.

---

237 Forslund, K. *et al*. Country-specific antibiotic use practices impact the human gut resistome. *Genome Res.* 23, 1163–1169

238 Biagi, E. *et al*. Gut microbiota trajectory in pediatric patients undergoing hematopoietic SCT. *Bone Marrow Transplant.* 50, 992–998

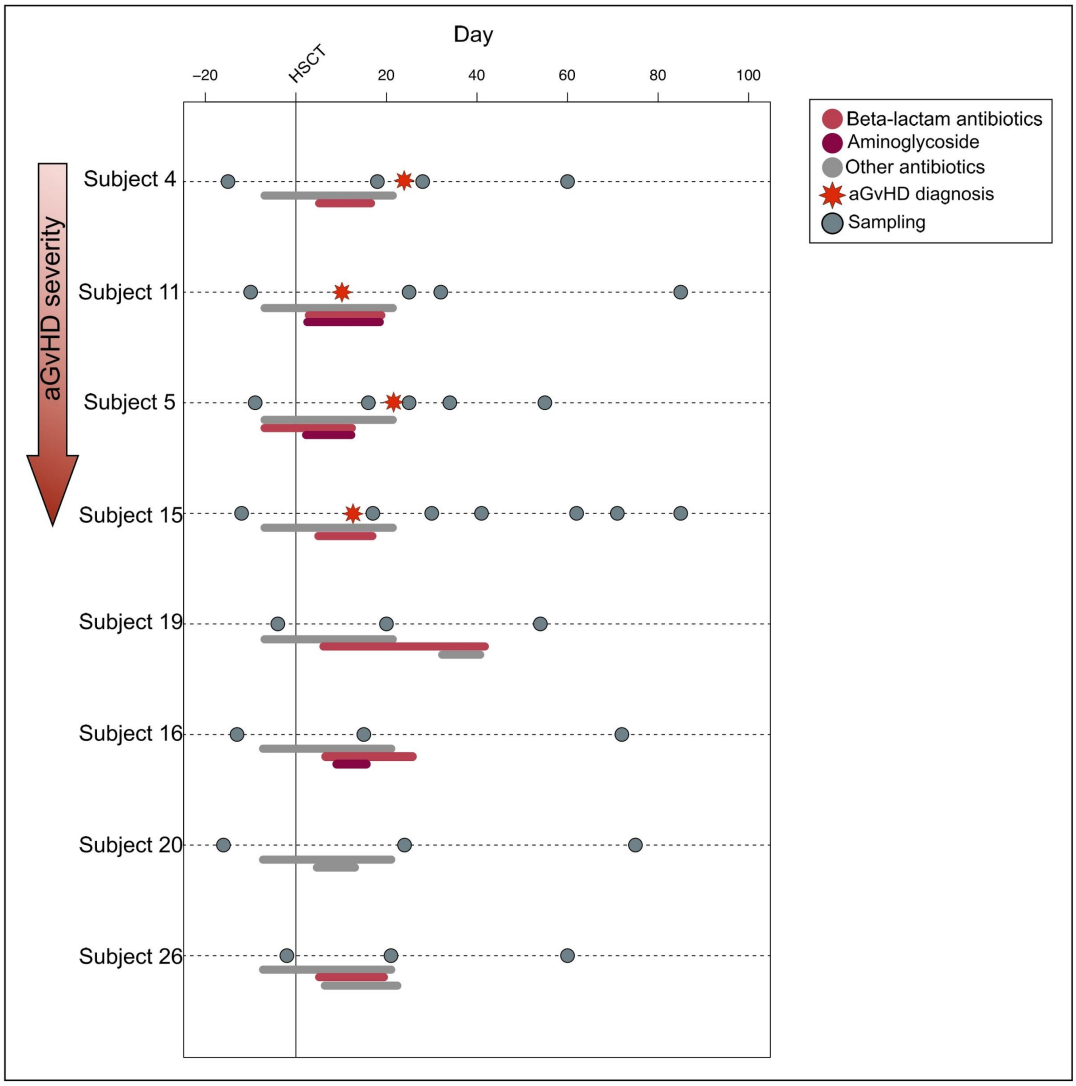


Figure 26 | Schematic representation of the sampling time for each enrolled patient. HSCT is represented as a vertical line in the graph, while the occurrence of aGvHD is highlighted with a red star on the subject timeline.

- Gut resistome analysis

Shotgun reads were quality filtered using the human sequence removal pipeline from the Human Microbiome Project, and filtered reads were assembled in contigs using the MetaVelvet tool[239]. Raw sequence reads were deposited in the National Center for Biotechnology Information Sequence Read Archive (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA525982). Protein sequences from the Antibiotic Resistance Genes Database (ARDB)27 were screened against the assembled metagenomes using tblastn[52]. Only alignments with identity ≥80% and alignment length of at least 200 residues were retained for further analysis. When multiple hits were present, the best one was selected

239 Namiki T, Hachiya T, Tanaka H, Sakakibara Y. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res. 2012 Nov 1;40(20):e155.

according to three criteria with the following priority: (I) percentage of identity and length of the alignment, (II) function showing the highest number of hit, and (III) presence of the corresponding microorganism in the respective gut ecosystem. For further analysis, the target resistance genes were normalized using the number of reads in the corresponding sample. Taxonomic classification of the identified sequences was retrieved from the results of tblastn. The amino acid sequences of the select proteins were clustered into Antibiotic Resistance Units (ARUs) at 30% identity level using UCLUST[139]. The most abundant sequence of each ARU was selected as a representative sequence and re-classified using BLASTP28 and ARDB27. ARU table containing resistance abundance across the samples was built using the script "make_otu_table.py" in QIIME[28] and used for further analysis as described below.

- Bioinformatics and statistical analysis

The ARU table was used as input for a Principal Coordinates Analysis (PCoA) based on Bray-Curtis distances between samples. PCoA graphs were generated using the "vegan" package (http://www.cran.r-project.org/package=vegan) in R studio version 1.0.153, and data separation was tested by permutation test with pseudo-F ratios (function "Adonis" in "vegan"). The ARU table was also used to build a heat map of the normalized ARU abundances before and after transplantation for all patients ("ggplot2" package). ARUs were superimposed on the bidimensional space using the function "envfit" of the "vegan" package and only AMR genes showing a significant correlation were plotted. Significant differences in ARU table between pre-HSCT patients and healthy controls were assessed by Wilcoxon signed rank-sum test. False discovery rate (FDR) < 0.05 was considered as statistically significant.

Results and discussion

In order to highlight the impact of previous therapeutic treatments on the gut resistome of HSCT pediatric patients, we firstly compared their pre-HSCT gut resistome configuration with the AMR gene composition of 10 healthy Italian subjects from Rampelli et al.[88]. According to findings, pre-transplant paediatric patients possess an overall gut resistome structure different from that of healthy individuals, possibly shaped by the previous prolonged exposure to health care settings and being enriched in AMR genes

providing for macrolide resistance[203-204] (Figure 27). However, it should be stressed that the comparison of the AMR composition was performed between children and adults, and therefore the data need to be taken with adequate caution as the gut microbiome structure is known to change with age. Afterward, have been analyzed the temporal variations of the gut resistome in the pediatric patients undergoing HSCT (Figure 28).

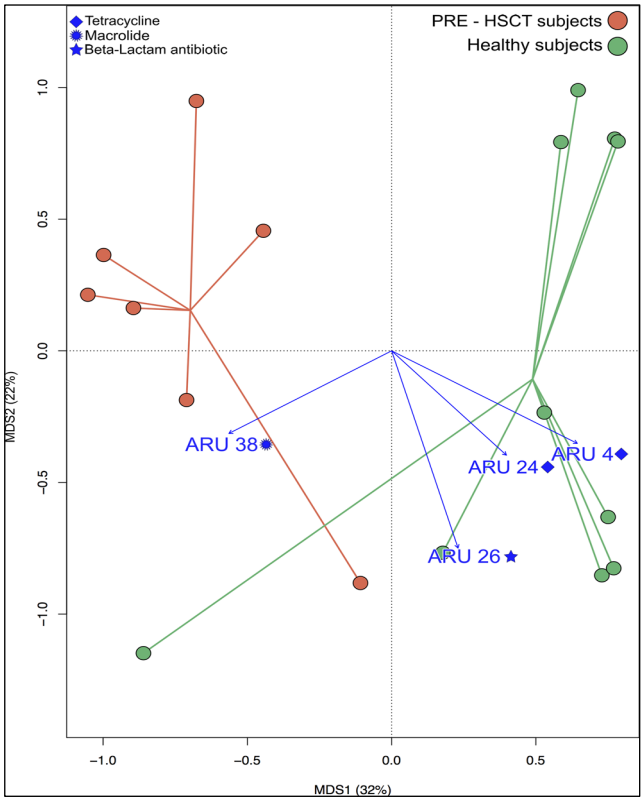


Figure 28 | Gut resistome structure of pre-HSCT pediatric patients and healthy subjects. Bray-Curtis distance-based Principal Coordinates Analysis showing separation between the gut resistome of pre-transplant pediatric patients and healthy controls. Permutation test with pseudo-F ratios (Adonis), p = 0.001. Antibiotic Resistance Units (ARUs) with a significant correlation with the bidimensional space are represented with a blue arrow.

Interestingly, the data highlight a distinctive gut resistome trajectory in patients developing aGvHD, involving not only the consolidation of AMR genes already present before transplanting, but also the acquisition of a vast number of new AMR genes following HSCT (Figure 29). Coding for multi drug, macrolide and aminoglycoside resistance classes, these newly acquired AMR genes were assigned to different bacterial families, including microorganisms of intestinal origin as *Bacteroidaceae*, *Enterobacteriaceae*, *Enterococcaceae*, *Eubacteriaceae* and *Streptococcaceae*, as well as cosmopolitan bacteria, such as *Pseudomonadaceae* and *Sphingobacteriaceae* (Figure 29 - 30). The gut resistome of aGvHD-positive patients was also found to be characterized by the increase in abundance of AMR genes already present before HSCT. In particular, a consistent post-HSCT bloom was detected only in aGvHD cases for AMR genes coding for a tetracycline inhibitor, β-lactamase CFXA3 and erythromycin resistance.

Interestingly, these AMR genes were assigned to GM components, including *Bacteroidaceae*, *Prevotellaceae*, *Lachnospiraceae* and *Streptococcaceae*. It is also worth noting that the bloom of β-lactamase CFXA3 and erythromycin resistance – prevalently

attributed to the major GM species of *Bacteroides* sp. and *B. fragilis* – was found to be associated with higher aGvHD severity (grade III and IV).

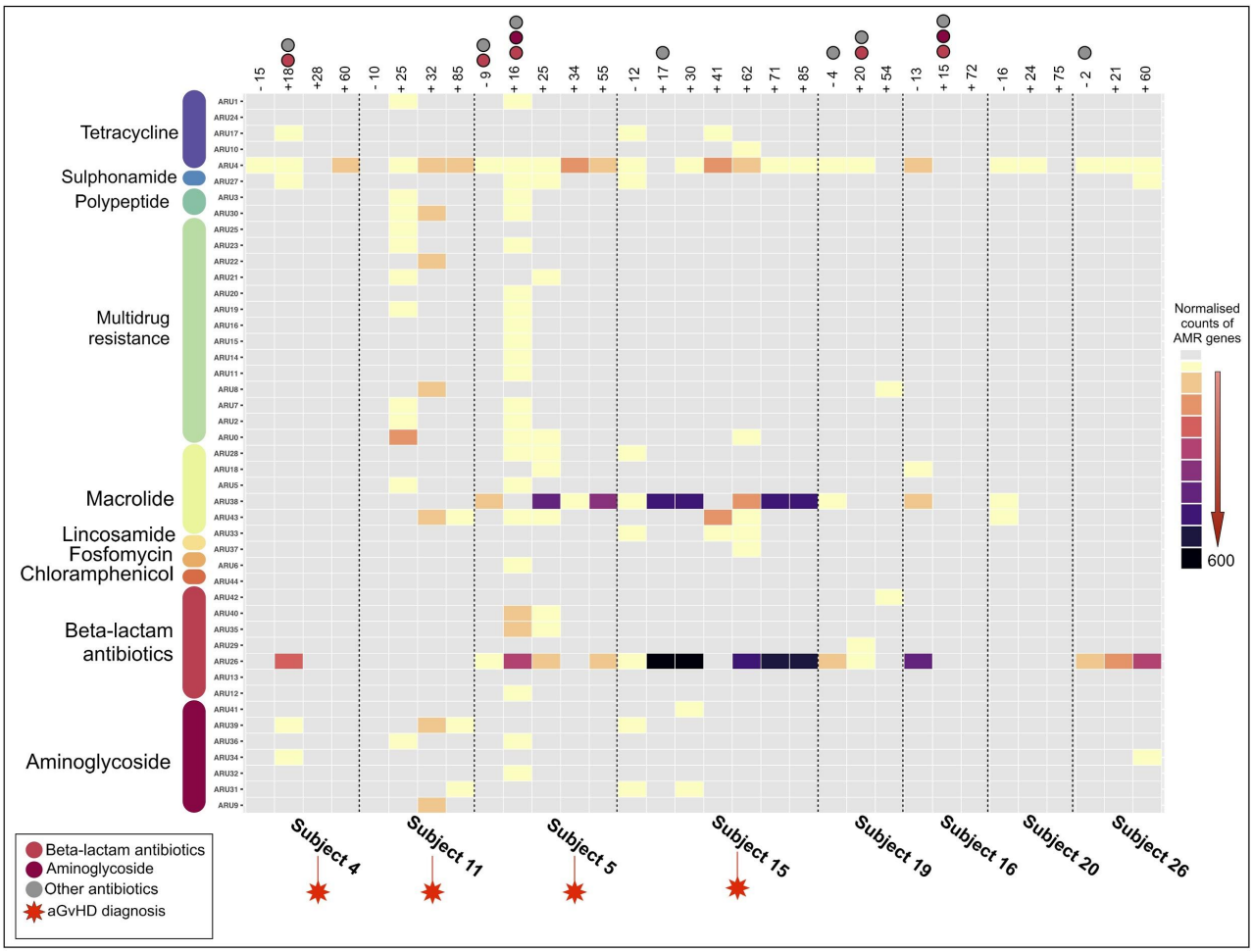


Figure 29 | ARUs trajectory over time in pediatric patients undergoing HSCT. Rectangles indicate the distribution of ARU abundances across time points for each subject, normalized by AMR gene count and represented with different colors, from gray (0 count) to black (600 counts). A black dotted vertical line is used to separate the sample sets of patients. aGvHD-positive subjects are highlighted with a red star. The presence of colored circles indicates the antibiotic intake during the specific time point (light red for beta-lactam antibiotics, dark red for aminoglycosides and gray for other antibiotics). Antibiotic Resistance Units (ARUs) are grouped by class of antibiotic, based on the assigned protein function.

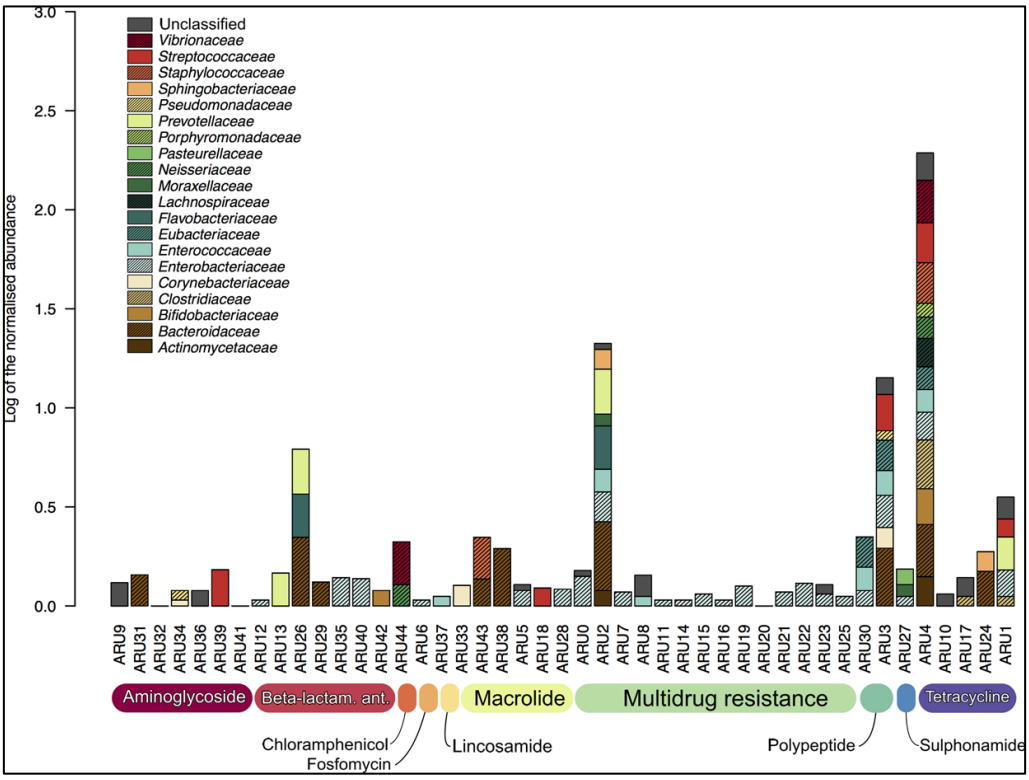


Figure 30 | Microbial ecology of Antibiotic Resistance Units (ARUs). For each ARU, total abundance and family-level distribution are represented. ARUs are grouped by class of antibiotic, based on the assigned protein function. The reported values were normalized using a logarithmic scale.

Conclusion

In conclusion, the assessment of the gut resistome dynamics in eight pediatric patients undergoing HSCT allowed to shed for the first time some light on the microbial ecology of ARB in HSCT, going beyond the limit of the traditional culture-dependent studies. Despite the low number of subjects, was indeed provided evidence that aGvHD onset is associated with a peculiar trajectory of the personal gut resistome following HSCT. Even if all patients received fluoroquinolone antibiotic prophylaxis from day −9 to day 21 and anti-infective therapy based on beta-lactam antibiotics after transplant, only aGvHD subjects showed an extremely diversified and rich gut resistome, with a pattern of AMR genes far exceeding the selective pressure due to the administered antibiotics. In particular, after HSCT, the resistome of pediatric patients developing aGvHD acquires a new and diversified pattern of AMR genes, either from enteric and environmental microorganisms, and including multi drug resistance, as well as resistances to macrolide and aminoglycoside antibiotic classes. However, in parallel with the acquisition of new AMR genes, the aGvHD development is also associated with a bloom of internal AMR

genes, already present in the individual gut resistome before the HSCT, and provided by major gut microbiome components such as *Bacteroides* sp. Particularly, this last element leads to the consolidation of AMR genes such as tetracycline inhibitor, β-lactamase CFXA3 and erythromycin resistances, the latter two associated with a high aGvHD severity grade The research indicate that the individual GM of HSCT patients can thus act as a dynamic reservoir of ARB, with the potential to implement the AMR gene pattern following HSCT. According to findings, this aGvHD-associated magnification process of the individual gut resistome involve variations in the abundance of endogenous gut microbiome ARB, as well as the acquisition of allochthonous ARB, of enteric or environmental nature. Even if these data must be confirmed on a larger cohort, in a recently published research[240], it has been assessed the gut resistome dynamics in 12 subjects exposed to an antibiotic therapy. Results highlighted a plastic resistome response which partially resembled observations. Indeed, according to the authors, four days post treatment it was observed an enrichment of AMR genes, not limited to the ones targeted to the administered antibiotics. The inherently plastic behavior of the human gut resistome supports the importance of WGS-based resistome surveys in pediatric HSCT patients, allowing a better comprehension of the ecological dynamics of antibiotic resistance in aGvHD-positive cases, with the final goal of allowing a better refinement of antibacterial therapies.

240 Palleja, A. *et al*. Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nat. Microbiol.* 3, 1255–1265

PART

Chapter 5 - Viruses

- Biology, ecology and study of viruses

Chapter 6 - Virome characterisation

- A new tool for profiling the virome and its charachterization across different populations

Chapter 7 - Fungi

- Biology, ecology and study of fungi

Chapter 8 - New insights in Mycobiome characterisation

- A new tool for profiling the fungal fraction in metagenomic studies

III FUNGI AND VIRUSES

CHAPTER 5 - Viruses

Introduction

Viruses are the most abundant biological entities present on Earth, representing one of the most successful forms of life on our planet[241-242-243]. Every known virus is an obligate genetic parasite, needing the support of the replicative machinery of cellular organisms to replicate and pack their genomes inside the viral particles (known as 'virions')[244]. Since the appearance of genetic parasites is theoretically inevitable in replicator systems[245], probably viruses have co-evolved with their hosts during the entire process of evolution[246-247]. In spite of their crucial importance in the context of the biosphere, little is known about the origin of viruses, mainly due to the lack of fossil remains or every other biological signatures. To date, three different hypotheses have been proposed regarding the rise of viral life on the planet[248-249-250-251-252-253]:

(I) basing on the 'primordial virus world' hypothesis, viruses are direct descendants of an early stage of life evolution on our planet in which was present a pre-cellular stage. During the evolution, life has organized in complex cellular structures, and viruses managed to exploit these structures for their advantage.

(II) By contrast, the 'regression scenario' hypothesizes the appearance of viruses as an event of evolutionary reduction, in which a protocell has over time lost its replicative ability, transitioning to obligate intracellular parasitism.

---

241 Danovaro, R. et al. Virus-mediated archaeal hecatomb in the deep seafloor. *Sci. Adv.*2, e1600492 (2016).

242 Cobián Güemes, A. G. et al. Viruses as winners in the game of life. *Annu. Rev. Virol.*3, 197–214 (2016).

243 Koonin, E. V. & Dolja, V. V. A virocentric perspective on the evolution of life. *Curr. Opin. Virol.* 3, 546–557 (2013).

244 Raoult, D. & Forterre, P. Redefining viruses: lessons from mimivirus. *Nat. Rev. Microbiol.* 6, 315–319 (2008).

245 Koonin, E. V., Wolf, Y. I. & Katsnelson, M. I. Inevitability of the emergence and persistence of genetic parasites caused by evolutionary instability of parasite-free states. *Biol. Direct* 12, 31 (2017).

246 Koonin, E. V. Viruses and mobile elements as drivers of evolutionary transitions. *Philos. Trans. R Soc. B Biol. Sci.* 371, 20150442 (2016).

247 Forterre, P. & Prangishvili, D. The major role of viruses in cellular evolution: facts and hypotheses. *Curr. Opin. Virol.* 3, 558–565 (2013).

248 Forterre, P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* 117, 5–16 (2006).

249 Sapp, J. The prokaryote-eukaryote dichotomy: meanings and mythology. *Microbiol. Mol. Biol. Rev.* 69, 292–305 (2005).

250 Flugel, R. M. The precellular scenario of genovirions. *Virus Genes* 40, 151–154 (2010).

251 Forterre, P. & Prangishvili, D. The origin of viruses. *Res. Microbiol.* 160, 466–472 (2009).

252 Koonin, E. V., Senkevich, T. G. & Dolja, V. V. The ancient virus world and evolution of cells. *Biol. Direct* 1, 29 (2006).

253 Morse, S. S. (ed.) in The Evolutionary Biology of Viruses 1–28 (Raven Press, 1994).

(III) Finally, the 'escaped gene hypothesis' proposes that during evolution several genes acquired a characteristic of selfish replication, disengaging from the original genome and becoming singular living entities. Viruses indeed use all the possible strategies for their genome replication and expression, whit different forms of nucleic acid involved: ssRNA, dsRNA, ssDNA and dsDNA. This great variance in genomic strategies collides with the homogeneity observed in cellular organisms and seems to give hints at the hypothesis of the viral origin in a pre-cellular 'primordial virus world'[254]. The 'regression scenario' as well has several clues in its direction: the discovery of giant viruses which genome encodes for part of the replicative system and their dimensions are comparable to bacterial cells[255] suggests that these genomic parasites have originated from a regression process of a 'complete organism'. Recently, Krupovic and colleagues[256] proposed an integrative hypothesis, the so-called 'chimeric origin'. In this scenario different types of primordial selfish replicons gave rise to viruses by recruiting host proteins for virion formation. New groups of viruses have likely repeatedly emerged at all stages of the evolution of life, often through the displacement of ancestral structural and genome replication genes.

Relations with other organisms

Although the origins of these organisms can only be hypothesized, their involvement within the biosphere is evident. Because of their stringent biology, viruses interact and infect almost all living forms for their replication and propagation. This interaction can bring different consequences to the host organism, ranging from mild infections to severe consequences or death. Seeing from another perspective, viral infections have probably represented a boost and a crucial factor in the evolution of life on Earth, representing nowadays the biggest reservoir of genetic diversity present on our planet[257]. Viruses, indeed, act as one of the main vectors of horizontal gene transfer between species, promoting genetic diversity and triggering evolutionary processes[258].

---

254 Holmes, E. C. What does virus evolution tell us about virus origins? *J. Virol.* 85, 5247–5251 (2011).

255 Abrahao, J. et al. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat. Commun.* 9, 749 (2018).

256 Krupovic, M; Dolja, VV; Koonin, EV (2019). "Origin of viruses: primordial replicators recruiting capsids from hosts". *Nature Reviews Microbiology*. 17 (7): 449–458.

257 Suttle CA, Marine viruses—major players in the global ecosystem, in Nature Reviews Microbiology, vol. 5, n° 10, 2007, pp. 801–12

258 Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brüssow H, Phage as agents of lateral gene transfer, in Current Opinion in Microbiology, vol. 6, n° 4, 2003

Among species, Humans are a well-known target of viral activities, and the study of the viral fraction in the human holobiont is a trivial objective to better understand their involvement in health and disease[259].

The study of viruses

Tobacco mosaic virus (TMV) was the first virus studied, although initially a transmission of the disease was suspected by means of toxins of bacterial origin and the presence of the viral particles was totally neglected . At the end of the 19th century the first scientist to introduce the term '*virus*' was the Dutch botanist and microbiologist Martinus Beijerinck, who repeated the experiment already carried out years before by the Russian biologist Dmitry Ivanovsky. In the experiment was shown that the filtrate extracted from infected tobacco leaves was able to infect other healthy plants. Despite Ivanovsky imputed this ability to unknown bacteria-originated toxins, Beijernck stated that the filtered water contained a new infective agent[260]. These events represented the first steps of modern virology paving the path for future discoveries: in 1935 Wendell Stanley examined the tobacco mosaic virus and found it was mostly made of proteins[261], while 4 years later was determined that the viral content was mainly represented by nucleic acid (in the case of TMV consisting of RNA)[262]. Later, in 1955 Rosalind Franklin finally resolved the structure of TMV, producing the first resolute representation of a viral particle[263]. Finally, the second half of the 20th century and the beginning of the 21st represented a golden era in virus discovery[264], in which the broader part of our knowledge about viruses was gained.

Nowadays, the introduction of the next generation sequencing techniques, largely discussed before, and metagenomic surveys opened a second golden era and new

259 Foxman EF, Iwasaki A. Genome-virome interactions: examining the role of common viral infections in complex disease. Nat Rev Microbiol. 2011;9(4):254–64.

260 Leppard, Keith; Nigel Dimmock; Easton, Andrew (2007). *Introduction to Modern Virology*. Blackwell Publishing Limited. pp. 4–5.

261 Stanley WM, Loring HS. THE ISOLATION OF CRYSTALLINE TOBACCO MOSAIC VIRUS PROTEIN FROM DISEASED TOMATO PLANTS. Science. 1936 Jan 24;83(2143):85.

262 Loring HS (1939). "Properties and hydrolytic products of nucleic acid from tobacco mosaic virus". Journal of Biological Chemistry. 130 (1): 251–258.

263 Franklin R.E. & Holmes K.C. (1956), "The Helical Arrangement of the Protein Sub-Units in Tobacco Mosaic Virus" (PDF), Biochimica et Biophysica Acta, 21 (2): 405–406,

264 Norrby E (2008). "Nobel Prizes and the emerging virus concept". Archives of Virology. 153 (6): 1109–23.

perspectives in the study of viruses, allowing the depiction of the viral communities associated to different environments and hosts[265-266-267].

265 Aguirre de Cárcer D, López-Bueno A, Pearce DA, Alcamí A. Biodiversity and distribution of polar freshwater DNA viruses. Sci Adv. 2015 Jun 19;1(5):e1400127.

266 Virgin HW. The virome in mammalian physiology and disease. Cell. 2014;157(1):142–50.

267 Pratama AA, van Elsas JD. The 'Neglected' Soil Virome - Potential Role and Impact. Trends Microbiol. 2018 Aug;26(8):649-662.

CHAPTER 6 - Virome characterization

Unlike for bacteria, the complete lack of universal phylogenetic marker genes in viruses increases the challenge in virome profiling. The viral fraction of a microbial community can be estimated from metagenomic shotgun sequencing and RNA-seq, through in-silico work aimed to isolate and assign the viral reads to the appropriate viral taxa. Metagenomic samples contain indeed nucleic acids derived from bacteria, archaebacteria, eukaryotes, and viruses, and this amount of information can be used to characterize the 'metavirome' by assembled or read-mapping approaches.

In this context, I have worked in two separate frameworks with the same objective: the depiction of the Human gut virome. In the first framework I produced a software, namely ViromeScan[268], focused on the characterization of the viral community in metagenomic samples. After its validation, the program was used to study different gut viromes across several Human populations.

Section 6.1 - ViromeScan: a new tool for metagenomic viral community profiling

Introduction

Even if the importance of the interplay among virome, microbiome and immune system is already evident, the available techniques for virome characterization usually underestimate the quantity and diversity of viruses in the samples[269]. For example, it is recognized that the methods for the viral isolation based on filtering procedures miss giant virus[270]. The viral taxonomic composition of a microbial community could be estimated from metagenomic shotgun sequencing and RNA-seq of the microbiota DNA/RNA, by detecting and assigning the viral reads to the appropriate viral taxa. Metagenomic samples contain indeed nucleic acids for bacteria, archaebacteria, host, phages and eukaryotic viruses. However, currently the most advanced experimental

268 https://sourceforge.net/projects/viromescan/

269 Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. Curr Opin Virol. 2012;2(1):63–77. doi: 10.1016/j.coviro.2011.12.004

270 Colson P, Fancello L, Gimenez G, Armougom F, Desnues C, Fournous G, et al. Evidence of the megavirome in humans. J Clin Virol. 2013;57(3):191–200.

procedures foresee to extract and isolate the encapsidated viral fraction[271 - 272] and only at a later stage, to characterize the metavirome by assembled or read-mapping approaches[273-274]. To sequence unprocessed samples and directly assign the obtained reads would instead allow a faster characterization of the virome in the context of the microbiome of origin avoiding the loss of giant virus in filtering procedures. It is here presented ViromeScan, a new tool that accurately profiles viral communities and requires only few minutes to process thousands of metagenomics reads. ViromeScan works with shotgun reads to detect traces of DNA and/or RNA viruses, depending on the input sequences to be processed. ViromeScan is available at the website http://sourceforge.net/projects/viromescan/.

Methods

- Workflow of the software

Once downloaded, ViromeScan locally processes the metagenome to search for eukaryotic viral sequences. Input files should be single-end or paired-end reads in .fastq format (for paired-end reads compressed files in .gzip, .bzip2 and .zip formats are also

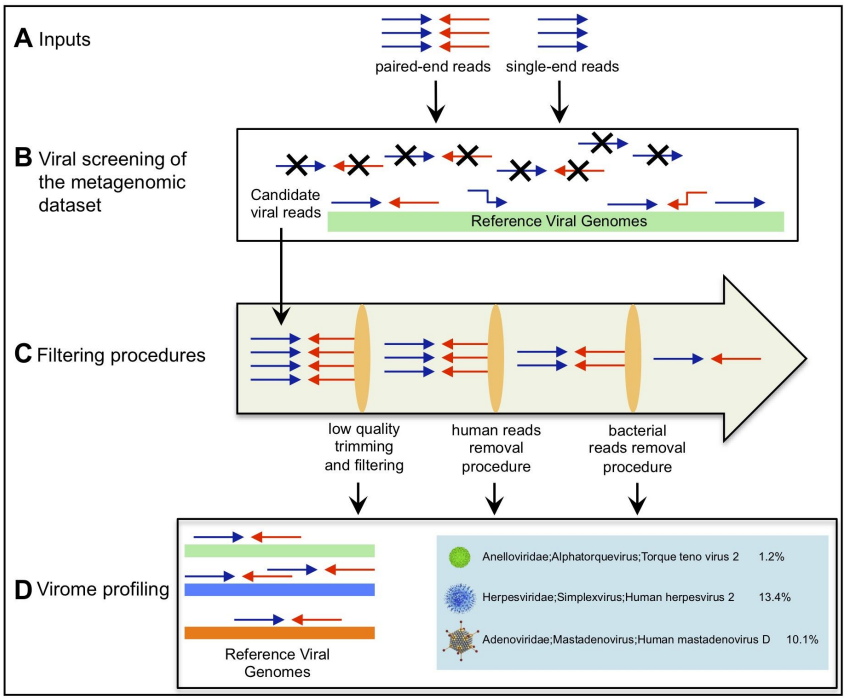


Figure 31 | Workflow of ViromeScan. **A** Inputs are single-end reads (fastq format) or paired-end reads (fastq or compressed fastq format). **B** Candidate viral reads are identified by mapping the sequences to the selected reference database. Unmapped reads are not contained in the resulting file. **C** Three filtering procedures to trim low quality reads and completely remove human and bacterial contaminations are computed. **D** The remaining viral sequences are assigned to appropriate taxonomy and the results are tabulated as both relative abundance and read counts

271 Duhaime MB, Sullivan MB. Ocean viruses: rigorously evaluating the metagenomic sample-to-sequence pipeline. Virology. 2012;434(2):181–6.

272 Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. Laboratory procedures to generate viral metagenomes. Nat Protoc. 2009;4(4):470–83.

273 Lorenzi HA, Hoover J, Inman J, Safford T, Murphy S, Kagan L, et al. TheViral MetaGenome Annotation Pipeline(VMGAP):an automated tool for the functional annotation of viral Metagenomic shotgun sequencing data. Stand Genomic Sci. 2011;4(3):418–29.

274 Wommack KE, Bhavsar J, Polson SW, Chen J, Dumas M, Srinivasiah S, et al. VIROME: a standard operating procedure for analysis of viral metagenome sequences. Stand Genomic Sci. 2012;6(3):427–39.

accepted) retrieved from shotgun sequencing or RNA-seq. Depending on the research strategy, ViromeScan gives users the option to choose from a range of ad-hoc built reference databases, including human DNA virus database, human DNA/RNA virus database, eukaryotic DNA virus database and eukaryotic DNA/RNA virus database. The human virus databases contain only viruses that have the human species as the natural host; on the other hand, the eukaryotic virus databases also include viruses for vertebrates, invertebrates, fungi, algae and plants, while excluding bacteriophages. All databases are based on the complete viral genomes available on the NCBI website[275]. The schematic description of the procedures of analysis computed by ViromeScan is provided in Figure 31. In detail, metagenomic reads are compared to the viral genomes of the selected database using bowtie2[49]. This first step is a complete and accurate screening of the sequences to select candidate viral reads. Performing this procedure before filtering processes allows a considerable gain of time in the subsequent parts of the pipeline, due to the reduction of the dataset to less than 1 % of the total amount of metagenomic reads. Afterwards, a quality filtering step of the candidate viral reads has been implemented as described in the processing procedure of the Human Microbiome Project (HMP)[276]. In brief, sequences are trimmed for low quality score using a modified version of the script trimBWAstyle.pl that works directly from BAM files. The script is utilized to trim bases off the ends of sequences, which show a quality value of two or lower. This threshold is taken to delete all the bases with an uncertain quality as defined by Illumina's EAMMS (End Anchored Max Scoring Segments) filter. Additionally, reads trimmed to less than 60 bp are also removed. Since the sequences analyzed are whole-genome or RNA-seq products, it is plausible that the candidate viral reads contain a small percentage of human reads. For this reason, it is necessary to subject the sequences to the control for human contamination. As reported in the HMP procedures[277], Human Best Match Tagger (BMTagger)[278] is an efficient tool that discriminates among human, viral and microbial reads. First, BMTagger attempts to discriminate between human reads and the other reads by comparing the 18-mers produced from the input file with those contained

275 The NCBI viral genome database. http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?opt=virus&taxid=10239.

276 Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449(7164):804–10.

277 NIH Human Microbiome Project website. http://www.hmpdacc.org

278 BMTagger. 2011. ftp://ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger/

in the reference human database. If this process fails, an additional alignment procedure is performed to guarantee the detection of all matches with up to two errors.

Human-filtered reads may also contain an amount of bacterial sequences, which need to be filtered out to avoid biases due to bacterial contamination. Bacterial reads are identified and masked using BMTagger, the same tool utilized for the human sequence removal procedure. In particular, in order to detect bacterial sequences, human-filtered reads are screened against the genomic DNA of a representative group of bacterial taxa that are known to be common in the human body niches. See Additional file 2 for the list of bacteria included in this process. Nevertheless, the user can customize the filtering procedure by replacing the bacterial database within the ViromeScan folder with the microbial sequences of interest, associated to environments other than the human body (e.g. microbiome associated with animals, soil or water). Finally, filtered reads are again compared to the viral genomes of the chosen hierarchical viral database using bowtie2, allowing the definitive association of each virome sequence to a viral genome. For each sample analyzed, the total amount of counts is summarized in a table as number of hits and relative abundance. Additionally, graphs representing the abundances at family, genus and species level are provided, using the "graphics" and "base" R packages.

- Validation of the tool and comparison with other existing methods

Five different mock communities each containing 20 human DNA viruses at different relative abundances were built and submitted to ViromeScan for its validation. The mock communities contained also human sequences and reads of other microorganisms to test the filtering steps of the pipeline. The simulated metagenomes were composed of sequences of 100 bp randomly generated from the chosen genomic DNAs by an in-house developed script. In order to compare the performance of ViromeScan with other existing tools, the same mock samples were analyzed using Metavir[279] and blastN[51]. In particular, in the Metavir pipeline, we determined the taxonomic composition using the number of best hits normalized by genome length through the GAAS metagenomic tool[280].

---

279 Roux S, Tournayre J, Mahul A, Debroas D, Enault F. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. BMC Bioinformatics. 2014;15:76.

280 Angly FE, Willner D, Prieto-Davó A, Edwards RA, Schmieder R, et al. The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. PLoS Comput Biol. 2009;5(12):e1000593.

- Case study: using ViromeScan to profile the eukaryotic DNA virome across different human body sites

Twenty metagenomic samples from HMP[244], belonging to four body sites, including stool, mid vagina, buccal mucosa and retroauricular crease, were used to illustrate the results that can be obtained by ViromeScan. These metagenomes have been sequenced using the Illumina GAIIx platform with 101 bp paired-end reads. The entire metagenomic dataset was utilized to study the differences in the composition of the viral communities across different body sites. No ethics approval was required for any analysis performed in this study.

Results and discussion

The ViromeScan software is specifically designed to the analysis of viromes. In particular, it can be used to determinate the viral fraction inside the microbiome from a given environment using raw reads, mostly in .fastq format generated by next-generation sequencing technologies. ViromeScan has the advantage of using a read-mapping approach that allows I) the characterization of the virome within a metagenome, including bacterial, eukaryotic and host sequences, without specific extraction/purification strategies, and II) the preservation of all the information retained in the input files, information that may be lost by an assembly approach[281]. Specifically in the context of a metagenomic dataset, the viral DNA could be under-sequenced due to the huge amount of bacterial and human DNA in the samples, making the assembly difficult or even impossible for viruses with a limited number of reads. However, as all the read-mapping approaches, ViromeScan is blind to viral sequences that are not closely related to viruses already present in the reference database. ViromeScan determines the taxonomic composition of the virome by sequence alignment of the reads to completely known viral genomes, and displays the results as raw number of hits or normalized hits (relative abundance). The ViromeScan classifier can be used for multiple analysis of the virome, in particular the normalized results describe the structure of the viral community in terms of relative abundance, and the read count output can be used as an indicator of the richness and diversity of such community in the context of the metagenome of origin. The initial

281 Davenport CF, Tümmler B. Advances in computational analysis of metagenome sequences. Environ Microbiol. 2013;15(1):1–5.

choice of the appropriate reference database is possible because the hierarchical databases built within ViromeScan contain sequences for DNA or DNA/RNA eukaryotic viruses, making the tool very adaptable to the needs of the user. Specifically, 92 genomes for the human DNA virus database, 664 for the human DNA/RNA virus database, 1646 for the eukaryotic DNA virus database and 4370 for the eukaryotic DNA/RNA virus database are contained in the tool. In addition, every ViromeScan user can create its own database for a customized analysis, including assembled sequences of unknown viruses, which could be useful to extend the taxon detection limit of the tool. Finally, another advantage of the tool is that the same metagenomic sample utilized to characterize bacterial, archaeal and eukaryotic fractions within the microbiome, can be used for the viral profiling, opening new perspectives in metagenomic characterization studies.

Was first evaluated ViromeScan performance in estimating the composition of viral communities using synthetic data. To this aim, 5 mock communities comprising reads from 20 different human DNA viruses, bacterial microorganisms and human genome were constructed, simulating metagenomes retrieved from the intestinal microbiota. ViromeScan correctly mapped the majority of the reads and identified all the 20 viruses in the synthetic communities, accurately estimating their relative abundance at different taxonomic levels (r.m.s. errors 0.04 at family level and 0.05 at species level), with 100 % of the viral species within 1 % deviation from expected value and the best overall prediction (Pearson $r > 0.999$, species level Pearson $P < 1 \times 10-22$). ViromeScan was more accurate on all tested synthetic metagenomes than the other existing methods, with blastN showing the closest performance but substantially slower running time (Figure 32). Several other tools for viral community characterization are available but they have been specifically designed to work with long sequences, or to detect open reading frames, which prevented their employment in our comparative analysis[240-241]. Furthermore, ViromeScan performed the classification at 140 reads per second on a standard single processor system, which was faster and more performing than other methods (Figure 32D).

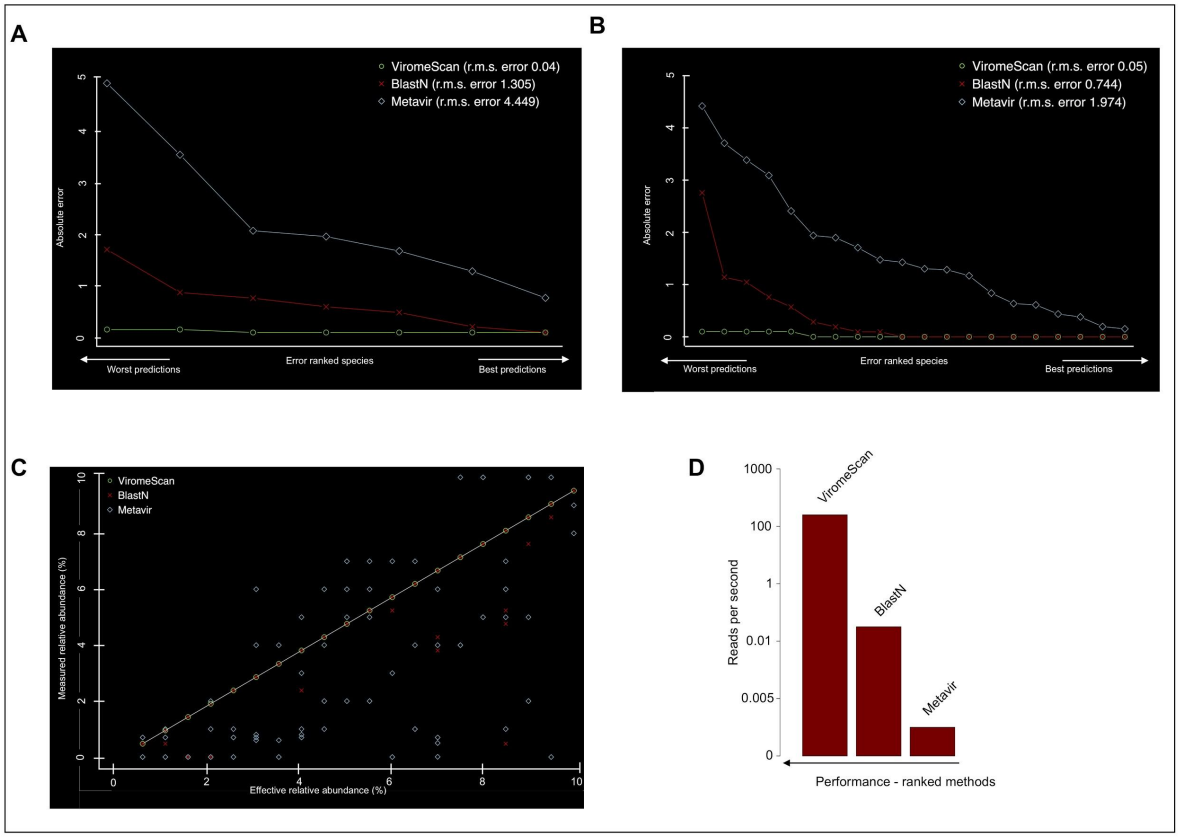


Figure 32 | Comparison of ViromeScan to other existing methods. A total of five synthetic viral communities were used in order to compare ViromeScan with Metavir and blastN. Absolute and r.m.s. errors in assigning taxonomy at (**A**) family and (**B**) species level are shown. (**C**) Correlation between predicted and real relative abundance for the 5 non-evenly distributed mock communities. (**D**) Read rate for the tested tools on single CPU.

The currently existing tools do not foresee filtering steps during the computational process, because they are designed to directly analyze viral reads. This fact constitutes a major limitation for the analysis of metagenomic samples, which usually contain a huge amount of bacterial and human reads. The strategy adopted by ViromeScan has been specifically studied to overcome this problem. In particular, two filtering steps, one for bacterial and one for human reads, have been adopted to reduce the dimensionality of the input dataset, saving computational time. Additionally, ViromeScan produced a better representation of the abundance of the mock communities when compared to the other methods (Figure 33). The better accuracy is probably due to bacterial and human reads not being filtered by the other approaches. By analyzing the assignment read by read, we deeply investigated how the non-filtering biases affected the performance of the other

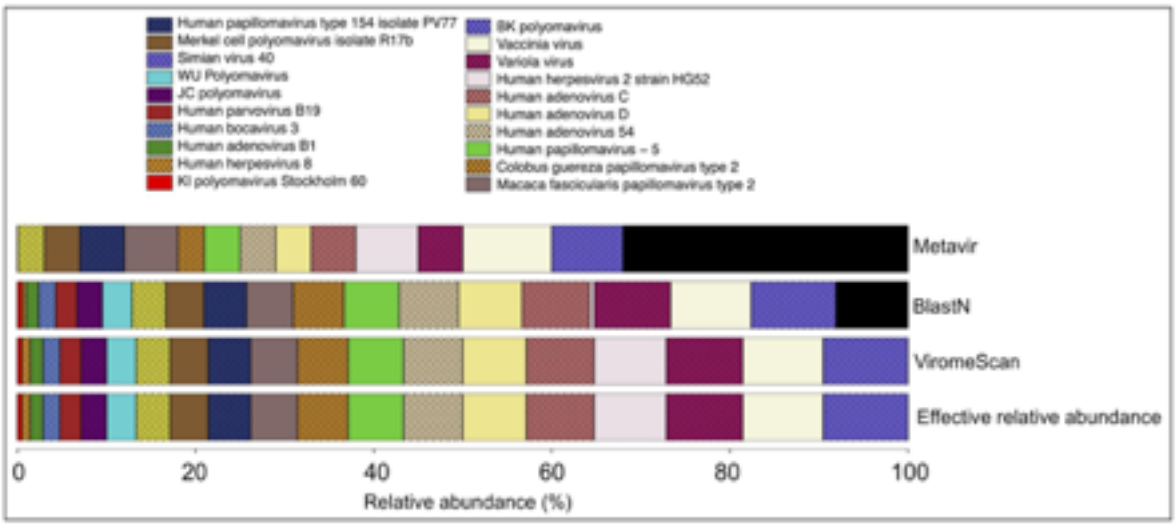


Figure 33 | Comparison between the relative abundances of a single non-evenly distributed mock community as detected using Metavir, blastN and ViromeScan, and its real composition. Black portions of the bars correspond to the unassigned viral fraction or erroneous viral assignment.

classification tools. Specifically, blastN failed to classify 50 % and 30 % of the reads belonging to Human herpesvirus 2 strain HG52 and Human adenovirus 54, respectively. Furthermore, it assigned to a different strain the majority of the reads of Human bocavirus 3 and Vaccinia virus. Analogously, Metavir failed to detect Human herpesvirus 2 strain HG52, Variola and Vaccinia virus, Human adenovirus C and D. Moreover, it assigned to a different species the reads for BK polyomavirus, and overestimated the reads for Parvoviridae and Polyomaviridae. In these cases, the superior accuracy of ViromeScan is probably due to the unique "two-step" assignment process in the pipeline, which involves two consecutive alignments of the reads to the reference database. The first one is computed at the very beginning of the analysis to detect viral candidate reads. The second one is computed after the filtering processes, as validation and final assignment of the viral reads to the correct taxonomy.

Notably, the "two-step" method is not used in the other existing tools. This uniqueness makes ViromeScan a very efficient tool in saving computational time, because it immediately skims the input reads, and at the same time permits a more accurate assignment of the viral sequences. Finally, by removing from the database the reference genomes closely related to those included in the mock communities, we evaluated the potential for viral discovery of the tool. According to our findings, ViromeScan was able to identify the correct genus of the Human adenovirus and Human papillomavirus species when their closest genome sequences were removed from the database, but it did not

assign any human DNA virus when all the related genomes up to family level were deleted. For these reasons, ViromeScan cannot be used as a classifier of viruses belonging to lineages that are completely missing in the database.

ViromeScan was next used to characterize the virome of metagenome samples from different body niches of people enrolled in the HMP[244], analyzing a total of 20 samples belonging to four human body sites: stool (representative of the gut ecosystem), mid vagina, buccal mucosa and retroauricular crease. ViromeScan detected 207 viral species from 22 viral families with abundance ≥ 0.5 % in at least one sample. The body site that showed the highest diversity was the retroauricular crease with 98 ± 10 (mean of viral species at ≥ 0.5 % ± sem), followed by gut (85 ± 3), buccal mucosa (48 ± 6), and vagina (42 ± 4). Looking at the genus-level diversity, we found a mean of 5.2 genera per sample, consistent with that detected in a previous study on 102 HMP samples (5.5 genera per sample)[282]. Thus, we investigated the hypothesis that different body sites reflect different virome profiles at family and species level through hierarchical clustering of the 20 samples (Figure 34). Interestingly, the gut virome was consistently different from that of the other body sites ($P < 0.05$, Fisher's exact test). In particular, it was characterized by *Geminiviridae*, *Phycodnaviridae*, *Asfarviridae, Iridoviridae, Mimiviridae, Adenoviridae, Nimaviridae, Baculoviridae, Anelloviridae, Nudiviridae, Marseilleviridae, Malacoherpesviridae, Parvoviridae, Circoviridae, Nanoviridae* and *Poxviridae* viral families. On the other hand, the other body sites shared some families, such as Polydnaviridae, *Herpesviridae, Polyomaviridae, Alloherpesviridae, Ascoviridae* and *Papillomaviridae* ($P < 0.05$, Fisher's exact test). The differences are also displayed in terms of relative abundance in the histograms and pie charts of Figure 35(A). *Mimiviridae* and *Poxviridae* dominated the human gut eukaryotic virome, while *Herpesviridae* and *Polydnaviridae* were the most represented viral families in the other body sites. Notably, the relative abundances of HMP samples as determined by ViromeScan, were consistent with the results obtained by applying blastN (data not shown), and the viral taxa identified confirm the little available literature. In particular, *Papillomaviridae*, *Herpesviridae*, and *Polyomaviridae* have already been detected in the microbiota of vagina, skin and mouth[251], and *Adenoviridae*,

282 Wylie KM, Mihindukulasuriya KA, Zhou Y, Sodergren E, Storch GA, Weinstock GM. Metagenomic analysis of double-stranded DNA viruses in healthy adults. BMC Biol. 2014;12:71.

Anelloviridae and *Circoviridae* in stool[251]. Additionally, our findings on the gut samples led to the detection of Megavirales and other giant viruses that were not found in previous analyses of the human gut virome, probably due to the filtering approach used for virus isolation[283-284], but have recently been isolated in human stool and other human samples through different approaches[285]. Taken together, all these data confirm the applicability of ViromeScan to microbial communities and its suitability to metagenomic samples.

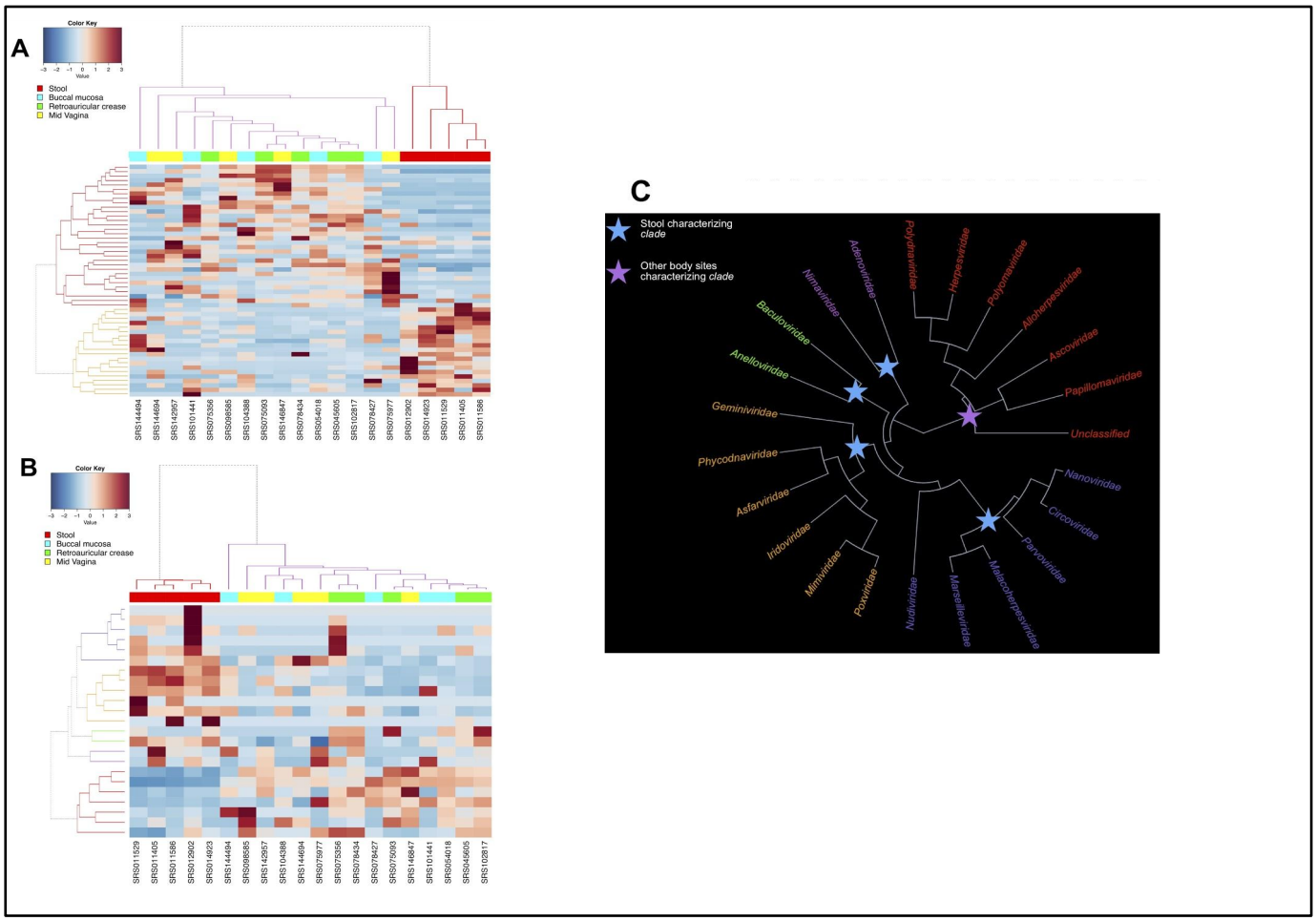


Figure 34 | Different body sites reflect different virome configurations. Species (**A**) and families (**B**) level hierarchical Ward-linkage clustering based on the Spearman correlation coefficients of the viral profiles of 20 HMP samples as determined by ViromeScan. Analysis was carried out considering all the families detected and species with at least 0.5 % of abundance in 25 % of samples. (**C**) Hierarchical Ward-linkage clustering of viral families generated characteristic clades, which discriminated the gut environment from the other body sites. The names of the families are colored according to the colors of the dendrogram (**B**)

283 Holtz LR, Cao S, Zhao G, Bauer IK, Denno DM, Klein EJ, et al. Geographic variation in the eukaryotic virome of human diarrhea. Virology. 2014;468–470:556–64.

284 Reyes A, Blanton LV, Cao S, Zhao G, Manary M, Trehan I, et al. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. Proc Natl Acad Sci U S A. 2015;112(38):11941–6.

285 Saadi H, Reteno DG, Colson P, Aherfi S, Minodier P, et al. Shan virus: a new mimivirus isolated from the stool of a Tunisian patient with pneumonia. Intervirology. 2013;56(6):424–9.

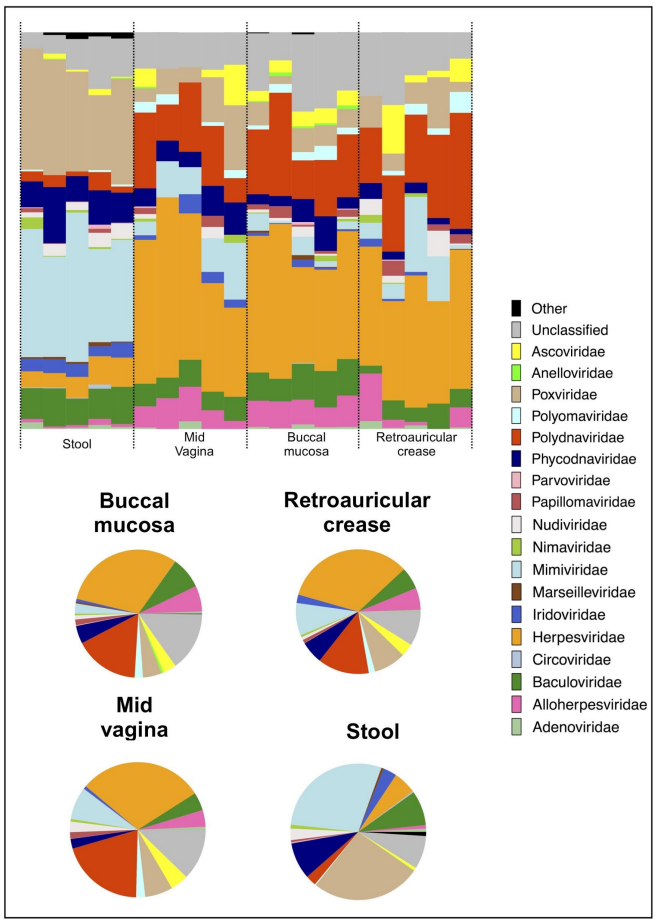


Figure 35 | The eukaryotic virome at family level in an asymptomatic Western population, as predicted by ViromeScan. Analysis was carried out on 20 HMP samples from 4 human body sites, including gut (stool), mouth (buccal mucosa), skin (retroauricular crease) and vagina (mid vagina). The relative abundance of viral families for each HMP sample and the mean relative abundance for each body site are reported in the histograms and pie charts, respectively.

Conclusion

ViromeScan provides new perspectives in the virome characterization analysis to end-users. Shotgun metagenomics and RNAseq techniques are rapidly decreasing in cost and already supply a community-wide profiling of the bacterial, archeal and eukaryotic microbiome. By enabling an efficient detection of the viral counterpart from shotgun sequencing, ViromeScan extensively integrates the analysis of the microorganisms that inhabit the human body. Furthermore, the pipeline can be applied to any environment as a tool for taxonomic profiling of the virome with resolution up to species level. An interesting and flexible aspect for users is that the pipeline of analysis can also be used with a customized database containing viral genomes of interest. However, this version of the tool remains blind to new viruses, which are not present in the database.

Section 6.2 - Characterization of the human DNA gut virome across populations with different subsistence strategies and geographical origin

Introduction

Humans intestinal virome consists of more than 109 viral particles per gram of feces[286]. This hidden part of our microbiota includes viruses infecting each domain of life (Bacteria, Archaea and Eukarya), including the human host. However, a great portion of this viral counterpart remains yet to be identified[287]. Studies of the human virome are just at the beginning, and only recently the retrieval of viral sequences in large sequence data sets has been possible thanks to bioinformatics tools, which can identify viral sequences within metagenomes[288]. Comparative studies of the gut microbiome between unindustrialized rural and hunter–gatherer communities from Africa and South America, and industrialized western populations from Europe and North America have revealed specific ecosystem adaptations to their respective lifestyles[50 -157-289]. It is here hypothesized an analogous variation in the human gut virome profile, possibly reflecting a unique response to the kind of subsistence. To date, only two studies have explored the human intestinal virome from different populations using next–generation sequencing approaches: a first study by Holtz and colleagues[290] investigated the intestinal eukaryotic virome of Australian rural and urban children from two different locations, with acute diarrhea. In a second study, Reyes and colleagues [291] characterized the gut virome from infants and toddlers up to 30 months of age from Malawi, with a focus on twins discordant for severe acute malnutrition. The main findings of these studies highlighted

---

286 Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013) Rapid evolution of the human gut virome. *Proc Natl Acad Sci USA* 110: 12450–12455.

287 Paez–Espino, D., Eloe–Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., *et al.* (2016) Uncovering earth's virome. *Nature* 536: 425–430.

288 Rampelli, S., Soverini, M., Turroni, S., Quercia, S., Biagi, E., Brigidi, P., and Candela, M. (2016) ViromeScan: a new tool for metagenomic viral community profiling. *BMC Genomics* 17: 165.

289 Gomez, A., Petrzelkova, K.J., Burns, M.B., Yeoman, C.J., Amato, K.R., Vlckova, K., *et al.* (2016) Gut microbiome of coexisting BaAka Pygmies and Bantu reflects gradients of traditional subsistence patterns. *Cell Rep* 14: 2142–2153.

290 Holtz, L.R., Cao, S., Zhao, G., Bauer, I.K., Denno, D.M., Klein, E.J., *et al.* (2014) Geographic variation in the eukaryotic virome of human diarrhea. *Virology* 468–470: 556–564.

291 Reyes, A., Blanton, L.V., Cao, S., Zhao, G., Manary, M., Trehan, I., *et al.* (2015) Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc Natl Acad Sci USA* 112: 11941–11946.

the presence of discriminatory viruses in malnourished children and in children with enteric diseases, compared to healthy controls, but compositional virome differences between healthy subjects related to lifestyle and community–based differences of sampled populations were not explored.

To fill these gaps in knowledge, it is here presented an analysis of the human DNA virome variation across the gut microbiome of human populations with different subsistence styles. To this aim, publicly available shotgun metagenomics sequences derived from two previous studies[51 - 107] were used, including five populations with different lifestyles: (I) the Hadza, traditional hunter–gatherers from Tanzania; (II) urban western residents of Bologna, Italy; (III) the Matses, hunter–gatherer population from the Peruvian Amazon; (IV) the Tunapuco, a rural agriculture community from Peru and (V) urbanized US people from Norman, Oklahoma.

Methods

- Sample collection and shotgun sequencing

The Illumina paired–end reads of the 96 metagenomes used in this study were previously generated by shotgun sequencing[51-107], and are publicly available at the National Center for Biotechnology Information – Sequence Read Archive (NCBI SRA database under the Bioproject ID PRJNA268964 and PRJNA278393).

- Bioinformatics and statistics

Raw sequences were processed using the ViromeScan software, that allows the user to taxonomically characterize the virome directly from metagenomic reads, denoising samples from reads of other microorganisms. In particular, was selected the option '–d human_DNA' to detect only DNA viruses that had the human being as a natural host. The relative abundance outputs of the software were used to perform statistical analyses and comparisons among samples. Bar plots and box plots were obtained and plotted by the graphics package of R (R version 3.1.3). Alpha–diversity was computed for each sample considering the number of viral species detected within each metagenome. Significance testing was performed using the R package stats. When appropriate, P values were adjusted for multiple comparisons using the Benjamini–Hochberg correction. A false

discovery rate < 0.05 was considered as statistically significant. Random Forests[121] was computed in R environment using the package 'randomForest'.

- Co−occurrence network analysis

The analysis was carried out using R (packages stats, made4 and vegan) and Cytoscape software. The co−occurrence between each pair of viral species was evaluated as percentage of subjects that showed both species at more than 0.5% relative abundance and displayed by hierarchical Ward−linkage clustering based on Spearman correlation coefficients. The results obtained with the entire pool of samples were used to define the CoOGs. Permutational multivariate analysis of variance was used to determine whether CoOGs were significantly different from each other.

- Detection of envelope glycoproteins through bioinformatics approach

Metagenomic sequences were aligned, using the bowtie2 tool[243], to the Virus−Host database[292] that contains 387 611 viral genes, with assigned function and taxonomy. Hits for envelope glycoproteins were picked up and their taxonomy was manually verified on the NCBI database. The results were normalized on the total amount of viral metagenomic reads. The mean and standard error of the mean were computed in R using the stats package. Wilcoxon test was used to assess the significance of differences among population groups.

Results and discussion

The analysis was performed using ViromeScan, a newly developed tool with a read−mapping approach, which uses a reference database containing type strain viral genomes, while remaining blind to unknown viruses, unrelated to those present within the database[257]. Results provided some glimpses into the human DNA virus specificity in populations with different lifestyles, highlighting peculiarities in the composition and prevalence of our viral inhabitants, which are possibly attributable to the specific pattern of environmental and community exposure. Taxonomic analysis of the metagenomic reads indicated *Herpesviridae, Anelloviridae, Adenoviridae, Papillomaviridae, Parvoviridae* and *Polyomaviridae* as the main families of the human intestinal virome (Figure 36). However,

292 Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., et al. (2016) Linking virus genomes with host taxonomy. Viruses 8: 66.

it should be pointed out that reads with taxonomic assignment to human DNA viruses were only a minority of the total (roughly tens of reads per million of metagenomic sequences). For this reason, was assumed that the analysis was blind to low abundant viruses, and that the viral traces detected were representative of the dominant species. In particular, *Herpesviridae* was the most represented family among samples (mean relative abundance, 54%), followed by *Anelloviridae* (13%), *Papillomaviridae* (10%) and *Adenoviridae* (10%). The core virome profiles observed are consistent with previous studies, which detected such viral families within fecal samples from healthy individuals[293-294-295], underscoring the notion that virus interactions with the host cannot be solely envisioned as pathogenic[233]. We hypothesize that these viruses are found in the feces mainly as a result of mucosal and epithelial exfoliation[296]. Indeed, the diffusion of *Torque teno* viruses in healthy human populations had already been observed as proven by their prevalence in serum samples of people of wide geographical origin[297-298-299].

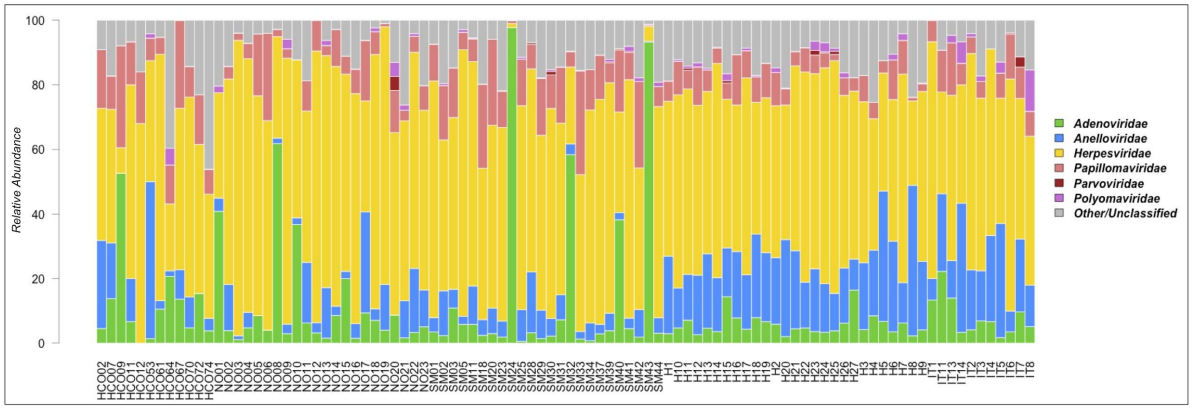


Figure 36 | The human DNA virome at family level in the gut microbiome of Tunapuco, Matses, Hadza, Italian and US people. Histograms show the relative abundances of viral families for each sample. Samples are named as reported in the original studies. HCO code for Tunapuco; NO for US people; SM for Matses; H for Hadza; IT for Italian people.

293 Wylie, K.M., Mihindukulasuriya, K.A., Zhou, Y., Sodergren, E., Storch, G.A., and Weinstock, G.M. (2014) Metagenomic analysis of double-stranded DNA viruses in healthy adults. *BMC Biol* 12: 71.

294 Di Bonito, P., Della Libera, S., Petricca, S., Iaconelli, M., Sanguinetti, M., Graffeo, R., *et al.* (2015) A large spectrum of alpha and beta papillomaviruses are detected in human stool samples. *J Gen Virol* 96: 607–613.

295 Vetter, M.R., Staggemeier, R., Dalla Vecchia, A., Henzel, A., Rigotto, C., and Spilki, F.R. (2015) Seasonal variation on the presence of adenoviruses in stools from non-diarrheic patients. *Braz J Microbiol* 46: 749–752.

296 Foxman, E.F., and Iwasaki, A. (2011) Genome-virome interactions: examining the role of common viral infections in complex disease. *Nat Rev Microbiol* 9: 254–264.

297 Takahashi, K., Hoshino, H., Ohta, Y., Yoshida, N., and Mishiro, S. (1998) Very high prevalence of TT virus (TTV) infection in general population of Japan revealed by a new set of PCR primers. *Hepatol Res* 12: 233–239.

298 Huang, L.Y., Jonassen, T.O., Hungnes, O., and Grinde, B. (2001) High prevalence of TT virus-related DNA (90%) and diverse viral genotypes in Norwegian blood donors. *J Med Virol* 64: 381–386.

299 Stelekati, E., and Wherry, E.J. (2012) Chronic bystander infections and immunity to unrelated antigens. *Cell Host Microbe* 12: 458–469.

According to findings, besides a shared gut virome fraction, well–defined lifestyle and population–associated virome structural peculiarities were evident. In particular, hunter–gatherers showed as discriminatory species *Human herpesvirus 7, Torque teno midi virus 1, Torque teno midi virus 2* and *Betapapillomavirus 3*, while *Betapapillomavirus NC015692.1, Human papillomavirus type 178* and *Human herpesvirus 8* were peculiar for rural Tunapuco, urban Italian and US people respectively. Well matching the gut microbiome behaviour, the Hadza and Mates hunter–gatherers were characterized by a higher gut virome diversity compared to urban US and Italians (Figure 37) .

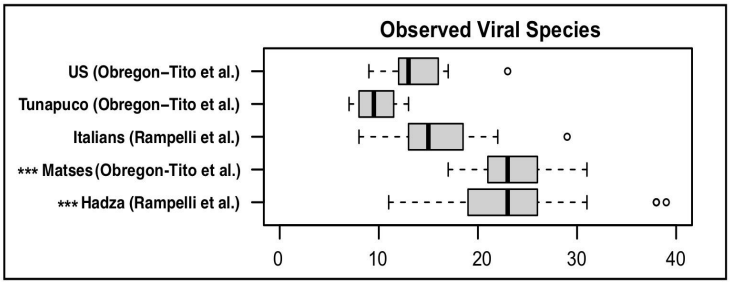


Figure 37 | The DNA virome biodiversity within the Matses, Hadza, Tunapuco, US and Italian gut metagenomes. Hadza and Matses show a higher level of alpha–diversity in terms of observed viral species compared to the other populations ($P < 0.001$, Wilcoxon test).

To identify patterns of viral community variation among Hadza, Matses, Tunapuco and western individuals, Italians and US people, we determined co–occurrence associations, meaning the frequency of concomitant detection of two taxa, between viral species and then clustered them according to the co–occurrence profile for the human gut microbiome. This analysis resulted in the identification of three co–occurrence groups (CoOGs), whose variation involved virome differences among the studied populations ($P < 0.001$, permutational test with pseudo F–ratio). According to results, the *Human herpesvirus 2* CoOG defined the core virome, as it included highly prevalent species in each population. Interestingly, the viral species included in this CoOG were also the most abundant across all samples, and were arranged in population–specific subsets, meaning that some specific viral taxa of this CoOG were present in more than 90% of the subjects of a given population (Figure 38). This could suggest the establishment of a population–specific persistent relationship with the human host.

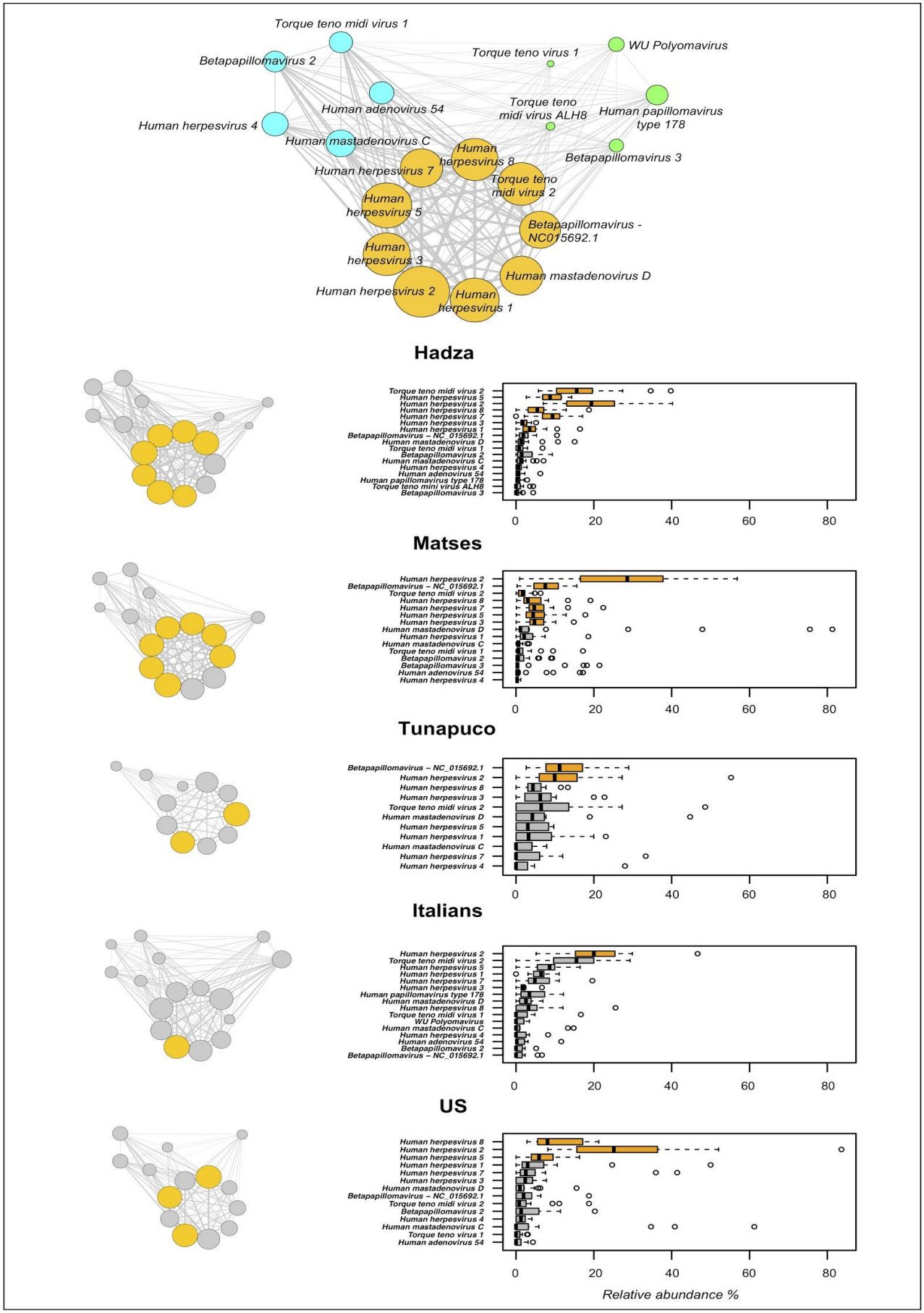


Figure 38 | Network plots describing the co-occurrence and prevalence of viral species in the gut microbiota of all samples (top) and of each population (bottom-left). Viral species with at least 0.5% relative abundance in at least 30% of the samples in each group were plotted, with the exception of the network plot including all samples (top) for which the species present in at least one of the other networks were plotted. CoOGs were named according to the dominant species as follows: *Human herpesvirus 2* (orange), *Human mastadenovirus C* (slate blue), *Betapapillomavirus 3*(green).

Notably, Tunapuco, Italian and US western individuals showed a lower number of high–prevalence virus species (Tunapuco = 2, Italians = 1, US = 3) compared to the two hunter–gatherer groups (Hadza = 7, Matses = 7), while *Human herpesvirus 2* was endemic in all populations (mean rel. ab., 24%). Remarkably, Hadza and Matses hunter–gatherers shared several high–prevalence viral species within the *Human herpesvirus 2* CoOG. In particular, *Human herpesvirus 2, Human herpesvirus 3, Human herpesvirus 5, Human herpesvirus 7, Human herpesvirus 8* and *Torque teno midi virus 2* were present in both populations (prevalence ranging from 92% to 100%), while *Human herpesvirus 1* was characteristic of the Hadza and *Betapapillomavirus* NC015692.1 of the Matses. It is important to note that the high–prevalence viral taxa within this CoOG corresponded to the species with the highest relative abundances within each group. Different subsistence strategies can result in a different exposure to environmental contamination, and risk of virus transmission, explaining, at least in part, some of the observed gut virome differences between human populations. For instance, the higher prevalence of *Torque teno midi* viruses in the Hadza and Matses may result from the habitual consumption of contaminated water. Indeed, in a study comparing the virome population between wastewater, stormwater, surface water, groundwater and drinking water samples from US, Italy and Australia, viral markers including *Torque teno virus*, *Adenovirus* and *Polyomavirus*, were more frequently found in environmental samples (occurrence frequency, 25%–100%), compared to waters used as drinking water sources (5%)[300] . Furthermore, another study reported high detection of *Polyomavirus* and *Papillomavirus* in water environments[301], supporting the hypothesis of a continuous inoculum of a wide range of environmental viruses in people that consume unpurified water. This aspect is particularly relevant in the context of hunter–gatherer populations, because they are fully enveloped within their natural environment in a way that is no longer possible in the western world[50 - 51]. Conversely, the viral contamination of food resources may also feed the human gut virome[302], which is particularly true for hunter–gatherer individuals that do not consume industrial food or use sterile cleaners.

---

300 Charest, A.J., Plummer, J.D., Long, S.C., Carducci, A., Verani, M., and Sidhu, J.P. (2015) Global occurrence of torque teno virus in water systems. *J Water Health* 13: 777–789.

301 Fratini, M., Di Bonito, P., and La Rosa, G.(2014) Oncogenic Papillomavirus and Polyomavirus in water environments: is there a potential for waterborne transmission?*Food Environ Virol* 6: 1–12.

302 Rzezutka, A., and Cook, N. (2004) Survival of human enteric viruses in the environment and food. *FEMS Microbiol Rev* 28: 441–453.

Supporting this hypothesis, *Papillomavirus* and *Herpesvirus*, which were found to be the most prevalent and abundant viruses in the gut microbiome of Hadza, Matses and Tunapuco, have been shown to persist up to one week in the environment at room temperature[303-304], thus they might be able to survive on the surface of untreated fruit, tubers and vegetables that hunter–gatherers daily consume.

Conclusion

In summary, this work constitutes a first important step in the characterization of the viral metacommunity within the fecal microbiome of human populations with different lifestyles. Findings highlight the presence of a complex viral community that shows both general and population–specific features. Even if genetic evidence of a potential pathogenic behavior was found in the DNA virome from all the study populations, the data also raise the question of the biological importance of the gut virome in human physiology and the possible role of our virome counterpart as a co–evolutionary partner. Further studies are necessary to identify the sources of the viral nucleic acids detected in stools, for example by sampling and characterizing viruses within gut tissues, blood and food, together with fecal samples. Particular attention will be devoted to shed light on the mechanisms of interaction between the virome and a healthy human host, and the resulting impact on the host immunological functions. Finally, it will be necessary to increase the number of subjects and populations to validate the findings of this work with greater statistical power.

303 Roden, R.B., Lowy, D.R., and Schiller, J.T.(1997) Papillomavirus is resistant to desiccation. *J Infect Dis* 176: 1076–1079.

304 Pesaro, F., Sorg, I., and Metzler, A. (1995) In situ inactivation of animal viruses and a coliphage in nonaerated liquid and semiliquid animal wastes. *Appl Environ Microbiol* 61: 92–97.

CHAPTER 7 - Fungi

Introduction and relation with other organisms

Fungi represent a high-diversified clade of eukaryotes, being found in almost every environment in our planet[305]. The first onset of Eukaryotic life on the land was indeed represented by a symbiosis between a fungus and a phototrophic organism[306], a mutualistic relationship that lasts also in our days in lichens. The colonization of land by the fungal lineage took place about 600 million years ago, in Late Precambrian period[307], but molecular clocks research fore-dated this event earlier in Precambrian period[308]. In spite of their 'late' appearance in Earth's evolutionary history fungi deeply impacted the evolutionary history of our planet. Their biological activity is indeed connected to the 'Snowball Earth' event, an extended glaciation event dated 800-750 million years ago[309], and in the Neoproterozoic rise in oxygen, a key-event for the consequent Cambrian explosion of animals[310]. Indeed, fungal activity can impact weathering[279], leading to lower CO_2 levels and air temperatures[311].

Fungi plays also a key-role in the nutrient cycling, acting as pillars in bio-geo-chemicals cycles of crucial mineral elements in the biosphere[312], allowing the conservation and perpetration of life on our planet. The 'Mycobiome', referring primarily to the fungal biota inside a determined environment, is also an important component of vertebrate's microbiome, populating several ecological niches inside and outside the organism[313]. Despite its involvement in Human diseases is proven since the 19th century[314], the

---

305 Richards, T. A., Leonard, G. & Wideman, J. G.(2017). What defines the "Kingdom" Fungi?*Microbiology Spectrum* 5, 1–21.

306 K. A. Pirozynski, D. W. Malloch, *Biosystems* 6, 153 (1975).

307 M. L. Berbee, J. W. Taylor, in *The Mycota*, vol. VIIB, *Systematics and Evolution*, D. J. McLaughlin, E. McLaughlin, Eds. (Springer-Verlag, New York, 2000), pp. 229–246

308 Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB. Molecular evidence for the early colonization of land by fungi and plants. Science. 2001 Aug 10;293(5532):1129-33.

309 P. F. Hoffman, A. J. Kaufman, G. P. Halverson, D. P. Schrag , *Science* 281, 1342 (1998).

310 A. H. Knoll, *Science* 256, 622 (1992).

311 T. J. Crowley, R. A. Berner, *Science* 292, 870 (2001).

312 Tedersoo L, Sánchez-Ramírez S, Kõljalg U, Bahram M, Döring M, Schigel D, et al. High-level classification of the Fungi and a tool for evolutionary ecological analyses. Fungal Divers. 2018;90:135–59.

313 Cui L, Morris A, Ghedin E. The human mycobiome in health and disease. *Genome Med.* 2013;5(7):63. Published 2013 Jul 30. doi:10.1186/gm467

314 Hassall A. On the development of Torulæ in the urine, and on the relation of these fungi to albuminous and saccharine urine. Med Chir Trans. 1853;5:23–78.

Mycobiome role in the holobiont still remains partly unclear, starting its exploitation in recent years[315].

The study of fungi

The mycological study started more than 160 years ago, with the first published paper released in 1853[314] by the British Arthur Hill Hassall. The first mycological studies were essentially observative, translating in the 1920's to cultural approaches based on the growth of the fungi in sterile liquid media[316]. Experimental results were evaluated basing on the growth structure and assessing their composition within the media. Culturing methods have been largely improved over the decades but, like for the vast majority of bacteria, scientists and technology are not able to reproduce the optimal conditions for culturing the most requiring and stringent components of the Mycobiome[317]. The introduction of the culture-indipendent high-throughput sequencing allowed nowadays to start bridging this gap in knowledge. Taking the gut niche as an example, comparative studies reported that culture-independent methods identified 37 different fungal groups compared to only 5 species found by culture-dependent analyses[318]. Despite the rise of this new genomic era in the Mycobiome field, a lack in metagenomic data exploitation tool is still present.

315 Sam QH, Chang MW, Chai LY. The Fungal Mycobiome and Its Interaction with Gut Bacteria in the Host. *Int J Mol Sci*. 2017;18(2):330. Published 2017 Feb 4. doi:10.3390/ijms18020330

316 Marloth RH. An apparatus for the study of matforming fungi in culture solutions. Science. 1929;5:524–525.

317 Beck JM, Young VB, Huffnagle GB. The microbiome of the lung. Transl Res. 2012;5:258–266.

318 Chen Y, Chen Z, Guo R, Chen N, Lu H, Huang S, Wang J, Li L. Correlation between gastrointestinal fungi and varying degrees of chronic hepatitis B virus infection. Diagn Micr Infec Dis. 2011;5:492–498.

CHAPTER 8 - New insights in Mycobiome characterization

In the attempt to produce a new tool useful for the Mycobiome characterisation, here I present HumanMycobyomeScan, a modular program devoted at the extraction and assignation of fungal reads originated from metagenomic surveys.

Section 8.1 - HumanMycobiomeScan: a new bioinformatics tool for the characterization of the fungal fraction in metagenomic samples

Introduction

The characterization of the Mycobiome structure can be done using both culture-dependent and independent methods[319]. Culture-dependent techniques, which generally combine methods such as microscopy[320], biochemical assays[321] and growth on selective media[322], represent a classical approach for the profiling of complex microbial ecosystems, and have the great advantage of allowing the determination of the viable fraction of the mycobiome. However, this is a time-consuming approach and, most importantly, blind to species that are obligate symbionts or have complex nutritional requirements or that are otherwise hard or impossible to raise in culture[323]. On the other hand, culture-independent methods basically rely on the amplification and sequencing of ITS (Internal Transcribed Spacer) or 18S rDNA phylogenetic markers[274], or on multi-gene metabarcoding[324], followed by dedicated bioinformatics pipelines for the inference of the community structure, such as QIIME[28] , CloVR-ITS[325], UPARSE[326], CONSTAX[327] and

319 Huseyin CE, O'Toole PW, Cotter PD, Scanlan PD. Forgotten fungi-the gut mycobiome in human health and disease. FEMS Microbiol Rev. 2017;41(4):479–511.

320 de Repentigny L, Phaneuf M, Mathieu LG. Gastrointestinal colonization and systemic dissemination by *Candida albicans* and *Candida tropicalis* in intact and immunocompromised mice. Infect Immun. 1992;60:4907–14.

321 Khatib R, Riederer KM, Ramanathan J, Baran J Jr. Fecal fungal fora in healthy volunteers and in patients. Mycoses. 2001;44:151–6.

322 Ouanes A, Kouais A, Marouen S, Sahnoun M, Jemli B, Gargouri S. Contribution of the chromogenic medium CHROMagar(®)Candida in mycological diagnosis of yeasts. J Mycol Med. 2013;23:237–41.

323 Hall RA, Noverr MC. Fungal interactions with the human host: exploring the spectrum of symbiosis. Curr Opin Microbiol. 2017;40:58–64.

324 Hebert PD, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proc Biol Sci. 2003;270(1512):313–21.

325 White JR, Maddox C, White O, Angiuoli SV, Fricke WF. CloVR-ITS: automated internal transcribed spacer amplicon sequence analysis pipeline for the characterization of fungal microbiota. Microbiome. 2013;1:6.

326 Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods. 2013;10:996–8.

327 Gdanetz K, Benucci GMN, Vande Pol N, Bonito G. CONSTAX: a tool for improved taxonomic resolution of environmental fungal ITS sequences. BMC Bioinformatics. 2017;18(1):538.

MICCA[328]. However, no gold standard approach for culture-independent mycobiome analysis has yet been developed, as highlighted by the variety of genomic regions and techniques used in different studies[329-330-331]. In this context, a pipeline specifically devoted to the characterization of the mycobiome based on metagenomic reads from whole genome sequencing of microbial communities is completely missing. In an attempt to bridge this gap, here is presented HumanMycobiomeScan, a new bioinformatics tool that taxonomically profiles the mycobiome within the original microbiome, requiring only a few minutes to process thousands of metagenomics reads. HumanMycobiomeScan works with shotgun reads to detect traces of fungal DNA and estimate the abundance profiles by filtering out human and bacterial sequences and mapping the remaining sequences onto a hierarchical fungal database. HumanMycobiomeScan is available at the website: http://sourceforge.net/projects/hmscan.

Methods

- Workflow of the software

HumanMycobiomeScan directly analyzes metagenomics reads to detect and extract fungal sequences without any pre-processing steps. Accepted input files are single- or paired-end reads in .fastq format (.bzip2, .gzip and .zip compressions are accepted as well) produced by shotgun sequencing. The HumanMycobiomeScan database is based on the complete fungal genomes available at the NCBI website (downloaded in February 2018). The database contains a total of 1213 entries, corresponding to 66 different fungal genomes (referred to as *Fungi_LITE* on the project website). A second database containing 38,000 entries (including "not complete" genome records), corresponding to 265 different fungal genomes, is available for download (referred to as Fungi_FULL), and can be obtained and formatted by following the instructions on the project web page

328 Albanese D, Fontana P, De Filippo C, Cavalieri D, Donati C. MICCA: a complete and accurate software for taxonomic profiling of metagenomic data. Sci Rep. 2015;5:9743.

329 Araujo R. Towards the genotyping of fungi: methods, benefits and challenges. Curr Fungal Infect Rep. 2014;8:203–10.

330 Tang J, Iliev ID, Brown J, Underhill DM, Funari VA. Mycobiome: approaches to analysis of intestinal fungi. J Immunol Methods. 2015;421:112–21.

331 Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L. Mycobiome diversity: high-throughput sequencing and identification of fungi. Nat Rev Microbiol. 2019;17(2):95–109.

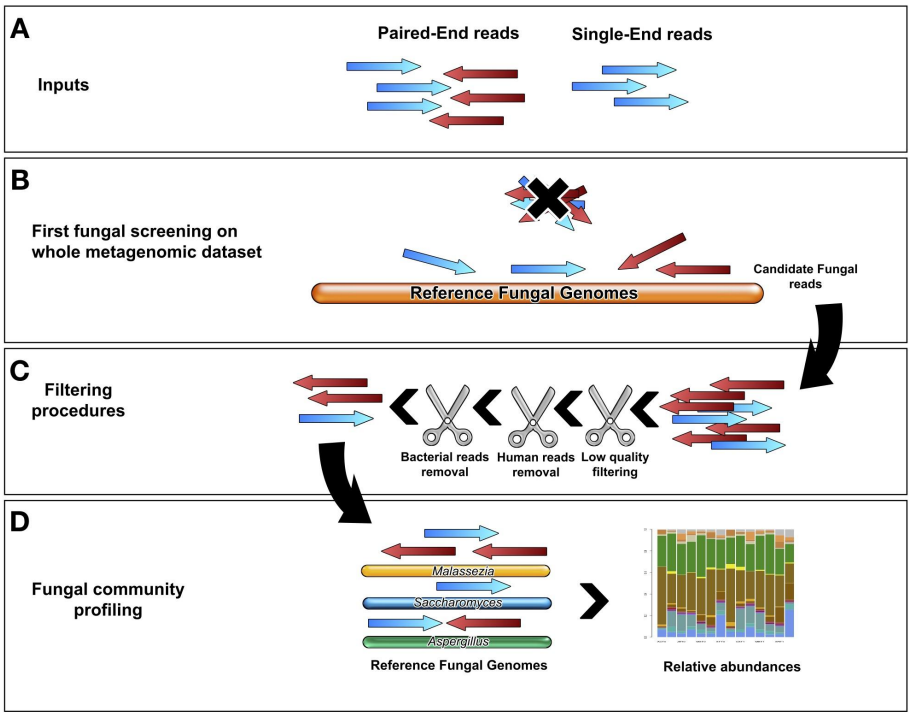


Figure 39 | Analysis workflow of HumanMycobiomeScan. **A** Inputs are single- (.fastq) or paired-end (.fastq or compressed .fastq) reads. **B** Candidate fungal reads are screened by mapping onto reference fungal genomes contained in a precompiled database. This allows for a first reduction of the sample size, lowering the number of sequences that will be subjected to further steps. **C** Three filtration steps are carried out to eliminate low quality reads as well as reads belonging to humans and bacteria. **D** The remaining sequences are realigned onto the fungal genome database for definitive taxonomic assignment of the reads. The results are tabulated as both abundance profiles and read counts, and represented by bar plots.

(https://sourceforge.net/projects/hmscan/). The schematic workflow of HumanMycobiomeScan is reported in Figure 39. In detail, metagenomic reads are aligned to the fungal genome database using bowtie2[39]. This first step is necessary to identify candidate fungal reads and reduce the sample size by filtering out sequences that do not match the reference database. It is important to note that performing this procedure at the beginning of the analysis allows for a significant decrease (~100X) in the time required for the subsequent parts of the pipeline. Afterwards, a quality-filtering step of putative fungal reads was implemented by modifying the processing procedure of the Human Microbiome Project (HMP)[244]. Briefly, sequences are trimmed for low quality scores (less than Q30) using a modified version of the script trimBWAstyle.pl directly on BAM files. Additionally, reads shorter than 60 bases are discarded. Since the input sequences may derive from human-associated samples, such as feces or tissues, it is plausible to expect a certain amount of contamination due to human and bacterial sequences. To remove these contaminations as accurately as possible, a double filtering step is performed using BMTagger[246].

BMTagger is a proficient tool capable of discriminating between human or bacterial and other reads by comparing short fragments of 18 bases (18-mers) originated from both the input sequences and the reference human or bacterial database. Specifically, we used the hg19 database for human sequences[332] and a custom bacterial database, also used for ViromeScan, including bacteria from human specimens and the archaeon that normally inhabits the human body and especially the intestine, i.e. *Methanobrevibacter smithii*[333]. The released version of HumanMycobiomeScan is thus functionally implemented to work with human-associated microbiota samples. Nevertheless, the databases can be customized by the user, making the program flexible and capable of working with datasets of various origin (e.g. mycobiomes associated with soil, water, air or other animals). As a final step of the workflow, filtered reads are matched again to the fungal database using bowtie2 for definitive taxonomic assignment. The taxonomic affiliation is deduced by matching the result of the taxonomic assignment with an annotated list of fungal species, containing the entire phylogenetic classification for each genome included in the database. At the end of the process, an additional pipeline step allows the user to normalize the results by the length of the references included in the database. The obtained relative abundance profiles and the normalized number of hits for each sample are reported in tab-delimited files, along with histograms representing the fungal community, generated using the 'base' and 'graphics' R packages. The fungal reads, as identified above, are also provided in a .fastq file.

- Validation of the tool and comparison with other existing methods

A synthetic sample containing 1 million random sequences was generated using the EMBOSS makenucseq utility and analysed to evaluate the HumanMycobiomeScan performance in avoiding the detection of false positives. Five additional mock communities composed of a set of 100-base reads were in silico generated. In particular, the latter contained a fungal fraction, consisting of 20 different species of varying abundance, 5 bacteria and the human genome, to simulate real metagenomes. The performance of HumanMycobiomeScan in correctly profiling the fungal community was compared with that of other available tools (i.e. the web-interfaces blastn[51] and MG-

332 Genome Reference Consortium Human Build 37 (GRCh37), hg19. https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/.

333 Rampelli S, Soverini M, Turroni S, Quercia S, Biagi E, Brigidi P, et al. ViromeScan: a new tool for metagenomic viral community profiling. BMC Genomics. 2016;17:165.

RAST). An evaluation dataset can be downloaded together with the tool at the project web page (https://sourceforge.net/projects/hmscan/).

- Case study: using the tool to profile the gut mycobiome of hunter-gatherers and Western subjects

Thirty-eight stool metagenomes from Rampelli et al[88]. , including 11 metagenomes from Italian adults and 27 from the Hadza hunter-gatherers, were downloaded from the Sequence Read Archive [NCBI SRA; SRP056480, Bioproject ID PRJNA278393] and used to illustrate the performance and results of HumanMycobiomeScan. These metagenomes had been sequenced using the Illumina GAIIx platform, obtaining 0.9 Gbp of 2 × 100 bp paired-end reads. The entire metagenomic dataset was used to explore differences in the composition of fungal communities between groups of individuals relying on different subsistence strategies. No ethics committee approval was required to perform the analysis included in this study.

Results and discussion

HumanMycobiomeScan was first applied to a synthetic sample containing random sequences to evaluate possible biases in the detection of false positives. As expected, no fungal hit was found but all sequences were filtered out in the first step of the procedure, when reads are screened against the database. Was then evaluated the performance of the tool in investigating the fungal composition of five mock communities simulating a human-associated metagenome (i.e. including fungi, bacteria and the human genome). HumanMycobiomeScan correctly identified the 20 fungal species within the synthetic communities and estimated their abundance at different taxonomic levels (average number of misassigned reads: at family level, 8.5 (0.8% of reads); at species level, 14.1 (1.34% of reads)). All the species contained in the mock communities were detected and 86% of the fungal ones were assigned within 1.5% deviation from the expected value with the best overall prediction (Pearson $r = 0.851$, species-level Pearson $P < 1 \times 10-07$) (Figure 2A-B). HumanMycobiomeScan was more accurate in profiling the mycobiome of synthetic metagenomes than other existing methods, with blastN showing the closest performance but being considerably slower (Figure 2C). In particular, HumanMycobiomeScan performed the characterization at 4.36 reads per second on a

standard single-processor, single-core system, which was several orders of magnitude faster than the other methods used for comparison. In addition, HumanMycobiomeScan showed a better prediction of fungal abundances (Figure 2D). I then analysed the results read by read, to understand how the approaches failed to assign the correct taxonomy. BlastN under- or over-estimated several fungal species, completely failed to detect 12 species (*Cryptococcus neoformans, Aspergillus fumigatus, Fusarium verticillioides, Komagataella phaffii, Saccharomyces arboricola, Candida albicans, Saccharomyces eubayanus, Magnaporthe oryzae, Saccharomyces kluyveri, Neurospora crassa, Encephalitozoon romaleae and Sporisorium scitamineum*), and assigned some reads to species that were not actually present in the mock community. The performance of MG-RAST was even more inaccurate, with nine reads out of 10 assigned to species not present in the mock samples. The greater accuracy of HumanMycobiomeScan and its computational speed in the assignment are probably due to the "two-step" process of the pipeline, which consists of two consecutive alignments of the reads to the reference database. The first alignment is performed at the very beginning, to identify candidate reads that are likely to belong to the fungal fraction of the ecosystem. The second alignment is subsequent to the filtering steps, as a validation and final assignment of the reads to the correct fungal taxonomy. Notably, this "two-step" approach, including filtering processes for bacterial and human reads, is the same as that used for the software ViromeScan, but designed, tested and optimized for mycobiome characterization. HumanMycobiomeScan was also able to assign the correct genus to reads for species not present in the databases, meaning that the tool is able to assign reads to the correct phylogeny when a related reference (i.e. belonging to the same genus) is present in the database.

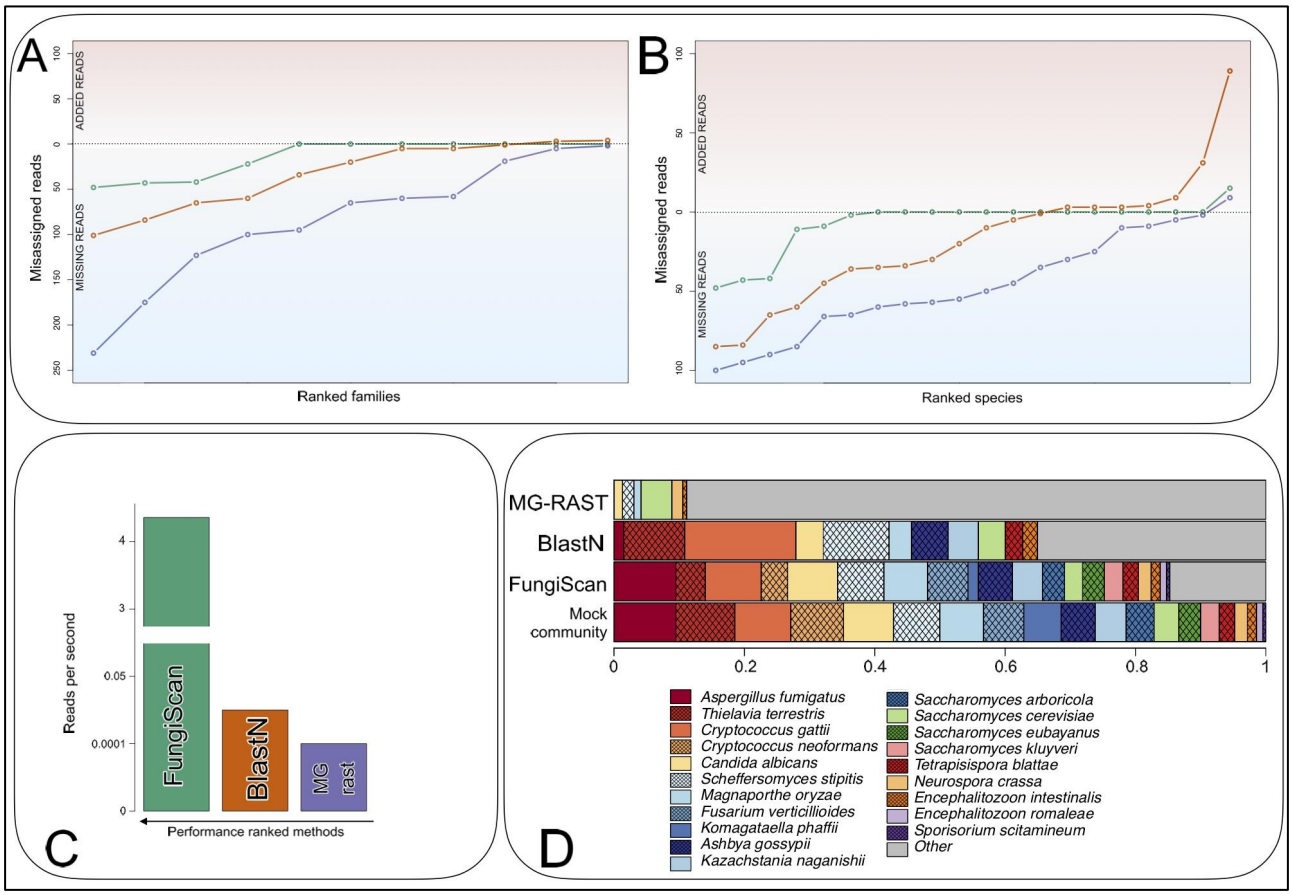


Figure 40 | Comparison of HumanMycobiomeScan with other existing assignment methods. Five synthetic fungal communities were used to compare HumanMycobiomeScan (HMS) with BlastN and MG-RAST. The actual number of misassigned reads, including those under- or over-assigned, is reported at family (**A**) and species (**B**) level. The horizontal line in the plots represents the "expected" value, meaning that all reads for a specific taxon were assigned to the correct reference genome. Points below or above the line indicate a lower or higher number of reads assigned to a specific taxon compared to the expected value. (**C**) The number of reads processed per second working on a single CPU is shown. (**D**) A comparison between the actual relative abundances of a mock community taken as an example and those reconstructed using the various methods of analysis was carried out. The gray portion represents the fraction of misassigned reads.

In the second part of the analysis, HumanMycobiomeScan was used to explore the gut mycobiome of 38 subjects adhering to different subsistence strategies: 27 Hadza hunter-gatherers from Tanzania and 11 Western individuals from Italy. One Hadza subject (H4) was excluded from statistical analysis and graphical representations as no fungal hits were retrieved from shotgun sequences. HumanMycobiomeScan characterized the fungal community at different phylogenetic levels, detecting a total of 19 families and 65 species. Hierarchical clustering, performed using the Spearman distance and the Ward linkage on the family-level relative abundance profiles of the samples, revealed two distinct groups ($p < 0.05$, Fisher's exact test) characterized by the dominance (relative abundance (rel. ab.) ≥ 30%) or not of the family Saccharomycetaceae (Figure 3 A-B). Interestingly, *Saccharomycetaceae* was almost the only fungal component detected in the feces of six subjects (rel. ab. > 90%). On the other hand, subjects with low abundance of

Saccharomycetaceae (rel. ab. < 30%) showed greater biodiversity, with the concomitant presence of several fungal families, such as Sclerotinaceae, Ustilaginaceae, *Hypocreaceae, Dipodascaceae* and *Schizosaccharomycetaceae.* In spite of the profoundly different lifestyles of Hadza and Italians, in terms of both diet and contact with the environment , no significant differences in taxon relative abundance were found between the two populations. Future studies on larger worldwide cohorts, possibly including subjects practicing varying subsistence strategies and/or diseased patients, are needed to unravel the biological role of the human mycobiome in health and disease.

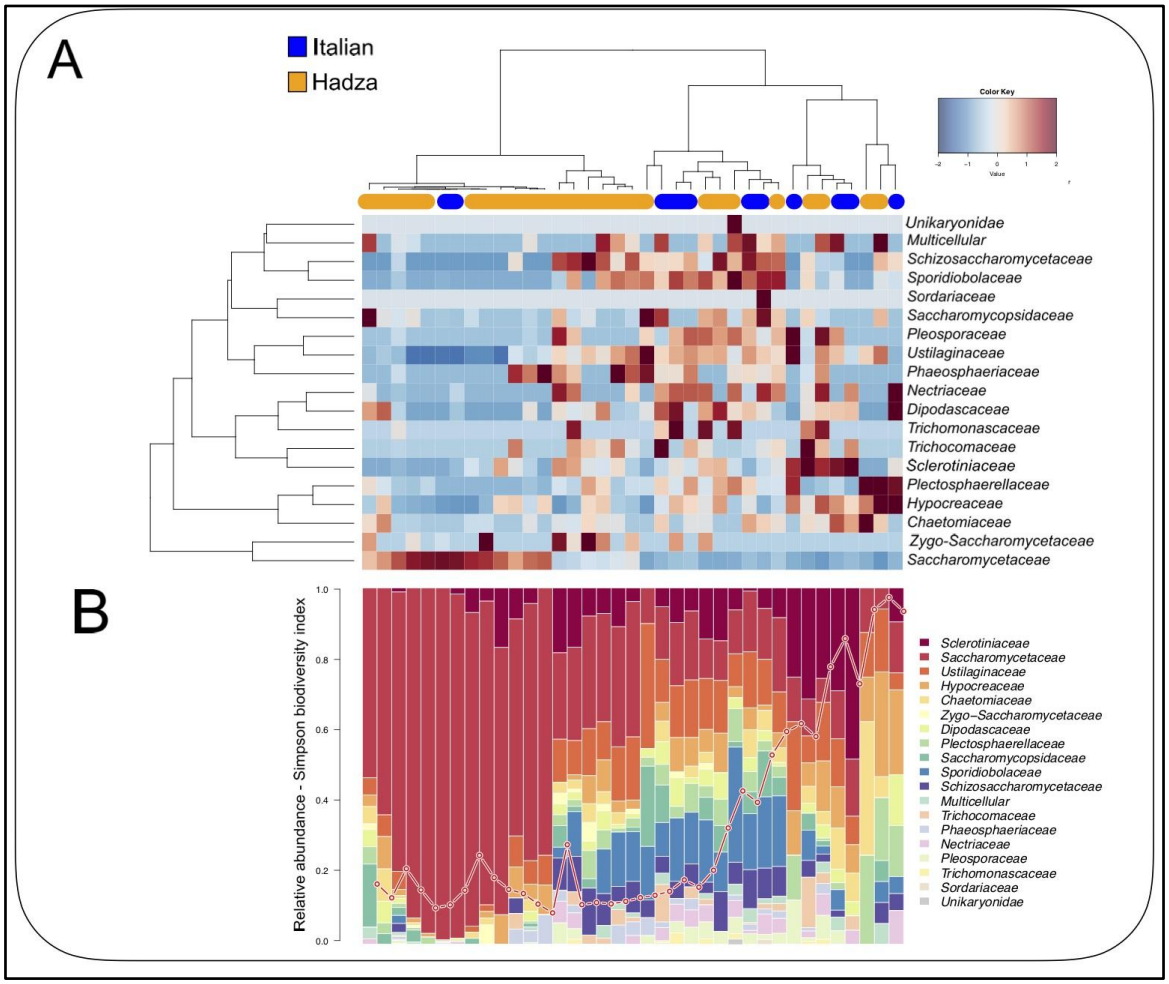


Figure 41 | Characterization of the fungal fraction of the gut microbiome of populations with different subsistence strategies. **A** Family-level hierarchical Ward-linkage clustering based on the Spearman correlation coefficients of the fungal profiles of 37 metagenomes, assigned using HumanMycobiomeScan. The study cohort includes 11 Italian subjects (in blue in the upper phylogenetic tree) and 26 Hadza hunter-gatherers from Tanzania (in orange). **B** The relative abundances of families are represented below the heatmap along with Simpson's biodiversity index for each subject (red line)

Conclusion

HumanMycobiomeScan opens to new possibilities in the metagenomics analysis of complex microbial ecosystems, extending in silico procedures to the characterization of the fungal component of microbiomes. By integrating the analysis with other tools already available to the scientific community, the user can profile the viral, bacterial and fungal counterpart of a microbial community using the same shotgun sequencing data, with a considerable gain in cost and time. Furthermore, such an integrated approach allows to obtain a more complete picture of the analysed microbiome, in terms of both microbial composition and richness of bacterial, viral and fungal sub-communities. A further advantage of HumanMycobiomeScan is the possibility of customizing the database by substituting or implementing the one supplied with the tool with fungal sequences of interest.

CHAPTER 9 – Overall conclusions

The presented dissertation aims to answer the prominent biological question about the possible involvement of the whole gut microbial community in the ecological and physiological processes of the holobiont. With this aim, I have explored the microbiome in different mammals and health conditions using several different bioinformatic approaches. In the studies reported in chapters 3 and 4 I focused my research on the role of the bacterial microbiome in the adaptive capacity of the host and its involvement in health and disease. To investigate these aspects, I used a combination of bioinformatic approaches, exploiting data produced both by marker-gene and metagenomic surveys. Applying taxonomic and functional assignment methods and elaborating the data obtained using biostatistics and multivariate ordinative or correlative methods, I obtained new correlative evidences about the possible involvement of the gut bacterial community in host fitness and health condition. Indeed my findings reinforced the evidences of the gut 'bacteriome' as a determinant factor in human holobiont adaptation, observing its taxonomical and metabolic shifts in the westernization process of the human species and its taxonomical and functional reconfigurations in other mammals who have, during evolutionary history, exploited new niches (like dolphins and naked-mole rats). The gut bacterial community has been deeply impacted by the westernization process, co-adapting along the trajectory of subsistence changes across human evolutionary history, from hunter-gatherers to rural agricultural to the most recent development of completely industrialized societies. In this context, humans have probably been subjected to a substitution of bacterial species in their gut, better tailoring the requirements imposed by new lifestyles and environments. The here-presented substitution of *Treponema* in spite of *Bifidobacterium* highlights the different metabolic requirements imposed by this shift: *Bifidobacterium* has a more flexible metabolic potential, better fitting the degradative tasks imposed by a typical Western diet generally poor in microbiota accessible carbohydrates, establishing a deeper interaction with the host, which may help support a continuous and abundant Bifidobacterial presence in adults, allowing this commensal to outcompete other opportunistic bacteria. The gut bacterial community plays a key role in eco-adaptive processes also in other mammalians, allowing the exploitation of new dietary niches and ecosystems. My work on dolphins give hints in this direction: even if

they are still retaining a multi-chambered foregut derived from their herbivorous terrestrial artiodactyls ancestors, these aquatic mammals possess a gut microbiota ecosystem similar to that of marine piscivores. This suggests the importance of the gut bacterial community as a dolphin adaptive partner, strategic for the occupation of new dietary niches in new environments. As a demonstration of the adaptive function to diet and environment provided by the gut microbial component, my work on Loggerhead sea turtles shows that this marine reptile shares more gut microbiota features with marine mammals (i.e., dolphins and seals) or other carnivorous terrestrial vertebrates, than with the phylogenetically close, but herbivorous, green turtles or other terrestrial tortoises. Similarly, naked mole-rats possess a peculiar gut microbiome composition, which appears to be the result of the host phylogeny and adaptation to its particular ecological niche. The bacterial microbiome layout I have depicted in this rodent has many compositional and functional peculiarities, suggesting the role of the gut microbiota as a universal contributor to mammalian health and fitness. Finally, findings seem to suggest a capacity of the naked mole-rat gut bacteria to use soil sulphate as a terminal electron acceptor to sustain an anaerobic oxidative metabolism in the gut, representing an unprecedented ecological equilibrium and giving a strong evidence of the importance of the gut microbiota in the eco-evolutionary processes.

The role and involvement of the gut microbial communities plasticity in health and disease has also been assessed in several studies, highlighting how microbiome reconfiguration plays a role in adaptive and maladaptive processes among and inside different human populations. In the work on the Italian 'pan-microbiome' results highlighted the presence of well distinguished bacterial functional groups in regards to carbohydrate degradation, allowing to separate a subsample of the Italian people in different activity clusters. Moreover, the analysis of rural Bassa and urbanized individuals in Nigeria provided insights into the complex host-microbiome relationships across subsistence strategies, increasing our understanding on the changes in gut microbial communities and metabolic networks that have accompanied modern human history. The plasticity of the bacterial community can also be involved in processes that are not favorable to host's fitness, as I reported in the study about the gut resistome in aGvHD (graft versus host disease) pediatric patients. The assessment of resistome dynamics

disentangled the microbial ecology underlying antimicrobial resistance in HSCT, overcoming the limits of the traditional culture-dependent studies. Despite the low number of subjects, was provided evidence that aGvHD onset is associated with a peculiar trajectory in the personal gut resistome following HSCT, highlighting a plastic resistome response. This behavior of the gut microbial community represents the maladaptive side of the plasticity of the bacterial counterpart, stressing the importance of WGS-based surveys in paediatric HSCT patients for a better comprehension of the ecological dynamics of antibiotic resistance in aGvHD-positive cases.

In chapters 7 and 8 I focused on the viral and fungal fractions of the holobiont. Given the lack in virome and mycobiome characterization tools suitable for metagenomic surveys, my activity has been mainly oriented to the creation, set-up and validation of two modular pipelines with this aim, merging the knowledge of sequences processing and alignment gained prior and during my PhD. The resulting tools operate in a similar way: multiple alignment and filtering steps are performed, in order to extract microbial reads and discard human or environmental contaminants. These platforms are designed to be used by the scientific community to advance the knowledge on this hidden part of the microbiome.

Taken together, this work consolidates the dynamic vision of the microbiome in ecological, adaptive and sometimes maladaptive phenomena in holobionts. In all the cases I studied, I observed a mutualistic microbiome that may follows adaptive strategies aimed at the conservation of the homeostasis of the total ecosystem. It is worthy of note that I had the opportunity to study a limited number of organisms, and further studies are needed to understand the interactions between the holobiont players in a more integrated way. Neverthless, my work contributes to enrich the overall knowledge on the holobiont, also exploring some peculiar ecosystems for the first time. The data presented here may form the basis for future developments in the field, in order to obtain a more comprehensive profiling of bacterial, viral and fungal fractions within complex ecosystems.

List of publications included in this thesis

1. D'Amico F, **Soverini M**, Zama D, Consolandi C, Severgnini M, Prete A, Pession A, Barone M, Turroni S, Biagi E, Brigidi P, Masetti R, Rampelli S, Candela M, *Gut resistome plasticity in pediatric patients undergoing hematopoietic stem cell transplantation* (2019) Scientific Reports 9 (1) art. no. 5649.
2. **Soverini M**, Turroni S, Biagi E, Brigidi P, Candela M, Rampelli S, *HumanMycobiomeScan: A new bioinformatics tool for the characterization of the fungal fraction in metagenomic samples* (2019) BMC Genomics 20 (1) art. no. 496
3. Biagi E, D'Amico F, **Soverini M**, Angelini V, Barone M, Turroni S, Rampelli S, Pari S, Brigidi P, Candela M, *Fecal bacterial communities from Mediterranean loggerhead sea turtles (Caretta caretta)* (2019) Environmental Microbiology Reports 11 (3) pp. 361-371. Cited 2 times.
4. Quercia S, Freccero F, Castagnetti C, **Soverini M**, Turroni S, Biagi E, Rampelli S, Lanci A, Mariella J, Chinellato E, Brigidi P, Candela M, *Early colonization and temporal dynamics of the gut microbial ecosystem in Standardbred foals* (2019) Equine Veterinary Journal 51 (2) pp. 231-237. Cited 2 times.
5. Ayeni F.A, Biagi E, Rampelli S, Fiori J, **Soverini M**, Audu H.J, Cristino S, Caporali L, Schnorr S.L, Carelli V, Brigidi P, Candela M, Turroni S, *Infant and Adult Gut Microbiome and Metabolome in Rural Bassa and Urban Settlers from Nigeria* (2018) Cell Reports 23 (10) pp. 3056-3067. Cited 7 times.
6. Rampelli S, Turroni S, Schnorr S.L, **Soverini M**, Quercia S, Barone M, Castagnetti A, Biagi E, Gallinella G, Brigidi P, Candela M, *Characterization of the human DNA gut virome across populations with different subsistence strategies and geographical origin* (2017) Environmental Microbiology 19 (11) pp. 4728-4735. Cited 3 times.
7. **Soverini M**, Turroni S, Biagi E, Quercia S, Brigidi P, Candela M, Rampelli S, *Variation of carbohydrate-active enzyme patterns in the gut microbiota of Italian healthy subjects and type 2 diabetes patients* (2017) Frontiers in Microbiology 8 art. no. 2079 . Cited 4 times.
8. Candela M, Biagi E, **Soverini M**, Consolandi C, Quercia S, Severgnini M, Peano C, Turroni S, Rampelli S, Pozzilli P, Pianesi M, Fallucca F, Brigidi P, *Modulation of gut microbiota dysbioses in type 2 diabetic patients by macrobiotic Ma-Pi 2 diet* (2016) British Journal of Nutrition 116 (1) pp. 80-93. Cited 28 times.
9. Debebe T, Biagi E, **Soverini M**, Holtze S, Hildebrandt T.B, Birkemeyer C, Wyohannis D, Lemma A, Brigidi P, Savkovic V, König B, Candela M, Birkenmeier G, *Unraveling the gut microbiome of the long-lived naked mole-rat* (2017) Scientific Reports 7 (1) art. no. 9590 . Cited 10 times.
10. **Soverini M**, Rampelli S, Turroni S, Schnorr S.L, Quercia S, Castagnetti A, Biagi E, Brigidi P, Candela M.*Variations in the post-weaning human gut metagenome*

profile as result of Bifidobacterium acquisition in the western microbiome (2016) Frontiers in Microbiology 7 (JUL) art. no. 1058 . Cited 8 times.

11. **Soverini M**, Quercia S, Biancani B, Furlati S, Turroni S, Biagi E, Consolandi C, Peano C, Severgnini M, Rampelli S, Brigidi P, Candela M, *The bottlenose dolphin (Tursiops truncatus) fecal microbiota* (2016) FEMS microbiology ecology 92 (4) p, fiw055. Cited 15 times.
12. Rampelli S, **Soverini M**, Turroni S, Quercia S, Biagi E, Brigidi P, Candela M, *ViromeScan: A new tool for metagenomic viral community profiling* (2016) BMC Genomics 17 (1) art. no. 165 . Cited 42 times.