NANOPORE SEQUENCING APPROACHES TO STUDY MICE GUT MICROBIOTA UNDER ETHANOL EXPOSURE IN UTERO

Cristiano Miguel Pedroso Roussado

A thesis submitted in partial fulfilment of the requirements of the University of Brighton for the degree of Doctor of Philosophy

June 2022

Abstract

The gut microbiome plays a vital role in host homeostasis and an improved understanding of its biology is essential for a better comprehension of disorders such as foetal ethanol spectrum disorder (FASD). While next-generation sequencing has been used to assess the role of the gut microbiome, differences in experimental approaches may induce undesired biases. In this dissertation, the capacity of nanopore sequencing to retrieve the microbial profile and relative abundance of fungal and bacterial communities were assessed. A fungal mock community comprised of *Candida glabrata*, *Candida parapsilosis*, *Clavispora lusitaniae*, *Meyerozyma guilliermondii*, and *Pichia kudriavezevii*, was used to optimize targeted nanopore sequencing approaches based on the internal transcribed spacer (ITS) genomic region. Additionally, using a FASD mouse model, one untargeted (shotgun/metagenomic) and two targeted (based on the full-length 16S rRNA gene and the 16S-ITS-23S region from the ribosomal RNA operon) nanopore sequencing approaches were performed to sequence PCR products amplified from DNA extracts of gut content of 12 infant mice exposed to ethanol *in utero*. *Faecalibaculum rodentium* and *Duncaniella* sp. were the two most prevalent taxa detected using targeted sequencing approaches, while bacterial taxa were more evenly represented when using the metagenomic approach. The targeting of the full-length 16S rRNA gene provided the most comprehensive results. The nanopore sequencing optimization results suggest that experimental and/or bioinformatics steps, can introduce biases, misrepresenting certain taxa depending on the genomic region used for sequencing. Afterwards, the optimized 16S rRNA gene-targeted nanopore sequencing was performed to sequence DNA extracts from the gut content of nine infant mice exposed to ethanol *in utero* and eight non-exposed infant mice and compared with Illumina sequencing to determine the differences in the composition of gut microbiota between ethanol treated and untreated mice pups. No fungi were detected in the infant mice gut which implicate that fungi failed to reach the PCR detection threshold. Nanopore and Illumina sequencing retrieved different microbial profiles, probably explained by the different experimental and bioinformatical setups employed. Specifically, *Prevotella* and *Bidifobacterium* were not detected by nanopore sequencing and *Akkermansia muciniphila* was not detected by Illumina sequencing. Only *Parabacteroides distasonis*, *Muribaculum intestinale*, and *Duncaniella* sp. were simultaneously detected by both sequencing platforms. Neither platform detected any relevant effect of ethanol exposure. Concluding, this dissertation highlights the experimental handicaps of nanopore sequencing while detailing sequencing biases that could be observed towards certain taxa, therefore nanopore targeted approaches should be used with caution to avoid inaccurate microbial profiles.

Table of Contents

List of Figures

List of Tables

Abbreviations

ANOSIM – Analysis of Similarities

BBB – Blood-brain Barrier

BHB – β-hydroxybutyrate

CNS – Central Nervous System

F/B – Firmicutes/Bacteroidetes

FAS – Foetal Alcohol Syndrome

FASD – Foetal Alcohol Spectrum Disorder

FMT – Faecal Microbiota Transplantation

IBD – Inflammatory Bowel Disease

ITS – Internal Transcribed Spacers

LDA – Linear Discriminant Analysis

LEfSE – Linear Discriminant Analysis Effect Size

LPS – Lipopolysaccharide

MAG – Metagenome Assembled Genome

MALDI-TOF MS – Matrix-assisted Laser Desorption Ionization-time Of Flight Mass Spectrometry

NMDS – Non-metric Multidimensional Scaling

ONT – Oxford Nanopore Technologies

OOB – Out Of The Bag

OTU – Operational Taxonomic Unit

PAE – Prenatal Ethanol Exposure

PCR – Polymerase Chain Reaction

PERMANOVA – Permutational Analysis of Variance

PERMDISP – Permutational Analysis of Multivariate Dispersions

rho – Greek Letter to denote Spearman's rank correlation coefficient

ROS – Reactive Oxygen Species

SCFA – Short-chain Fatty Acid

TLR – Toll-like Receptor

WIMP – What's In My Pot?

Acknowledgements

I would like to thank the supervisory team, Dr Joao Silva, Dr Lucas Bowler, and Dr Fergus Guppy, for all the support during these last years. I would like also to thank Dr Nigel Brissett who kindly contributed with relevant biological material for the work here presented. The experimental work could not be possible without their efforts, trust, and *bone fide*. I also want to extend my thanks to Dr Peter Cragg and Dr Andrew Overall for the relevant contribution in crucial moments during the fellowship.

My family was the backbone that supported me during the very challenging times of my PhD fellowship. As always, thank you.

"The world produces waves. Surf or drown. You decide." Virgil Abloh

Declaration

I declare that the research contained in this thesis, unless otherwise formally indicated within the text, is the original work of the author. The thesis has not been previously submitted to this or any other university for a degree, and does not incorporate any material already submitted for a degree.

Signed CRISTIANO M. PEDROSO-ROUSSADO

Dated JUNE 10, 2022

Chapter 1 - Introduction

1.1 Background

The gut microbiota is a complex ecosystem that has an important role in many physiological functions. While bacteria dominate the gut composition, other microorganisms such as fungi, bacteriophages, archaea, and protists are also present. In the past decades, most studies focused on gut bacteria (commonly termed the 'bacterial microbiota' or 'bacterial microbiome') and the exploration of the fungal component ('mycobiota' or 'mycobiome'), is at an early stage, revealing many specific technical challenges in its analysis. Recent studies have highlighted that the gut mycobiome is much more diverse and abundant than previously thought, and there is increasing evidence of significant roles that fungi play in host homeostasis, and of their interactions with gut bacteria (Zhang *et al*., 2021). The mother is considered the most probable source of the infant microbiota (Ward *et al*., 2018). Newborns are initially colonized with mothers' microbes through vertical transmission (Ward *et al*., 2018). However, recently several questions have been raised about how the foetus gets colonized (Blaser *et al*., 2021). Other factors such as mothers' diet during pregnancy can impact the infant health and development, and ethanol consumption during gestation periods may result in several health-related complications in the newborn (Chun *et al*., 2017, Durack *et al*., 2018, Blaser *et al*., 2021). Foetal Alcohol Spectrum Disorder represents a cluster of physical and cognitive impairments resulting from ethanol consumption during gestation (Petrelli *et al*., 2018). In this dissertation, nanopore sequencing methodologies were used to microbially profile the infant mice gut when exposed to ethanol *in utero*. Nanopore sequencing is a third-generation sequencing approach recently developed which has the potential to provide a fast, cheaper, portable, and eventual real-time genomic analysis of biological samples (Deamer *et al*., 2016, Jain *et al*., 2016, Wang *et al*., 2021). Throughout targeted and untargeted nanopore sequencing optimization attempts, the fungal and bacterial gut microbiota of infant mice exposed to ethanol *in utero* were analysed and the derived data compared with that obtained from Illumina sequencing, which is considered the golden standard in terms of high-throughput microbial profiling analyses (Breitwieser *et al*., 2019).

1.2 Gut microbiota

1.2.1 Bacterial and fungal gut microbiota

Microbes are everywhere. Scientists have studied with increasing depth the diversity and function of microbes since the 19th century (Cavaillon and Legout, 2016). Now it is clear that microbes can colonize many niches, including the human body. Microbial colonizers of humans are considered determinant players in many aspects of health and disease and key regulators of host's physiology. Inside our bodies there are more microbial cells than human cells in a ratio of 1.3:1 (Sender *et al*., 2016). At the genetic level, more than 99% of the genes in human body are from microbial origin (Cryan *et al*., 2019). Additionally, the gut microbiome has been linked to many aspects of neurodevelopment, neuroinflammation, and behaviour, through a bidirectional communication pathway interlinking the gut and the Central Nervous System (CNS) – termed the microbiota-gut-brain axis (Collins *et al*., 2013, Sherwin *et al*., 2018, Cryan *et al*., 2020).

1.2.1.1 Bacterial gut microbiota

It is estimated that around 50% of human gut bacterial species currently lack a reference genome (Sunagawa *et al*., 2013, Nayfach *et al*., 2016). Despite being studied in increasing depth, the human gut microbiota still poses a variety of technical and practical challenges to isolate and sequence its constituents. Usually, metagenomic approaches assign taxonomical classifications to taxa that have not being cultured nor fully characterized. Still, the functional diversity remains underexplored as well the relevance for disease and pathogenicity of target phenotypes. Both culture-dependent and independent approaches have inherent biases towards the most abundant microorganisms, which increases the chance of missing less abundant ones. On the other hand, metagenomic approaches, and single-cell sequencing, are applied to overcome the need of prior isolation through cultivation (Nayfach *et al*., 2019). Whenever these approaches rely on assembly of sequencing reads into contiguous sequences (contigs) – set of sequence reads that are related to one another by an overlapping of their sequences –, which are then binned into metagenome assembled genomes (MAGs). Binning is the process of grouping contigs that are assumed to belong to the same species. MAGs are assembled with individual or multiple co-assembled metagenomes based on nucleotide frequency abundance and/or co-variation of abundance across a group of samples (Alneberg *et al*., 2014, Stewart *et al*., 2018). Finally, MAGs are further analysed to evaluate the extent of genomes' completeness and contamination, presence of marker

genes, and generated contiguity (Parks *et al*., 2015, Bowers *et al*., 2017, Sczyrba *et al*., 2017). Irrespectively of their prevalence in the gut, many human gut microbial colonizers lack a sequenced genome and have not been grown under laboratory conditions so far (Human Microbiome Jumpstart Reference Strains Consortium 2010, Fodor *et al*., 2012, Browne *et al*., 2016, Lagier *et al*., 2016). Various studies have been applying "culturomics" approaches to close this gap (Browne *et al*., 2016, Lagier *et al*., 2016). Culturomics' approaches rely on the diversification of culture growth conditions aiming the increase of the microbial repertoire harvested. During the last decade, other technologies such as Matrix-assisted Laser Desorption Ionization-time of Flight Mass Spectrometry (MALDI-TOF MS) have been coupled with culture-dependent approaches to deepen the characterization power of the cultivated isolates (Lagier *et al*., 2015).

It is common to use the mouse model to uncover key insights to be later translated for human physiology comprehension. Investigation of the mouse gut microbiota can be helpful for study of human counterpart because both gut microbiotas show functional similarity and resemblances at the genus level, although typically revealing different proportions in prevalence of taxa (Xiao *et al*., 2015, Liu *et al*., 2020). Although, many conclusions from mice studies have subsequently been confirmed in clinical human studies (Wang *et al*., 2019). Recent studies showed that mice have a core gut bacteriome – the shared bacteria, bacterial genes, or functional capabilities observed in the gut of all or the majority of mice (Gu *et al*., 2013, Xiao *et al*., 2015, Wang *et al*., 2019, Yang *et al*., 2021). The composition of mice core gut bacteria has been shown to be closely related to the age, sex, and healthy state of the host, playing an important role in host gut homeostasis and health (Wang *et al*., 2019). Wang and colleagues (2019) studied the gut microbiota of 101 healthy mice and found the following bacteria genera as members of the core-microbiota: *Anaerotruncus*, *Parabacteroides*, *Oscillibacter*, *Clostridium*, *Flavonifractor*, *Bacteroides*, *Barnesiella*, *Alistipes*, *Helicobacter*, *Saccharibacteria_genera_incertae_sedis*, *Prevotella*, *Lachnoanaerobaculum*, *Lactobacillus*, *Intestinimonas*, *Roseburia*, *Alloprevotella*, *Rikenella*, *Enterorhabdus*, *Erysipelotrichaceae_incertae_sedis*, *Eggerthella*, *Allobaculum*, *Lachnospiracea_incertae_sedis*, *Pseudoflavonifractor*, *Bifidobacterium*, *Marvinbryantia*, *Mucispirillum*, *Blautia*, *Anaerofilum*, *Parasutterella*, *Odoribacter*, *Olsenella*, *Turicibacter*, *Gordonibacter*, *Ruminococcus*, and *Acetatifactor*. Although *Bacteroides* spp. are widely studied due to their common presence in the gut of western populations (Marcobel *et al*., 2011, Human Microbiome Project Consortium, 2012, Nielsen *et al*., 2014), its intra-species diversity might be poorly understood (Pasolli *et al*.,

2019). Thus, the global human bacterial microbiome diversity may be greater than previously expected. Additionally, regional initiatives that cover non-western populations and different lifestyles are increasing both microbiota and microbiome comprehension, shedding light on hidden functional and phylogenetic diversity of microbial gut members. Recently, various research groups have tried to curate human gut microbiome species databases through high-throughput assembly of metagenomes (Almeida *et al*., 2019, Nayfach *et al*., 2019, Pasolli *et al*., 2019). The focus has been on finding the highest taxonomical discrimination as possible within human gut population and across individuals (Greenblum *et al*., 2015, Garud *et al*., 2017, Zhao *et al*., 2019). The relative lack of human gut microbiome studies from underrepresented regions such as Africa and South America constitutes a handicap and increases the biases towards regional-enriched taxa. In order to have a clear and true picture of what is going on in the human gut landscape, both compositionally and functionally, is imperative to expand microbial profiling and subsequent functional analysis to large cohorts worldwide. Pasolli and colleagues (2019) analysed 9,429 metagenomes and were able to detect more than 154,000 microbial genomes that covered many body sites, ages, countries, and lifestyles. The authors found microbial signatures that might be associated and/or enriched regionally. They found that *Succinatimonas* spp., *Eggerthella* gen. *Dialister* spp., *Treponema succinafaciens*, *Clostridium* spp., among others, were enriched in the Malagasy population, which significantly contrasts with the microbial composition enriched in westernized regions (Pasolli *et al*., 2019). Additionally, *Elusimicrobia* phylum, a less typically human-associated phyla, was detected, showing that a wider sample representation is necessary to fully understand the human gut microbiome. In the western population sampled the authors detected *Bacteroides uniformis*, *Alistipes putredinis*, *Parabacteroides distasonis*, among others, with enriched profiles. Additionally, the prevalence and abundance of these taxa is quite distinct between Malagasy and Western populations. Moreover, most of the taxa detected in the Malagasy population are unknown to the existing databases or contain only isolate sequences previously assembled by the authors in other studies (Pasolli *et al*., 2019).

In another study, Almeida and colleagues (2019) analysed 13,133 human gut metagenomic datasets from around the globe derived from both unhealthy and healthy adults. Three of the most frequently derived genomes were assigned to species known to colonize the human intestine, *Ruminococcus bromi*, *A. putredinis*, and *Eubacterium rectale* (Almeida *et al*., 2019). A more robust approach to assign the most likely taxonomic lineage to each metagenome was also tried. It consisted of complementing the

phylogenetic inference method of CheckM with protein searches against the UniProt Knowledgebase (UniProtKB) (Uniprot, 2017). Based on this approach the authors found that 94% of the metagenomes did not match any genome present in their customized database, thus these represented previously uncultured taxa (Almeida *et al*., 2019). As of August 2019, 74% of those uncultured taxa corresponded indeed to novel genomes. Almeida and colleagues (2019) were also able to assign almost all metagenomes to either phylum, class, or order levels. This reveals that there is a poor assignment success to lower taxonomical levels, however the effect may also be explained by the increase of the database size (Nasko *et al*., 2018).

Among family ranks, the three most frequently assigned were *Coriobacteriaceae* (21%), *Ruminococcaceae* (10%), and *Peptostreptococcaceae* (7%) (Almeida *et al*., 2019). Most frequent genera were *Collinsella* (18%), *Clostridium* (7%), and *Prevotella* (4%). These data reveal that even known clades have not been extensively cultured and sequenced, as exemplified by the polyphyly of the *Clostridium* genus, which phylogenetic estimates suggest may span 121 genera belonging to 29 families (Parks *et al*., 2014). Therefore, the detection of uncultured species from MAGs assigned to high taxonomical ranks may result from a lack of updates to the reference databases re-imposing a classification limitation. These limitations are exacerbated when analysing taxa possessing high polyphyletic profiles. Almeida and colleagues (2019) also determined the prevalence and abundance of the uncultured species found in each gut microbiome. Similar to what was previously reported, the most frequent assignments were represented by *Ruminococcaceae* family, *Faecalibacterium* genus, and members from *Clostridia* class (Almeida *et al*., 2019).

In another attempt to catalogue human gut bacteria, Nayfach and colleagues (2019) reconstructed 60,664 draft prokaryotic genomes from 3,810 faecal MAGs from a diverse human population sample covering different geographies and backgrounds. The authors formed the Integrated Gut Genomes Database which now comprises 156,478 genomes (Nayfach *et al*., 2019). They discovered 2,058 newly identified species-level operational taxonomic units (OTUs), which is a 50% increase in comparison to the most recent data for gut bacterial microbiota phylogenetic diversity. The newly identified OTUs were enriched in samples from rural populations. Moreover, the authors suggested that this group have undergone genome reduction, loosing certain biosynthetic pathways (Nayfach *et al*., 2019). This observation may guide future attempts to improve culture-based approaches. Despite a broad taxonomical assignment distribution observed, newly identified OTUs tend to represent *Firmicutes* orders' *Lachnospirales* and *Oscillospirales*

(Nayfach *et al*., 2019). Within *Bacteroidetes*, almost 400 OTUs have been discovered (Nayfach *et al*., 2019). Surprisingly, no high-quality genome belonging to *Clostridia* clade was represented, which may be explained by genome reduction or other unknown factors that hampered genome assembly attempts (Nayfach *et al*., 2019).

More recently, Almeida and colleagues (2021) published a unified catalogue of the human gut microbiome comprising 203,938 reference genomes. In accordance with other reports, more than 70% of the catalogued species have not been cultured (Almeida *et al*., 2021). The most represented taxa with nonredundant genomes detected per species were *Agathobacter rectalis* (recently reclassified from *E. rectale* (Rosero *et al*., 2016)), *Escherichia coli, Bacteroides uniformis*, *A. putredinis*, and *P. distasonis*, both represented by more than 3,000 genomes (Almeida *et al*., 2021).

Merging culture-dependent and independent approaches, Poyet and colleagues (2019) generated a library of 7,759 human gut bacteria isolates collected from healthy Faecal Microbiota Transplantation (FMT) donors living in the Boston area of the USA. To test the extent of the library, the authors applied a culture-dependent approach and observed that this biobank comprehensively captured the bacterial diversity to the genus-level (Poyet *et al*., 2019). This biobank contained isolates belonging to the most common phyla detected in the human gut – *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Fusobacteria*, *Proteobacteria*, and *Verrucomicrobia*. In total, the biobank has representatives from 11 classes, 16 orders, 40 families, and 133 genera (Poyet *et al*., 2019). Additionally, the authors isolated 132 different *Akkermansia* strains from which 45 *Akkermansia muciniphila* genomes were sequenced. Moreover, they also sequenced 67 previously unknown *Akkermansia* strains, which may represent a new species (Poyet *et al*., 2019). Interestingly, the authors showed that microbial populations within the human gut maintain stable sizes (Poyet *et al*., 2019).

Evidently, studying commensal microbiota in their original hosts is the most appropriate approach when considering the global genetic and ecological impact of microbe-microbe and host-microbe interactions. Despite the extensive use of mouse models, human and mouse gut microbiota have compositional dissimilarities which limits the translation research between them (Beresford-Jones *et al*., 2021). One of the limitations is the known fact that few bacterial species are shared between the gastrointestinal tracts of the two species (Xiao *et al*., 2015, Chung *et al*., 2021). Additionally, a great extent of microbial species detected in the mouse gut remain unidentified and poorly characterized. To overcome the translational handicap, studies have been carried out with a reasonable degree of success which tested mice colonized

with human microbiota (Chung *et al*., 2012, Surana *et al*., 2017, Lundberg *et al*., 2020, Park *et al*., 2020). Even though not many species are shared between human and mice, the estimation using gene-level functional analyses is that microbial functions are shared almost completely (Xiao *et al*., 2015). Beresford-Jones and colleagues (2021) recently presented a catalogue that features 26,640 non-redundant, high-quality bacterial genomes including 1,094 species and a collection of 223 cultured isolates from 132 species. The authors successfully grew 62 species for the first time (Beresford-Jones *et al*., 2021). This resource was specifically designed to enable the identification of the closest functionally bacterial species between humans and mice. The authors developed a toolkit that allows researchers to query any target function annotated in KEGG, COF, eggnog, and GO (https//github.com/BenBeresfordJones/MGBC-Toolkit) (Beresford-Jones *et al*., 2021). The toolkit also allows the screening of gene products at the sequence level, thus improving identification of nonannotated genes and predicted functions of interest. Interestingly, the authors observed that microbial functionality is not necessarily associated with the same taxonomic profile between human and mice (Beresford-Jones *et al*., 2021).

Due to the overrepresentation of bacterial genomes in more comprehensive reference databases and well-established standard and tools, most knowledge about the human gut whole microbiome is limited for the remaining gut colonizers such as archaea, eukaryotes, and viruses. More reliable approaches need to be developed in order to elucidate the whole compositional and functional profile of the gut microbiome (Uritskiy *et al*., 2018).

1.2.1.2 Fungal gut microbiota

While gastrointestinal bacterial microbiota represents 10^{12} microorganisms per gram of content, fungi had been cultivated from faeces representing between 10^2 and 10^6 colony forming units per gram (Simon and Gorbach, 1984). Therefore, the fungal microbiome is the minority in terms of composition, abundance, and the genetic landscape of the human gut, covering less than 0.1% of the whole gut metagenome (Qin *et al*., 2010). Fungi present in the gut are known to interact with the bacterial microbiome promoting major shifts in its population ecology and affecting innate and adaptive immune development in mice (van Tilburg Bernardes *et al*., 2020). Moreover, neglecting the gut mycobiome – the fungal component of the gut microbiome – may exclude relevant effects that fungi exert on the bacterial component. As an example, fungal cell wall components are crucial for fungal-bacteria interactions (Iliev *et al*., 2012), as illustrated

by the influence that mannan has on *Bacteroides thetaiotaomicron*, a commonly found human gut bacterium (Cuskin *et al*., 2015). Also, the composition of the anaerobic bacterial population is known to be much affected by fungal chitin (Charlet *et al*., 2018). Fungal communities have lower diversity and greater unevenness when compared with bacterial gut microbiotas. The accepted wisdom is that the observed dissimilarities between subjects and across time within the same subject suggest that most fungi are not stable colonizers of the human gut (Raimondi *et al*., 2019). Fungi can shift from transitory to persistent states when a combination of specific genetic features is present in both microbe and host, for instance if permeability alterations in the gastrointestinal barrier take place (Sokol *et al*., 2017). Additionally, the direct contributions of either bacterial or fungal components on host disease onset, and the relevance of fungi-bacteria interactions on the appearance of disease-promoting phenotypes is still poorly understood (Sokol *et al*., 2017).

There is some data reporting that mycobiome alpha-diversity decreases while bacterial microbiome increases during the first year of life (Laforest-Lapointe and Arrieta, 2017). This dynamism elucidates the importance of fungal-bacteria interactions in the whole microbiome establishment and host response to stresses such as antibiotic-related alterations (van Tilburg Bernardes *et al*., 2020).

Commonly, fungi have a commensal or mutualistic interaction with mammalian hosts. However, several taxa can be considered pathobionts (opportunistic microorganisms that develop as a result of perturbations in the healthy microbiome), such as *Candida albicans*. One troublesome fact is that most studies only focus on the fungal pathogenicity of certain species (Mayer *et al*., 2013, Merseguel *et al*., 2015, Cavalieri *et al*., 2018, Markey *et al*., 2020). Despite recent efforts, human fungal composition is not well studied (Huseyin *et al*., 2017) and the gut fungi impact on host homeostasis have still to be further investigated (Seed, 2015, Suhr and Hallen-Adams, 2015, Strati *et al*., 2016, Nash *et al*., 2017, Auchtung *et al*., 2018, Borges *et al*., 2018). The circumstances that lead to the fungal pathogenicity are related to the lack of fungi-specific immune mechanisms in the mucosa and its immune system components. If the recognition and interaction between fungal community and mucosal immune system is disrupted, overgrowth and systemic diseases can occur (Cao *et al*., 2015, Corvilain *et al*., 2018). Nevertheless, fungi colonizing the colon also have protective actions towards the host epithelial layer. As an example, fungi are necessary for the formation of balloon-like projections formed by subepithelial macrophages in the distal colon (Chikina *et al*.,

2020). These projections protect the epithelial layer from damage caused by harmful fungal metabolites.

Research on human gut mycobiota has found *Aspergillus*, *Candida*, *Debaryomyces*, *Cladosporium*, *Clavispora*, *Cyberlindnera*, *Galactomyces*, *Malassezia*, *Penicillium*, *Pichia*, *Rhodotorula*, and *Saccharomyces* genera as the most prevalent taxa (Hoffmann *et al*., 2013, Seed, 2015, Surh and Hallen-Adams, 2015, Strati *et al*., 2016, Surh *et al*., 2016, Nash *et al*., 2017, Auchtung *et al*., 2018, Borges *et al*., 2018, Raimondi *et al*., 2019, Wu *et al*., 2020). However, only *Saccharomyces* spp. and *Candida* gen. seem to be stable colonizers. Overgrowth of *Candida* and elevated serum anti-*Saccharomyces cerevisae* immunoglobulin G (IgG) antibodies have been associated with alcoholic liver disease (Lang *et al*., 2020, Yang *et al*., 2017). Candidalysin – a *C. albicans* virulence factor –, has been implicated in exacerbation of alcoholic liver disease (Chu *et al*., 2020), which highlights the relevance of *C. albicans* to this disease. Interestingly, *C. albicans* was found to alter the caecal lipid profile in a non-obese state (Markey *et al*., 2020). This fact implies that fungi may have a relevant role in metabolic dysfunction. Fungal dysbiosis caused by antibiotic treatment in early life may be also associated with childhood obesity alongside bacterial dysbiosis (Markey *et al*., 2020), and fungi in general, and yeasts in particular, have demonstrated the capability of taking up lipids from their surroundings, and this should be a factor considered with respect to the hosts' physiology when lipid availability is abundant (Klug *et al*., 2014, Gutierrez *et al*., 2021). Recently, it was also observed that *C. albicans* colonization in mice altered hepatic gene expression which was related to an increase availability of long chain polyunsaturated fatty acids (Mackey *et al*., 2020). These observations constitute a new perspective on the fungal-host interaction and metabolic gene modulation.

Other *Candida* species can also be interpreted as commensals of the human body, such as *Candida dubliniensis*, *Candida glabrata*, *Candida guilliermondi* (syn. *Meyerozima guilliermondi*), *Candida kefyr*, *Candida krusei* (syn. *Pichia kudriavzevii*), *Candida tropicalis*, and *Candida parapsilosis* (Muadcheingka and Tantivitayakul, 2015). The increase in clinical infections caused by non-albicans *Candida* species is of rising health concern, mostly as a consequence of their observed intrinsic resistance to conventional antifungal drugs (Merseguel *et al*., 2015, Wu *et al*., 2017). Therefore, further research on the ecological and phenotypical relevance of intestinal fungi is necessary to disclose both target and clinically relevant interactions with the host (Huseyin *et al*., 2017).

The gut mycobiome is shaped from birth, and ongoing analysis from this early stage provides a window for study of its development and maturation. Early events, such as breast-feeding, are becoming seen as crucial to our understanding of how the gut whole microbiome gains its shape. Many fungi have been detected in breast milk (Boix-Amorós *et al*., 2019, Heisel *et al*., 2019, Moossavi *et al*., 2020). Boix-Amorós and colleagues (2019) recently proposed a core breast-milk mycobiome represented by *Malassezia*, *Daviddiella*, *Sistotrema*, and *Penicillium*, however there is a risk that some fungi, such as spore-forming species, may be environmental contaminants and not truly present in the samples. Fungal dysbiosis causing fungal overgrowth will modulate gastrointestinal and systemic immune system early in life, posing developmental pressures on host immunity and metabolism (van Tilburg Bernardes *et al*., 2020). Fungal overgrowth, promoted by antibiotic-induced perturbations, increases the interaction with intestinal epithelia and anti-fungal responses are initiated thereafter (van Tilburg Bernardes *et al*., 2020). Alterations in fungal composition detected during early years have been linked to atopic asthma, with detected overgrowth of *Candida* spp., *Rhodotorula* spp., and *Issatchenkia orientalis* (syn. *P. kudriavzevii*) (Fujimura *et al*., 2016, Arrieta *et al*., 2018, van Tilburg Bernardes *et al*., 2020). These changes were found to exhibit stronger associations with disease than bacterial dysbiosis in an Ecuadorian study (Arrieta *et al*., 2018).

The main limitations of the study of fungal microbiome through culture-independent approaches are related to the non-standardization of DNA extraction and handling protocols, primer sets options for targeted Polymerase-Chain Reaction (PCR) amplification, the consistency, curation, and completeness of fungal reference databases, among other factors (Nash *et al*., 2017). Similar to bacterial 16S rRNA gene profiling approaches, Internal Transcribed Spacer regions (ITS's) within eukaryotic rRNA genes allows fungal microbiota profiling in the intestine (Markey *et al*., 2020). In a recent study, Raimondi and colleagues (2019) investigated the faecal mycobiome in healthy adults for 12 months combining culture-based approaches with ITS1 high-throughput sequencing. The authors concluded that fungi cannot easily colonize the human gut and most samples were dominated by one or two genera (Raimondi *et al*., 2019). Additionally, fungal populations were found to have uneven compositions due to the large diversity detected between and within subjects across the 12 months of the experiment. This may be explained by the transitory state of most fungi in the human gut, or by inter-variability between each donor (Raimondi *et al*., 2019). Such observation is contrary to what has been observed in the bacterial human gut microbiota community (Raimondi *et al*., 2019). Raimondi and colleagues' observations are consistent with previous data, where

Aspergillus, *Candida*, *Debaryomyces*, *Malassezia*, *Penicillium*, *Pichia*, and *Saccharomyces* are the most common genera detected (Hoffmann *et al*., 2013, Strati *et al*., 2016, Nash *et al*., 2017, Auchtung *et al*., 2018, Borges *et al*., 2018). However, the authors reported the unusual detection of *Mucor*, and the absence or scarcity of *Cladosporium*, *Clavispora*, *Cyberlindnera*, and *Galactomyces*. These discrepancies may be explained by regional microbial patterns associated with different lifestyles or different experimental approaches. Besides, metagenomic analysis detected few non-albicans *Candida* species (*Candida alcoholica*, *Candida inconspicua*, *C. parapsilosis*, *C. tropicalis*, and *Candida zeylanoides*) (Raimondi *et al*., 2019). Interestingly, the authors found using fingerprinting techniques on the derived isolates that *Geotrichum candidum* and *Rhodotorula mucilaginosa* are stable colonizers of the gastrointestinal tract even though *Rhodotorula* and *Geotrichum* genera were not detected using a metagenomics approach (Raimondi *et al*., 2019). Moreover, metagenomics revealed that *Debaryomyces udenii* may be a putative colonizer, because it was detected in every sample of the same subjects (Raimondi *et al*., 2019). In the same study, the authors reported that *C. albicans* was the sole species showing the ability to colonize the murine intestine (Raimondi *et al*., 2019). This feature might be in part explained by the phenotypical arsenal present in *C. albicans* that seems to play a role in stable colonization – transition to hyphal form, biofilm formation, adhesion to intestinal epithelial cells, *etc* (Raimondi *et al*., 2019). Still, further research is necessary to elucidate the colonization and survival potential of other species inside the human and murine gut, such as *D. udenii*, *C. zeylanoides*, *G. candidum*, *S. cerevisae*, *R. mucilaginosa*, among others.

After the early development years, the gut mycobiome stabilizes although possessing a lower alpha-diversity in comparison with the bacterial microbiota (Nash *et al*., 2017, Raimondi *et al*., 2019). The most abundant and observed phyla at this stage are *Ascomycota*, *Basidiomycota*, and *Mucoromycota*, the last one much less abundant than the rest (Nash *et al*., 2017, Raimondi *et al*., 2019, Wu *et al*., 2020). Even though more stable, the mature gut mycobiome is still characterized by relative instability with both inter- and intra-individual differences occurring over time (Nash *et al*., 2017, Raimondi *et al*., 2019). Transient gut colonization can potentially explain the fungal composition instability because some fungi reach the gut as a result of environmental exposure and via dietary intake (Nash *et al*., 2017, Raimondi *et al*., 2019). Above all, the composition of fungal populations detected in the human gut through culture-independent approaches are mostly dominated by environmental or food-associated taxa. As Yeung and colleagues (2020) demonstrated, exposure to environmental fungi in mice may trigger a more robust

immune response (Yeung *et al*., 2020). Using a rewilded mice model the authors observed an increase of T cell activation and circulating granulocytes when filamentous fungi were more abundant (Yeung *et al*., 2020). Nevertheless, *C. albicans* and other described yeasts can colonize the murine gastrointestinal tract in a stable fashion (Raimondi *et al*., 2019, van Tilburg Bernardes *et al*., 2020). Biofilm formation and prominent adherence phenotypes can explain the persistence of such species over time (Raimondi *et al*., 2019). Gastrointestinal fungal colonization is therefore both persistent and transient and such patterns need further analysis to discover the various levels of impact to host physiology. For instance, it is not clear whether the transitory state of some detected genera, such as *Aspergillus* and *Penicillium*, is due to an intrinsic inability to colonize or death during gastrointestinal transit (Strati *et al*., 2016, Raimondi *et al*., 2019). Since other species were only detected by culture-independent approaches, it is possible that they cannot survive the gastrointestinal conditions. As *Aspergillus* and *Penicillium*, among others, are potential sources of mycotoxins or can behave as opportunistic pathogens targeting mucosal tissues (Hedayati *et al*., 2007, Perrone and Gallo 2017, Perrone and Susca, 2017) it is imporant to detail their functional role while inside the human gut.

Specific Pathogen-Free mice models are commonly used to test the effect of fungal colonization on the host (van Overbeek and Saikkonen, 2016, Deveau *et al*., 2018). However, these approaches also rely on the use of antimicrobial treatment that could affect the bacterial microbiome as well, which may be sufficient to explain the observed effects (van Overbeek and Saikkonen, 2016, Deveau *et al*., 2018). Recently, van Tilburg Bernardes and colleagues (2020) reported the role of fungi in the absence of bacteria in the mouse gut. The authors found that fungi strongly impact microbiome dynamics in mice despite being outnumbered by bacteria, promoting vigorous local and systemic immunological changes both in the gut and in the lungs (van Tilburg Bernardes *et al*., 2020). Interestingly, the authors were able to consistently recover fungi through cultivation from gnotobiotic mice faecal samples weeks after inoculation (van Tilburg Bernardes *et al*., 2020). Such observations mean that fungi can inhabit the gut in absence of bacteria and provided clues to the potential detrimental effects of fungal-bacterial interactions. Interkingdom competition or antagonistic interactions between fungi and bacteria have been described in the other natural habitats (Boer *et al*., 2005). However, van Tilburg Bernardes and colleagues (2020) observed that fungal population alpha-diversity increased alongside various species´ abundances in the presence of bacteria. Interkingdom synergistic relationships as depicted by the authors are not new since they have also been reported in the fungi detected in agricultural food-chains (Frey-Klett *et*

al., 2011, van Tilburg Bernardes *et al*., 2020). In summary, research so far has showed the intricacies and diversified ecological relationships between fungi, bacteria, and host inside the gut.

Despite recent advances in mycobiome studies, associations with the onset of some human diseases remains to be further explored. As per bacterial microbiome, fungal microbiome's variability between individuals presents further challenges in identifying patterns between mycobiome and remaining host physiology (Hallen-Adams *et al*., 2015). For example, it is not clear how dynamic the mycobiome is in relation with ageing. Prolonged exposure to a processed diet and high isolation periods in pathogen-free environments has been shown to reduce fungal gut diversity (Mims *et al*., 2021). Additionally, this diversity reduction is associated with high levels of adiposity and physiologic alterations as is also observed in Metabolic Syndrome (Mims *et al*., 2021). Therefore, diet may impose key baseline alterations in the mycobiome composition that follows metabolic changes in the host. Obesity alters gut mycobiome in humans (Rodriguez *et al*., 2015) and mice (Heisel *et al*., 2017, van der Merwe *et al*., 2020). However, studies inducing obesity through diet have limitations in terms of the response observed in the fungal populations. Neither high-fiber diets, high-fat time restricted feeding, high-fat alternate day fasting, nor high-fat calorie restrictions had any relevant effect in altering murine gut mycobiota (van der Merwe *et al*., 2020). Such observations are contrary to what has been observed in bacterial gut microbiota (van der Merwe *et al*., 2020). However, Mims and colleagues (2020) recently reported that a low-fat processed diet caused alterations in mouse gut mycobiota, and the observed alterations were correlated with metabolic features such as weight gain, serum triglycerides and fasting ghrelin. In inflammatory bowel disease (IBD), there is an increase in the fungi/bacteria diversity ratio and *C. albicans* appeared to be more abundant in diseased than in healthy patients (Cryan *et al*., 2019). Another example that sustains the hypothesis of fungal overgrowth during inflammation is the presence of *Malassezia restricta* in most IBD patients carrying the risk allele CARD9 – which is a molecule associated with fungal innate immunity (Limon *et al*., 2019). However, the hypothesis that gut fungi directly influence the inflammatory tone of the mesenteric adipose tissue depot is still unclear (Limon *et al*., 2019). Although gut fungi translocation to the mesenteric adipose tissue depot in IBD was recently reported (Ha *et al*., 2020), it is unclear if the translocation phenomenon only occurs in intestinal inflammation or under any other inflammatory state. Allergic lung inflammation is another example that shows the relevance of fungi in diseases' onset, as antibiotic use during exacerbation of the disease in mice promotes

overgrowth of *C. albicans*, *C. parapsilosis* (Noverr *et al*., 2005, Kim *et al*., 2014), or even expansion of filamentous fungi after antifungal therapy (Wheeler *et al*., 2016). However, further research is necessary to elucidate more details about the causal role of fungi in these and related diseases. Although lacking evidence on the exact mechanisms, there are some clues from the perceived relevance of gut fungi in other diseases, such as *Candida* spp. and *Saccharomyces* spp. showing an association with type 1 diabetes risk in children (Honkanen *et al*., 2020). One other feature that is currently neglected is the influence of fungi in the overall bile acids content in the gut (Kollenov *et al*., 2016). Specifically, *Fusarium* gen. has been reported to metabolize deoxycholic acid, and *Aspergillus* gen. and *Penicillium* gen. have been found to produce secondary bile acids (Kollenov *et al*., 2016). Bile acids' importance is also revealed by the way they influence the stability of other fungal metabolites, such as the lipase produced by *Thermomyces* (reported as the most significant taxon associated with weight gain on a processed diet (Kollenov *et al*., 2016)). Metabolic hormones such as leptin, resistin, and ghrelin may also be influenced by the mycobiome (Kollenov *et al*., 2016). Bile acids' signalling has gaining momentum as more studies uncover the role of fungi in host metabolic processes (Kollenov *et al*., 2016). Gut fungi have also been found in the pancreas (Raimondi *et al*., 2019). High levels of *Malassezia* spp. have been found in pancreatic ductal adenocarcinomas both in human patients and mice (Raimondi *et al*., 2019). *Malassezia* spp. overgrowth promotes tumorigenesis in a mannose binding lectin (MBL)-C3 axis dependent manner (Raimondi *et al*., 2019). Further research on the stability of gut mycobiome is necessary to validate the hypothesis of their relevance for host immune education. It is not that clear what the impact of gut fungi transiency or persistency is in triggering immunological effects, therefore modulating the robustness of local and systemic immune responses and associated phenotypes (Gutierrez *et al*., 2021).

To further investigate the impact of environmental and pharmacological agents on the whole gut microbiome it is relevant to uncover in more detail the full fungi-bacteria interactive landscape in the gastrointestinal tract and their effects on host homeostasis (Gutierrez *et al*., 2021). Study of the role of fungal-derived molecules such as β-glucans, extracellular vesicles, and candidalysin is important in order to better understand the gut mycobiome functionally. Gnotobiotic models might be helpful in research projects which aim to go beyond taxonomic gut profiling. As a relevant initial development window, it remains to explain the early mechanisms that allows microbial colonization and permanence to host maturity.

1.2.2 Foetal gut microbiome controversy

Recently, the sterile hypothesis of the prenatal environment has been challenged. New data obtained from placenta, amniotic fluid, meconium, and foetal tissue samples analysed through next-generation sequencing, cultivation, microscopy, and quantitative PCR has helped to address questions regarding the dogma of a sterile womb (Blaser *et al*., 2021). In January 2021, The Microbiome journal released an entire issue dedicated to the emerging subfield around the existence and implications of the prenatal microbial communities (Fricke and Ravel, 2021). One of the factors in favour of the sterility hypothesis is that xenobiosis – the 'germ-free' state – has been experimentally achieved in many organisms, such as mice, ungulates, pigs, and even humans (Steel *et al*., 2005, Aagaard *et al*., 2014, Collado *et al*., 2016). Therefore, xenobiosis occurrence contradicts the existence of an indigenous placental microbiota. However, contradictory results have been reported that challenges the acceptance of a microbial presence at some body sites. Evidence supporting the microbial presence in placenta and amniotic fluid was gathered through culture-independent methods (Steel *et al*., 2005, Aagaard *et al*., 2014, Collado *et al*., 2016). However, those results have been challenged by authors who claimed that microbial detection signals could be attributed to false-positives and external contamination factors (Lauder *et al*., 2016, Lim *et al*., 2018, de Goffau *et al*., 2019). Therefore, further research is necessary to disclose the controversy and to better understand the influence that early microbial colonizers, transient or persistent, have on mucosal immune development (Rackaityte *et al*., 2020). These necessary research efforts should target perinatal sites, since identifying microbial taxa does not immediately mean that they are colonizers (Rackaityte *et al*., 2020). Colonization may happen over either short or longer time periods, but a true niche occurs when complex interactions arise between niche members and the host, resembling a commensalism-related environment. In any case, the influence of the mother's microbiota on the developing foetus is relevant. The presence of microbes or microbial products in the placenta may modulate the foetus immunological development (Blaser *et al*., 2021). Even though these products and microbes may show no or low viability, their source is the mother's microbiota and the foetal exposure to such microbes and microbial products deserves further research (Blaser *et al*., 2021). Mother-to-infant microbial transfer and the potential effect of other environmental factors happens perinatally (Ward *et al*., 2018). Recent links observed between the early developing gut microbiome and disease conditions such as whole food allergies, asthma, and autism, are increasing the speculation about any eventual causal relationship (Fujimura *et al*., 2016, Levan *et al*., 2019, Sharon *et al*., 2019, Sanidad and

Zeng, 2020). Maternal exposure to determined environmental factors during pregnancy are associated with higher risk of chronic inflammatory disease during childhood, like asthma (Yu *et al*., 2018), and an abnormal neonatal meconium (*i.e.*, the first stool newborns generate) is associated with an increased risk of inflammation (Fujimura *et al*., 2017, Durack *et al*., 2018). The vulnerability of the neonate gut microbiome to environmental and dietary changes is apparent, and a healthy neonatal gut microbiome is essential to avoid risks of infections and for a normal activation and development of immune cells later in life (Chu *et al*., 2017, Durack *et al*., 2018). Nevertheless, neonatal meconium consists of amniotic fluid ingested during gestation and a simple microbial presence and products (Chu *et al*., 2017, Durack *et al*., 2018). Mode of delivery and diet influence neonatal gut postnatal colonization (Blaser *et al*., 2021). However, neonatal gut microbiome patterns directed by the mode of delivery waned after 6 weeks, which imply that other factors are in play to modulate the neonatal gut microbiome development (La Rosa *et al*., 2014, Chu *et al*., 2017). *Bifidobacterium* and *Lactobacillus* are found in the more acidic intestinal environment when infants are breast-fed (Yoshioka *et al*., 1983, Martin *et al*., 2003). Such gut microbiome and biophysical landscape exert a powerful defence against common enteric pathogens in a critical window for maturation of the immune system (Martin *et al*., 2003, Chu *et al*., 2017).

Despite the recent controversy around the microbe-free perinatal state, it is not yet clear if viable microbes persist *in utero* nor how they interact with the intestinal immune system (Blaser *et al*., 2021). Since mucosal immunity develops from week 11 to 14 of gestation in the human foetal intestine, even a reduced microbial presence might have a relevant impact (Rackaityte *et al*., 2020). An adult gut is characterized by high diversity and richness in its microbial composition with overrepresentation of facultative anaerobes (Singer *et al*., 2019). However, the current understanding of the neonatal gut microbiome is many levels below the comprehension of the adult gut microbiome (Sanidad and Zeng, 2020). Neonates gut microbiome seems to show more plasticity and adaptability than the adult one, which is reasonably stable (Sanidad and Zeng, 2020). The transition from first colonizers to obligate anaerobes in the neonatal gut is mediated by the decrease of luminal oxygen (Sanidad and Zeng, 2020). In the first month of life, neonatal gut is typically colonized by facultative anaerobes such as *Lactobacillaeceae* and *Enterobacteriaceae* members and shifts to obligate anaerobes in the following months, comprising for instance *Bifidobacterium*, *Bacteroides*, and *Clostridium* (Robertson *et al*., 2019). The foetal microbiota, if present, is thought to firstly originate through maternal vertical transmission or other inflammatory conditions that allow translocation of microorganisms

to the foetal intestine (Ward *et al*., 2018). The neonatal gut environment contains other selective factors for microorganisms, such as pH, availability of digestive enzymes, and the presence of the host's carbohydrates, proteins, and lipids (Dallas *et al*., 2012). However, it is poorly understood how the foetal intestine restricts bacterial colonisation and function.

Toll-like receptors (TLRs) and short-chain fatty acids, among other maternal bacterial products and metabolites, can circulate to the foetus through the placenta and therefore possibly impact the foetal epithelium and immune cellular landscape (Sanidad and Zeng, 2020). This means that the maternal gut microbiome plays a pivotal role in influencing the development of foetal immunity without any microbial translocation event. Although not clear what is behind the foetal gut microbial colonization, recent reports observed dendritic and memory T cells in the foetal gut, whose presence may be induced by microbial antigens (Li *et al*., 2019). This observation strengthens the hypothesis of the utero bacterial colonization. However, in healthy pregnancies, microbial colonization *in utero* is unlikely due to the restricted communication between mother and the foetus. Thus, underlying maternal conditions such as gestational diabetes, urinary tract or vaginal infections may increase the chance of vertical transmission and colonization of the placenta (Sanidad and Zeng, 2020).

Rackaityte and colleagues (2020) recently performed a combined approach using molecular, immunologic, microscopy, cultivation, and *ex vivo* experiments, and observed sparse but viable bacterial presence in the human foetal intestine. The authors also reported that this microbial community had the capacity to limit inflammatory potential through modulation of foetal immune T cell populations. (Rackaityte *et al*., 2020). This modulation event was detected at the mid-gestation period. Specifically, the authors observed enrichment of 18 taxa in foetal meconium belonging mostly to *Micrococcaceae* and *Lactobacillus*. Moreover, when *Micrococcaceae* were abundant, foetal intestines demonstrated distinctive T cell composition and epithelial transcription patterns (Rackaityte *et al*., 2020). To support the authors' claim, it was observed that *Micrococcus luteus* could only be isolated in the presence of monocytes, showing growth in the presence of placental hormones even within antigen presenting cells. Reduced inflammation was observed *ex vivo*, and *M. luteus* was found to harbour genetic markers associated with survival in the foetus (Rackaityte *et al*., 2020). After an experimental approach aiming contamination removal, Rackaityte and colleagues (2020) detected only 17 taxa in 50 foetal intestinal samples, resulting in the conclusion that the foetal gut is a very challenging growth environment for bacteria, making it very difficult to further

investigate their potential mechanist role in foetal development. Specifically, the authors found that *Lactobacillus*, *Micrococcus*, *Gardnerella*, *Megasphaera*, *Corynebacterium*, *Limnohabitans*, *Peptonophilus* and *Rhizibiaceae* are the most enriched taxa in meconium (Rackaityte *et al*., 2020). However, it is possible that even these taxa did not represent meaningful biological signal, and their detection could be explained by another contamination source not investigated in the study. The most probable route of transmission to the foetal gut is from the mothers, as the example of *Micrococcus*, since it is present in the maternal cervicovaginal microbiota (Martin *et al*., 2008, Chen *et al*., 2017). Interestingly, the micrococcal strains isolated by Rackaityte and colleagues (2020) showed strain-specific genes not detected in the genome of vaginal *Micrococcus*. This observation reveals that foetus may harbour niche-specific phenotypes that allow strains to survive the severe conditions found in the foetal intestine. Since *M. luteus* can resist starvation by living in dormant, viable but not cultivable state (Mukamolova *et al*., 2002), it may also survive under conditions characterized by limited nutrition and pregnancy hormone exposure in the foetal intestine. In accordance with this hypothesis, Rackaityte and colleagues (2020) observed that *M. luteus* could still grow in the presence of progesterone and β-estradiol in low-nutrient conditions. Such achievement may be possible as a result of *M. luteus* encoding a ketoisomerase putatively involved in steroid metabolism. Moreover, among other potentially relevant *Micrococcus* activities, it is able to induce expression of LLT1 (the natural ligand for the foetal-specific inhibitory C-type lectin CD161) on antigen presenting cells, which is a mechanism unique in the foetal adaptive immunity (Halkias *et al*., 2019). Therefore, it can be hypothesized that immunological memory to foetal *Micrococcus* starts in utero. In general, these observations reinforce the hypothesis presented in other recent studies that foetal microbiota detected in utero trigger an adaptive immune response (McGovern 2017), such as T cell activation (Halkias *et al*., 2019, Li *et al*., 2019, Schreurs *et al*., 2019). Nevertheless, it is still to be further characterized the full range of mechanisms *Micrococcus* spp., and other species, perform in the harsh environment of foetal intestine and in utero.

In another recent study, de Goffau and colleagues (2019) analysed placental biopsies from 537 pregnant women and concluded that there is no resident microbiome in the human placenta. Additionally, the authors reported that the role of bacterial placental infection on placental-related complications during pregnancy is not significant (de Goffau *et al*., 2019). Among the collected samples was included 318 samples from adverse pregnancies as well as samples from healthy ones (de Goffau *et al*., 2019). The

authors showed that nearly all placental samples harboured no bacterial presence irrespective of pregnancy health status (de Goffau *et al*., 2019). The bacterial signal detected were explained either by contamination during labour and delivery, or contamination acquired through laboratory reagents (Figure 1).

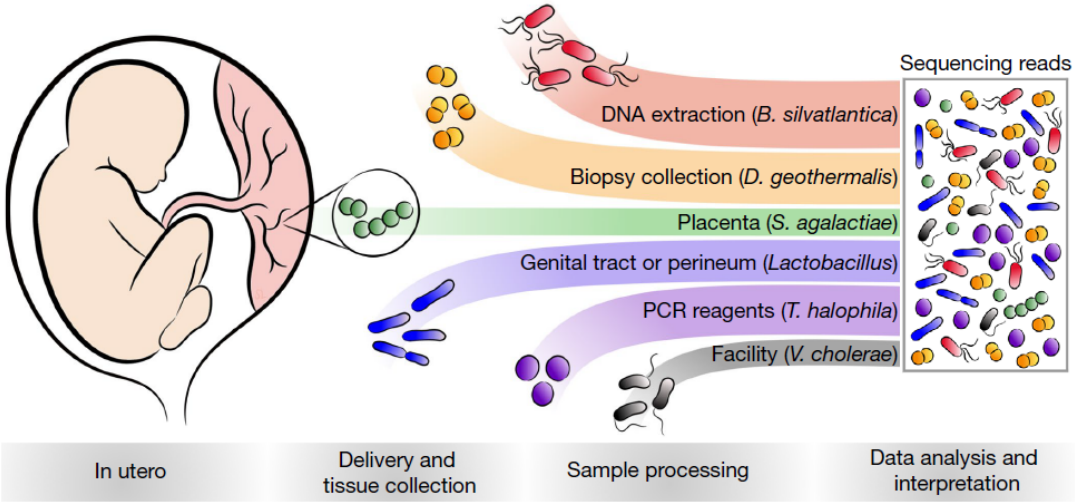


Figure 1 – Overview of the potential sources of bacterial signals detected in human placental samples (de Goffau et al., 2019).

However, 5% of the samples collected before the onset of labour held bacterial presence of *Streptococcus agalactiae*, which signal was not explained by contamination (de Goffau *et al*., 2019). Vertical transmission of *S. agalactiae* to the newborn through passage in mother genital tract during delivery is dramatic because it can lead to fatal sepsis in the infant. Interestingly, de Goffau and colleagues (2019) also found an association between spontaneous preterm birth and the presence of *Ureaplasma* and *Streptococcus anginosus*, and between pre-eclampsia and *Lactobacillus iners*. However, detection of *Ureoplasma* is probably explained by ascending uterine infection as reported before (Abele-Horn *et al*., 2000).

As happens in other body sites such as brain, the placenta represents a barrier to microbes, protecting its innermost biophysical environment and evidence supporting the presence of microorganisms in the foetal gut is not totally accepted. Technical issues such as contamination, and the presence of bacterial DNA in blood may explain the low-biomass microbial content observed. Moreover, there remains a degree of scepticism around colonization of the uterus, the microbiome assembly in C-section deliveries, and the current possibility to generate true germ-free mammals (Blaser *et al*., 2021).

1.2.3 Effect of ethanol on gut microbiota

According to the World Health Organization, 283 million people worldwide (237 million men and 46 million women) had a diagnose related to alcohol use disorders in 2018 (WHO, 2018). Ethanol is typically consumed by humans as alcoholic liquors and our understanding of the effect of the consumption of these type of beverages on the gut microbiota and subsequently on human health and alcohol-induced diseases is limited (Lee and Lee, 2021). One or two daily drinks are considered heavy ethanol use for women and men, respectively, and this level of consumption routine is associated with higher risk of cirrhosis, cardiovascular diseases, and liver (Morgan *et al*., 2004), colon-rectum (Cai *et al*., 2014), and pancreas, malfunction (Jiao *et al*., 2009). Different alcoholic beverages have been found to affect the gut microbial composition in different ways (Queipo-Ortuño *et al*., 2012, Lee *et al*. 2013, Fang *et al*., 2019). For example, red wine increased the alpha-diversity of gut microbiota, with observed increase in abundances of *Prevotella*, *Bifidobacterium*, *B. uniformis*, and *Enterococcus*, which was not observed when beer, cider, spirits, or white wine was consumed (Le Roy *et al*., 2020). Long-term excessive alcoholic drinking can cause alcoholic liver disease in humans manifestations of which are hepatic steatosis, alcoholic hepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma, among others (Gao *et al*., 2011, Wang *et al*., 2016, Seo *et al*., 2020). Alcohol-associated diseases can also comprise direct tissue damage caused by ethanol and its metabolites and by modulatory effects of gut microbiome as well. Still, the effect of alcohol consumption in humans is not completely understood, and further research is needed since dysbiosis, further injuries, and malfunctions in the gut, liver, and brain have not yet been fully described. Preliminary studies in humans indicate that the gut microbiome is dysregulated by the action of psychostimulants, ethanol, and opioids (de Timary *et al*., 2015, Kiraly *et al*., 2016, Wang and Roy, 2017, Meckel and Kiraly, 2019). In the gastrointestinal tract, ethanol is absorbed and then transferred to the liver and lungs where it is metabolized. Throughout its route inside the human body, ethanol and ethanol-related metabolites cause direct toxicity, oxidative stress, and accumulation of fatty acid ethyl esters (Guo *et al*., 2010). Specifically, acetaldehyde is a major factor for promoting ethanol-induced diseases (Salaspuro *et al*., 2017). Thus, ethanol consumption causes a cascade of events. Firstly, dysbiosis happens in gastrointestinal tract with observed increase of Gram-negative bacteria (Malaguarnera *et al*., 2014 Meroni *et al*., 2019), decrease of short-chain fatty acids (SCFA)-producing bacteria (Bjørkhaug *et al*., 2019). Secondly, endotoxins produced by Gram-negative bacteria compromise intestinal barrier integrity causing higher permeability (Patel *et al*., 2015, Tang *et al*., 2015). Enhanced permeability allows

bacterial cells and metabolites to cross and enter the portal and the systemic circulation system, causing damage elsewhere in the body (Stärkel *et al*., 2018). However, ethanol-related morbidity in humans has also been linked to alterations occurring in the gut microbiota (Mutlu *et al*., 2012, Tsuruya *et al*., 2016). Therefore, the effect of ethanol has consequences on the gut landscape and on the microbial composition. Thus, chronic ethanol abuse alters the composition and function of the gut microbiota (Mutlu *et al*., 2012, Wang *et al*., 2018, Bjørkhaug *et al*., 2019, Samuelson *et al*., 2019, Leclercq *et al*., 2020), and has a prejudicial effect on the brain, and cause liver lesions by the disruption of the intestinal barrier (Ferrere *et al*., 2017). However, there are few clinical studies about the impact of gut microbiota in ethanol-dependent humans (Hillemacher *et al*., 2016).

Infants acquire their microbiome most probably from their mothers during pregnancy, then it develops until relative maturation in the first three years of life (Wang *et al*., 2018). Prenatal environment and early postnatal colonization are crucial for a normal general and organ-specific development (Warner, 2019). However, despite the increasing amount of data, there is still no strong evidence about the potential influence of the gut microbiota in alcohol use disorder and neonatal immune, physiological, and neurological development, with the remaining issues surrounding both physical and biological nuances still to be uncovered (Giménez-Gomez *et al*., 2019). Studies in humans showed another aggravated outcome of ethanol intake – the observed increase in *Enterobacteriaceae* abundance while *Lachnospiraceae* and *Ruminococcaceae* decreased. The problem is that the first may produce dangerous endotoxins and the second produces beneficial SFCAs (Bajaj *et al*., 2014, Bajaj *et al*., 2017). Additionally, ethanol consumption was reported to cause a significant decrease of *Bacteroidetes* (Bull-Otterson *et al*., 2013, Qin *et al*., 2014), and an increase in *Corynebacteriaceae* and *Clostridiales* (Harnisch *et al*., 1989, Cericco *et al*., 1996). Despite their pathological relevance, *Bacteroides* species are important for the maintenance of a beneficial interaction with intestinal homeostasis. *Bacteroides* species may also control other competing pathogens by influencing the host immune system (Samuelson *et al*., 2017). Samuelson and colleagues (2017) were able to characterize an ethanol-induced dysbiostic state described as comprising changes in the relative abundances of certain species, and in particular an increase in the abundance of *Bacteroides acidifaciens*, *Bacteroides eggerthii* and *Oscillospria* spp, (both Gram-negative bacteria), and a decrease in *B. uniformis*, *Parabacteroides gordonii*, and *A. muciniphila* (Samuelson *et al*., 2017). In a recent study, Gurwara and colleagues (2020) analysed 97 snap-frozen colonic biopsies from polyp-free individuals. The individuals were categorized by their drinking habits: never drinkers,

former drinkers, light drinkers, and heavy drinkers (Gurwara *et al*., 2020). However, the sampled individuals were all adult men, and the authors performed microbial profiling analyses using the hypervariable region 4 (V4) of the 16S ribosomal RNA, which represent experimental limitations. The selected hypervariable region has shown higher alpha-diversity compared to V1-V2 and V3-V4 regions (Chen *et al*., 2019). Therefore, the results obtained cannot be generalized to women and younger individuals nor compared to microbial profiling data obtained through other hyper variable regions (Rintala *et al*., 2017). In accordance with other reports, Gurwara and colleagues (2020) found that heavy drinkers had the lowest and abnormal/impaired microbial diversity, richness and evenness (Ciocan *et al*., 2018). These drinkers also had the least relative abundance of *Subdoligramulum*, *Roseburia*, *Akkermansia*, *Desulfovibrio*, *Faecalibacterium*, *Sutterella*, and *Lachnospiraceae* spp., and the highest abundance of *Escherichia*, *Haemophilus*, *Tyzzerella*, *Erysipelatoclostridium*, and *Lachnospiraceae* spp. (Gurwara *et al*., 2020). Light drinkers showed the highest relative abundance of *Akkermansia*, *Lachnosporidium*, *Desulfovibrio*, and *Lachnospiraceae* spp. (Gurwara *et al*., 2020). Both never and former drinkers presented similar microbial profiles (Gurwara *et al*., 2020). Additionally, never drinkers and heavy drinkers were in the opposite spectrum of relative abundances of certain species. *Faecalibacterium*, *Subdoligranulum*, *Sutturella*, *Alistipes*, and *Roseburia* were abundantly detected in never drinkers and sparsely in heavy drinkers (Gurwara *et al*., 2020). Never drinkers showed the highest relative abundance of *Bilophila* among all groups (Gurwara *et al*., 2020). Therefore, the authors concluded that all major bacteria significantly followed abundance profiles based on ethanol use (Gurwara *et al*., 2020). In contrast to the reported gut dysbiosis and inflammation in the gut caused by ethanol, some studies reported increase in biodiversity in the faecal microbiota of human drinkers (Kosnicki *et al*., 2019). Inconsistent microbial alterations trends have been spotted in *Proteobacteria*, *Bacteroides*, *Firmicutes*, and *Faecalibacterium*, however they may be explained by different experimental approaches in ethanol concentration, model system, among other experimental factors (Barr *et al*., 2018, Bjørkhaug *et al*., 2019, Kosnicki *et al*., 2019, Zhang *et al*., 2019, Lee *et al*., 2020). However, two independent reports concluded that ethanol alone is not sufficient for the development of alcoholic liver disease, and thus the importance of the gut microbiota in such diseases may be greater than expected (Canesso *et al*., 2014, Llopis *et al*., 2016). In another recent study, Seo and colleagues (2020) analysed faecal samples collected from 212 Korean twins and found an association between butyrate-producing *Roseburia* spp. and alcohol consumption. Additionally, the authors showed that administration of

Roseburia spp. in mice improved alcoholic liver disease (Seo *et al*., 2020). Ethanol consumption seems to reduce *Roseburia* spp. abundance and a shift from a *Ruminococcaceae*-based community to *Veillonellaceae*-based community was observed (Seo *et al*., 2020), which was in accordance with other reports (Hartmann *et al*., 2015). The authors concluded that *Roseburia* spp. alone could significantly improve hepatic steatosis, inflammation, recover gut barrier integrity, and restore the gut microbiota (Seo *et al*., 2020). Moreover, Seo and colleagues (2020) demonstrated that *Roseburia* spp. can improve the overall gut ecosystem, preventing the chances of leaky gut. These results showed the potential of *Roseburia* spp. for application as an alcoholic liver disease biomarker (Hartmann *et al*., 2013, Mir *et al*., 2016, Patterson *et al*., 2017). Thus, clinical trials may be implemented to test the efficacy of a new *Roseburia*-based therapeutics targeting alcoholic liver disease. Recently, Bjørkhaug and colleagues (2019) published a report about a 10-year follow-up on the gut microbiota and metabolic function in patients with ethanol overconsumption. In accordance with other reports, the authors observed a decrease of *Faecalibacterium* and increase of genera *Suturella*, *Holdemania*, *Clostridium*, and *Proteobacteria* phylum in the alcoholic group (Bjørkhaug *et al*., 2019). Moreover, the authors also predicted that genes related to invasion of epithelial cells were more prevalent in overconsumers (Mutlu *et al*., 2012, Bull-Otterson *et al*., 2013, Bjørkhaug *et al*., 2019). The trend detected is supported by the worse results in the nutritional assessment from the alcoholic group, which represents a health risk of increased inflammation (Shin *et al*., 2015, Litvak *et al*., 2017, Bjørkhaug *et al*., 2019). In the case of alcohol use disorder, metagenomic studies on human patients showed that the gut microbiome of affected individuals had a higher proportion of functions related to ethanol metabolism (Dubinkina *et al*., 2017). Even in the absence of ethanol exposure, alcohol use disorder patients still have higher ethanol levels in faeces, confirming the presence of adapted microbes, an ethanol-dependent microbiota (Leclercq *et al*., 2020). Ethanol-dependent microbiota comprises species such as *Clostridium* spp., *Lactococcus* spp., *Turicibacter* spp., and *Akkermansia* spp., and they all share the functional ability to produce and/or metabolize ethanol (Leclercq *et al*., 2020). Therefore, it seems that the gut microbiome plays a role in the development of ethanol addiction.

One of the first indicators that the gut microbiota could be relevant to alcohol-related liver diseases was the observation that disease remission in mice was recapitulated by changes in microbial composition (Hillemacher *et al*., 2016, Ferrere *et al*., 2017). Similar to what has been observed in humans, *Firmicutes* and *Lactobacillus* were reduced, while *Enterococcus*, *Corynebacteriumi*, and *Alcaligenes* spp. increased in the

gut of mice fed with ethanol (Bull-Otterson *et al*., 2013, Sanuelson *et al*., 2017). Additionally, antibiotic treatment in an ethanol-exposed rodent model led to decrease in Gram-negative bacteria and lipopolysaccharides (LPS) which triggered disease remission (Adachi *et al*., 1995). Nevertheless, compounding data from animal studies are describing the preventive effects of gut microbiota on alcoholic liver disease susceptibility (Ferrere *et al*., 2017, Grander *et al*., 2018). Recently, Grander and colleagues (2018) observed improvements in hepatic inflammation and in gut barrier integrity when *A. muciniphila* was administered to ethanol-fed mice. Interestingly, there are different effects on ethanol-induced liver injury at the strain-level, as seen by the protection exerted by genetically engineered interleukin 22-producing *Lactobacillus reuteri* which was not seen with wild-type *L. reuteri* administration (Hendrikx *et al*., 2018). Recently, Li and colleagues (2021) tested the protective effect of *Lactobacillus plantarum* KLDS1.0344 and *Lactobacillus acidophilus* KLDS1.0901 supplementation in a Lieber-DeCarli liquid diet using a chronic alcoholic liver lesion C57BL/6J mouse model. The authors reported a modulatory effect of both strains on gut microbiota, with increase of SCFA producers, decrease of Gram-negative bacteria, and improvement of intestinal epithelial permeability (Li *et al*., 2021). Moreover, ethanol-induced liver inflammation was reduced due to less LPS entering the portal vein (Li *et al*., 2020). In another study, Jiang and colleagues (2020) tested a Lieber-DeCarli diet plus ethanol and supplementation with the newly isolated probiotic strain *Pediococcus pentosaceus* CGMCC 7049 in an experimental alcoholic liver disease mice model. The authors found that ethanol caused gut dysbiosis-related barrier disruption (Jiang *et al*., 2020). Additionally, the authors found that SCFA-producers (*e.g*., *Prevotella*, *Faecalibacterium*, and *Clostridium*) were depleted and pathobiont taxa, such as *Escherichia* spp. and *Staphylococcus* spp., had increased relative abundances (Jiang *et al*., 2020). In the mice supplemented with *P. pentosaceus*, it was observed higher microbial diversity, and stable and normal levels of *Lactobacillus*, *Pediococcus*, *Prevotella*, *Clostridium*, and *Akkermansia* (Jiang *et al*., 2020). In general, *P. pentosaceus* supplementation decreased and restored the *Firmicutes*/*Bacteroides* ratio (Jiang *et al*., 2020). Additionally, the authors hypothesized that the increase of SCFA-producing bacteria promoted by the presence of *P. pentosaceus* could also improve the gut barrier function pointing the strain as a potential probiotic in alcoholic liver disease treatment (Jiang *et al*., 2020). In contrast to other reports, a recent study by Wang and colleagues (2018) observed a slight increase in the gut microbiota diversity after active and forced drinking in ethanol-withdrawal mice models, while a few commensal residents reduced in abundance. In the same study, it was observed an increase in *Firmicutes*/*Bacteroidetes*

ratio in the groups exposed to ethanol (Wang *et al*., 2018). However, the study could not explain the differences found in the gut microbiota observed between the active and forced drinking. For instance, despite a higher microbial diversity detected, some taxa abundances varied (Wang *et al*., 2018). Specifically, members of the *Allistipes*, *Odoribacter* groups (both *Bacteroidetes*) and *Bifidobacterium* (*Firmicutes*) changed their abundance level between active and forced drinking mice models (Wang *et al*., 2018). The discrepancies found in microbial composition in gut microbiome studies when mice models are analysed may be explained by the different ethanol administration regimen performed, as well other factors, such as subjacent liver disease (Bull-Otterson *et al*., 2013, Samuelson *et al*., 2017). Undisputed is the fact that ethanol abuse induces gut dysbiosis in rodent models (Meckel and Kiraly, 2019). Even in cases where FMT was performed, it did not affect alpha-diversity in the gut albeit preventing the gut microbial profile imbalance, namely by depleting *Bacteroides* (Ferrere *et al*., 2017). One possible causistic explanation for the protection against ethanol in ethanol-resistant mice is the overexpression of two defensins in the colon, Reg3β and Reg3γ. Both FMT and pectin-treated mice showed this overexpression (Ferrere *et al*., 2017). However, further research must be performed to elucidate the phenotypical, genetic, and metabolic landscape the gut microbiota members have in different ethanol use and disease scenarios. In alcoholic liver disease settings, an association was found between ethanol intake and fungi overgrowth, in particular *Candida* spp., promoting liver injury (Bajaj *et al*., 2018). Moreover, alcoholic liver disease stimulates a stronger host immune response in humans against fungi, and recent studies in mice have been relating fungi presence with ethanol-induced diseases, like cirrhosis (Yang *et al*., 2017). In any case, the experimental setup, *e.g.*, breeding facilities (Tomas *et al*., 2012), also influences the gut microbiota composition, and some results can be explained by chance, although, ethanol-induced liver lesions are not observed in all cases (Ferrere *et al*., 2017). Additionally, correlation was previously found between an enhanced intestinal permeability and physiological symptoms like depression, anxiety, and ethanol craving (Leclercq *et al*., 2014). Microbial changes also revealed correlation with altered intestinal permeability, however, germ-free mice revealed few depression and anxiety behaviours, and the same was observed in antibiotic-treated mice (Bercik *et al*., 2011, Gacias *et al*., 2016, Wong *et al*., 2016). Previous reports noted that antibiotic treatment in mice after ethanol consumption reduced gut bacterial number, consequently reducing endotoxin levels, and diminished liver inflammation (Chen *et al*., 2015, Lowe *et al*., 2017). The reduction of neuroinflammation derived by antibiotic treatment clearly revealed the importance of the

gut microbiota abundance and pathogen-associated molecular patterns in the gut-brain axis in many contexts, such as ethanol consumption (Lowe *et al*., 2018). In the gut-brain axis context, chronic-to-acute models of ethanol consumption showed that ethanol causes major prejudice in brain and gut, namely neuroinflammation in CNS, and cytokines overexpression in the small intestine (Lowe *et al*., 2018). The relevance of the gut microbiota load on these processes is shown by the decrease in inflammatory levels during antibiotic treatment. However, antibiotic treatment also increased the levels of inflammasome components and cytokines processed by the inflammasome found in CNS and gut, which may be considered contradictory events (Lowe *et al*., 2018).

1.2.3.1 Effect of ethanol on gut microbiota and cancer onset

Generally, alcohol consumption is considered an integral part of the western lifestyle, and it is a major risk factor for colon carcinogenesis, increasing the chance of colorectal cancer by about 60% (Slattery, 1997, Bishehsari *et al*., 2017). Individuals with habitual heavy ethanol-drinking behaviours are more prone to cancers, particularly in the upper aerodigestive tract (combination of organs and tissues present in the respiratory tract and the upper part of the digestive tract), liver, and colorectum (Seitz and Becker 2007, Mizoue *et al*., 2008, Salaspuro, 2009). Colorectal cancer is the fourth most common cancer globally, accounting for 600,000 deaths per annum (Roswall *et al*., 2015, Kolligs, 2016) and it is the third and second most diagnosed cancer worldwide among men and women, respectively (Kolligs, 2016).

Ethanol stays in the blood after drinking before being metabolized (Halsted *et al*., 1973, Nosova *et al*., 2000, Zakhari 2006) and it is responsible for much of the increased reactive oxygen species (ROS) levels found in ethanol consumers since it diffuses with oxygen from the colorectal submucosa to mucosa cells (Yokoyama *et al*., 2008, Tsuruya *et al*., 2016). Higher ROS levels have been extensively linked to cell death, via oxidation of DNA, proteins, lipids, and many other cellular components (Patlevič *et al*., 2016). Acetaldehyde is one of the major ethanol-related metabolites and it represents a risk factor due to its low minimum mutagenic concentration, being considered a Group-1 carcinogen (Chadwick and Goode, 2007, Seitz and Stickel, 2010, IARC, 2012). Colon microbiota was recently described as a mediator in colon carcinogenesis (Rossi *et al*., 2018). Bacterial catalase, and ethanol and aldehyde dehydrogenase are examples of the activity regulation factors that bacteria such as *Enterobacteriaceae* members transcribe under ethanol stress (Elamin *et al*., 2013). Therefore, when ethanol is ingested, acetaldehyde accumulation in colon occurs, mainly due to the differential activity and mediation of

bacterial enzymes (Salaspuro, 1997, Nosova *et al*., 1998, Elamin *et al*., 2013, Tsuruya *et al*., 2016). Therefore, acetaldehyde-accumulating bacteria living in the colorectal mucosa surface can directly modulate the pathogenesis of ethanol-related colorectal cancers (Nosova *et al*., 1996, Salaspuro *et al*., 1999). However, it is not yet clear if dominant and obligate gut anaerobes produce or metabolize acetaldehyde, nor what is happening under aerobic conditions (Tsuruya *et al*., 2016). Ultimately, acetaldehyde and other metabolites generated by ethanol metabolism in the gut promote a series of 'cancer-inducing cascades', such as DNA-adduct formation, epigenetic alterations, epithelial barrier dysfunction, immune modulatory effects, and oxidative stress and lipid peroxidation (Rossi *et al*., 2018). To study in more depth the role of the gut microbiota when ethanol is consumed, Tsuruya and colleagues (2016) used a culture-dependent approach with faecal samples from Japanese patients under treatment for ethanolism and found that the genera *Ruminococcus*, *Collinsella*, *Prevotella*, *Coriobacterium*, and *Bifidobacterium* increased the production of acetaldehyde in the presence of ethanol. Moreover, *Ruminococcus* was also able to produce acetaldehyde under aerobic conditions in the colorectal mucosa (Tsuruya *et al*., 2016). These results are in accordance with what has been previously reported (Swidsinski *et al*., 2002, Swidsinski *et al*., 2005, Vesterlund *et al*., 2005, Macfarlane and Dillon, 2007, Chen *et al*., 2012, Mutlu *et al*., 2012). Tsuruya and colleagues (2016) also observed enhanced ethanol oxidation and acetaldehyde production by *Ruminococcus* and *Collinsella* in the presence of oxygen and ROS in the colorectal mucosa. Thus, it might be speculated that similar metabolic and phenotypic patterns are present in other species inside gastrointestinal tract, as exemplified elsewhere by the association between lactic acid bacteria and colorectal mucosa (Swidsinski *et al*., 2002, Swidsinski *et al*., 2005, Vesterlund *et al*., 2005, Macfarlane and Dillon, 2007, Chen *et al*., 2021). Intriguingly, these observations focusing on lactic acid bacteria pose questions about their potential use as probiotics for heavy ethanol drinkers. Summarizing the global ethanol-cancer interface potentially modulated by the gut microbiome, Song and colleagues (2020) published a review recently where they clustered the major gut microbial compositional alterations associated with colorectal cancer (Table 1).

Table 1 – Differential changes observed in taxa detected in the microbial profiles of colorectal cancer patients (adapted from Song et al., 2020).

Country of the study	**Taxa/indexes higher than control groups**	**Taxa/indexes lower than control groups[1]**	**Reference**
Belgium	*Clostridium leptum* and *Clostridium coccoides* subgroups' diversity	Nd	Scanlan *et al*., 2008
France	*Bacteroides*/*Prevotella*	Nd	Sobhani *et al*., 2011
USA	*Fusobacterium*	Nd	Kostic *et al*., 2012
USA	87 different taxa, including potential pathogens *Pseudomonas*, *Helicobacter*, and *Acinetobacter*	5 different OTUs: *Butyrivibrio* (twice), *Oligotropha*, *Algibacter*, *Succinispira*	Sanapareddy *et al*., 2012
USA	*Fusobacterium*, *Porphyromonas*	α-diversity, *Clostridia*	Ahn *et al*., 2013
USA	*Bacteroides fragilis*, *Fusobacterium*, *Porphyromonas*	Butyrate-producing bacteria	Zackular *et al*., 2014
France	*Bacteroidetes*, *Fusobacteria*, *Providencia*, virulence-related genes	Nd	Zeller *et al*., 2014
Austria	*Bacteroides dorei*, *Bacteroides vulgatus*, *Escherichia coli*, *Fusobacterium*	*Lactobacillus*, *Bifidobacterium*	Feng *et al*., 2015
USA and France	*Fusobacterium*, *Porphyromonas*, *Clostridia*	nd	Vogtmann *et al*., 2016
USA, Canada	*Porphyromonas assaccharolytica*, *Peptostreptococcus stomatis*, *Parvimonas micra*, *Fusobacterium nucleatum*	*Lachnospiraceae*	Baxter *et al*., 2016
USA	*Bilophila*, *Desulfovibrio*, inflammatory members of *Mogibacterium* gen.	*Veillonella*, *Clostridia* class, *Bifidobacteriales* order	Hale *et al*., 2017
Denmark, France, Austria	*Peptostreptococcus stomatis*, *Fusobacterium nucleatum*, *Parvimonas micra*, *Sofobacterium moorei*	nd	Yu *et al*., 2017

Table 1 – (continued)

Country of the study	Taxa/indexes higher than control groups	Taxa/indexes lower than control groups	Reference
China	*Fusobacterium nucleatum*, *Clostridia hathewayi*	*Bacteroides clarus*	Liang *et al*., 2017

[1] – nd, not defined

Moreover, epidemiological evidence has been reported for certain taxa in terms of their association with colorectal cancers, such as *Fusobacterium nucleatum*, enterotoxigenic *Bacteroides fragilis*, *pks*+ *E. coli* (*i.e.*, *E. coli* encoding a set of enzymes that synthesize colibactin through the pathogenicity island *pks* (Nougayrède *et al*., 2006)) and SCFA-producing bacteria (Arthur *et al*., 2012, Castellarin *et al*., 2012, Kostic *et al*., 2012, Donohoe *et al*., 2014, Mehta *et al*., 2017, Chung *et al*., 2018, Makki *et al*., 2018, So *et al*., 2018, Brennan and Garrett 2019, Wilson *et al*., 2019, Yachida *et al*., 2019, Sears *et al*., 2020). Therefore, many taxa have been reported as implicated in cancer onset, however with geographical differences (*Bacteroides* spp. decrease in abundance in a Chinese study but appear abundant in studies from France, USA, and Austria) and underrepresentation of studies from non-westernized countries (Table 1). In conclusion, the gut microbiome can be analysed as predictive biomarkers for colorectal cancer cases since modifications in its composition have been associated with the disease onset (Song *et al*., 2020).

1.2.3.2 Neurodevelopment and behavioural impact of ethanol intake and gut dysbiosis

The most vulnerable period for CNS development is during the first half and the last two-thirds of the first gestational trimester, and for brain growth, the third trimester (Tiew *et al*., 2020). Studies using rodent models showed that acceleration of brain growth occurs in the early postnatal period (Dobbing and Sands, 1979, Bockhorst *et al*., 2008). During these periods, ethanol can have increasingly teratogenic effects. In the perinatal period, gut colonization and neuronal organization are very dynamic, representing a crucial window for the normal neurodevelopment of the organism (Bockhorst *et al*., 2008). Although recent evidence has linked gut dysbiosis to IBD (Hold *et al*., 2014), and certain cancers, it is not completely understood what the relationship between dysbiosis and neurological disorders is (Bell *et al*., 2018). However, there is evidence of the importance of an early neurodevelopment window where the influence of the gut microbiota is evident (Heijtz *et al*., 2011).

Neurogenesis and neural activity are among the neurodevelopmental processes affected by the gut microbiome (Monteggia *et al*., 2004, Sudo *et al*., 2004, Neufeld *et al*., 2011, Clarke *et al*., 2013). They are affected by the aberrational pattern levels of neurotrophic factors, neuropeptides, and neurotransmitters (Neufeld *et al*., 2011, Clarke *et al*., 2013, Desbonnet *et al*., 2015, Fröhlich *et al*., 2016). Moreover, the gut microbiome can further impact host neurological conditions through changes in signalling pathways (Tiew *et al*., 2020). One such case is the exaggerated glucocorticoid response explained by the hypothalamic-pituitary-adrenal dysregulation in germ-free and antibiotic treated mice (Tiew *et al*., 2020). Ultimately, this dysregulation is associated with observed behavioural settings related to social activity (Desbonnet *et al*., 2014, Buffington *et al*., 2016, Leclercq *et al*., 2017), anxiety (Neufeld *et al*., 2011, Luo *et al*., 2018), and cognitive function and depression (Desbonnet *et al*., 2015, Fröhlich *et al*., 2016, Leclercq *et al*., 2017, Ceylani *et al*., 2018, Luo *et al*., 2018, Zhai *et al*., 2018, Lukić *et al*., 2019). Surprisingly, germ-free mice did not reveal equivalent protection for alcohol-induced liver damage as antibiotic treatment mice (Chen *et al*., 2015). Although this contradiction may be explained by the pivotal role gut microbiota may have during the early development, exerting a beneficial influence in the immunologic responses to ethanol intake.

Ethanol can largely contribute to the breakdown of the intestinal barrier integrity by inducing inflammatory signalling, causing gut bacterial dysbiosis, disturbing luminal homeostasis (Hartmann *et al*., 2013, Hartmann *et al*., 2018), increasing enterocyte cellular stress, and by shifting structural proteins regulation (Rao, 2008). Ethanol addiction impacts several neurological processes, like myelination, neurotransmission, inflammation, as well metabolic alterations related to behaviour, for instance depression and abnormal sociability. Ethanol exerts influence on TLRs, both directly, by affecting correct TLR signal transduction (Szabo *et al*., 2007, Fernandez-Lizarbe *et al*., 2013), and indirectly, by recognizing endogenous (Akira *et al*., 2004, Yang *et al*., 2010) and exogenous danger signals (Park and Lee, 2013). Studies reported that absence of TLRs in knockout and knockdown experiments in mice resulted in inflammatory protection in the liver (Uesugi *et al*., 2001) and brain (Lippai *et al*., 2013, Pascual *et al*., 2015, Alfonso-Loeches *et al*., 2016) after ethanol intake. Similar results were obtained when mice were treated with antibiotics with the intent to reduce the amount of bacterial LPS (an exogenous signal), suggesting that the gut microbiota has a role in triggering tissue inflammation in the liver and brain when ethanol is consumed (Uesugi *et al*., 2001). The connection between gut dysbiosis and ethanol addiction is that the former is also

associated with relevant metabolic alterations and neurological processes (Leclercq *et al*., 2020). Besides, ethanol can easily cross the blood brain barrier (BBB) (Tiew *et al*., 2020). Morphological changes in microglia and proinflammatory gene expression are examples of neuroinflammation events caused by persistent ethanol intake (Tiew *et al*., 2020), and it seems that gut bacteria are relevant in anxiety-like behaviours associated with ethanol withdrawal where the intestinal permeability plays a crucial role (Bala *et al*., 2014, Xiao *et al*., 2018,). Recently, Cryan and colleagues (2019) published a meticulous review on the gut-brain axis topic. Recent advances in the field of the gut-brain axis have shed light into the importance of gut microbiota in many brain and/or neurological disorders such as autism, anxiety, obesity, schizophrenia, Parkinson's, and Alzheimer's disease (Cryan *et al*., 2019). Despite the increase of studies about the gut-brain axis, translational knowledge in further development of microbial-based and conventional interventions is still deficient to provide practical solutions for neuropsychiatric disorders (Cryan *et al*., 2019).

So far, different microbiota manipulations have been tried whose aim was to investigate the microbiota-gut-brain axis. These complementary approaches include germ-free rodent models (Luczynski *et al*., 2016, Luk *et al*., 2018), antibiotic-induced gut microbiota depletion (Desbonnet *et al*., 2015, Staley *et al*., 2017, Guida *et al*., 2018), prebiotic and probiotic supplementation (Burokas *et al*., 2017, Fukui *et al*., 2018, Grimaldi *et al*., 2018, Kao *et al*., 2018, Kazemi *et al*., 2019, Tabouy *et al*., 2018), gastrointestinal targeted infection (Harris *et al*., 2017, Zuo *et al*., 2018), and faecal microbiota transplantation (Cryan and Dinan, 2015, Sherwin *et al*., 2016, Zhou *et al*., 2017, Singh *et al*., 2018). These approaches have been tried in attempts to study the relevance of gut microbiota in neurotransmitter signalling, synaptic plasticity, myelination, neurogenesis, as well as cognition and social impairments (Diaz Heijtz *et al*., 2011, Neufeld *et al*., 2011, Ogbonnaya *et al*., 2015, Hoban *et al*., 2016, Luczynski *et al*., 2016).

Gut microbiota can communicate with the brain through the action of SCFA, branched chain amino acids, peptidoglycans, among other microbial metabolites (Cryan *et al*., 2019). Communication happens via diverse routes including the immune system, tryptophan metabolism, the vagus nerve, and the enteric nervous system (Cryan *et al*., 2019). The enteric nervous system is an interface network of neurons between the gut microbiota and the host that respond both directly and indirectly to the microbiota and its metabolites. The enteric nervous system is also responsible for gut motility and fluid movement, among other gut functions (Furness, 2012). Communication between enteric

nervous system and gut-brain with CNS happens through intestinofugal neurons to sympathetic ganglia with sensory information traveling via extrinsic primary afferent neurons that follow spinal and vagal afferent routes (Furness, 2012). Therefore, signalling and molecular factors derived from the gut lumen, and eventually from the gut microbiota, may impact the nearby landscape and CNS (Cryan *et al*., 2019) (Figure 2).

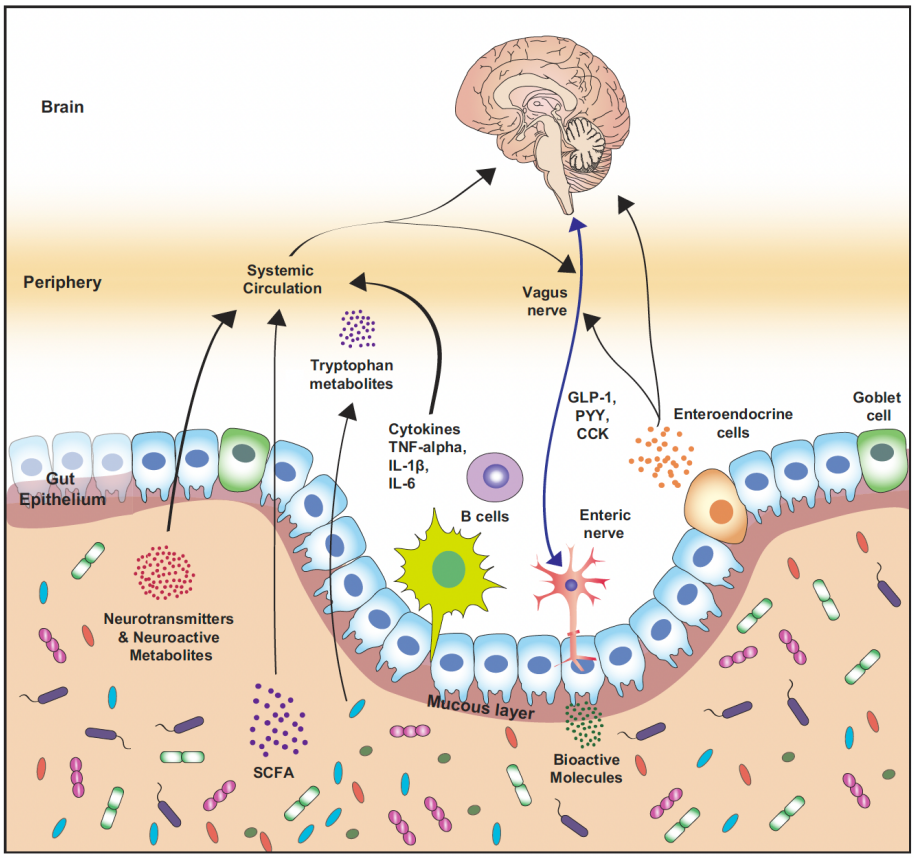


Figure 2 – Scheme displaying the bidirectional communication pathways known in the gut-brain axis (Cryan et al., 2019). CCK, cholecystokinin; GLP-1, glucagon-like peptine-1; IL, interleukin; PYY, peptide YY; TNF, tumor necrosis factor; SCFA, short-chain fatty acid.

The vagus nerve is the 10th cranial nerve and the fastest and most direct route that connects the gut to the brain (Cryan *et al*., 2019). It conducts fundamental information originating in the gastrointestinal tract, respiratory, and cardiovascular system (bottom-up), and transmits feedback to the viscera (top-down) through afferent (80%) and efferent (20%) fibers (Agostini *et al*., 1957, Prechtl and Powley, 1990, Cryan *et al*., 2019). Microglia are described as differentiated macrophages residing inside the brain that constitute between 5-15% of total brain cells (Pelvig *et al*., 2008). Mounting evidence is giving attention to the role microglia have in brain development and homeostasis as well as in the pathogenesis of many neurodegenerative and neurodevelopmental disorders

(Salter and Stevens, 2017). Initially, the gut microbiome critically affects microglial development and function (Mayer *et al*., 2015, Fung *et al*., 2017), thus influencing microglial maturation, identity, and function. However, the gut microbiome can also impact other features in the CNS (Abdel-Haq *et al*., 2019). Indeed, the gut microbiome has recently been studied in more detail for further identification and mapping of direct and indirect associations it may has with microglia (Abdel-Haq *et al*., 2019). Still, the genetic cellular formation and maturity development of microglia is not well understood (Abdel-Haq *et al*., 2019). Among other sources, the gut can potentially alter microglia activity in the CNS despite the protection provided by the BBB (Abdel-Haq *et al*., 2019). Recently, it has been observed that signals produced by maternal gut microbiota may affect the development of foetal microglia around the delivery period (Thion *et al*., 2018). Erny and colleagues (2015) reported that microglia show an abnormal response to several immunostimulants, such as LPS and the lymphocytic choriomeningitis virus, in germ-free mice and in mice only colonized with *Bacteroides distasonis*, *Lactobacillus salivarius*, and *Clostridium* cluster XIV. The authors hypothesized that gut microbiota absence or dysbiotic/reduced state triggers an immature gene expression profile thus explaining the deficient microglia phenotype (Erny *et al*., 2015). Therefore, the gut microbiota may be necessary for the microglia-mediated immune response against both microbes and microbial signals targeting the CNS (Abdel-Haq *et al*., 2019). In summary, the gut microbiome influences microglial complete maturation and function since early ages, and it communicates with CNS through microglia. There are two main facts that supports these observations. First, the gut microbiome regulates microglial function, and it can trigger pathogenic microglial phenotypes increasing the chance of neurological disease onset. And second, gut dysbiosis is thought to be sufficient to disturb microglial function and it is a feature significantly observed in several neurological conditions in which microglial abnormal function is a contributing factor, including autism spectrum disorder and Parkinson's disease (Hsiao *et al*., 2013, Sampson *et al*., 2016).

The impact of the gut microbiome in humans goes beyond contributions to neurological disorders and can be observed in social and behavioural settings (Leclercq *et al*., 2020). Recently, Leclercq and colleagues (2020) performed preclinical studies and showed that there was a link between "leaky gut" (increased intestinal permeability) and social behaviour impairments in humans. This link goes along with the known relation between the gut microbiome and social behaviour in multiple animal models, including mice (Sherwin *et al*., 2019). The presence of bacteria in mesenteric lymph nodes is indicative of disrupted intestinal permeability and it happens right after chronic ethanol

consumption is interrupted (Giménez-Gómez *et al*., 2019). None of these events were observed during the binge ethanol consumption pattern. Moreover, there is a relation between gut alterations and impaired sociability in alcohol use disorder (Leclercq *et al*., 2020). Studies showed that transplanting gut microbiota from alcohol use disorder' patients to mice blocks the neuroprotective effect of β-hydroxybutyrate (BHB) (Leclercq *et al*., 2020). This happens because certain bacterial genera start to produce ethanol and a reduction in lipolysis is observed. These two events are associated with a lower hepatic synthesis of BHB. Moreover, inoculation of dysbiotic ethanol-dependent microbiota from ethanol-dependent patients was also found to lead to a reduced synthesis of BHB (Leclercq *et al*., 2020). This reduced BHB availability may be the first cause of alterations in brain functions and behaviour in mice. Additionally, the role of gut microbiota in metabolic and behavioural disorders is evident as observed when microbiota transplantations from human to mice recapitulate neurological disorders' phenotypes (Leclercq *et al*., 2020). Recent evidence from FMT studies has shown that alterations in the gut microbiota is sufficient to explain aggravated neurological and psychological symptoms in a range of diseases, such as multiple sclerosis (Berer *et al*., 2011), Parkinson's (Sampson *et al*., 2016), Alzheimer's (Fujii *et al*., 2019), depression (Zheng *et al*., 2016), schizophrenia (Zheng *et al*., 2019), and others (Needham *et al*., 2020). However, FMT studies with humanized gut microbiota in mice remain speculative if they are kept in a preclinical setting. Thus, it is necessary to further investigate the clinical relevance of these type of findings.

Regarding the gut mycobiome, it represents a minor part of the gut microbiota, but it has also implications on the gut-brain axis (Tiew *et al*., 2020). Unusual high *Candida* abundance occurs in schizophrenia, autism and Rett Syndrome, and observations of fungal gut dysbiosis have been reported in various neurological disorders (Severance *et al*., 2012, Severance *et al*., 2016, Strati *et al*., 2016, Strati *et al*., 2017). Nevertheless, further research is necessary to decrease the speculative observations about the biological mechanisms that gut fungi present to the neurology of the host (Tiew *et al*., 2020). Markey and colleagues (2020) recently investigated the impact of *C. albicans* on the gut-brain axis of female adolescent mice. The authors observed that *C. albicans* increased anxiety-like behaviours and dysregulation of basal stress hormone corticosterone production (Markey *et al*., 2020), which resembles the neuroendocrine phenotypes found when mice are infected with the pathogen *Citrobacter rodentium* (Lyte *et al*., 2006).

In conclusion, recent observations within the field of the gut-brain axis have uncovered the possible communication route where gut microbes modulate the

pathogenesis of neurological diseases, for instance by shaping microglial identity and function (Abdel-Haq *et al*., 2019).

1.2.3.3 Foetal Alcohol Spectrum Disorder

Foetal alcohol spectrum disorder (FASD) is a term that has been used for at least 20 years and its consensus definition is: "*an umbrella term describing the range of effects that can occur in an individual whose mother drank alcohol during pregnancy (…) These effects include physical, mental, behavioural, and/or learning disabilities with possible lifelong implications (…) The term FASD encompasses all other diagnostic terms, such as Foetal Alcohol Syndrome (FAS), and is not intended for use as a clinical diagnosis*." (Petrelli *et al*., 2018). This definition was revised in 2016 in order to turn FASD into a diagnostic term. However, to diagnose FASD it is required a multidisciplinary team with physical and neurodevelopment expertise. FASD has now two diagnostic categories: "*i) FASD with sentinel facial features* (Figure 3) *and evidence of impairment in three or more identified neurodevelopmental domains, with prenatal alcohol exposure (PAE) either confirmed or unknow, and; ii) FASD without sentinel facial features, with evidence of impairment in three or more identified neurodevelopmental domains, and confirmed PAE*" (Petrelli *et al*., 2018). A third category is also specified in a way that describes children "*at risk for neurodevelopmental disorder and FASD associated with PAE*" (Szabo, 2015).

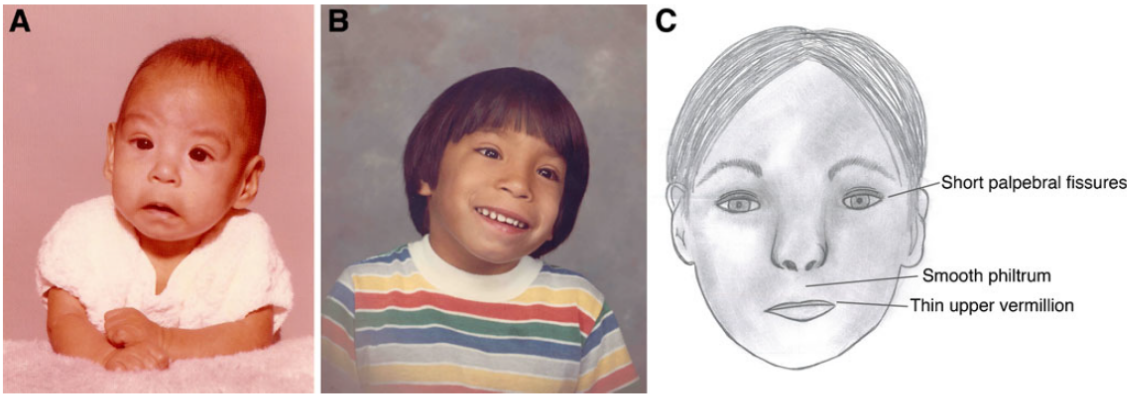


Figure 3 – Photo of a 4-month-old infant with FAS (A). Photo of same child at 5 years of age (B). Cartoon showing the major facial characteristics used in the diagnosis of FAS (C) (Riley et al., 2011).

The effects of PAE on neurogenesis manifest during all stages of gestation and last throughout life, predominantly affecting specific areas of the brain with the detrimental effects on a variety of processes including neurogenesis, differentiation, and

synaptogenesis (Dobbing and Sands, 1979, West *et al*., 1994, Guerri 1998, Almeida *et al*., 2020). Dosage, duration, and timing of PAE, the genetic setting provided by the mother, nutrition and metabolism, epigenetic factors (Kaminem-Ahola *et al*., 2010, Kleiber *et al*., 2014), foetal exposure to stress (Glavas *et al*., 2007, Uban *et al*., 2013, Raineki *et al*., 2014), and the mothers' capacity to metabolize ethanol (Ramchandani *et al*., 2001, Riley *et al*., 2011) are all possible causes of FASD (Petrelli *et al*., 2018). There is a possible association between PAE and neurodevelopmental disorders, as the evidence show that, in PAE mice models, on first-to-third trimester there were observed delayed developmental features, such as genitive geotaxis (*i.e.*, disoriented genito-pelvic movement away from the gravitational force), auditory startle (*i.e.,* reflexive and sudden response to noise), cliff aversion (*i.e.*, increased avoidance of floor edges driven by fear), and air righting (*i.e.*, a complex motor function that promotes general stability mediated by vestibular pathways, spinal interneurons, proprioceptive afferents and motor neurons) (Kleiber *et al*., 2011, Fletcher and Mentis, 2017). Similar observations have been made in children with FASD where brain maturation is impaired, besides their smaller brain and basal ganglia sizes (Inkelis *et al*., 2020). Moreover, an additional association between such small volume of the basal ganglia and abnormal neurobehavioral phenotypes, such as impaired visuospatial abilities and executive functioning, was previously established (Mattson *et al*., 1996, Riley *et al*., 2005). Additionally, mice PAE models showed affected hypothalamus, hippocampus, and cerebellum, and pro-inflammatory signalling, and abnormalities at the neuronal development level (Petrelli *et al*., 2018). The same brain regions are also affected in individuals presenting FASD (Volgin, 2008, Coleman Jr *et al*., 2012, Drew *et al*., 2015, Smiley *et al*., 2015). Microglial cell death is a common consequence mediated by ethanol consumption in mice (Kane *et al*., 2011). The importance of microglia for normal brain development is seen for instance in processes related to brain synaptic function (Paolicelli *et al*., 2011). Ethanol can directly activate microglia through mediation from TLR2 and TLR4 signalling. Afterwards, ROS and nitric oxide production, and a cascade of cytokine levels are triggered (Fernandez-Lizarbe *et al*., 2013). Interestingly, pro-inflammatory microglia that had been activated by ethanol appear to have a role in facilitating the clearance of developing hypothalamic and Purkinje neurons after PAE-related apoptosis (Kane, *et al*., 2011, Kleiber *et al*., 2011).

In summary, altered brain development and behavioural changes are observed consistently as result of inflammation and microglial activation in third trimester PAE mice models. These prejudicial effects in the foetal brain during development are similar to those observed in individuals with FASD. Moreover, acute ethanol exposure during

the third trimester specifically can cause reduced brain volumes (microcephaly), significant loss of GABAergic and pyramidal neurons in the cortex, and behavioural effects like increased hyperactivity later in life (Wilson *et al*., 2016).

Expression of tumour necrosis factor-α and interleukin-1β is increased in the cerebral cortex and cerebellum (Alfonso-Loeches *et al*., 2010). Simultaneously, microglial cell proliferation also occurs (Pascual *et al*., 2014, Pradier *et al*., 2018). Both events are triggered by increased and prolonged ethanol consumption. Therefore, such activation of microglia and astrocytes will typically result in a higher level of neuronal cell death. Overall, the effects of PAE on the neurodevelopment can be observed in a myriad of neurological processes, for instance involving microglia and astrocytes, promoting above normal levels of inflammatory response during the CNS development (Kane *et al*., 2021).

1.3 Nanopore sequencing

1.3.1 Technology overview

The first nanopore sequencer was deployed in 2014 by Oxford Nanopore Technologies (ONT) and thereafter a wide variety of different experimental and data analyses approaches using this technology have been tested (Ip *et al*., 2015, Deamer *et al*., 2016, Jain *et al*., 2016, Wang *et al*., 2021). Nanopore sequencing has revolutionized basic and applied research by enabling long and ultralong reads from DNA or RNA molecules in real time, far more cheaply and easily than with remaining standard sequencing platforms (Wang *et al*., 2021). However, despite the recent advances and growth of the number of scientific publications using nanopore sequencing, the technology still faces several challenges, including a relatively high error-rate (Wang *et al*., 2021). The basic concept behind nanopore sequencing chemistry is the engineering of biological nanopores embedded in an electrically resistant polymer membrane. These nanopores allow the passage of ions and molecules in solution when voltage is applied across the membrane. Nanopore sequencing works by measuring electric current changes produced when DNA or RNA strands pass through nanopores embedded in the membrane. Each set of base imposes a particular current variation which can be detected and defined in real time – for the R9 flowcells the set is composed of 5 bases. An event is called every time a current variance is measured, *i.e.*, when a base was identified passing the nanopore, and it is defined by the mean and variance of the current and by the event duration (Boža *et al*., 2017, Edwards *et al*., 2019). DNA tend to flow naturally through nanopores due to the potential differential signal of the membrane, causing a

disruption in the current, and careful library preparation is necessary to prepare the DNA templates in such a way that it can pass in a mediated way through the nanopores, allowing accurate detection and measurement. Currently, DNA can pass nanopores at a speed of 425 bases per second, and, in contrast to other sequencing technologies, there is no theoretical upper maximum fragment length size for sequencing (Brown, 2015, Brown and Clarke, 2016). The changes in measured current are meticulously determined, and the observed differences are associated with each single base. Changes in the current are detected, measured, and recorded, forming the so-called "squiggles". Squiggles are the raw compositional signal nanopore sequencing generate. Squiggles are recorded as FAST5 files by the MinKNOW software, proprietary to ONT, where the first sequencing stages happen. The process that translates every event (squiggles) into biological signal is called basecalling. Basecalling is not a trivial computational process, and it is crucial for generation of sequence reads quality (Santos *et al.,* 2020). In practice, basecalling can be summarily explained as the conversion of FAST5 files containing sequencing raw data into FASTQ files composed of sequenced bases, which provides a more convenient format for downstream analyses. The first ONT basecallers used Hidden Markov Models for basecalling, but currently machine-learning strategies are applied in all basecallers, including *guppy* (which is the standard basecaller for ONT), and *deepnano* (Boža *et al*., 2017, Teng *et al*., 2018). Such machine-learning strategies rely on neural networks with the possibility to be trained with real sequencing data (Santos *et al*., 2020). These modern approaches have improved nanopore sequencing reads' quality and minimizing errors related to base modifications or indels, commonly present in raw data (Wick *et al*., 2019).

ONT platforms generate long reads which is the hallmark of the platform, and in general, sequencing longer reads allows higher level of taxonomic and phylogenetic resolution since it is assumed that more information is thereby collected (Schneider and Dekker, 2012, Feng *et al*., 2015, van Dijk *et al*., 2018, Bahram *et al*., 2019, Pearman *et al*., 2020). Pacific Biosciences sequencing platform can also generate longer reads (Nilsson *et al*., 2019). However, PacBio sequencers can also generate multi-kilobase reads, but are large, expensive devices requiring power and network infrastructure. It thus has the requirement of sending samples for sequencing in fully equipped centralized research facilities. Still, Pacific Biosciences Single-Molecule, Real-Time technology has a substantially lower error-rate in comparison with ONT (Nilsson *et al*., 2019). Given the novelty of ONT platforms, the protocols are still under development and/or optimization. For researchers who are exploring this technology it is expected a greater effort to

undergo library preparation and data analysis for obtaining readable and comparable results.

The most convenient feature of ONT platforms is the portability of some devices, such as Flongle and MinION. This feature allows the execution of nanopore sequencing protocols with relatively ease in remote locations, including in the wild (Wawina-Bokalanga *et al*., 2019) and clinical settings (Imai *et al*., 2018, Charalampous *et al*., 2019), with the possibility to acquire real time data to allow faster and improved decision-making. Therefore, nanopore sequencing is in many ways providing a revolution in the ecological and biomedical fields.

The main steps for high-throughput sequencing approaches are the following: sample preparation; selection of genetic barcoding markers, primer selection and PCR conditions setup (in targeted approaches); controls and technical replication; sequencing library preparation; quality-filtering of data generated; sequence clustering and Operational Taxonomic Units generation (or other equivalent approaches); sequence-based taxonomic identification; and data processing and analysis pipelines (Nilsson *et al*., 2019). Common biases are mostly introduced in the following experimental steps: DNA extraction; PCR cycle; sequencing library preparation; basecalling; and further bioinformatics pipelines (Nilsson *et al*., 2019). In every pipeline node there are many bioinformatic tools to be considered depending on the biological problem being addressed and the level of curation and maintenance of each particular tool (Table 2) (Santos *et al*., 2020).

Table 2 – Summary of the most used bioinformatic tools performed during nanopore sequencing approaches.

Bioinformatic pipeline process	Bioinformatic tool	Reference
Basecalling	guppy	ONT proprietary
	deepnano	Boža *et al*., 2017
	chiron	Teng *et al*., 2018
Sequencing report	nanoplot	De Coster *et al*., 2018
	poRe	Watson *et al*., 2015
	pauvre	github.com/conchoecia/pauvre
	Poretools	Loman and Quinlan 2014
	pyroQC	https://github.com/a-slide/pycoQC
Demultiplexing	guppy	ONT proprietary
	qcat	github.com/nanoporetech/qcat
	porechop	github.com/rrwick/Porechop

Table 2 – (continued).

Bioinformatic pipeline process	Bioinformatic tool	Reference
Filtering/Trimming	nanofilt	De Coster *et al*., 2018
	filtlong	github.com/rrwick/Filtlong
	porechop	github.com/rrwick/Porechop
Taxonomic assignment	minimap2	Li, 2018
	What's In My Pot? (WIMP)	ONT proprietary
	centrifuge	Kim *et al*., 2016
	LASTZ	github.com/lastz/lastz
	BLAST	Eric *et al*., 2014
Clustering	nanoclust	Calus *et al*., 2018
	CARNAC-LR	Marchet *et al*., 2019
Data exploration	pavian	Breitwieser and Salzberg, 2020
	PHINCH	Bik, 2014
	krona	Ondov *et al*., 2011
	MEGAN6	Huson *et al*., 2016
	MicrobiomeAnalyst	Dhariwal *et al*., 2017

EPI2ME is a cloud-based platform developed by ONT to provide multiple analyses pipelines. The pltaform contains various end-to-end workflows for different applications, such as taxonomical identification and structural variant calling. EPI2ME is only available for ONT clients through a web platform, and thus parts of the in-built pipelines cannot be customized by the end user. One of the main challenges is the EPI2ME data output format, which is often not compatible with other public available bioinformatic tools for downstream analyses. In EPI2ME, *What's In My Pot?* (WIMP) is an in-built workflow that allows the taxonomical assignment of nanopore sequencing data previously basecalled (Juul *et al*., 2015). However, in recent years several bioinformatic tools have been developed to perform numerous tasks for nanopore sequencing data analysis (Table 2). In addition to independent bioinformatic tools that perform specific tasks within the same pipeline, there is also efforts in developing entire bioinformatic pipelines as stand-alone softwares. In 2021, Fan and colleagues (2021) developed *BugSeq*. BugSeq is a cloud-based metagenomic classifier for long reads (Fan *et al*., 2021). Although commercially available, it is free to use for academic purposes. Other existing tools are either optimized for short reads, demanding in terms of time, or fall short for quality performance (Fan *et al*., 2021), therefore BugSeq probably represents the most accessible, reliable, and affordable platform for researchers who lacks bioinformatic expertise. BugSeq's pipeline relies on several other bioinformatic tools such as *fastp* – for quality control of sequenced reads with a minimum average read quality Phred score

of 7, minimum read length of 100 bp, and the default low complexity filter (Chen *et al*., 2018); *minimap2* – to map reads to an index in RefSeq reference database containing all microbes, the human genome, and a database of contaminants (Li, 2018); *pathoscope* – to reassign aligned reads using a Bayesian statistical framework under default parameters (Francis *et al*., 2013); *recentrifuge* – to calculate and input the lowest common ancestor from the reassigned reads, setting the minimum required taxa to 1, and generic input mode for summarization and visualization (Marti, 2019); and *multiqc* – to summarize quality control results, using a custom configuration with Phred threshold of 7 (Ewels *et al*., 2016). All dependencies are packed in Docker images, and tasks are performed on Amazon Web Services Batch to assure a safe and private environment. The authors used Nextflow for packaging and made BugSeq available online for easy cloud analyses (https://bugseq.com/free) (Fan *et al*., 2021).

Two of the most used taxonomic classification algorithms currently utilised to analyse nanopore sequencing data are *centrifuge* (Kim *et al*., 2016) and *minimap* (Li, 2018). *Minimap* was developed specifically for mapping nanopore long reads while *centrifuge* has a broader purpose in metagenomic analyses, mostly due to mapping against full genomes databases (Santos *et al*., 2020). In terms of taxonomical assignments, database composition is crucial for a strong and discriminatory analysis (Balvočiūtė and Huson, 2017, Escobar-Zepeda *et al*., 2018). Currently, the most used databases for 16S rRNA gene marker microbial identification are SILVA (Quast *et al*., 2012), Greengenes (McDonald *et al*., 2012), RDP (Cole *et al*., 2014), and NCBI (Federhen, 2012). Additionally, beyond bacterial sequence data curation, the SILVA database also includes taxonomic information for Archaea and Eukarya (Quast *et al*., 2012). The NCBI database is the most used in 16S rRNA gene-targeted nanopore sequencing approaches (Greninger *et al*., 2015, Mitsuhashi *et al*., 2017, Turner *et al*., 2018, Edwards *et al*., 2019, Kai *et al*., 2019). However, NCBI includes some level of redundancy, where unique organisms are represented by different taxonomic identifiers (Santos *et al*., 2020). For bacterial 16S rRNA amplicon analysis using the MinION nanopore sequencer the basic steps are: basecalling, sequence clustering, taxonomic/diversity/correlational analyses, statistical analyses, and data visualization (Santos *et al*., 2020). Depending on the details of the experimental approach, various other steps may also be included in the bioinformatic pipeline, such as demultiplexing, one or more quality control filters, chimera removal, reads re-alignments, *etc*. Nonetheless, taxonomic profiles may change depending on the gene marker used, as exemplified by using different portions of the 16S rRNA gene (Kuczynski *et al*., 2012). When studying unknown ecological communities the challenges

are greater since there is a higher chance that representative taxa are absent or misrepresented in databases. Moreover, in some nanopore sequence approaches, longer reads are generated and databases are mostly composed of standard and shorter gene markers, such as fragments of the 16S rRNA gene or the ITS region. Therefore, nanopore sequencing has limited reliability in profiling unknown microbial communities, decreasing the confidence on the taxonomic output obtained (Santos *et al*., 2020). A recent study showed that nanopore sequencing platform provides a rapid way for initial diagnosis of plant diseases caused by pathogenic viruses, bacteria and fungi (Chalupowicz *et al*., 2019). Another study performed rapid identification of pathogens from positive blood culture bottles with MinION nanopore sequencer (Ashikawa *et al*., 2018). The results showed that *Candida* spp. identified by MinION sequencing agreed with the results obtained by MALDI-TOF (Ashikawa *et al*., 2018). Impressively, the real-time reads obtained within about 10 minutes of sequencing were enough to identify the pathogens. Recently, nanopore sequencing was employed for the first time in the investigation of *Candida auris* epidemiology during a fungal outbreak in the United Kingdom (Rhodes *et al*., 2018).

1.3.2 Human gut microbiome through nanopore sequencing

Current high-throughput sequencing platforms such as Illumina and PacBio often take two or more days to run and to analyse the data, which makes rapid profiling an impossibility (Leggett *et al*., 2020). Thus, nanopore sequencing represents a crucial methodology to perform rapid and more disseminated microbial profiling studies. One of the handicaps of short read sequencing is the limited capacity to analyse microbial composition at the species level due to the high similarity of 16S rRNA target marker sequence regions (Shin *et al*., 2016). Indeed, the variety of clinically relevant phenotypes increases the importance of discriminating at lower taxonomical levels (Zhang *et al*., 2012). Therefore, it is necessary to identify and characterize the gut inhabitants at the species-level to understand the functional impact of the microbiota on host health (Human Microbiome Project Consortium, 2012). Since intra-species similarity is related to the genetic distance found among subregions in the full-length 16S rRNA gene (Schloss, 2010), the sequencing of the entire marker may provide higher taxonomical discrimination since it covers longer genomic sequences, increasing the chance of a unique taxonomical match in the reference databases.

The human gut microbiome has been associated with several human diseases (Kuczynski *et al*., 2012), such as IBD (Perez-Lopez *et al*., 2016), type 2 diabetes (Qin *et*

al., 2012), and autism spectrum disorder (Hsiao *et al*., 2013). Microbiome studies so far have been reporting sequencing biases that miss, misclassify, or overrepresent taxa depending on the gene marker used to profile the microbial composition (Shin *et al*., 2016, Bukin *et al*., 2019, Johnson *et al*., 2019). Additionally, some bacterial species seem to be only observed and classified if the entire 16S rRNA gene is targeted for sequencing (Matsuo *et al*., 2021).

In a recent study, Matsuo and colleagues (2021) observed that short- and long-read sequencing provided similar results in terms of dominant bacterial genera relative abundances. The authors experimented three sequencing approaches (Matsuo *et al*., 2021). For nanopore sequencing, they tested sequencing the V1-V9 and V3-V4 region of the 16S rRNA gene. For Illumina, they tested the standard V3-V4 region. The authors found similar microbial compositional profiles between the three approaches at the genus level (Matsuo *et al*., 2021). Indeed, V1-V9 full-length nanopore sequencing showed comparable results to the high-quality V3-V4 Illumina short-read sequencing data (Matsuo *et al*., 2021). However, at species level, long-reads generated by nanopore sequencing showed more robust and discriminatory resolution, as seen in the analyses of *Bifidobacterium*, *Clostridium*, *Enterococcus*, and *Staphylococcus* genera. (Matsuo *et al*., 2021). Specifically, the V1-V9 full-length dataset showed lower number of ambiguous and unassigned reads in comparison with V3-V4 short-reads data (Matsuo *et al*., 2021). The authors observed that V3-V4 short-read analysis was not capable of correctly assigning *Bifidobacterium* spp., and when it could, most reads were assigned to *Bifidobacterium* sp. coming from non-human origin using the direct read mapping approach and the relatively shallow Genome-Sync reference database (Matsuo *et al*., 2021). Moreover, the authors also observed that 20,000 reads were enough to fully determine the bacterial composition of the human gut microbiota through nanopore sequencing (Matsuo *et al*., 2021). Therefore, nanopore sequencing shows advantages when species-rank analytical resolution is necessary.

In another recent study, Wei and colleagues (2020) examined the gut microbiota of 132 clinical subjects, including 36 patients with occult blood in the gut, and 43 cases in which adenomatous polyps were observed. The authors analysed the V3-V4 region of the 16S rRNA gene using nanopore sequencing with a PCR-free protocol (Wei *et al*., 2020). The authors found decreased relative abundance levels of *Bacteroides*, *Lactobacillus*, and *Prevotella* in the gut microbiota of individuals with blood presence in stool (Wei *et al*., 2020). Additionally, they found dominant presence of *Blautia*, *Faecalibacterium*, *Bacteroides*, and *Prevotella* in healthy individuals and individuals

with occult blood in the gut (Wei *et al*., 2020). In the group of individuals showing occult blood in the gut, short-read analyses revealed decreased relative abundance levels of *Enterococcus*, *Lactobacillus*, *Prevotella*, and *Bacteroides*; with increase of *Klebsiella*, *Streptococcus*, *Clostridium*, and *Citrobacter*, in the group of individuals with blood detected in stool (Wei *et al*., 2020). Generally, both short-read and long-read approaches consistently retrieved similar microbial taxonomical composition between study participants (Wei *et al*., 2020). These results are in accordance with previously reported associations between *Fusobacterium* spp. and different stages of colorectal cancer (Ohkusa *et al*., 2009, Sekizuka *et al*., 2017, Rezasoltani *et al*., 2018, Liang *et al*., 2020), and with the beneficial effects of *Prevotella* spp. Thus, *Prevotella* spp. were hypothesized as a biomarker for colorectal cancer, with the potential to be monitored in healthy individuals to detect disease onset (Ericsson *et al*., 2015, Kang *et al*., 2017, Zhang *et al*., 2018, Yu *et al*., 2019, Ervin *et al*., 2020).

Recently, Leggett and colleagues (2020) tested a bioinformatic pipeline that allowed the identification of pathogenic taxa – *Klebsiella pneumoniae* and *Enterobacter cloacae* –, and their resultant antimicrobial resistance gene profiles in about five hours of experimentation (four hours for sample preparations and one hour for sequencing). The authors confirmed their observations by culture isolation, whole-genome sequencing, antibiotic susceptibility testing, and use of a mock community (Leggett *et al*., 2020). This mock community was previously used in the Human Microbiome Project (HM-277D; BEI Resources) and was comprised of genomic DNA from 20 bacterial strains containing staggered RNA operon counts. Additionally, the authors benchmarked their pipeline against Illumina sequencing and reported a high correlation between the two sequencing platforms, specifically in terms of abundance levels, although some differences to lower taxonomical ranks were also observed (Leggett *et al*., 2020). Nanopore longer reads could therefore improve the specificity in assigning species, but its lower accuracy may reduce differentiation between closely related species (Leggett *et al*., 2020). To further explore their procedure, Leggett and colleagues (2020) tested a metagenomic approach using nanopore sequencing on 8 preterm infants, including 5 diagnosed with sepsis or necrotizing enterocolitis. Principal coordinates analysis showed clusters comprising of beneficial *Bifidobacterium* and possible pathogenic *E. cloacae* or *K. pneumoniae* (Chen *et al*., 2009, Leggett *et al*., 2020). Moreover, the authors also monitored the gut microbiota of a preterm infant for 28 days using a shotgun/metagenomics approach with nanopore and Illumina sequencing (Leggett *et al*., 2020). They found that nanopore sequencing can perfectly retrieve the samples' species diversity (Leggett *et al*., 2020). Indeed, the

taxonomic assignments obtained by the two sequencing approaches were similar, as observed by the associations between probiotic intervention, or sepsis periods, and presence of *Klebsiella*, *Enterobacter*, *Enterococcus*, *Veillonella*, *Staphylococcus*, and *Bifidobacterium* (Leggett *et al*., 2020). Thus, the authors reiterated the already reported potential of nanopore sequencing to profile the gut microbiota using shotgun/metagenomics approaches (Anand *et al*., 2007, Brown *et al*., 2017, Leggett *et al*., 2020). In summary, Leggett and colleagues (2020) built a convenient and cost-effective pipeline with the Flongle nanopore sequencer and NanoOK RT bioinformatic tool and showed that nanopore sequencing could be usefully exploited in future clinical studies to provide relevant data in about five hours (Leggett *et al*., 2020). Nevertheless, the error-rate of nanopore sequencing is still too high to give total confidence in the approach. However, further development both in sequencing chemistry and bioinformatic toolset will undoubtedly resolve these issues by improving the accuracy of reads.

1.3.3 Mouse gut microbiome through nanopore sequencing

In order to decipher the mouse gut microbial composition, Shin and colleagues (2016) studied the mouse gut microbiota by sequencing the full-length 16S rRNA gene using nanopore sequencing approaches and compared the results with short reads generated by sequencing the V3-V4 hypervariable region using the next-generation Illumina sequencing platform. Despite minor changes at one phylotype and three taxonomic units, nanopore sequencing outputs showed no significant differences in comparison to Illumina short reads sequencing data (Shin *et al*., 2016). Specifically, *A. muciniphila* was overrepresented in the short-read approach and *Blautia* gen., *Bifidobacterium pseudolongum* and *Bacteroides ovatus* were overrepresented in the long-read approach. Moreover, the taxonomic data retrieved by nanopore sequencing at the genus level was in accordance with that previously observed in mice gut microbiota profiling studies (Ravussin *et al*., 2012, Hilderbrand *et al*., 2013, Langille *et al*., 2014, Seekatz *et al*., 2014). Indeed, nanopore sequencing achieved a higher number of taxa at the species level than Illumina sequencing (Shin *et al*., 2016). This might be explained by the capacity of nanopore sequencing reads to span multiple 16S rRNA variable regions (Shin *et al*., 2016). For instance, *Bifidobacterium* spp. was only discriminated in nanopore sequencing data, whereas *Bacteroides* spp. could be separated by Illumina short-reads only (Shin *et al*., 2016). The authors observed 9 variants of V3-V4 regions between *B. acidifaciens*, *B. ovatus*, and only 4 were detected between *Bifidobacterium animalis* and *B. pseudolongum* (Shin *et al*., 2016). Additionally, these later variants were mostly in the

V1-V2 region. Therefore, targeting the V3-V4 hypervariable regions alone might overlook relevant genetic information necessary for further taxonomical separation at lower ranks, which could be solved by sequencing longer fragments encompassing multiple variable regions of 16S rRNA gene (Shin *et al*., 2016). In 2022, Jeon and colleagues performed a metagenomic experiment using full 16S-23S rRNA gene region nanopore sequencing to study the effect of potential probiotic strains in cognitive ability of healthy mice. The study aim was to detail the strain-level effects of *L. acidophilus*, *Lactobacillus paracasei*, and *Lacticaseibacillus rhamnosus* (Jeon *et al*., 2022). Jeon and colleagues (2022) found that *L. acidophilus* EG004 positively impacted the cognitive ability of healthy mice. In terms of microbial composition between control and probiotic-intervention mice groups, the authors found that *Bacteroidota* was the most abundant phylum detected, followed by *Firmicutes* in both groups (Jeon *et al*., 2022). Additionally, *Firmicutes* and *Proteobacteria* were enriched in the *L. acidophilus* group and *Bacteroidia* was found as the most abundant class in both groups (Jeon *et al*., 2022). *Bacilli* and *Gammaproteobacteria* classes were enriched in the *L. acidophilus* group, *Bacteroidales* order was overrepresented in both groups, and *Lactobacillales* and *Enterobacterales* orders were overrepresented in *L. acidophilus* group. The authors also observed that *Muribaculaceae* was the most abundant family found in both groups and *Lactobacillaeceae* and *Enterobacteriaceae* were significantly increased in *L. acidophilus* group (Jeon *et al*., 2022). At the genus level 12 taxa showed differences between groups, however, *Muribaculum* gen. was the most abundant between both groups (Jeon *et al*., 2022). In the control group the overrepresented genera were *Bacteroides*, *Desulfotomaculum*, *Lachnobacterium*, *Bittaerella*, *Agathobacter*, *Roseburia*, *Bariatricus*, and *Lachnospirarea* (Jeon *et al*., 2022). On the other hand, in the *L. acidophilus* group, *Lactobacillus*, *Staphylococcus*, and *Escherichia* genera were significantly increased (Jeon *et al*., 2022). At the species level, the authors found that *Muribaculum intestinale* was accountable for more than 50% of relative abundances in both groups, followed by *L. acidophilus*, *Lactobacillus johnsonii*, *Lactobacillus murinus*, and *L. reuteri* in the *L. acidophilus* group; and followed by *L. murinus*, *Bacteroides vulgatus*, *Faecalibaculum rodentium*, and *Kineothrix alysoides* in the control group (Jeon *et al*., 2022).

In summary, third-generation technologies such as nanopore sequencing whilst showing considerable promise still need to be further developed to improve the reliability of their sequencing outputs. So far, scientists have been employing a two-way sequencing approach to validate and/or complement nanopore sequencing studies together with more

mature sequencing technologies, such as short-read Illumina sequencing. These experiments will gain from the high-throughput resolution obtained by Illumina sequencing while increasing the sequencing coverage by the longer sequenced reads obtained by nanopore sequencing.

However, there are several technical challenges to be addressed in nanopore sequencing approaches, mostly related to the non-standardization of experimental and bioinformatics protocols. Thus, the application of targeted and untargeted metagenomic nanopore sequencing approaches to the study of the gut microbiota is still in the early stages. Indeed, our knowledge of the composition and function of the whole gut microbiota lacks a comprehensive description, with particular significance in a clinical context where appropriate diagnostics and therapeutics may be lacking, such as FASD. Since there is no consensus regarding foetal and neonatal gut microbiota transmission and colonization, the impact of stressors like alcohol consumption occurring during pregnancy needs an urgent analysis. Moreover, starting at early ages, the gut microbiota mediates a wide range of processes relevant to the host, including neurodevelopment through the microbiota-gut-brain axis, imposing additional challenges towards the characterization of relevant microbial phenotypes.

In this thesis, targeted and untargeted nanopore sequencing protocols were developed and optimized to study the effect of ethanol exposure *in utero* on the infant mice gut whole microbiota. In Chapter 3, a fungal mock community was used to assess and discuss the accuracy and precision of targeted nanopore sequencing approaches. In Chapter 4, two targeted and one untargeted (shotgun/metagenomic) nanopore sequencing approach were developed to analyse and discuss the microbial composition in the gut of infant mice exposed to ethanol *in utero*. Lastly, in Chapter 5, an optimized targeted nanopore sequencing approaches was used to analyse and discuss the effect of ethanol on the gut microbiota of infant mice exposed *in utero* compared with non-exposed control mice. These results were compared with short-read Illumina sequencing to assess the accuracy and precision of the nanopore sequencing approach developed approach. Finally, in Chapter 6, the most important findings of the study are integrated and discussed.

Chapter 2 – Methodologies

2.1 Preparation of a fungal mock community for optimization of nanopore sequencing targeted approaches

2.1.1 Fungal growth conditions

A fungal mock community was prepared with 5 strains, representing species *Meyerozyma guilliermondii*, *Candida glabrata*, *Pichia kudriavzevii*, *Candida parapsilosis*, and *Clavispora lusitaniae*, that are usually found as members of the human gut mycobiome. All stock isolates were grown in 5 mL of Sabouraud dextrose broth in universal flasks for 2 days at 37° C under shaking (270 rpm). After growth was observed, isolates were plated and streaked twice to achieve purification in Potato Dextrose Agar and incubated at 30 °C for 18-24h until DNA extraction.

2.1.2 Fungal DNA extraction and amplicons generation

2.1.2.1 Fungal isolates identification confirmation

For DNA extraction, a custom protocol was performed. From fresh culture plates (typically 18 to 24 h), a small amount of yeast cells was collected and transferred to a 1.5 mL microtube containing 500 µL sterile water. The microtube was vortexed at 10,000× g for 5 min and the supernatant discarded. Then, 300 µL of lysis buffer (100mM Tris-HCl [pH 8.0], 1 mM EDTA, 2% Triton X-100) were added to the pellet and mixed with the equivalent of 100 µL glass beads (ø = 0.5 mm). The mixture was briefly vortexed and incubated at 95 °C for 10 min before cells' disruption in the *Savant Bio 101 FastPrep FP120* (Savant Instruments, India) (maximum speed = 6.5; 40 sec). A final centrifugation step was performed for 5 min at 10,000× g. The supernatant containing the extracted DNA was stored at -20 °C. Polymerase Chain Reaction (PCR) was performed to amplify the ITS region with diluted genomic DNA (1:10) from each isolate, with the following primer sequences: ITSF1 – CTTGGTCATTTAGAGGAAGTAA; ITS4 – TCCTCCGCTTATTGATATGC (Op De Beeck *et al*., 2014). The PCR mixture (25 µL total volume) contained between 1 and 4 ng of DNA template (or 6 µL of unquantified initial DNA), 1X Standard *Taq* Reaction Buffer, 0.2 mM of dNTPs, 0.4 µM of ITS1F and ITS4, and 0.625 U of *Taq* DNA Polymerase (New England Biolabs, United Kingdom). The PCR thermal profile consisted of an initial denaturation of 1 min at 95 ºC, followed by 35 cycles of 30 s at 95 ºC, 40 s at 50 ºC, 60 s at 68 ºC, and a final step of 5 min at 68 ºC. The PCR products were visualized in an agarose gel 1.1% (w/v). After positive band check, the PCR products were purified with GeneJET purification kit according to the

manufacturer's protocol (Thermofisher, United States of America) and sequenced by Sanger sequencing using the same primers (Eurofins, Germany).

2.1.2.2. Fungal mock community amplicons generation for library preparation

To achieve the best quality gDNA, purified isolates' DNA extraction was performed with MasterPure™ Yeast DNA Purification Kit, according to the manufacturer's protocol (Lucigen, United States of America). Extracted DNA was stored at -20° C until further use. The PCR mixture and thermal profile performed was described above. PCR products were visualized in an agarose gel 1.1% (w/v). Then, PCR products were purified with GeneJET purification kit, according to the manufacturer's protocol and stored in the fridge at 4 ºC until further use. To build the mock community, equimolar concentrations of the purified PCR products originating from the 5 strains were pooled together for library preparation, *i.e.*, 40 fmol per strain, accordingly to ONT protocols.

2.2 Infant mice DNA extraction and amplicons' generation targeting the full-length 16S rRNA gene and the 16S-ITS-23S region of the *rrn* operon

2.2.1 Mice experimental design

2.2.1.1 Animal husbandry

The mice experimental design was performed in collaboration with Dr Nigel Brissett from University of Brighton (UK) under the project entitled "Neurochemical effects of prenatal alcohol exposure". The project was internally funded by the Centre for Stress and Age-Related Diseases from the University of Brighton, UK. Animal Welfare and Review Board ethically approved all protocols and procedures. The work is licensed under the Animals (Scientific Procedures) Act, 1986), with license number PPL P5686A14D. All experimental work was performed in accordance with the ARRIVE guidelines published by the National Centre for the Replacement Refinement & Reduction of Animals in Research. Three male and three female C57BL/6J mice (6 and 8 weeks old, respectively) were randomly paired into three groups. Each male randomly paired with one female and placed into an Optimice® cage, according to the breeding protocol for the pilot study. Three monogamous breeding pairs were achieved: BOX1 ethanol (n = 2), BOX2 ethanol (n = 2) and BOX3 ethanol (n = 2). Controls for non-ethanol exposure were mice that were cousins of the treated mice undergoing growth conditions as the treated minus the ethanol (Figure 4). A first batch of 12 infant mice exposed to

ethanol *in utero* were used for the nanopore sequencing approaches optimization experiments. Weights of the offspring were noted before sacrifice and it was observed that ethanol-exposed mice were smaller compared to the control group (data not shown).

Figure 4 – Family tree representation of mice husbandry experiments. 'A' represents the ethanol-exposure experiment. '0', '1', and '2', represent the generations. Yellow mice are brothers. Orange mice (n = 9) in A2 represent infant mice exposed to ethanol *in utero* from which faecal samples were collected and analysed in this dissertation. Blue mice (n = 7) in A2 represent non-exposed mice from which faecal samples were collected and analysed in this dissertation. Black mice represent unrelated mice.

Cage and holding room environment were maintained constantly with controlled light intensity (200lx, 1m from floor level and 25lx at cage level), relative humidity (41%-60%), room (19 °C-21 °C) and cage temperature (≤ 37 °C). To promote crepuscular behaviour, a 24 h lighting cycle of 12 h light followed by 12 h dark photoperiod was implemented in order to regulate both circadian rhythm and reproductive cycles. Enviro-Dry® fresh nesting was aided and preserved by microenvironment temperature. RM No.3 Special Diet Services (SDS) (5 g/day/mouse) was supplied to each group.

2.2.1.2 Drug formulation and administration

Formulation of 5% ethanol solutions (1000 mL) were formed prior to drug administration and stored at 4 °C. To mimic average human ethanol consumption, for each group, 5% ethanol solution was orally administered from colony set-up onwards, by means of Optimice® internal bottle assembly and allowed fresh daily consumption. The daily 5% ethanol solution intake was measured by means of a digital measuring scale, before and after refilling the solution bottle. All measurements were taken between 8 am and 11 am, thus limiting irregular disruption to animal subjects. Each breeding box consumed between 18-22 g of 5% ethanol per day. Offspring pups were weaned 23-28

days postnatal, separated and placed into gender specific boxes categorised with date of birth, wean date, and box origin.

2.2.1.3 Sacrifice and harvest

Animals were sacrificed in the context of the abovementioned Dr Brissett's project. Two independent experiments were performed. From the first batch, twelve infant mice (six males and six females) exposed to alcohol *in utero* were analysed with no controls. From the second batch, nine infant mice (five males and four females) exposed to alcohol *in utero* and seven controls (three males and four females) were analysed. To avoid subjective bias, the first eight female and eight male offspring born amongst the three mating pairs were sacrificed roughly at 18-20 week and harvested. All mice, including controls, were subjected to primary method of sacrifice (cervical dislocation), followed by spinal cord severing and first extraction of the brain, to limit distress. Four organs (brain, heart, liver, and kidney), the large intestine, a single posterior limb and faecal matter were harvested from each animal. Apart from the large intestine, the remaining materials were collected only for Dr Brissett's project and not used in the present work. Following the extraction, all animal tissue was placed into labelled tissue culture plate wells, stood immediately on ice and stored at -80 °C. Around 300 mg of faecal material were collected from each large intestine as an additional sample to be used in this dissertation. All manipulations involving animal husbandry, the sacrifice of mice, and gut sample collection from the same animals were performed by Dr Brissett.

2.2.2 Infant mice gut DNA extraction and amplicons generation

In order to develop an optimized nanopore sequencing approach to microbial profile infant mice gut samples, twelve faecal samples were collected from mice infants exposed to ethanol *in utero*. After nanopore sequencing data generation and analyses, additional nanopore sequencing assays were carried out using nine faecal samples originated from infant mice exposed to ethanol *in utero* and seven control faecal samples originated from non-exposed infant mice.

DNA extraction was performed using around 300 mg of biological faecal material with QIAmp PowerFecal Pro DNA Kit, according to the manufacturer's protocol (Qiagen, UK). Briefly, less than 300 mg was transferred to a 1.5-2 mL tube containing beads and vortexed with solution CD1 for 10 min at maximum speed. The mixture was centrifuged at 15,000× g for 1 min and the supernatant transferred to a clean tube. Additional centrifugation step with solution CD2 (previously vortexed for 5 s). Addition

of 600 µL from solution CD3 to the supernatant and vortexed for 5 s. A succession of 2 centrifugation steps at 15,000× g for 1 min were performed to ensure that all lysate passed the spin column. Addition of 500 µL of solution EA and Centrifugation at 15,000 x g for 1 min. In a new 2 mL tube, 500 µL of solution C5 was added and the tube centrifuged at 16,000× g for 2 min. In a new 2 mL tube, 50-100 µL of solution C6 was added and the tube centrifuged at 15,000 x g for 1 min. The supernatant containing the extracted DNA was used immediately or stored at -20 °C until further use. DNA quantity and quality ($A_{260/280}$ and $A_{260/230}$) were measured using a Qubit 4 Fluorometer (Thermofisher, United States of America), and with DeNovix DS-11 Spectrophotometer (DeNovix, United States of America), respectively. To target the fungal infant mice gut microbiota, PCR was performed to amplify the ITS region as previously described in the fungal mock community nanopore sequencing approaches optimization (see section 2.1.2.1).

To target the bacterial infant mice gut microbiota, PCR was performed to amplify two targets, the full-length of the 16S rRNA gene (~1,500 bp), and the whole *rrn* operon (16S rRNA gene-ITS-23S rRNA gene, with ~4,500 bp). The primers pairs used were the16S-27F – AGAGTTTGATCCTGGCTCAG, and the 16S-1492R – GGTTACCTTGTTACGACTT for the 16 rRNA gene. For the 16S-ITS-23S operon region., the same abovementioned 16S-27F was used together with the 23S-2241R – ACCRCCCCAGTHAAACT (Cusco *et al*., 2018).

The PCR mixture for the full-length 16S rRNA gene (25 µL total volume) contained between 0.2 and 1.6 ng of DNA template (or 3 µL of unquantified initial DNA extract), 1× Standard *Taq* Reaction Buffer, 0.5 mM of dNTPs, 0.4 µM of 16S-27F and 0.8 µM 16S-1492R, and 0.625 U of *Taq* DNA Polymerase. The PCR thermal profile consisted of an initial denaturation of 30 s at 95 ºC, followed by 30 cycles of 20 s at 95 ºC, 20 s at 51 ºC, 2 min at 68 ºC, and a final step of 5 min at 72 ºC.

The PCR mixture for the full-length 16S-ITS-23 region of the *rrn* operon (25 µL total volume) contained between 0.2 and 1.6 ng of DNA template (or 3 µL of unquantified initial DNA), 1× Standard *Taq* Reaction Buffer, 0.5 mM of dNTPs, 1 µM of 16S-27F and 23S-2241R, and 0.625 U of *Taq* DNA Polymerase. The PCR thermal profile consisted of an initial denaturation of 30 s at 95 ºC, followed by 30 cycles of 20 s at 95 ºC, 20 s at 51 ºC, 4 min and 30 s at 68 ºC, and a final step of 5 min at 72 ºC. The amplicons were purified with the AMPure XP beads (Beckman Coulter) using a 0.5× and 0.40× ratio for the full-length 16S rRNA gene and the 16S-ITS-23S region of the *rrn* operon, respectively, according to manufacturer's protocol. The quantity and quality of the purified DNA was measured by Qubit 4 Fluorometer, and with DeNovix DS-11 Spectrophotometer,

respectively. If the PCR purified amplicons did not show $A_{260/280}$ between 1.8-2.0 and $A_{260/230}$ between 2.0-2.2, they were discarded, and the PCR reaction repeated.

2.3 Nanopore sequencing approaches

2.3.1 Library preparation

Targeted nanopore sequencing approaches were used in this work to analyse ITS amplicons from a mock fungal community (see Chapter 3), 16S rRNA gene and 16S-ITS-23S operon amplicons from a batch of 12 infant mice gut samples exposed to ethanol *in utero* (see Chapter 4), and 16S rRNA gene amplicons from one additional batch of mice gut samples comprising nine samples of mice exposed to ethanol *in utero* and seven samples of control mice not exposed to ethanol (Chapter 5). A shotgun/metagenomic nanopore sequencing approach was also attempted with the same batch of 12 infant mice gut samples exposed to ethanol *in utero* above mentioned (see Chapter 4).

2.3.1.1 Library of ITS amplicons from a fungal mock community

For the nanopore sequencing approach optimization using a fungal mock community, the Ligation Sequencing Kit (SQK-LSK 109, ONT, United Kingdom) was applied. A mixture comprising ITS amplicons previously generated by PCR were used as template for DNA library preparation (see section 2.1.2). Two independent DNA sequencing libraries were prepared for separate sequencing runs. The pooled ITS amplicons concentration was adjusted to reach 1 µg in a volume of 47 µL, with 1 µL of DNA CS (sequencing positive control containing DNA obtained from lambda phage), 3.5 µL of NEBNext FFPE DNA Repair Buffer (New England Biolabs, United Kingdom), 2 µL of NEBNext FFPE DNA Repair Mix (New England Biolabs, United Kingdom), 3.5 µL Ultra II End-prep reaction buffer, and 3 µL of Ultra-prep enzyme mix (New England Biolabs, UK), completed with molecular biology grade water. The mixture was incubated in a thermal cycler at 20 °C for 5 min and 65 °C for 5 min. A purification step using 1× Agencourt AMPure XP beads was performed. Briefly, the mixture was incubated at room temperature with gentle mixing for 5 min, washed twice with 200 µL fresh 70% ethanol, with pellet allowed to dry at room temperature for 1 min before DNA being eluted in 61 µL nuclease-free water, and incubated at room temperature for less than 5 min. A 1 µL aliquot was quantified by fluorometry using Qubit 4 Fluorometer. For the adapter ligation step, the total 60 µL of end-prepped DNA from the previous step were added in a mix containing 25 µL of Ligation Buffer LNB (NEBNext Quick Ligation Module, New England Biolabs, United Kingdom), 10 µL NEBNext Quick T4 DNA Ligase (New

England Biolabs, United Kingdom), and 5 µL of Adapter Mix (ONT, United Kingdom), and incubated at room temperature for 10 min. A purification step using 0.4× Agencourt AMPure XP beads was performed. Briefly, the mixture was incubated at room temperature with gentle mixing for 5 min, washed twice with 250 µL of Short Fragment Buffer (ONT, United Kingdom) (to allow the recovery of fragments of all sizes), pellet allowed to dry at room temperature for 1 min before DNA being resuspended in 15 µL of Elution Buffer (ONT, United Kingdom), and incubated at room temperature for 10 min or more. The mixture was retained and coined as 'DNA sequencing library'. A 1 µL aliquot was quantified by fluorometry using Qubit 4 Fluorometer and quality measured with DeNovix DS-11 Spectrophotometer to ensure no contaminants were present before the library loading step. If the DNA sequencing library did not show $A_{260/280}$ between 1.8-2.0 and $A_{260/230}$ between 2.0-2.2, it was discarded, and the protocol repeated.

2.3.1.2 Libraries of 16S rRNA gene and 16S-ITS-23S operon amplicons

For the targeted approaches targeting the 16S rRNA gene and the 16S-ITS-23S region of the *rrn* operon, the Ligation Sequencing Kit (LSK 109) was used in combination with PCR-Barcoding Expansion Kit (EXP-PBC001, ONT, United Kingdom) to tag each of the previously generated sample specific PCR amplicons and prepare the sequencing libraries, following the manufacturer's protocol. The forward barcoding primer tag used was F – TTTCTGTTGGTGCTGATATTGC, and the reverse was R – ACTTGCCTGTCGCTCTATCTTC.

Amplicons were quantified using the Qubit 4 Fluorometer and the volume was adjusted to begin the barcoding PCR with 100-200 fmol (0.5 nM) of the first PCR products, according to ONT instructions. PCR mixture for barcoding PCR (100 µL total volume) contained 100-200 fmol of each PCR product to be barcoded in 48 uL, 1× LongAmp® *Taq* 2× Master Mix (New England Biolabs, United Kingdom), and 2 µL of a unique barcode. The PCR thermal profile consisted of an initial denaturation of 3 min at 95 °C, followed by 13 cycles (for the 16S rRNA gene) or 15 cycles (for the 16S-ITS-23S region of the *rrn* operon) of 15 s at 95 °C, 15 s at 62 °C, 1 min and 30 s at 65 °C (for the 16S rRNA gene) or 4 min and 30 s at 65 °C (for the 16S-ITS-23S region of the *rrn* operon), and a final step of 1 min 30 s at 65 °C (for the 16S rRNA gene) or 4 min and 30 s (for the 16S-ITS-23S region of the *rrn* operon) at 65 °C. The barcoded amplicons were purified with the AMPure XP beads using a 0.5× and 0.40× ratio for the full-length 16S rRNA gene and the 16S-ITS-23S region of the *rrn* operon, respectively, according to the manufacturer's protocol. Every tagged DNA sample were combined in a pool to reach 1

μg in a volume of 47 μL, 1 μL of DNA CS (sequencing positive control containing DNA obtained from lambda phage), 3.5 μL of NEBNext FFPE DNA Repair Buffer, 2 μL of NEBNext FFPE DNA Repair Mix, 3.5 μL Ultra II End-prep reaction buffer, and 3 μL of Ultra-prep enzyme mix, and completed with molecular biology grade water. The mixture was incubated in a thermal cycler at 20 °C for 5 min and 65 °C for 5 min. A purification step using 1× Agencourt AMPure XP beads was performed. Briefly, the mixture was incubated at room temperature with gentle mixing for 5 min, washed twice with 200 μL fresh 70% ethanol, pellet allowed to dry at 37 °C until a light brown colour was observed (more time is needed if the mixture is more complex, *i.e.*, containing more DNA samples pooled together, or if longer fragments are to be sequenced) before DNA being eluted in 61 μL nuclease-free water, and incubated at room temperature for 30 min or more (increased incubation times allow a better recovery of the DNA from the beads). A 1 μL aliquot was quantified by fluorometry using Qubit 4 Fluorometer. For the adapter ligation step, the total 60 μL of end-prepped DNA from the previous step were added in a mix containing 25 μL of Ligation Buffer LNB, 10 μL NEBNext Quick T4 DNA Ligase, and 5 μL of Adapter Mix, and incubated at room temperature for 10 min. A purification step using 0.4× Agencourt AMPure XP beads was performed. Briefly, the mixture was incubated at room temperature with gentle mixing for 5 min, washed twice with 250 μL of Short Fragment Buffer (to allow the recovery of fragments of all sizes), pellet allowed to dry at 37 °C until a light brown colour was observed before DNA being resuspended in 15 μL of Elution Buffer, and incubated at room temperature for 30 min or more. The mixture was retained and coined as 'DNA sequencing library'. A 1 μL aliquot was quantified by fluorometry using Qubit 4 Fluorometer and quality measured with DeNovix DS-11 Spectrophotometer to ensure that no contaminants were present before the library loading step. If the DNA sequencing library did not show $A_{260/280}$ between 1.8-2.0 and $A_{260/230}$ between 2.0-2.2, it was discarded, and the protocol repeated.

2.3.1.3 Library for the shotgun/metagenomic nanopore sequencing approach

For the untargeted shotgun/metagenomic nanopore sequencing approach the Ligation Sequencing Kit (LSK 109) was used in combination with Native Barcoding Kit (EXP-NBD104, ONT, United Kingdom) to tag each DNA template extracted from the gut samples with a unique barcode, following the manufacturer's protocol. The forward barcoding primer tag used was F – TTTCTGTTGGTGCTGATATTGC, and the reverse was R – ACTTGCCTGTCGCTCTATCTTC. Every tagged DNA sample were pooled

together to reach 1 µg in a volume of 48 µL containing 3.5 µL of NEBNext FFPE DNA Repair Buffer, 2 µL of NEBNext FFPE DNA Repair Mix, 3.5 µL Ultra II End-prep reaction buffer, and 3 µL of Ultra-prep enzyme mix, and completed with molecular biology grade water. The mixture was incubated in a thermal cycler at 20 °C for 5 min and 65 °C for 5 min. A purification step using 1×Agencourt AMPure XP beads was performed. Briefly, the mixture was incubated at room temperature with gentle mixing for 5 min, washed twice with 200 µL fresh 70% ethanol, pellet allowed to dry at 37 °C until a light brown colour was observed before DNA being resuspended in 25 µL nuclease-free water, and incubated at room temperature for 1 h or more. A 1 µL aliquot was quantified by fluorometry using Qubit 4 Fluorometer to ensure that at least 700 ng of DNA was retained.

For barcode ligation, 500 ng of end-prepped DNA was added separately to 2.5 uL of a unique Native Barcode, and 25 uL of NEB Blunt/TA Ligase Master Mix (New England Biolabs, United Kingdom). The mixture was incubated at room temperature for 10 min. A purification step using 1× Agencourt AMPure XP beads was performed. Briefly, the mixture was incubated at room temperature with gentle mixing for 5 min, washed twice with 200 µL fresh 70% ethanol, pellet allowed to dry at 37 °C until a light brown colour was observed before DNA being resuspended in 26 µL nuclease-free water, and incubated at room temperature for 1 h or more. A 1 µL aliquot was quantified by fluorometry using Qubit 4 Fluorometer. Equimolar concentrations of each native barcoded sample were pooled together assuring that a total 700 ng of DNA was guaranteed. A 1 µL aliquot was quantified by fluorometry using Qubit 4 Fluorometer. For the adapter ligation step, 65 µL containing 700 ng of pooled end-prepped native barcoded DNA from the previous step were added in a mix containing 5 µL of Adapter Mix II (AMII, ONT, United Kingdom), 20 µL NEBNext Quick Ligation Reaction Buffer (5×, New England Biolabs, United Kingdom), and 10 µL of Quick T4 DNA Ligase, and incubated at room temperature for 10 min. A purification step using 0.5× Agencourt AMPure XP beads was performed. Briefly, the mixture was incubated at room temperature with gentle mixing for 5 min, washed twice with 250 µL of Short Fragment Buffer (to allow the recovery of fragments of all sizes), pellet allowed to dry at 37 °C until a light brown colour was observed before DNA being eluted in 15 µL of Elution Buffer, and incubated at room temperature for 30 min or more. The mixture was retained and coined as ‘DNA sequencing library’. A 1 µL aliquot was quantified by fluorometry using Qubit 4 Fluorometer and quality measured with DeNovix DS-11 Spectrophotometer to ensure no contaminants were present before the library loading

step. If the DNA sequencing library did not show $A_{260/280}$ between 1.8-2.0 and $A_{260/230}$ between 2.0-2.2, it was discarded, and the protocol repeated.

2.3.2 MinION sequencing and basecalling

Nanopore sequencing was performed using MinION MK1B sequencer (ONT, United Kingdom) and flowcells R9.4.1 (FLOW-MIN106D, ONT, United Kingdom) using MinKNOW™ software v.21.02.1 (ONT, United Kingdom). After the MinION Flow Cell was setup accordingly to manufacturer's instructions, the pre-sequencing mix was prepared by adding 12 µL of the DNA sequencing library containing at least 50 fmol, 37.5 µL of Sequencing Buffer (ONT, United Kingdom), and 25.5 µL of Loading Beads (ONT, United Kingdom). The flowcell was primed with a mixture of 30 µL Flush Tether introduced directly into the Flush Buffer tube (Flow Cell Priming Kit, EXP-FLP002), according to manufacturer's protocol (ONT, United Kingdom). Five minutes after the priming of the flowcell, the DNA nanopore sequencing library was mixed by pipetting and loaded in a dropwise fashion using the SpotON port, according to manufacturer's protocol. Once the DNA sequencing library was loaded, the 24 h sequencing run was initiated using MinKNOW™ software v.21.02.1.

For the targeted sequencing approach using the full-length 16S rRNA gene, two MinION Flow Cells were used so that 6 barcoded DNA samples were sequenced per sequencing run, thus maximizing sequencing output and avoiding carryover of the same barcode to the next sequencing run. In the targeted sequencing approach using the 16S-ITS-23S operon region and the shotgun/metagenomic nanopore sequencing approach, a single flowcell was used with 12 barcoded DNA samples. All nanopore sequencing experiments ran for 24 consecutive hours.

Basecalling was performed with *guppy* standalone tool (https://nanoporetech.com) from the command line using the *fast basecalling* mode, with the quality score filter of *8*, and adapters' trimming mode *on*. Quality controls were performed using *nanoplot* tool (De Coster *et al*., 2018) with the length filters parameters of maximum length at *1300* and minimum at *300* for the fungal mock community, maximum length at *2000* and minimum at *1000* for the 16S rRNA gene-targeted approach, maximum length at *5000* and minimum at *1000* for the 16S-ITS-23S region of the *rrn* operon-targeted approach, and minimum at *300* with no maximum length filter for the shotgun/metagenomic approach.

2.3.3 Taxonomical classification with WIMP and BugSeq

2.3.3.1 BugSeq

For the fungal mock community nanopore sequencing experiment BugSeq online software (Fang *et al*., 2021) was employed. BugSeq algorithm performs *minimap2* alignments using a nucleotide database, followed by Bayesian reassignment and Lowest-Common Ancestor identification (in some cases, an abundance calculation can also be performed (Fan *et al*., 2021)). In detail, reads were quality controlled with *fastp* (v0.20.1), using a minimum average Phred score (Qscore) of 7, a minimum read length of 100 bp, and the default low complexity filter (Chen *et al*., 2018). Phred or Qscores are defined by: $Q = -10\log_{10}(e)$; Where *e* is the estimated probability of a particular base call being wrong. Higher Q scores indicates a smaller probability of error, and lower Q scores means the reads have insufficient quality to be further used. In practice, a Q score of 10 means that 10 out of 100 bases may be wrong or falsely called, a Q20 means 1 in 100, and Q30 means 0.1 in 100. Then, reads were mapped with *minimap2* (v2.17) (*map-ont* mode on, containing the flag *a*, and the parameter for secondary alignments set to *10*) to an index containing all microbes in RefSeq reference database, the human genome, and a contaminants' database, regardless of their completeness status (Li, 2018). After generation of aligned reads, they were reassigned based on a Bayesian statistical framework using *pathoscope* (version 2.0.7) and default parameters (Francis *et al*., 2013). Lastly, reassigned reads were used for Lowest-Common Ancestor calculation to guarantee precision when closely related sequences are present in the reference database and stored into *recentrifuge* (v1.1.1), with the minimum required taxa set to *1*, and generic input mode for summarization and visualization *on* (Marti, 2019).

Quality control results were computed using *multiqc* using a custom configuration (Phred score > 7) (Ewels *et al*., 2016). For comparative analyses, non-normalized relative frequencies were used as input for MicrobiomeAnalyst software (Chong *et al*., 2020), and only taxa with relative abundances above 1% were considered.

2.3.3.2 WIMP

WIMP (version 2021.11.26) was used for the classification assignments in the infant mice gut microbiota nanopore sequencing profile experiments. WIMP is a workflow available in the EPI2ME Desktop Agent™ (ONT, United Kingdom), for the taxonomic classification of basecalled sequences (reads) generated by nanopore sequencing. Firstly, WIMP filtered FASTQ files with a quality Qscore below a minimum threshold (default is 7). The sequencing read quality was also measured by the Qscore.

For reads that pass the quality threshold, Centrifuge software (Kim *et al*., 2016) was used to assign a taxon to each read and thus a score is attributed to every assigment. These scores are calculated by the sum of squared lengths of segments subtracting 15; $[\text{length(hit)} - 15]^2$. During Centrifuge computing, false-positives assignments for high error-prone reads can be reduced by removing assignments scores < 300 and match lengths ≤ 50 bps (Kim *et al*., 2016). Then, Centrifuge classification results were filtered and aggregated to calculate the number of counts at the most discriminatory rank possible. If no assignment could be attributed, *i..e*, quality and *centrifuge* score below a scoring threshold, the sequence was termed "Unclassified". EPI2ME uses a customized and normalised version of the Centrifuge assignment score, therefore enabling a more comparable filtering for matches by score.

For comparative analyses, non-normalized relative frequencies were used as input for MicrobiomeAnalyst software (Chong *et al*., 2020), and only taxa with relative abundances above 1% were considered.

2.4 Microbial profiling of infant mice gut samples exposed to ethanol *in utero* through Illumina sequencing

Seven control and nine test gut samples from mice infants exposed to ethanol *in utero* were used for Illumina sequencing (the same samples were formerly analysed with nanopore sequencing). Genomic DNA was extracted from infant mice faecal samples as described in Section 2.2.2) and sent to Eurofins (Germany) for sequencing by Illumina MiSeq platform (Illumina, United States of America) targeting the V3-V4 region of the 16S rRNA gene (2×300 bps). Then, reads were demultiplexed and primers clipped before reads were merged. All reads underwent a quality filter step before microbial profiling analyses (see Appendix I). FASTQ files with quality filtered reads were used as input for microbial profiling. Chimeric reads were identified and removed based on the *de novo* algorithm of UCHIME (Edgar *et al*., 2011) as implemented in the VSEARCH package (Rognes *et al*., 2016). The remaining set of high-quality reads was processed using minimum entropy decomposition (Eren *et al*., 2013, Eren *et al*., 2015). Minimum Entropy Decomposition provides a computationally efficient means to partition marker gene datasets into OTUs. To assign taxonomic information to each OTU, DC-MEGABLAST alignments of cluster representative sequences to the sequence database were performed. The most specific taxonomic assignment for each OTU was then transferred from the set of best-matching reference sequences (lowest common taxonomic unit of all best hits). Sequences identities of 70% across at least 80% of the representative sequence was set as

a minimal requirement for considering references sequences (reference database: /dbdir/nt. ltered.fa; Release 2020-02-03). Further processing of OTUs and taxonomic assignments was performed using QIIME software package (version 1.9.1). Abundances of bacterial taxonomic units were normalized using lineage-specific copy numbers of the relevant marker genes to improve estimates (Angly *et al*., 2014).

For comparative analyses between the experimental mice groups, non-normalized relative frequencies were used as input for MicrobiomeAnalyst software (Chong *et al*., 2020), and only OTUs with relative abundances above 0.1% were considered.

2.5 Visualization and data exploration

Data exploration, statistics, and visualization were performed using MicrobiomeAnalyst (Chong *et al*., 2020). Cluster analyses were employed since relationships between objects are expected to be discontinuous as well as defined categories or groups of objects are expected (Ramette *et al*., 2007). Heat tree analysis leveraging the hierarchical structure of taxonomic classifications was performed using the median abundances of taxa found in the microbial profiles and statistically compared using the non-parametric Wilcoxon Rank Sum Test (Foster *et al*., 2017, Chong *et al*., 2020).

2.5.1 Diversity index analyses

Alpha-diversity measure was employed to analyse the diversity within a sample or an ecosystem using Shannon-Wiener and/or Simpson indexes. Beta-diversity was employed to measure the diversity between samples using non-metric multidimensional scaling (NMDS) as ordination technique. For community analyses, Bray-Curtis and/or Jaccard dissimilarity statistic were performed as distance methods.

2.5.2 Testing for significant differences between groups

After representing objects in ordination plots and assess eventual clustering of similar objects, a statistical test was performed to assess differences between similarities within and between groups. Permutational analysis of multivariate dispersions (PERMDISP), Permutational analysis of variance (PERMANOVA), and Analysis of similarities (ANOSIM) were employed. Since neither PERMANOVA nor ANOSIM clearly suppose common variance among experimental groups, both approaches are sensitive to differences in multivariate dispersion (however, PERMANOVA is more sensitive than ANOSIM) (Giatsis *et al*., 2014). Consequently, PERMDISP was performed

in the first place to test the hypothesis of equality within group dispersion (*i.e.*, functional similarity). As PERMANOVA assumes homogeneity of multivariate dispersion between groups, its use is only recommended if PERMDISP previously accepted the null hypothesis – no dispersion between groups' dispersions (see Appendix II).

2.5.3 Clustering and correlation

Cluster analyses were performed in MicrobiomeAnalyst software (Chong *et al*., 2020) using Bray-Curtis dissimilarity distance measure and Ward as the clustering algorithm. Pattern search was used to identify possible taxa correlations with taxonomical gene marker (Chapter 4), sequencing platform (Chapter 5), and ethanol-exposure (Chapter 5). Although both Spearman and Pearson correlation coefficients might be employed to microbiome data – due to independence of variables under test – the non-parametric Spearman correlation coefficient (*rho*) was employed with an alpha-level of 0.05.

2.5.4 Comparison and classification

Non-parametric Linear Discriminant analysis (LDA) effect size (LEfSe) was employed for biomarker discovery and data explanation in the Illumina and nanopore sequencing microbial profiling approaches (Segata *et al*., 2013) incorporating the non-parametric Kruskal-Wallis test with an alpha-value of 0.05 (see Appendix III). Random Forest analysis was also employed with 1000 trees tested and 7 predictors attempted, to identify taxa that could works as classification tree predictors in mice experimental groups or microbial profiling sequencing approaches (see Appendix III).

Chapter 3 – Optimization of nanopore approaches using a fungal mock community

3.1 Brief introduction

The global fungal diversity is underestimated (Taylor *et al*., 2001, Garcia-Solache and Casadevall, 2010, Tedersoo *et al*., 2014, Pal *et al*., 2017). While next-generation sequencing approaches have been used to assess the diversity and role of the gut mycobiome, differences in experimental design and downstream analyses may induce bias and differential outcomes (Nilsson *et al*., 2019). Nanopore sequencing is a recently developed third-generation sequencing technology allowing the development of fast, cost-effective, and portable approaches to study complex ecosystems, such as, the gut microbiome, and it represents a promising tool to further unveil the fungal component of these microbial communities (Wang *et al*., 2021). Here a fungal mock community comprised of *Clavispora lusitaniae*, *Candida parapsilosis*, *Meyerozyma guilliermondii*, *Pichia kudriavzevii*, and *Candida glabrata* was used to assess the capacity of nanopore sequencing to retrieve its original microbial profile and relative abundance accurately and precisely. The approach was implemented using tge MinION system and involved the analysis of sequencing libraries made from equimolar mixtures of the PCR-amplified ITS genomic regions of the target species.

3.2 Results and discussion

3.2.1 *Candida glabrata* pure culture nanopore sequencing

3.2.1.1 Nanopore sequencing quality control assessment

In the nanopore sequencing targeted approach optimization, ITS amplicons previously generated by PCR from a pure *C. glabrata* culture were used as template for library preparation and sequencing. The nanopore sequencing general output parameters are provided in Table 3. A lower than anticipated number of reads were generated, though read lengths and quality were within the expected in the ITS fragment size from *C. glabrata* and nanopore sequencing reported error-rate (Kim *et al*., 2016, Nilsson *et al*., 2019). The low number of reads obtained reflects that the sequencing protocol used needs optimization. Despite the low number of reads, and consequently the low number of total bases sequenced, the remaining quality parameters were in accordance with expectations.

Table 3 – Nanopore sequencing performance overview of a *C. glabrata* pure culture.

Sample ID	**No. analysed reads**	**No. unclassified reads**	**No. classified reads w/ *centrifuge* score < 300 (%)**	**No. bases (Mbases)**	**Mean read length (bp)**	**Mean read quality (Qscore)**
Candida glabrata ITS amplicons	39,294	650	803 (2.0%)	36.0	916	9.95

In Figure 5 it can be observed that the number of nanopores sequencing was low, probably due to non-optimized library preparation. This observation explains the low number of sequenced reads. Additionally, most reads were obtained during the first hours of the experiment, and after 10 hours, their generation was negligible, supporting the remark that nanopore sequencing is highly sensitive to experimental factors, due to initial DNA quality, and inaccuracies inserted during library preparation, among others (Figure 6 and 7). The most probable cause of these results regards either DNA quality or DNA concentration used for library preparation, as seen in Figure 7, where the basecall quality decreased over time. Thus, there is a chance that poor and or/insufficient quality DNA was uploaded in the flowcell, originating the underperformance of the nanopores. Indeed, nanopores were not underperforming due to inherent bad quality but because there was not enough DNA in their surroundings to be sequenced. It was concluded that the amount of DNA needed to be increased despite the ONT protocol recommendations. However, this assumption needs further clarification, for instance by using a TapeStation to analyze the quantity and size of the DNA fragments.

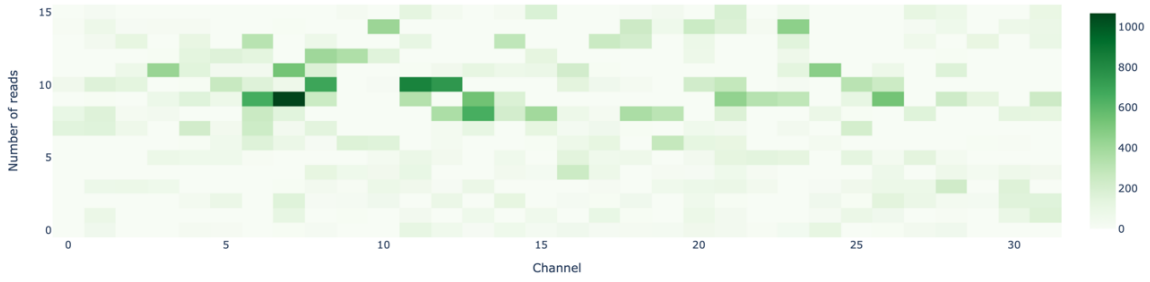


Figure 5 – Pore utilization map. The flow cell is comprised of 512 channels, each able to sequence one molecule through one of 4 nanopores. One rectangle represents one channel. The sensor chip panel is converted into a heatmap showing the number of reads generated per channel during the *C. glabrata* nanopore sequencing experiment. Darker green means that nanopores belonging to target channel sequenced a high number of reads. Light green and white mean that nanopores in those channels performed poorly.

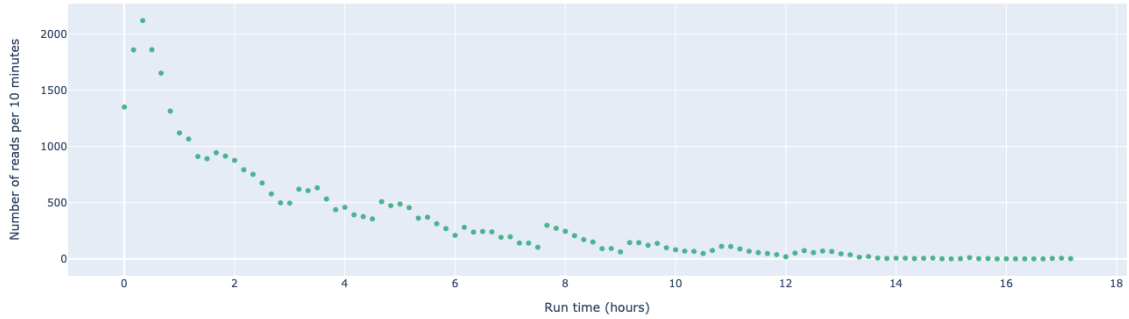


Figure 6 – Line graph showing the number of reads generated over time during the *C. glabrata* nanopore sequencing experiment. Each green dot represents the number of reads generated in the previous 10 minutes of sequencing. The x-axis shows the runtime when the data was collected, and the y-axis shows the number of reads generated in the last 10 minutes.

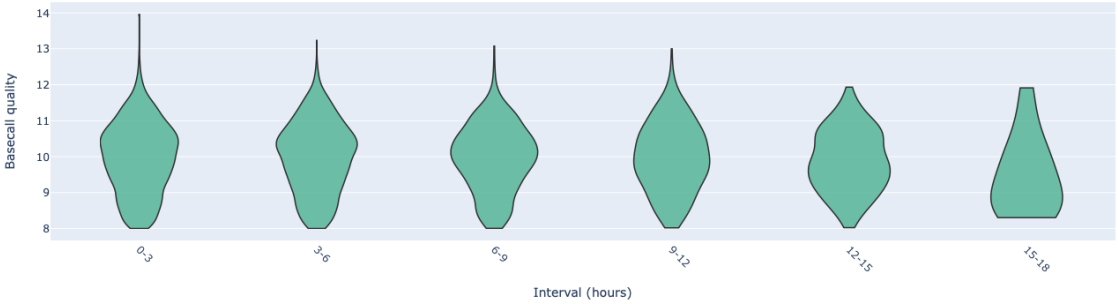


Figure 7 – Violin plot showing the nanopore sequencing basecall quality over time during the *C. glabrata* nanopore sequencing experiment. The x-axis shows the timepoint when the data was collected; data was merged to represent time intervals of 3 hours. The y-axis whos the basecall quality (Phred score) of each time interval.

Despite the low number of sequencing nanopores, which motivated the interruption of then sequencing run after 18 hours instead of the previous planned 24 hours, nanopore sequencing was able to retrieve the correct assignment of the pure culture present in the library. In Table 4 is shown the five most abundant taxa retrieved by WIMP.

Table 4 – Main five taxa retrieved by WIMP with their respective relative abundances and number of reads quality filtered due to low centrifuge assignment score.

TaxaID	**No. reads (%)**	**No. reads w/*centrifuge* score < 300 (%)**
Candida glabrata	35,615 (92.2%)	592 (1.7%)
Escherichia coli	991 (2.6%)	4 (0.4%)
Schizosaccharomyces japonicus yFS275	229 (0.6%)	5 (2.2%)
Paracoccidioides lutzii Pb01	105 (0.3%)	3 (2.9%)
Human mastadenovirus C	91 (0.2%)	16 (17.6%)

The majority of the classified reads were correctly assigned to *C. glabrata* in spite of the high basecalling error-rate (~10%). Almost 80% of the reads showed a *centrifuge* classification score above 300, and more than 90% of the classified reads was assigned to *C. glabrata* with the expected ITS fragment size (Table 4).

However, a total of 31 unrelated taxa were assigned by WIMP. Nevertheless, only one of non-*C. glabrata* taxa exceeded 1% of the classified reads – *Escherichia coli* (2.6%). The long list of 'taxonomical noise' obtained is explained by the high error-rate of the sequencing run (~10%), which originates a high number of sequenced read mismatches, possible reference databases limitations, and the underperformance of the Centrifuge algorithm to classify the reads (Ammer-Herrmenau *et al*., 2021). However, it can be observed that the number of reads with *centrifuge* scores below 300 increases with the decrease of the relative abundance, which supports the hypothesis of the high taxonomical noise retrieved by nanopore sequencing approaches, where a long list of low prevalent taxa was detected (Kim *et al*., 2016). However, other explanations cannot be excluded, such as human error that may had contaminated samples throughout one or more steps of library preparation, the eventual internal contamination present in the DNA extraction kits – 'DNA kitome' (Olomu *et al*., 2020), possible reference databases' limitations, and the underperformance of the Centrifuge algorithm to classify the reads (Ammer-Herrmenau *et al*., 2021). Specifically, DNA extraction of fungi is challenging due to their rigid cell wall, so genomic DNA from some fungi may be easier to extract than others. One example is the recent study from Frau and colleagues (2019) that reported method-based limitations in extract genomic DNA from fungal pure cultures of *C. albicans* and *Aspergillus fumigatus*.

Despite the 31 taxa found in the *C. glabrata* pure culture experiment, headed by *E. coli* with a relative abundance less than 3%, nanopore sequencing gave a good confidence to proceed with the fungal mock community experiment. The relevance of these observations poses questions when interpreting low prevalent taxa, since they can result either from basecalling errors occurring during nanopore sequencing, or from misclassifications in bioinformatics taxonomical assignment steps.

3.2.2 Fungal mock community nanopore sequencing

3.2.2.1 Nanopore sequencing quality control assessment

To continue the nanopore sequencing optimization attempts, a fungal mock community was tested. The mock community was generated by amplifying the ITS regions of five fungal species - *C. glabrata*, *C. parapsilosis*, *C. lusitaniae*, *M.*

guilliermondii, and *P. kudriavzevii*. After amplicons' generation and purification, they were mixed and pooled together in equimolar concentrations for library preparation. Two independent DNA sequencing libraries were prepared and sequenced in separate experiments. One first classification attempt was tried with WIMP that originated an inaccurate taxonomical assignment (see Appendix IV). The incongruent output retrieved by WIMP is related to the increased representation of closely related sequences in the database, allowing increased specificity of assignments at the family and genus level, but which might blur the line between lower levels of taxonomical classification, such as genus and species (ONT technical team, personal communication). Despite less curated entries in the reference databases, the bioinformatic pipeline behind WIMP might not be the best choice to analyse our type of data. Although the advantage of *centrifuge* is that it is optimized for solving metagenomic classification problems and it is capable of accurate identification of reads even when databases containing multiply highly similar references genomes are used (ONT technical team, personal communication), it has limitations compared with other algorithms such as *minimap2* (Ammer-Herrmenau *et al*., 2021). Thus, it can be hypothesized that a different approach based on BLAST/MEGAN can surpass these limitations (Bağcı *et al*., 2019).

WIMP attributes a score after determining the most reliable place in the taxonomy tree for those reads (Kim *et al*., 2016). On the other hand, BugSeq pipeline assigns taxonomical classifications based on the *recentrifuge* algorithm instead of *centrifuge* performed by WIMP, and it relies on a three steps pipeline node to assign taxonomical identifiers to every read (Fan *et al*., 2021). The first step is the reads mapping using *minimap2*, the second is the reassignment of mapped reads based on a Bayesian statistical framework using *pathoscope* (Francis *et al*., 2013). This reassignment step optimizes the multiple strains coverage, mediating the absence of target taxa in the reference database, surpassing the need of multiple alignments, extensive homology searches, or genome assembly steps (Francis *et al*., 2013). The last step in the BugSeq pipeline is the calculation of the Lowest-Common Ancestor as input for *recentrifuge* (Marti, 2019), which facilitates vigorous contamination removal (such as crossovers) and quality comparative analysis of multiple samples, giving a confidence level score for every result that propagates to further downstream analyses. Therefore, BugSeq's pipeline may be more curated than WIMP's, since it includes a post-classification step to remove and quality control the assignments which is absent in WIMP pipeline (Marti, 2019, Fan *et al*., 2021). Therefore, BugSeq pipeline was employed for nanopore sequenced reads data analyses (Table 5).

Table 5 – Nanopore sequencing performance overview of two independent sequencing runs from the same fungal mock community targeting ITS amplicons.

Sample ID	Quality control[1]		Alignment parameters			
	No. reads	**Average Qscore**	**No. analysed reads**	***recentrifuge* score average**	**Average read length (bp)**	**No. taxa detected**
Run A	1,314,571	12.03	1,308,263	56.9	512	146
Run B	4,879,747	10.65	4,655,632	58	543	255

[1] – Quality control was obtained through *multiqc* tool and alignment's quality through *recentrifuged* tool in BugSeq pipeline.

As previously observed, nanopore sequencing tends to assign many taxa even when templates from a pure culture were being sequenced (Table 5). Both average Qscore and *recentrifuge* scores revealed a high basecalling error-rate (~6-9%) and a low alignment score (56.9-58.0, with 57-58% of the sequenced reads matched to one entry in the reference databases). Indeed, nanopore sequencing could retrieve the correct assignments of the mock community in both sequencing runs (Table 6). BugSeq could attribute the expected taxa identifiers even when the *recentrifuge* assignment score barely surpass the 50% of matching. However, the presence of ITS sequences in the references databases is satisfactorily represented and thus a 50% match is considered sufficient to assign the correct taxon using BugSeq bioinformatic pipeline. Still, the relative abundances of these species in the fungal mock community did not yield the expected value (Table 6). *C. lusitaniae* was consistently overrepresented in the mixtures, showing relative abundances of 40% to 47% of the assigned reads instead of the expected 20%, while *C. glabrata* was underrepresented, with only 4% to 8% of the assigned reads.

Table 6 – Nanopore sequencing performance overview of two independent runs (termed 'Run A' and 'Run B') from the same fungal mock community targeting ITS amplicons. Taxonomic assignments were retrieved through *recentrifuge* tool in BugSeq pipeline.

Taxa ID	ITS fragment size (bp)	No. reads		Relative abundance (%)		Normalized average *recentrifuge* score	
		Run A	**Run B**	**Run A**	**Run B**	**Run A**	**Run B**
Clavispora lusitaniae ATCC 42720	552	607,695	1,842,887	46	40	53.5	56.7
Candida parapsilosis	690	237,533	532,047	18	11	62.4	65.3
Meyerozyma gulliermondii ATCC 6260	775	207,528	773,358	16	17	60	63.1
Pichia kudriavzevii	679	179,138	859,268	14	18	60.6	62.6
Candida glabrata	1050	54,718	376,756	4	8	61.3	64

The two independent fungal mock community nanopore sequencing runs showed high precision. Despite a difference of around 3.5 million reads between the two sequencing runs, the average Qscore (12.03 *vs* 10.65), *recentrifuge* alignment score (56.9 *vs* 58), and average read length (512 *vs* 543) were similar. The five species were detected and were the most abundant taxa observed between the assigned taxa. The overall performance here described revealed an expected nanopore sequencing accuracy, with relatively high error-rate and a long list of 'taxonomical noise' (146 total taxa found in Run A, and 255 in Run B). Indeed, a higher number of taxa detected was observed in Run B, the sequencing run where the most sequenced reads were generated. In the Run A, 1% of the total reads were not assigned, and less than 1% was assigned to bacteria (0.2%), synthetic constructs (0.2%), viruses (0.2%), and archaea (less than 0.01% assigned to *Sulfolobus acidocaldarius*). In the Run B, 5% of the total reads were not assigned, and less than 1% was assigned to bacteria (0.6%), synthetic constructs (0.1%), viruses (less than 0.01%), and archaea (less than 0.01% assigned to *Euryarchaeota*). However, the average classification score and the high error-rate observed were not sufficient to hamper the fungal mock community taxonomical assignments since the original fungi mock community's members were correctly assigned (Table 6).

Recently, D'Andreano and colleagues (2020) were successful in developing a nanopore sequencing approach to assess the fungal composition from clinical samples. The authors targeted a 3.5 kb region (V3, 18S-ITS1-5.8S-ITS2-28S D2) and a 6 kb region (V1, 18S-ITS1-5.8S-ITS2-28S D12) for amplicons' generation, performed taxonomical assignments using WIMP, and compared the performance with a known fungal mock community – ZymoBIOMICS™ (D'Andreano *et al*., 2020). Besides using longer amplicons than the approach here tested, the authors also removed barcodes and adapters using *porechop* (which is unsupported since October 2018). However, it is unlikely that the inclusion of barcodes and adapters to downstream analyses could explain such significant discrepancies found in the fungal mock community analysed in this dissertation, because *C. parapsilosis* (Table 6, run A), *M. guilliermondii* (Table 6, run B), and *P. kudriavzevii* (table 6, run B) showed relative abundances as expected (Table 6). Moreover, there was a difference of 1-7 percentage points between the two independent runs attributed to the same taxa, which was more probably explained by chance or other factors, such as basecalling error-rate, or *centrifuge* indexation, than to the presence of barcodes and adapters. In another study, Hu and colleagues (2021) evaluated a nanopore sequencing approach to study a fungal mock community comprised of 43 species. The

authors generated amplicons by amplifying the ITS1 regions of the rRNA gene with the universal fungal primers ITS1F and ITS2 and performed adapters and barcodes' trimming with *qcat* after a barcoded nanopore sequencing run (Hu *et al*., 2021). The authors compared classification and community analysis pipelines for nanopore sequencing data in amplicon and shotgun approaches from fungal mock communities comprised of 43 species (Hu *et al*., 2021). Their approach consisted in the generation of a custom reference database comprised solely with the species from the mock community, *i.e.*, all the genomes of the 43 species in the mock community were downloaded from the NCBI (Hu *et al*., 2021). Then, the authors used *kraken2* to perform a search to identify potential contaminated regions in the concatenated FASTA and masked those regions using *bedtools* (Quinlan *et al*., 2010). Low complexity regions were also masked using *dustmasker* from the BLAST+ package (Camacho *et al*., 2009). Lastly, they used *makeblastdb* to construct the mock community custom database (Hu *et al*., 2021). In regards to the amplicon-based approach, the authors used *kraken2* as the k-mer based-algorithm and *minimap2* as the alignment-based algorithm. In the *minimap2* analysis, the accessions of the best hits were extracted from the output files and their taxonomic correspondence were searched in the NCBI taxonomic map using python *pandas* module (Hu *et al*., 2021). Finally, the authors merged information from different output files and performed *ete3* module again to assign taxonomic information to each read (Hu *et al*., 2021). The authors concluded that combining BLAST with the fungal specific RefSeq fungal database was the best approach to obtain the most precise classifications for the fungal mock community data sets (Hu *et al*., 2021). The classification can also be improved if cutoffs on query coverage were applied, including flow on effects on downstream community composition analysis sourced from complementary shotgun/metagenomics data sets (Hu *et al*., 2021). Such curated nanopore sequencing approach decreased the chance of misclassification and mismatches in the classification. However, the nanopore sequencing approach here tested to sequence a fungal mock community revealed a correct classification although with dubious relative abundances results despite the low curation of the bioinformatic pipeline performed (Table 6). These results cannot be explained by PCR amplification biases since the mock community was built with equimolar concentrations of each species ITS amplicons after PCR (Hu *et al*., 2021). In detail, PCR amplification biases derive mainly from inconsistent amplification of the genomic barcoding region that differs in every species, caused by copy number variations and different primer binding specificities (Nearing *et al*., 2021, Silverman *et al*., 2021). Additionally, in fungi, variations in barcode genomic regions lengths are the

most probable cause to explain biases in recording fungal community compositions, with longer amplicons being underrepresented (Castaño *et al*., 2020). Thus, these results give evidence that we can treat nanopore sequence data with confidence despite the high error-rate. Since nanopore sequencing is a novel technology there is not enough experimental and analytical standards and so it is challenging to define what could be accepted or not as a correct approach (Leggett *et al*., 2017). Additionally, recent studies have observed that the extra length of nanopore reads overcome the accuracy limitations (Pearman *et al*., 2020). Still, it is expected that ONT or other researchers develop better sequencing chemistry and bioinformatic tools to improve these challenges, as it is already expected by the super accurate basecalling model (announced as < 2% error-rate).

In conclusion, these results confirm that nanopore sequencing is a very sensitive approach, and the protocol execution must be performed with care and optimized to the specific biological sample and scientific question under investigation. Impairments may occur either upstream during sample collection, DNA extraction, PCR amplicons' generation, *etc*, or downstream, during or post-sequencing, such as basecalling errrors, and further issues in the bioinformatic pipeline performed.

Chapter 4 – Targeted and untargeted nanopore sequencing approaches to profile the gut microbiota of mice infants exposed to ethanol in utero

4.1 Brief introduction

The study of hosts' microbiomes has various experimental and data analyses challenges. They are due to the multidisciplinary nature of the field with input from microbiology, genomics, bioinformatics, epidemiology, among others. Nonetheless, the consistency and commonality of procedures are not well established (Simoneau *et al*., 2021). Indeed, nanopore sequencing approaches rely on longer reads, while reference databases are often composed shorter gene markers, such as fragments of the 16S rRNA gene or the ITS region (Markey *et al*., 2020, Matsuo *et al*., 2021). Therefore, nanopore sequencing may have limited reliability in profiling unknown microbial communities, decreasing the confidence on the taxonomical composition obtained (Santos *et al*., 2020). Here, two targeted and one untargeted nanopore sequencing approaches were optimized to profile the gut microbiota of 12 infant mice exposed to ethanol *in utero*. The sequencing was implemented using MinION system and the targeted approaches involved the sequencing of the full-length 16S rRNA gene and 16S-ITS-23S region of the *rrn* operon PCR-amplified amplicons. In the untargeted approach, a sequencing library was prepared directly from barcoded DNA templates extracted from the same infant mice samples.

4.2 Results and discussion

4.2.1 Nanopore sequencing quality control assessment

In this study, total DNA was extracted from 12 infant mice faecal samples exposed to ethanol *in utero*. The aim was to discern the best bacterial genomic region for nanopore sequencing targeted approaches. Two targets were selected and compared – the full-length of the 16S rRNA gene (~1,500 bp), and the whole *rrn* operon (16S rRNA gene-ITS-23S rRNA gene, with ~4,500 bp). Additionally, the results obtained using these two targets were compared with a shotgun/metagenomic microbial profiling approach (untargeted). For the full-length 16S rRNA gene, two sequencing runs were performed, with barcodes tagging six samples per run. For the 16S-ITS-23S operon region and the metagenomic approach, a single sequencing run was performed, with one barcode tagging each of the 12 unique samples. The nanopore sequencing runs' outputs are summarized in Table 7. Despite ONT possess a sequencing kit tailored for the 16S gene (*eg*. SQK-RAB-204 with the 27F and 1249R primer set), all sequencing experiments were performed with the Ligation Sequencing kit (LSK109) because the approach targeting the

16S-ITS-23S and the untargeted approach couldn't be analysed with the 16S sequencing kit.

Table 7 – Nanopore sequencing performance overview depicting the total number of reads generated, the total number of bases sequenced, N50, average read quality, and percentage of reads with less than 10% (Q10), and 8.8% (Q12) error-rate.

Sequencing run[1]	**No. reads**	**No. bases (Mbp)**	**Mean read length (bp)**	**Read length N50[2] (bp)**	**Mean read quality (Qscore)**	**Percentage of reads above quality cutoff (%)**	
						Q10	**Q12**
Targeted full-length 16S rRNA gene 1-6	1,453,818	2,300	1,594	1,466	10.65	54.8	2.5
Targeted full-length 16S rRNA gene 7-12	374,050	573	1,532	1,476	10.12	34.5	0.4
Targeted 16S-ITS-23S region 1-12	701,930	2,600	3,708	4,005	10.75	58.8	2.2
Untargeted/ metagenomic 1-12	952,483	1,500	1,555	6,353	10.73	52.3	8.5

[1] – 1-6, 7-12, 1-12, represent the samples that were barcoded with a unique tag.

[2] – In nanopore sequencing experiments, N50 is defined as the weighted median length of the sequence in a set for which all sequences of that length or greater sum to 50% of the set's total size (Shafin *et al*., 2020).

As previously observed and discussed (Chapter 3), nanopore sequencing showed similar sequencing performance between the 4 runs. One of the targeted approaches revealed an expected mean sequenced read length (1,500 bps for the full-length 16S rRNA gene). Interestingly, the average read length for the 16S-ITS-23S region of the operon was below the expected (3,708 bps *vs* 4,500 bps). Since the overall sequencing parameters were within the expected values, it can be concluded that the shorter sequence read length observed was a direct consequence of a shorter DNA fragment.

In this study, we explored a nanopore sequencing metagenomic approach to microbial profile infant mice gut samples obtained from mice exposed to ethanol *in utero*. One advantage of such approaches is the removal of upstream experimental steps, such as amplicon amplification by PCR, which may introduce amplification biases prior of library preparation, as previously discussed here and elsewhere (Nilsson *et al*., 2019,

Nearing *et al*., 2021). In the experimental nanopore metagenomic approach tested here, it was attempted the most simple and fastest protocol. The results were compared with the two targeted approaches afterwards.

Despite the N50 of 6,353 bps, the average read length was 1,555 bps, which indicates that the DNA was fragmented throughout library preparation, leaving few longer fragments ready to be sequenced (Table 7). As previously noted, this is a direct observation that supports the hypothesis of nanopores' preference for shorter fragments. The remaining nanopore sequencing quality parameters were in accordance with what was seen in the targeted approaches (Table 7).

In the metagenomic approach, several long reads were detected (Figure 8), which is a feature of nanopore sequencing. In Table 8 are provided the 10 longest reads with their basecalling (Qscore) and alignment scores (WIMP – *centrifuge*, and BugSeq - *recentrifuge*). Interestingly, clear discrepancies were observed between the two microbial profiling pipelines, where most of the longest reads taxonomically assigned by WIMP were not assigned by BugSeq. The exceptions were *Muribaculum intestinale*, *Duncaniella* spp. – although assigned to different strains (*Duncaniella* sp. B8 and *Duncaniella* sp. C9, recently reclassified as members of the same species, *Duncaniella muris* (Miyake *et al*., 2020a)) –, *Bacteroides caecimuris*, and *Bifidobacterium animalis*. Indeed. *M. intestinale* (*recentrifuge* score = 23) and *B. caecimuris* (*recentrifuge* score = 28) showed a low classification *recentrifuge* score. This observation means that the correspondent matched long reads were enough to retrieve the same taxonomical assignment by both classification algorithms. Additionally, and as observed in the fungal mock community experiment, BugSeq *recentrifuge* score around 50 (or as low as 23) is enough to probably assign the correct species identification, matching WIMP assignment, revealing the robustness of BugSeq pipeline (Table 8). The differences detected in the two bioinformatic pipelines are explained by the different approaches employed. One of the major differences is the classification algorithm. BugSeq uses *recentrifuge* in order to overcome the limitation of *centrifuge* used by WIMP (Fan *et al*., 2021). Additionally, BugSeq employs a better curated bioinformatic pipeline than WIMP (quality control with *fastp*, mapping with *minimap2*, reassignment with *pathoscope*, lowest common ancestor calculation with *recentrifuge*, quality control visualization with *multiqc*). Therefore, WIMP represents a much rawer pipeline if applied alone.

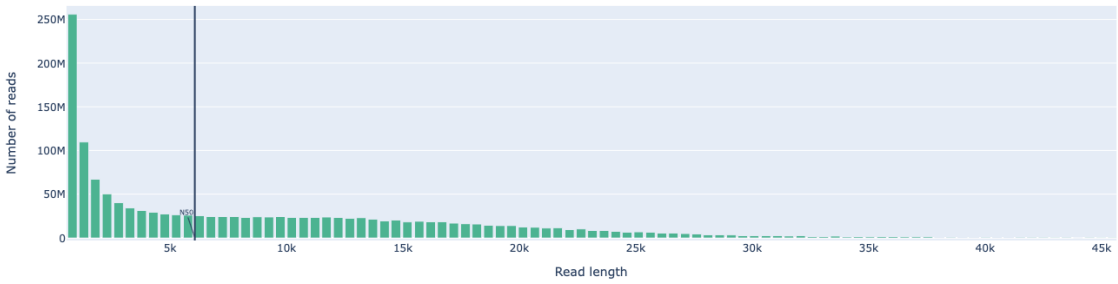


Figure 8 – Metagenomic/ shotgun nanopore sequencing weighted histogram showing the number of reads generated, and their respective read lengths, when the 12 DNA templates extracted from infant mice faecal samples were analysed. In the x-axis is displayed the sequence read length in bases. The vertical black line represents the N50 (the weighted median length of the sequence in a set for which all sequences of that length or greater sum to 50% of the set's total size).

Table 8 – Ten longest reads generated by shotgun/metagenomic nanopore sequencing of the 12 DNA templates extracted from infant mice faecal samples.

WIMP taxa assignment	**Read length (kb)**	**Read score (Phred)**	**WIMP *centrifuge* score**	**BugSeq taxa assignment**	**Read length (kb)**	**BugSeq *recentrifuge* score**
Duncaniella sp. B8	70.8	10.2	14086	Not assigned	70.7	5
Bacteroides fragilis 638R	68.6	8.4	169	Not assigned	68.4	5
Selenomonas sp. oral taxon 136	64.1	8.4	289	Not assigned	63.8	5
Duncaniella sp. B8	60.9	10.4	6542	Not assigned	60.8	5
Muribaculum intestinale	60.7	9.9	67,028	*Muribaculum intestinale*	60.2	23
Duncaniella sp. B8	55.9	10.0	202,964	*Duncaniella* sp. C9	55.7	57
Prevotella jejuni	55.0	11.5	144	Not assigned	54.9	5
Bacteroides caecimuris	53.8	11.9	131,389	*Bacteroides caecimuris*	53.6	29
Bifidobacterium animalis subsp. *animalis*	52.6	11.2	294,529	*Bifidobacterium animalis*	52.5	56
Muribaculum intestinale	52.1	10.9	77,822	Not assigned	52.0	5

Each row corresponds to a single read.

In the metagenomic approach, most reads were not classified by WIMP (76.5%) and 35% of those classified were assigned to *Homo sapiens* (8.2% of total analysed reads) (Table 9). This observation might be explained by contamination throughout the sample handling and/or library preparations (Eisenhofer *et al*., 2019). However, due to the high basecalling error-rate of nanopore sequencing the hypothesis that reads assigned to *Homo*

sapiens were misassigned, truly belonging to the mouse host, cannot be excluded. Differently from the targeted approaches, the metagenomic approach showed a majority of *Homo sapiens* reads in those that were classified (35%) (Table 9). Most probably due to a contamination in library preparation, or mice sequences wrongly attributed to *Homo sapiens*, those reads were removed before further analyses were made. Since *Homo sapiens* reads were absent or negligent in the targeted approaches, the contamination of the samples was not introduced during library preparations steps. Besides the "DNA kitome" (Salter *et al*., 2014, Thoendel *et al*., 2017, Olomu *et al*., 2020), there are other type of contamination, such as reagents (Kulakov *et al*., 2002, Olm *et al*., 2017), host material (Ames *et al*., 2015), and surrounding environment after sampling (e.g., airborne particles, and crossovers between DNA under analysis or from remaining DNA kept unattended in the flowcell from previous sequencing runs) (Lusk *et al*., 2014, Gruber *et al*., 2015). These contaminants would be sequenced alongside the desired sequencing library, potentially leading to biases in taxa abundances, sequencing coverage, specifically in library prepared with low biomass DNA (Lu *et al*., 2018). Additionally, the hypothesis of cross-contamination installed in the metagenomic reference databases cannot be fully discarded (Ames *et al*., 2015, Gruber *et al*., 2015, Lu *et al*., 2018). Still, it cannot be discarded that these reads were misassigned from the host mice material.

Table 9 – Shotgun/metagenomic nanopore sequencing performance parameters of 12 DNA templates extracted from infant mice faecal samples, displaying the total number of analysed, classified, and unclassified reads with their respective average read length and WIMP basecalling score.

Barcode ID	**Total reads analysed**	**No. reads classified**	**No. reads unclassified**	**Mean read length (bp)**	**Mean read quality (Qscore)**	***Homo sapiens* reads (%)**
All combined[1]	952,483	223,397	728,886	1,555	10.73	78,236[3] (35.0)
Sample 1	56,815	17,115	39,700	2,037	10.76	5,278[3] (30.8)
Sample 2	49,893	9,069	40,809	1,606	10.70	1,050[5] (11.6)
Sample 3	37,491	8,205	29,286	2,485	10.73	1,488[3] (18.1)
Sample 4	33,600	10,706	22,894	3,106	10.73	2,684[3] (25.1)
Sample 5	119,285	23,616	95,669	1,144	10.76	6,826[3] (28.9)
Sample 6	120,685	26,080	94,605	1,258	10.70	5,527[4] (21.2)
Sample 7	93,369	25,441	67,928	1,268	10.93	14,159[3] (55.7)
Sample 8	77,840	12,983	64857	1,776	10.65	1,265[5] (9.7)

Table 9 – (continued).

Barcode ID	Total reads analysed	No. reads classified	No. reads unclassified	Average read length (bp)	Mean read quality (Qscore)	No. *Homo sapiens* reads (%)
Sample 9	55,777	11,886	43,891	1,829	10.66	1,002[4] (8.4)
Sample 10	125,584	23,497	102,087	984	10.84	8,480[3] (36.1)
Sample 11	55,724	14,242	41,482	2,128	10.81	4,710[3] (33.1)
Sample 12	73,044	24,389	48,655	1,507	10.92	16,774[3] (68.8)
No barcode[2]	53,191	16,168	37,023	1,657	9.94	8,993[3] (55.6)

The total number of reads assigned to *Homo sapiens* is shown. Each barcode corresponds to one infant mouse faecal sample.

[1] – all infant mice faecal samples combined.

[2] – fragments sequenced with no barcode detected.

[3] – most abundant taxon.

[4] – second most abundant taxon.

[5] – third most abundant assigned taxon.

To further evaluate the microbial profiles retrieved by the metagenomic approach, and to perform abundance and correlations' comparisons with the two targeted approaches all *Homo sapiens* reads were removed.

4.2.2 Influence of targeted and untargeted approaches in the microbial profiles obtained by nanopore sequencing

In Table 10 it is shown the combined relative abundances in the pooled samples of the three most abundant taxa at phylum, genus, and species level when different approaches – targeted *vs* untargeted – were used to microbially profile the same infant mice faecal samples.

Table 10 – Top three most abundant taxa classified by WIMP at the phylum, genus, and species level of the 12 DNA templates extracted from infant mice faecal samples, for each sequencing approach.

Taxonomical rank	16S rRNA gene (relative abundance)	16S-ITS-23S operon region (relative abundance)	Metagenomic (relative abundance)
Phylum	*Firmicutes* (74%) *Bacteroidetes* (22%) *Verrucomicrobia* (3%)	*Firmicutes* (59%) *Bacteroidetes* (41%)	*Bacteroidetes* (57%) *Actinobacteria* (25%) *Proteobacteria* (12%)
Genus	*Faecalibaculum* (28%) *Duncaniella* (11%)	*Faecalibaculum* (50%) *Duncaniella* (24%)	*Muribaculum* (29%) *Curtobacterium* (21%)

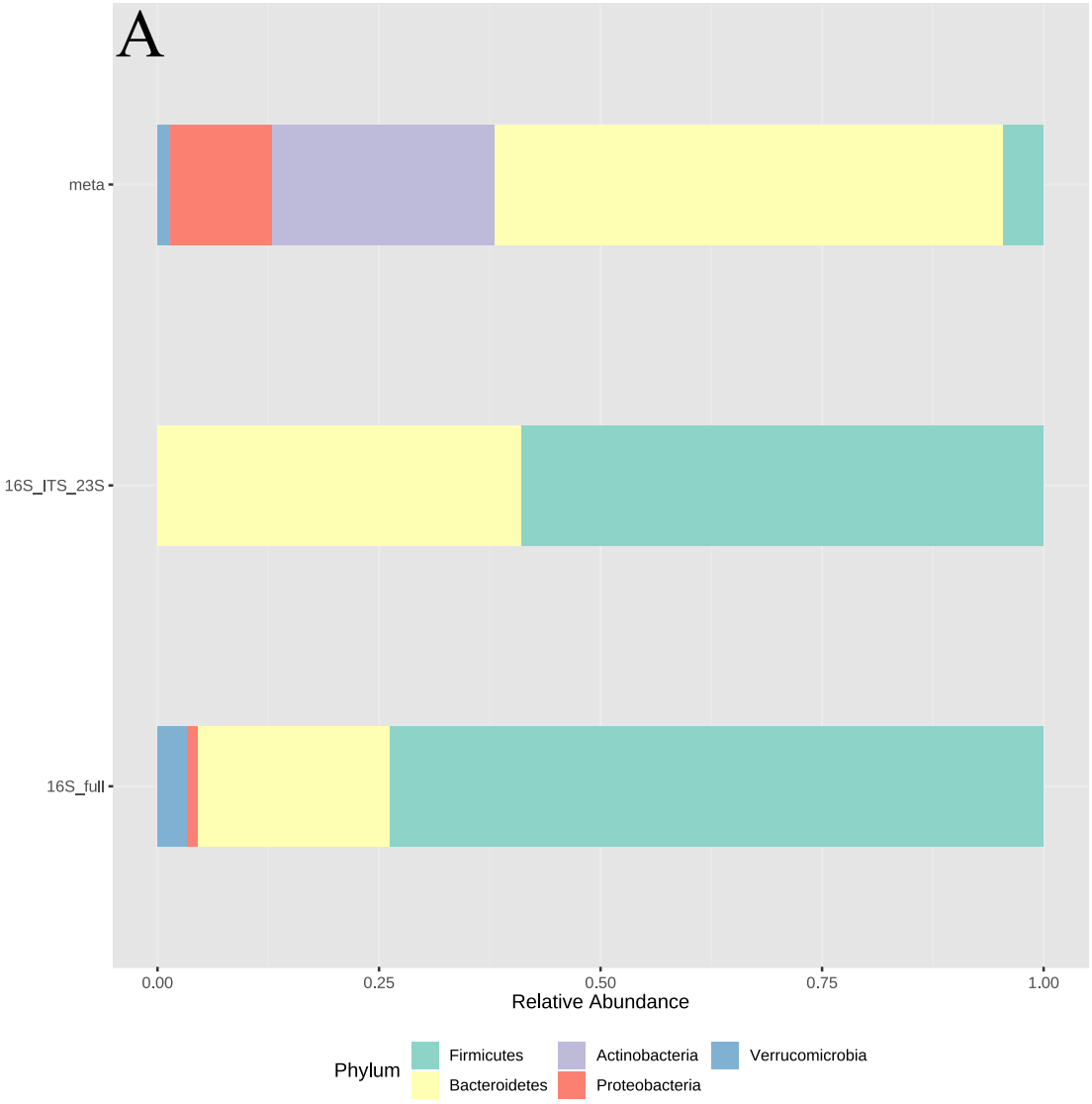
Table 10 – (continued).

Taxonomical rank	16S rRNA gene (relative abundance)	16S-ITS-23S operon region (relative abundance)	Metagenomic (relative abundance)
Genus	*Muribaculum* (10%)	*Muribaculum* (16%)	*Duncaniella* (18%)
Species	*Faecalibaculum rodentium* (28%)	*Faecalibaculum rodentium* (50%)	*Muribaculum intestinale* (29%)
	Duncaniella sp. B8 (11%)	*Duncaniella* sp. B8 (24%)	*Curtobacterium flaccumfaciens* (21%)
	Muribaculum intestinale (10%)	*Muribaculum intestinale* (16%)	*Duncaniella* sp. B8 (18%)

Relative abundances of all sequenced reads were combined to calculate the percentual relative abundances.

Based on the WIMP classification, the three approaches were able to assign a different number of taxa showing the distinctive microbial profiling capacity of the genomic targets. At the phylum level, the full-length 16S rRNA gene-targeted nanopore sequencing approach assigned a total of 4 taxa, while the 16S-ITS-23S region-targeted approach assigned 2, and the metagenomic approach assigned 5 (Figure 9). At the genus level, the full-length 16S rRNA gene-targeted approach assigned a total of 26 taxa, while the 16S-ITS-23S region-targeted approach assigned 16, and the metagenomic approach assigned 13 (Figure 9). Finally, at the species level, the full-length 16S rRNA gene-targeted approach assigned a total of 31 taxa, while the 16S-ITS-23S region-targeted approach assigned 17, and the metagenomic approach assigned 13 (Figure 9). The low number of taxa detected by the 16S-ITS-23S region-targeted approach is most probably explained by the lower prevalence of sequences spanning the 16S-ITS-23S operon in the reference databases (Martijn *et al*., 2019). The same observation can be made for the metagenomic approach. One explanation may be the less reliable number and curation of whole-genomes present in the reference databases (Breitwieser *et al*., 2019). Other explanations are possible like the reliability of the species assignments in the targeted approaches and other software-based limitations, copy number variation, intragenic variation, non-linked 16S-ITS-23S sequences presence in the databases, *etc* (Breitwieser *et al*, 2019, Johnson *et al*., 2019, Brewer *et al*., 2020). However, since longer reads were sequenced, there is a chance that some part of the read covered and matched a more represented sequence in the reference database (for instance, one or more regions of the 16S rRNA gene). Additionally, recent studies have been reporting a higher prevalence of taxa showing non-canonical (non-linked) 16S-ITS-23S DNA sequences (Ahn *et al*., 2020, Brewer *et al*., 2020), and these non-linked arrangements have the potential to disturb

taxonomical assignments when the 16S-ITS-23S region of the *rrn* operon is targeted for sequencing approaches (Brewer *et al*., 2020). Still, some of these taxa were detected with a low prevalence and their detection may also be explained by basecalling errors occurred during nanopore sequencing.



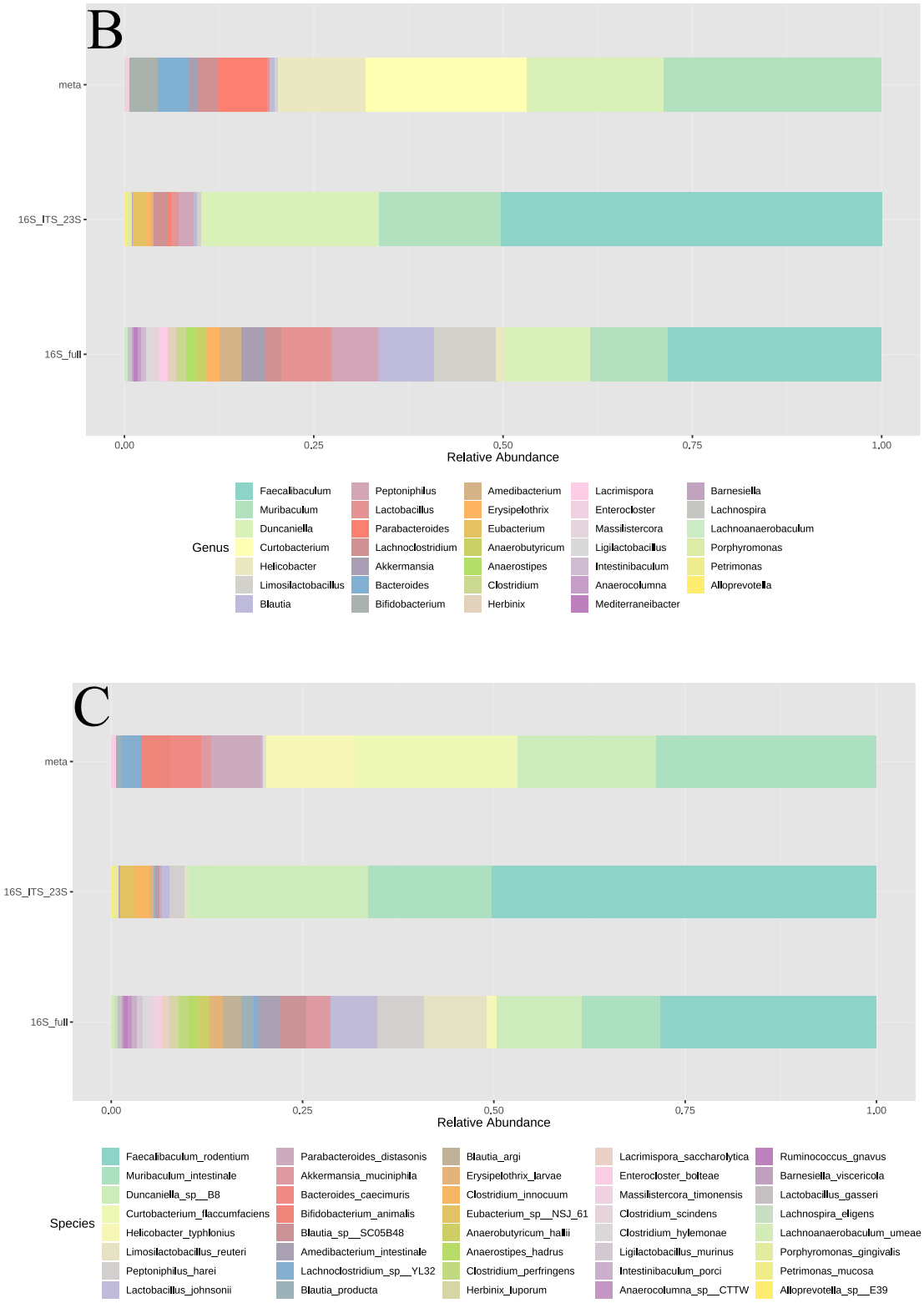


Figure 9 – Bar chart showing the relative abundance of the combined taxa at the phylum (A), genus (B), and species level (C) detected by the 16S rRNA gene-, 16S-ITS-23S region-targeted, and untargeted shotgun/metagenomic nanopore sequencing approaches when 12 DNA templates extracted from infant mice faecal samples were pooled together.

The *Firmicutes*/*Bacteroidetes* (F/B) ratio differed from the sequencing approaches, with the targeted approach based on the full-length 16S rRNA gene having

the highest ratio (7.33), followed by 16S-ITS-23S region-targeted approach (2.45), and then the metagenomic approach with the lowest ratio (0.07). Since the number and distribution of taxa detected between the targeted and untargeted sequencing approaches may be due to other factors than the true microbial composition of the samples (like high basecalling error-rate and misrepresentation in the reference databases), the F/B ratio may give a misleading conclusion regarding the biological meaning of the microbial profiles found.

Some differences were observed between the nanopore sequencing approaches. The phylum *Verrucomicrobia* was only detected when the full-length 16S rRNA gene was used as genomic sequencing target. The only species assigned to this phylum was *Akkermansia muciniphila*, and the expected relative abundance of this taxon in the 16S-23S rRNA operon was less than 1% (Seol *et al*., 2022). There is a lack of 16S-ITS-23S rRNA operon sequences representation in the databases and they may have dubious quality (Bowers *et al*., 2017, Shaiber and Eren, 2019). These two observations and the possibility of occurring variants in the primer binding site (Seol *et al*., 2019) can explain why the targeted nanopore sequence approach based on the 16S-ITS-23S operon region was not able to assign *A. muciniphila* to any read. Moreover, the use of the 16S-ITS-23S operon region as a barcoding genetic marker is not always appropriate for microbial profiling (Seol *et al*., 2019). Since non-linked rRNA genes have been increasingly detected in environments that facilitates slower-growing or symbiotic taxa (Ahn e *et al*., 2020, Brewer *et al*., 2020), caution must be taken when choosing these type of barcodes for microbial profiling. Therefore, there is a relevant number of taxa that will be neglected if present in the samples under study, such as *Helicobacter*, *Rickettsia*, *Wolbachia*, *etc* (Seol *et al*., 2019). Up until now, unliked rRNA genes have been mostly reported in environments populated with slow-growing taxa like soils and sediments (Vieira-Silva and Rocha, 2010, Brown *et al*., 2016). However, they might be more frequent and present in free-living taxa as well, including members of *Verrucomicrobia* (Brewer *et al*., 2020). Interestingly, these taxa do not show slow growth patterns, low rRNA copy numbers, mutations' fixation by genetic drift, small effective population sizes, nor reduced genome sizes (Brewer *et al*., 2020). Indeed, these taxa are commonly abundant and ubiquitous in their environments and having non-linked rRNA genes could eventually confer a selective advantage by facilitating the production of heterogenous ribosomes with a diverse array of features (Brewer *et al*., 2020). This speculative hypothesis is supported by recent reports that observed specialization in some rRNA loci detected in these taxa, which were translating genes involved in adaptation phenotypes related to temperature

and nutrient shifts (Song *et al*., 2019). Generally, it can be speculated that the different microbial profiles given by the two targeted nanopore approaches may be explained at least in part by the presence of non-linked 16S-ITS-23S DNA sequences as some taxa have separated 16S and 23S rRNA genes across the genome (Brewer *et al*., 2020). Since there is an imbalance in the public databases that favours fast-growing and easily cultivated organisms (Vartoukian *et al*., 2010, Zhi *et al*., 2012), it is expected that targeted sequencing approaches could find biases towards those taxa, neglecting others in possess of non-linked 16S-23S rRNA arrangements. However, this speculative explanation needs to be further studied in other organisms, such as in mice, since most, if not all, bacterial rRNA genes found in the human gut are linked (Brewer *et al*., 2020).

In the metagenomic approach, no reads were assigned to *Firmicutes*, and it was the only approach that detected *Proteobacteria* (Table 10). Moreover, *Curtobacterium flaccumfaciens* was only detected in the metagenomic approach with a relatively high abundance (21%). This species is widely recognized as the causative agent of bacterial wilt of edible dry beans worldwide (Osdaghi *et al*., 2020, Osdaghi *et al*., 2022) and its presence cannot be easily explained. Since all mice received the same diet, a feed contamination is unlikely but cannot be entirely ruled out. There is also a chance that the longer reads generated by the shotgun/metagenomic nanopore sequencing approach could be incorrectly assigned (average reads error-rate = 8.5%, Table 8). However, further research must be necessary to clarify the global taxonomical annotations within the genus *Curtobacterium* (Evseev *et al*., 2022). Of note is the higher number of longer reads generated by the shotgun/metagenomic nanopore sequencing approach. This approach retrieved a total of 34,150 reads with 10 or more kilobases. As discussed before, only *M. intestinale*, *Duncaniella* sp., *B. caecimuris*, and *B. animalis* had the same match between the taxonomy assignment algorithms from WIMP and BugSeq (Table 8). The observed mismatches between the classification algorithms supports the need of the development of standard protocols to analyse nanopore sequencing data, thus allowing further comparisons between independent studies. Since the pipelines have different correction and classification bioinformatics' steps, it is expected some degree of mismatches. For instance, BugSeq has a final filter step using *recentrifuge*, which performs vigorous contamination removal (Marti, 2019). Therefore, it can be extrapolated that using different classification bioinformatics' tools could retrieve different assignments to the same reads, imposing the need of a much more curated pipeline than the one here tested with WIMP, as the abundant prevalence of *Homo sapiens* reads, the high number of unclassified reads, and the relatively high error-rate show. Nevertheless, the presence of

such longer reads are relevant for annotation of whole genomes and proceeding to further hypothesis testing related to the functional analysis. Although not here performed, whole-genome and functional analysis represent an advantage of exploring untargeted shotgun/metagenomic nanopore sequencing approaches (Anand *et al*., 2007, Schneider and Dekker, 2012, Feng *et al*., 2015, Brown *et al*., 2017, van Dijk *et al*., 2018, Bahram *et al*., 2019, Leggett *et al*., 2020).

Still, there is a chance that the bioinformatic pipeline here employed for the 16S-targeted approach was not suitable for the reliable detection of bacteria, and therefore it would be better to employ one that specifically employ a 16S reference database rather the database used by WIMP.

4.2.3 Assessment of potential biases and correlations found by targeted and untargeted approaches

Microbial profiles showed significant diversity differences within each infant mice faecal samples at the phylum, genus, and species level (Mann-Whitney U Test; p-value < 0.05). However, diversity similarities were also found within samples when comparing taxonomical genetic targets, specifically between full-length 16S rRNA gene and the 16S-ITS-23S operon region at the phylum level, and between the full-length 16S rRNA gene and the metagenomic approach at the genus and species level (Mann-Whitney U Test; p-value > 0.05, Figure 10).

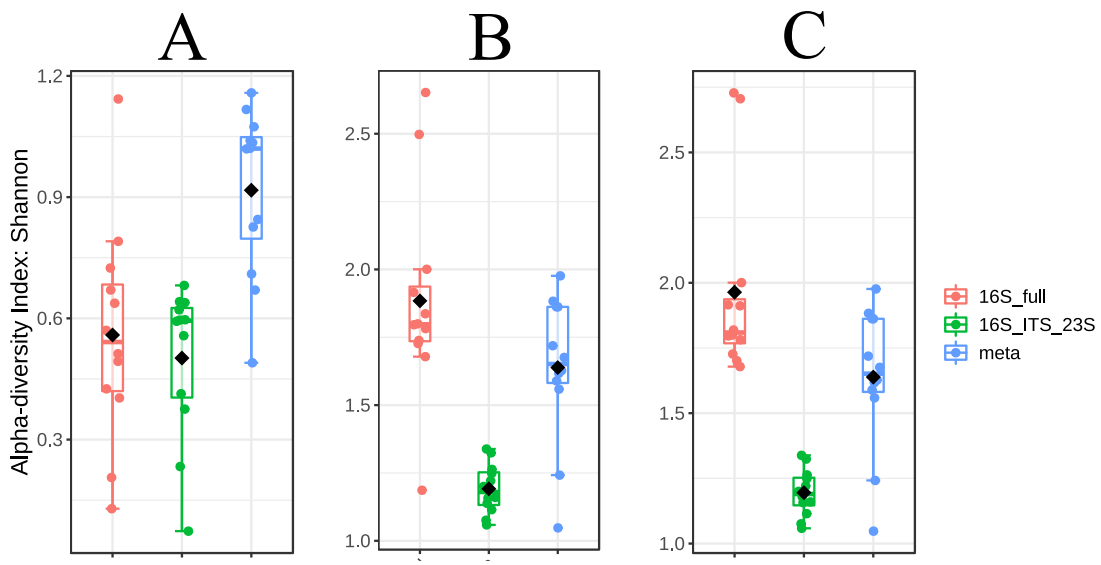


Figure 10 – Boxplots displaying the combined alpha-diversity index at the phylum (A), genus (B), and species level (C), in the 16S rRNA gene-, 16S-ITS-23S region-targeted, and untargeted metagenomic/ shotgun nanopore sequencing approaches when 12 DNA templates extracted from infant mice faecal samples were pooled together.

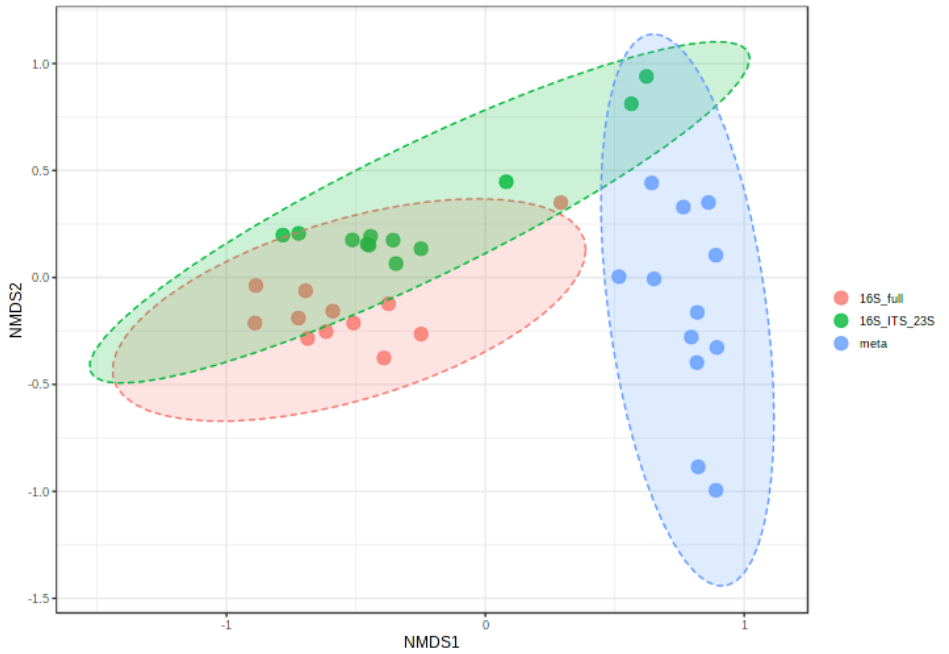
Even though the microbial compositional differences observed between the 16S rRNA gene- targeted approach and the shotgun/metagenomic approach, the diversity found within their respective samples were not significantly different at the genus and species level. Due to the underrepresentation of whole genomes in the databases, and the decreased chance of matching reads with the most represented genomic regions in the reference databases, it was expected that the metagenomic approach was unable to discriminate reads to lower taxonomical ranks than phylum level (Breitwieser, *et al*., 2019). Indeed, at the phylum level, the alpha-diversity index in the metagenomic approach was higher than the 16S rRNA gene- targeted approach, mostly due to the *Actinobacteria,* only detected in the metagenomic approach, and the higher abundance of *Proteobacteria* in comparison with the targeted approaches (Figure 10A). At the genus and species level, the metagenomic approach revealed higher alpha-diversity indices when to compared to the 16S-ITS-23S region-targeted approach (Figure 10B). This observation reveals that the metagenomic approach was more discriminatory and it was able to assign more taxa to the genus level than the mentioned targeted approach. This observation may be explained by the underrepresentation of 16S-ITS-23S operon regions sequences in the databases (Bowers *et al.,*2017, Shaiber and Eren, 2019). Regarding the diversity found within the samples analysed by the 16S-ITS-23S region-targeted approach, the low prevalence of correspondent sequences in the reference databases may undervalue the true diversity present in the infant mice gut. Indeed, targeting the 16S-ITS-23S operon region may miss many phylogenetic groups due to non-linked rRNA genes, and the results presented here did not seem to improve the species level identification in comparison to the nanopore sequencing approach targeting the full-length 16S rRNA gene. This observation is supported by the least number of genera and species identified and, therefore, the smaller alpha- and beta-diversity indices found between the two targeted approaches (Figures 9 and 10). Thus, it is unlikely that the 16S-ITS-23S operon region can increase the taxonomic resolution, allowing strain level identification (Zheng *et al*., 2013, Cusco *et al*., 2018). Specifically, the distance between non-linked 16S and 23S rRNA genes in the genomes present in the datasets can span to 41 kilobases pairs or more, which constitutes a handicap for PCR amplification (Brewer *et al*., 2020). Other attempts to microbial profile environmental samples with longer taxonomical barcodes have been successfully tried with other sequencing platforms, such as PacBio (Martijn *et al*., 2019, Jamy *et al*., 2020) and Illumina synthetic long read sequencing, although limited to ~2000 base pairs (Karst *et al*., 2018), but also with ONT

(Karst *et al*., 2020). Karst and colleagues (2020) recently performed a combined PacBio and nanopore sequencing approach which was optimized to minimize the error-rate of assigned reads. This curated pipeline could improve the strain-level taxonomical discrimination of the 16S-ITS-23S operon region barcode (Karst *et al*., 2020). Although the 16S-ITS-23S region-targeted nanopore sequencing approach here tested included the ITS region, the microbial profile obtained did not show increased species- nor strain-level composition and diversity, in accordance with the report by Karst and colleagues (Karst *et al*., 2018), but contrary to others (de Oliveira Martins *et al*., 2020). Additionally, since some taxa (such *Lactobacillus reuteri* and *B. animalis* subsp. *animalis*) show dissimilarities within ITS copies of their strains and between each ITS copies in the same strain, the hypothesis of a better discriminatory power using this genomic barcode for targeted sequencing approaches is tempting but unlikely (Martijn *et al*., 2019).

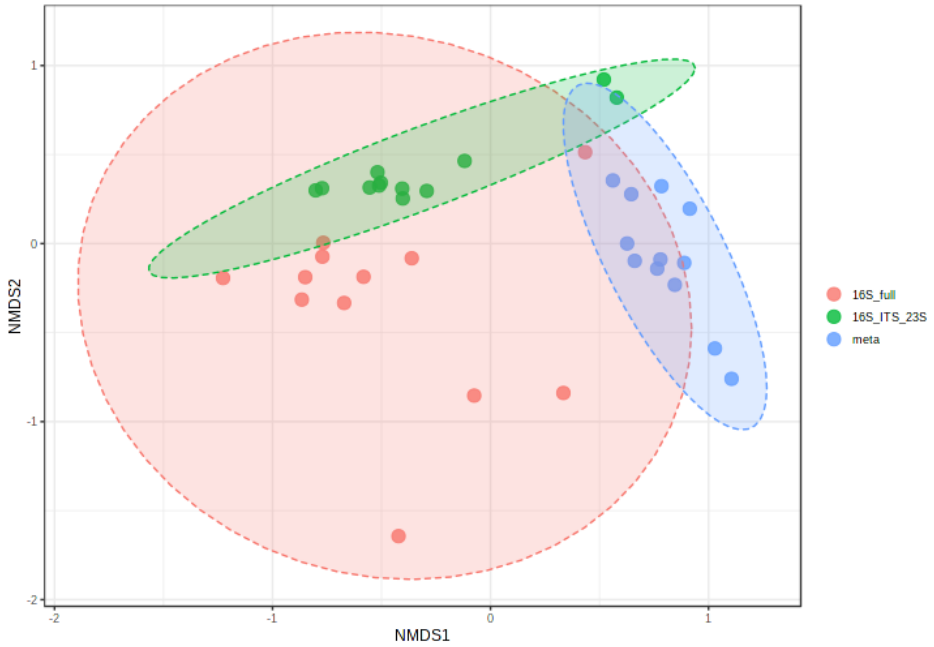
The untargeted and targeted approaches showed significant differences in the beta-diversity indices, representing the different microbial profiles obtained depending if the approach is targeted or untargeted. In the targeted approaches, nanopore sequencing was able to differentiate the diversity between samples at the genus and species level, however the level of dissimilarity is low (PERMANOVA R^2 = 0.16-0.17; ANOSIM R = 0.25-0.27; p-value < 0.05). The dissimilarities in the beta-diversities are greater between the targeted and untargeted approaches (PERMANOVA R^2 = 0.45-0.70; ANOSIM R = 0.69-0.81; p-value < 0.05) (Figure 11). Therefore, the nanopore sequencing approaches here tested retrieved different microbial diversities between each infant mice faecal sample.

However, there is no difference between groups' dispersions in any of the approaches (PERMDISP F = 0.1-2.6; p-value > 0.05), and all groups showed significantly differences between themselves when analysed together (PERMANOVA R^2 = 0.47-0.59; p-value < 0.05). Thus, these observations represent evidence of potential nanopore sequencing biases towards the experimental approach chosen for target nanopore sequencing.

A


NMDS2
NMDS1
16S_full
16S_ITS_23S
meta

B


NMDS2
NMDS1
16S_full
16S_ITS_23S
meta

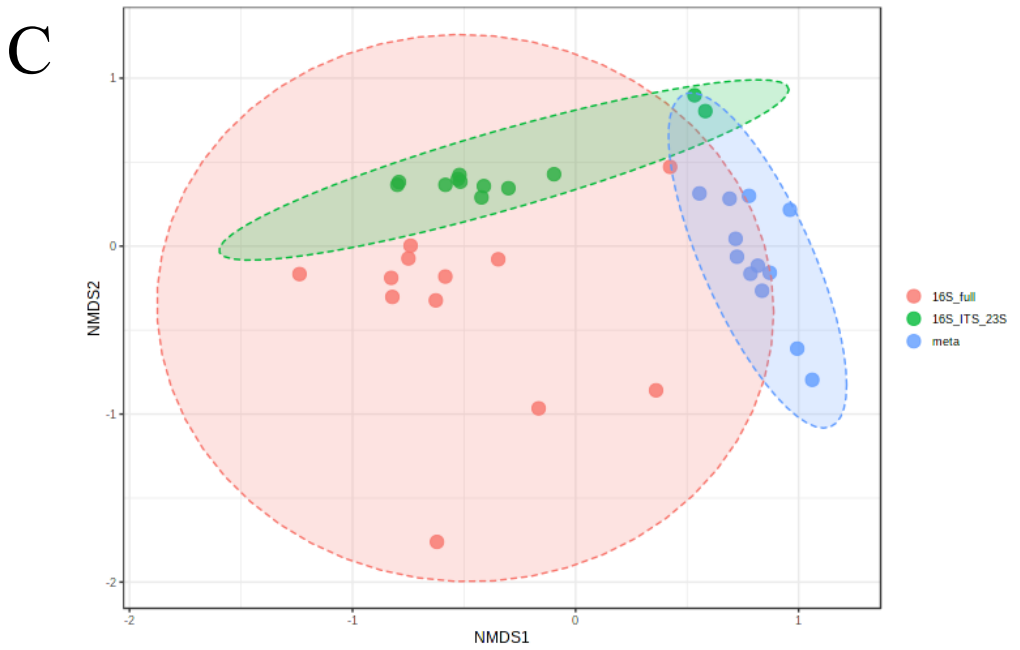
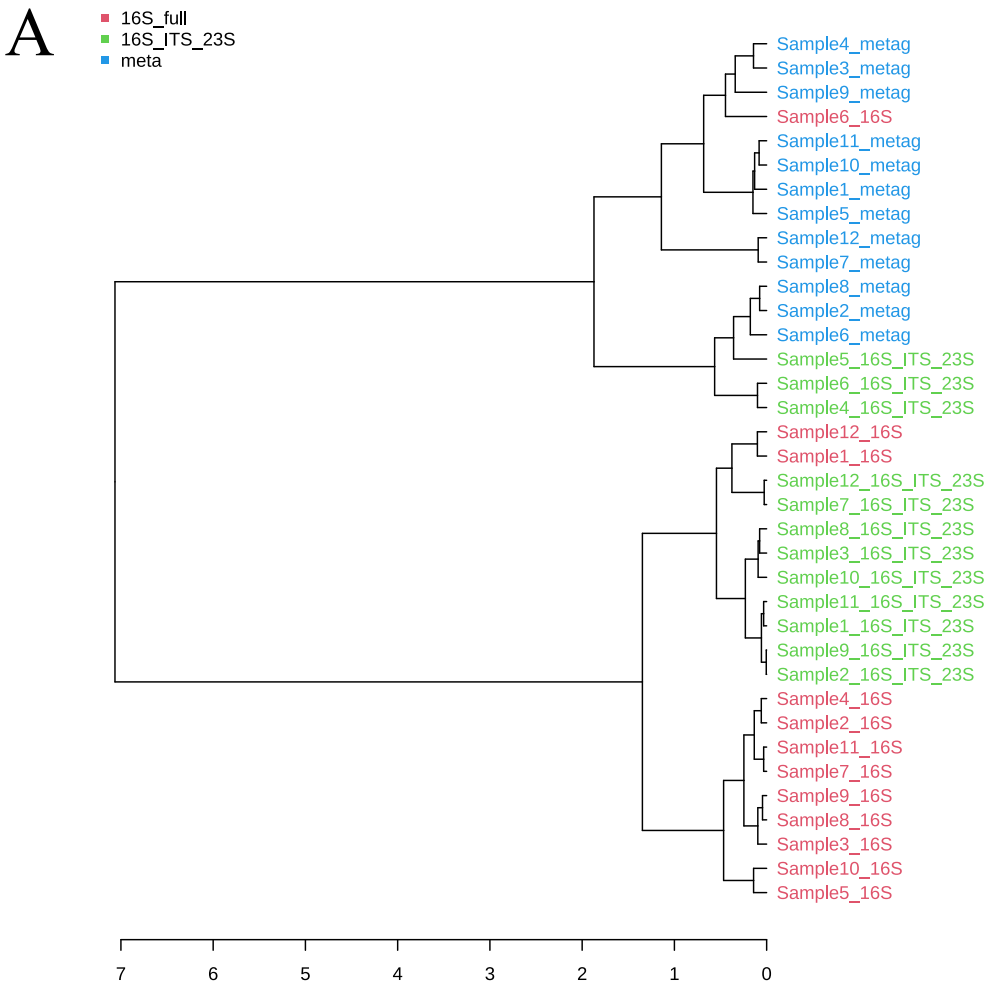
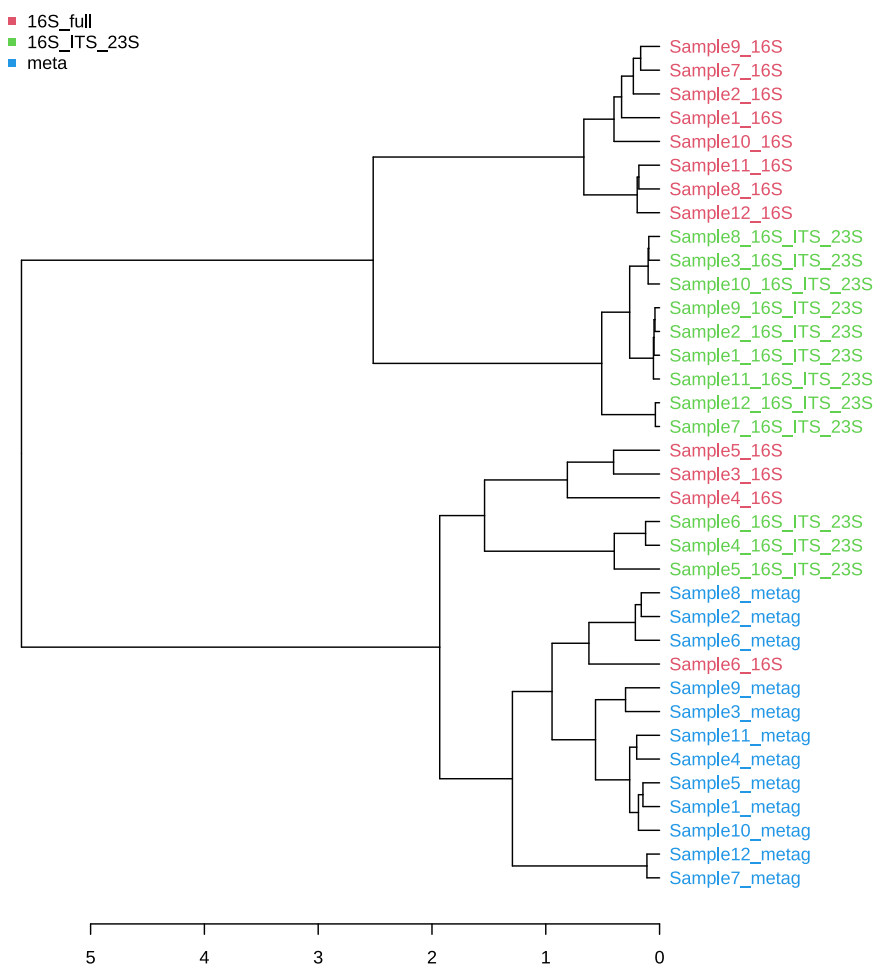


Figure 11 – Two-dimensional ordination NMDS graph displaying beta-diversity index compared between the 16S rRNA gene-, 16S-ITS-23S region-targeted, and untargeted metagenomic nanopore sequencing approaches analysed using PERMANOVA stats using the Bray-Curtis dissimilarity index (and corroborated with Jaccard index). A, phylum level; B, genus level; C, species level.

Multivariate statistical methods can be employed to analyse complex microbiome communities taking in consideration the environmental context (Xia *et al*., 2017). Such statistical methods test the strength and statistical significance of sample grouping based on distance matrices (Chong *et al*., 2020) (see Appendix III). The microbial profiles obtained by the taxonomical genomic targets seemed to cluster well, showing that choosing the genomic target for nanopore sequencing can potentially influence the microbial profiles detected at different taxonomical ranks (Figure 12). The clustering observed indicates that the genomic target used for nanopore sequencing can reveal biases in terms of the taxonomical assignments retrieved by the classification algorithms. Samples were grouped in clades representing the genomic targets used for sequencing. These three clades were observed at the phylum, genus, and species level (Figure 12). This observation constitutes reinforced evidence supporting the potential sequencing biases associated towards the genomic target selected for nanopore sequencing (Figure 12).


A
16S_full
16S_ITS_23S
meta
Sample4_metag
Sample3_metag
Sample9_metag
Sample6_16S
Sample11_metag
Sample10_metag
Sample1_metag
Sample5_metag
Sample12_metag
Sample7_metag
Sample8_metag
Sample2_metag
Sample6_metag
Sample5_16S_ITS_23S
Sample6_16S_ITS_23S
Sample4_16S_ITS_23S
Sample12_16S
Sample1_16S
Sample12_16S_ITS_23S
Sample7_16S_ITS_23S
Sample8_16S_ITS_23S
Sample3_16S_ITS_23S
Sample10_16S_ITS_23S
Sample11_16S_ITS_23S
Sample1_16S_ITS_23S
Sample9_16S_ITS_23S
Sample2_16S_ITS_23S
Sample4_16S
Sample2_16S
Sample11_16S
Sample7_16S
Sample9_16S
Sample8_16S
Sample3_16S
Sample10_16S
Sample5_16S
7
6
5
4
3
2
1
0

B


16S_full
16S_ITS_23S
meta
Sample9_16S
Sample7_16S
Sample2_16S
Sample1_16S
Sample10_16S
Sample11_16S
Sample8_16S
Sample12_16S
Sample8_16S_ITS_23S
Sample3_16S_ITS_23S
Sample10_16S_ITS_23S
Sample9_16S_ITS_23S
Sample2_16S_ITS_23S
Sample1_16S_ITS_23S
Sample11_16S_ITS_23S
Sample12_16S_ITS_23S
Sample7_16S_ITS_23S
Sample5_16S
Sample3_16S
Sample4_16S
Sample6_16S_ITS_23S
Sample4_16S_ITS_23S
Sample5_16S_ITS_23S
Sample8_metag
Sample2_metag
Sample6_metag
Sample6_16S
Sample9_metag
Sample3_metag
Sample11_metag
Sample4_metag
Sample5_metag
Sample1_metag
Sample10_metag
Sample12_metag
Sample7_metag
5
4
3
2
1
0

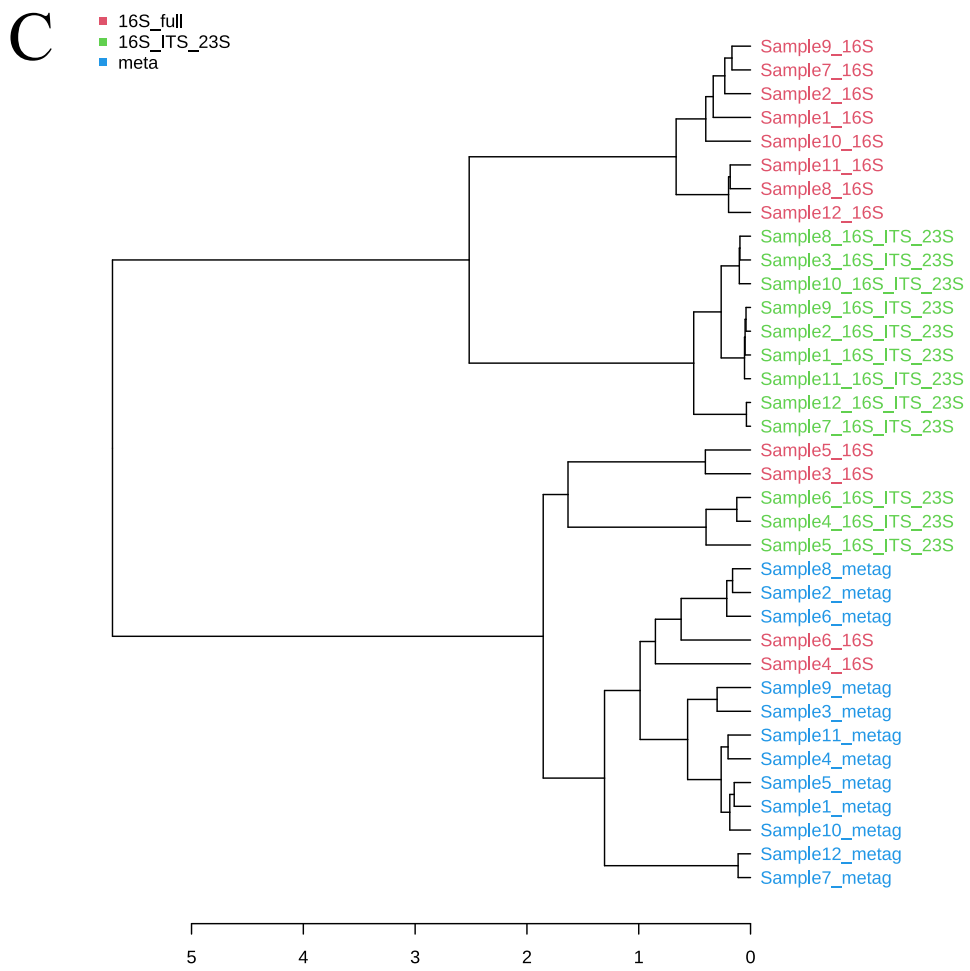


Figure 12 – Dendrogram analyses performed using Bray-Curtis index distance measure and Ward clustering algorithm on the16S rRNA gene-, 16S-ITS-23S region-targeted, and untargeted metagenomic/ shotgun nanopore sequencing approaches analysed. A, phylum level; B, genus level; C, species level.

The different nanopore approaches detected a different microbial profile, and beta-diversity index was significantly different between them. To assess sequencing approach-based biases, correlations between assigned taxa and the16S rRNA gene-, 16S-ITS-23S region-targeted, and untargeted shotgun/metagenomic nanopore sequencing approaches were analysed (Table 11). *Limosillactobacillus* gen., and *L. reuteri*, and *Blautia argi* were correlated with the full-length 16S rRNA gene marker, and *Bacteroidetes* phylum, genera *Duncaniella*, *Eubacterium*, *Muribaculum*, *Faecalibaculum*, *D. muris*, and *Clostridium inocuum*, *Eubacterium* sp. NSJ 61, *M. intestinale*, and *Faecalibaculum rodentium,* correlated with the nanopore sequencing approach based on the 16S-ITS-23S operon region genomic target. Therefore, the potential nanopore sequencing biases may be related to the genomic target used for the

amplicon generation and be significantly observed at phylum, genera, and species level alike (Table 11). Since nanopore sequencing is being increasingly explored for microbial profiling complex ecosystems, such as clinical and environmental samples, caution should be taken when targeted approaches are performed. The genomic target chosen may impose biases in the microbial composition consequently confounding the conclusions obtained by the experiments. Nanopore sequencing is known to confer biases based on electrical signal measurements obtained from the passage of DNA or RNA molecules through the nanopores (Delahaye and Nicolas, 2021). In a recent report, Delahaey and Nicolas (2021) described in detail the scope of nanopore sequencing biases. The authors analysed a set of human and bacterial datasets and evaluated the overall performance of nanopore sequencing. They concluded that updated basecallers, such as *guppy*, are continuously improving the error-rate; error related to base transitions are more common than transversions; and low-GC species and reads have improvement quality scores in general (Delahaye and Nicolas, 2021). Moreover, the authors found that nearly half of sequencing errors are due to homopolymers and the sequence of homopolymers is dependent on length and GC content (Delahaye and Nicolas, 2021). However, Browne and colleagues (2020) performed next-generation and third-generation sequencing approaches to assess biases and found that nanopore sequencing was not significantly affected by GC bias. The authors only studied two taxa, *Fusobacterium* (GC content: 28.9%), and *Aminobacter* (GC content: 63.0%) (Browne *et al*., 2020), which could skew their observations for related taxa, not representing a reasonable sample representativeness for supporting their claims. However, nanopore sequencing poses several challenges to microbial profiling such as high error-rate, different classification assignments depending on the classification algorithm used, and genomic target-dependent biases.

Table 11 – Taxa showing significant Spearman correlations (rho > 0.5) with the16S rRNA gene-, 16S-ITS-23S region-targeted, and untargeted shotgun/metagenomic nanopore sequencing approaches' pairs, at the phylum, genus, and species level.

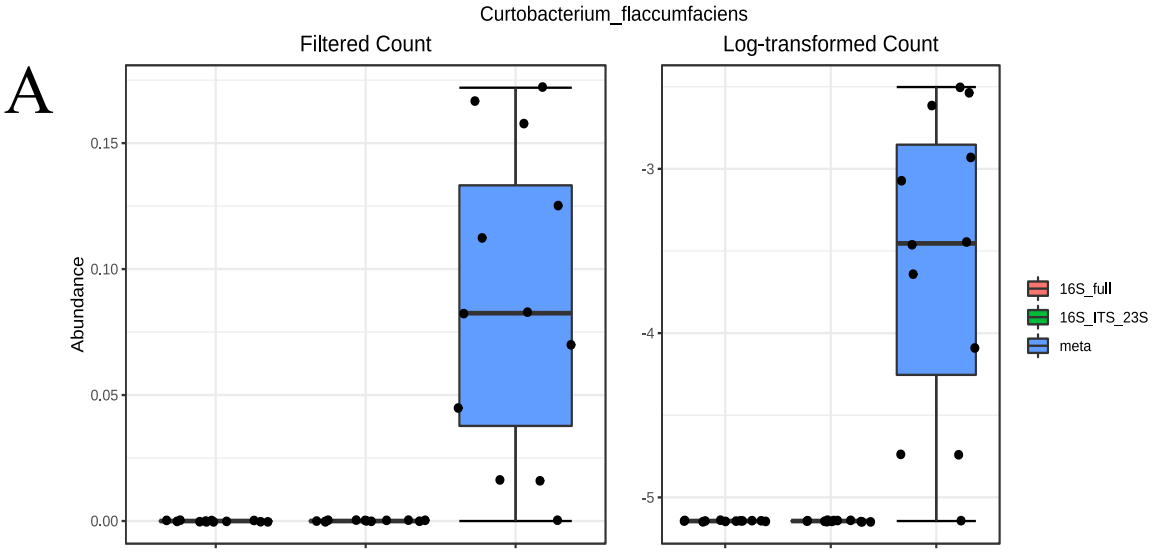
Comparison of interest	Taxa correlated with[1]								
	Full-length 16S rRNA gene			16S-ITS-23S region			Metagenomics		
	Phylum	**Genus**	**Species**	**Phylum**	**Genus**	**Species**	**Phylum**	**Genus**	**Species**
16S rRNA gene *vs* 16S-ITS-23S region	NS	*Limosilactobacillus*,	*Limosilactobacillus reuteri* *Blautia argi*	*Bacteroidetes*	*Duncaniella* *Eubacterium* *Muribaculum* *Faecalibaculum*	*Duncaniella* sp. B8 *Clostridium innocuum* *Eubacterium* sp. NSJ 61 *Muribaculum intestinale* *Faecalibaculum rodentium*	NP	NP	NP
16S rRNA gene *vs* metagenomic	*Firmicutes*	*Faecalibaculum* *Peptinophilus* *Limosilactobacillus* *Lactobacillus* *Amedibacterium* *Erysipelothrix* *Clostridium*	*Faecalibaculum rodentium* *Peptinophilus harei* *Limosilactobacillus reuteri* *Lactobacillus johnsonii* *Amedibacterium intestinale* *Erysipelothrix larvae* *Blautia argi* *Clostridium perfringens*	NP	NP	NP	*Actinobacteria* *Proteobacteria* *Bacteroidetes*	*Curtobacterium* *Parabacteroides* *Bacteroides* *Bifidobacterium* *Helicobacter* *Muribaculum*	*Curtobacterium flaccumfaciens* *Parabacteroides distasonis* *Bacteroides caecimuris* *Bifidobacterium animalis* *Helicobacter typhlonius* *Muribaculum intestinale*
16S-ITS-23S region *vs* metagenomic	NP	NP	NP	*Firmicutes*	*Faecalibaculum* *Peptinophilus* *Eubacterium* *Duncaniella*	*Faecalibaculum rodentium* *Peptinophilus harei* *Clostridium innocuum* *Eubacterium* sp. NSJ 61 *Duncaniella* sp. B8	*Actinobacteria* *Proteobacteria*	*Curtobacterium* *Helicobacter* *Bacteroides* *Parabacteroides* *Bifidobacterium*	*Curtobacterium flaccumfaciens* *Helicobacter typhlonius* *Bacteroides caecimuris* *Parabacteroides distasonis* *Bifidobacterium animalis* *Lachnoclostridium* sp. YL32

[1] – *rho* > 0.5; p – value < 0.05; FDR < 0.05

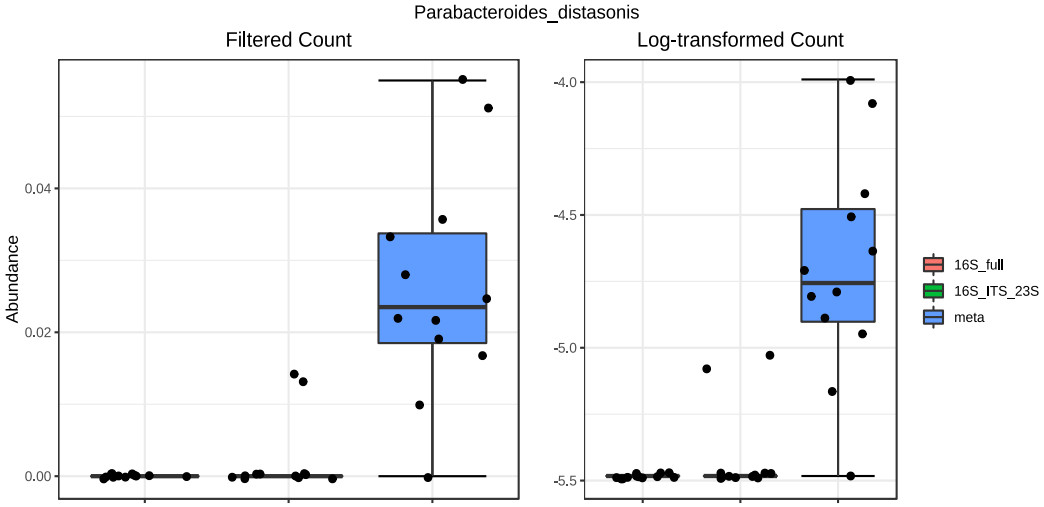
NP – not performed in this row

NS – p-value > 0.05, not significant

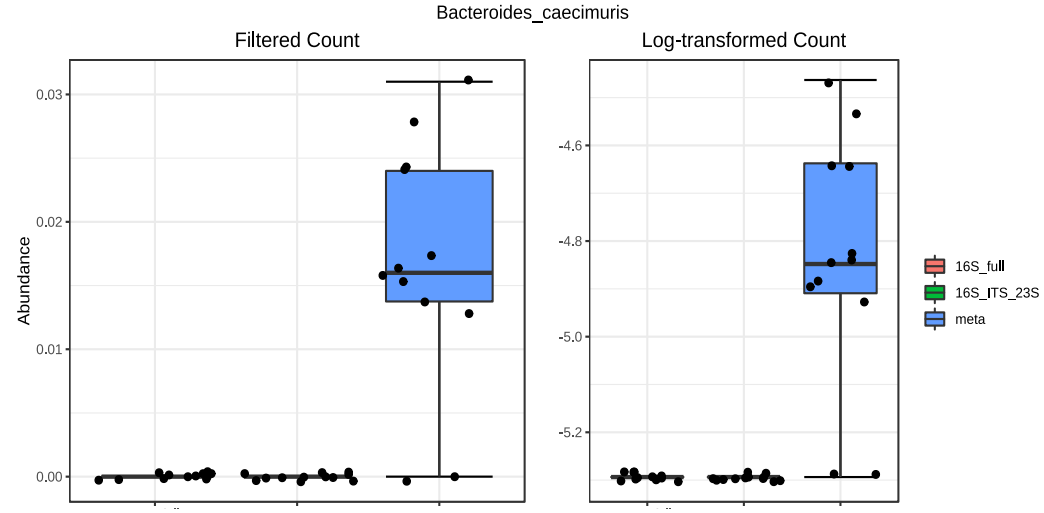
When 16S rRNA gene-, 16S-ITS-23S region-targeted, and untargeted shotgun/metagenomic nanopore sequencing approaches were analysed together, six taxa were significantly correlated exclusively with one sequencing approach (Figure 13). *C. flaccumfaciens*, *Parabacteroides distasonis*, and *B. caecimuris* were significantly correlated with the untargeted shotgun/metagenomic nanopore sequencing approach (Figure 13A, 13B, 13C, and 13D), while *Peptinophilus harei* and *Limosilactobacillus reuteri* were significantly correlated with the full-length 16S rRNA gene nanopore sequencing approach (Figure 13E and 13F). Therefore, the genomic target used for microbial profiling can induce biases at the compositional/population level, seen in the diversity indices, and at the individual level when nanopore sequencing approaches are attempted. Consequently, the correlations found between these taxa and specific genomic targets lead to the conclusion that *C. flaccumfaciens*, *P. distasonis*, and *B. caecimuris*, can be considered "bias markers" for shotgun/metagenomic nanopore sequencing approaches, and *P. harei* and *L. reuteri* for the 16S rRNA gene-targeted nanopore sequencing approaches. Nevertheless, further investigation must be performed to elucidate the true meaning of the presence of these taxa and in which cases they may represent taxa erroneously assigned.



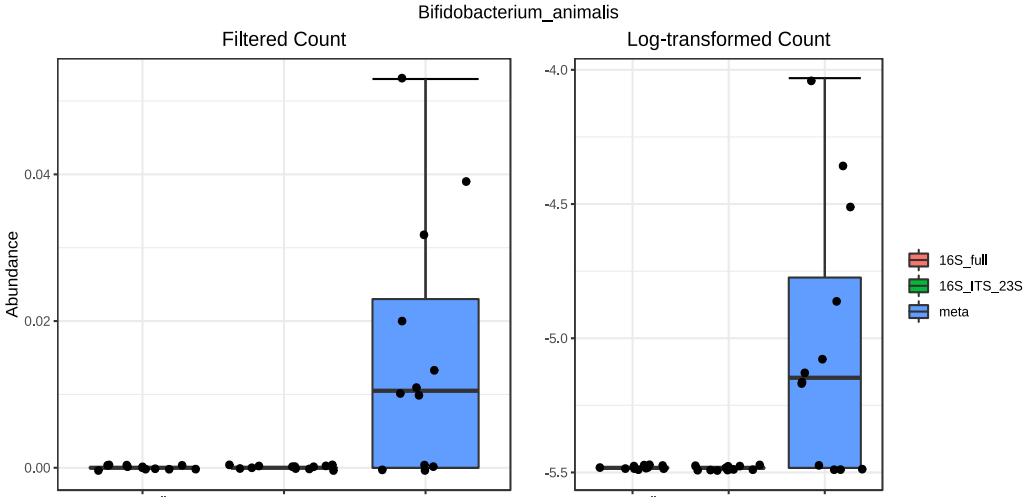
B


Parabacteroides_distasonis
Filtered Count
Log-transformed Count
Abundance
0.04
0.02
0.00
-4.0
-4.5
-5.0
-5.5
16S_full
16S_ITS_23S
meta

C


Bacteroides_caecimuris
Filtered Count
Log-transformed Count
Abundance
0.03
0.02
0.01
0.00
-4.6
-4.8
-5.0
-5.2
16S_full
16S_ITS_23S
meta

D


Bifidobacterium_animalis
Filtered Count
Log-transformed Count
Abundance
0.04
0.02
0.00
-4.0
-4.5
-5.0
-5.5
16S_full
16S_ITS_23S
meta

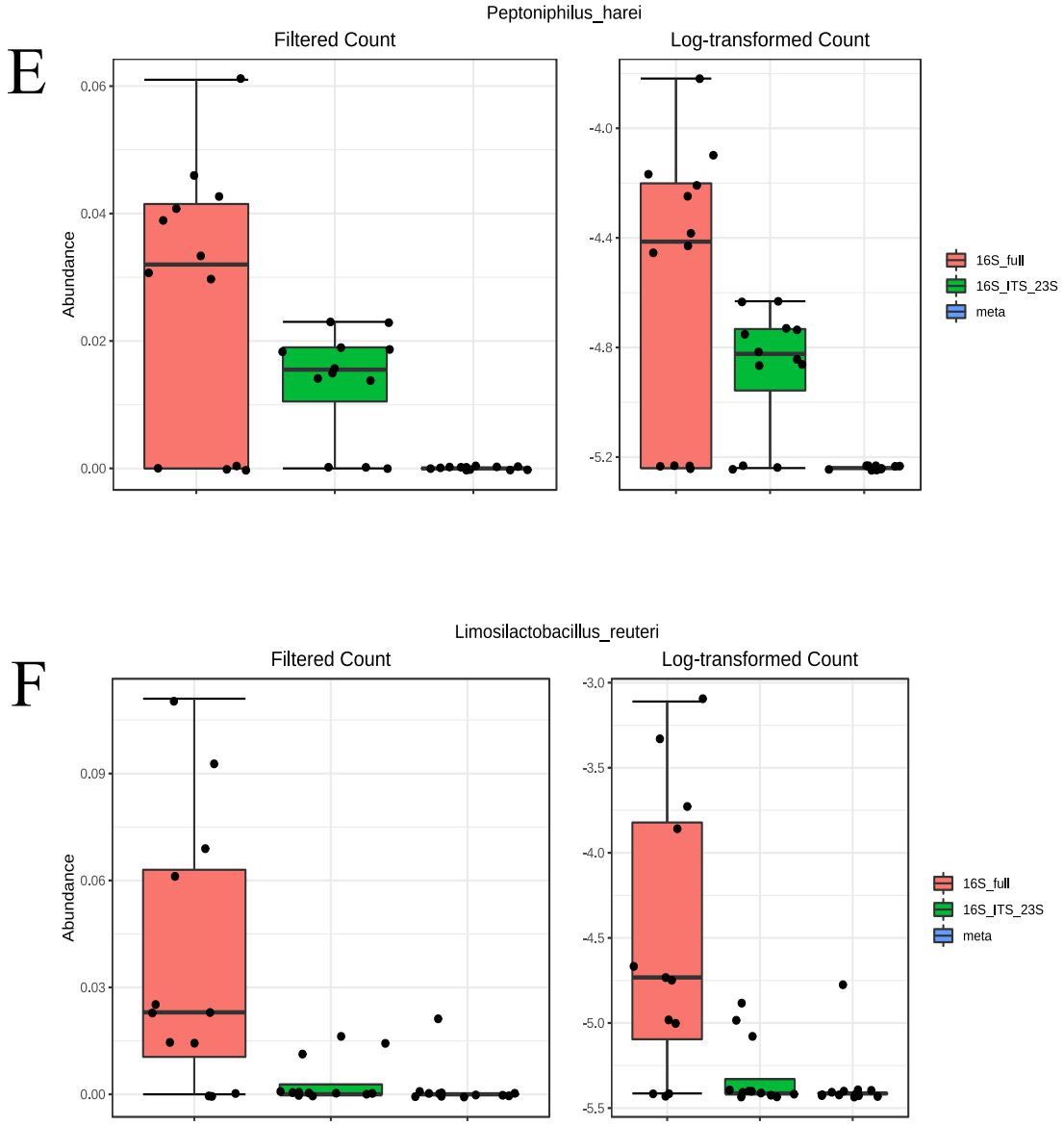


Figure 13 – Boxplots displaying the relative abundances of taxa significantly correlated with one of the nanopore sequencing approaches tested. A, B, C, and D show the taxa correlated with the metagenomic/ shotgun nanopore sequencing approach; and E and F show the taxa correlated with the 16S rRNA gene-targeted nanopore sequencing approach. Left: raw relative abundances; Right: log-transformed relative abundances.

In accordance with the observations in the targeted approaches, there was also observed a clustering pattern between samples at the phylum, genus, and species level (Figure 12), thus evidencing dependent biases in the microbial profile during the untargeted approach as well. The biases towards the metagenomic approach were supported by the correlations found at the phylum, genera, and species level in comparison with the other two targeted approaches (Table 11). Phyla *Actinobacteria* and *Proteobacteria* were correlated with the metagenomic approach, and *Bacteroidetes* was also correlated with the metagenomic approach when compared with the full-length 16S

rRNA gene-targeted approach but not with the 16S-ITS-23S region-targeted approach (Table 11). At the species level, *C. flaccumfaciens*, *P. distasonis*, *B. caecimuris*, *B. animalis*, and *Helicobacter typhlonius* were correlated with the metagenomic approach when compared with both targeted approaches (Table 11). Additionally, *M. intestinale* and *Lachnoclostridium* sp. were also correlated with the metagenomic approaches when compared with the full-length 16S rRNA gene approach and with the 16S-ITS-23S operon region approach, respectively (Table 11). Although statistically significant, these correlations may be related to the previously discussed biased genomic regions' representativeness in the reference databases, as well by the non-linked 16S-ITS-23S rRNA gene presence in some species that may skew the read assignments in the sequencing approach that target that region (Brewer et al 2020).

In conclusion, different gut microbiota profiles were obtained from the same samples when using different nanopore sequencing approaches. Both targeted approaches, based on the 16S rRNA gene and the 16S-ITS-23 region of the *rrn* operon, and untargeted approaches yielded preferences towards specific taxa thus evidencing sequencing biases. In particular, *C. flaccumfaciens*, *P. distasonis*, *B. caecimuris*, and *B. animalis*, were correlated with the untargeted shotgun/metagenomic approach, and *P. harei*, and *L. reuteri* were correlated with the 16S rRNA gene-targeted approach. These results show that either upstream experimental steps, downstream bioinformatics analyses, or both, can introduce biases that are confounding the true microbial composition and abundances obtained from infant mice gut samples.

Chapter 5 – Nanopore and Illumina sequencing approaches to profile the gut microbiota of mice infants exposed to ethanol *in utero*

5.1 Brief introduction

The gut microbiome plays a vital role in host homeostasis and understanding of its biology is essential for a better comprehension of the etiology of disorders such as foetal alcohol spectrum disorder (FASD). FASD represent a cluster of abnormalities including growth deficiencies and neurological impairments, which are not easily diagnosed nor treated (Riley *et al*., 2011, Petrelli *et al*., 2018). Recent advances in sequencing technologies have emphasized questions about the microbial profiles obtained when different approaches are tested. Such questions are mainly due to the genomic target used for sequencing and differences in the bioinformatic pipeline used (Breitwieser *et al*., 2019, Brewer *et al*., 2020, Markey *et al*., 2020, Santos *et al*., 2020, Matsui *et al*., 2021). Here the effect of ethanol exposure *in utero* on the gut microbial profiles of 16 infant mice (nine exposed *in utero* and seven non-exposed) was assessed by targeted nanopore sequencing and Illumina sequencing approaches. The nanopore sequencing was implemented using MinION system targeting PCR-amplified amplicons made from the full-length 16S rRNA gene.

Despite ONT possess a sequencing kit tailored for the 16S gene (*eg*. SQK-RAB-204 with the 27F and 1249R primer set), all sequencing experiments were performed with the Ligation Sequencing kit (LSK109) because the approach targeting the 16S-ITS-23S and the untargeted approach couldn't be analysed with the 16S-based sequencing kit. Additionally, as also observed in the last Chapter, there is a chance that the bioinformatic pipeline here employed for the 16S-targeted approach was not suitable for the reliable detection of bacteria, and therefore it would be better to employ one that specifically employ a 16S reference database rather the database used by WIMP (Kai *et al*., 2019, Fujiyoshi *et al*., 2020).

The Illumina sequencing was performed using Miseq system targeting the V3-V4 region of the 16S rRNA gene.

5.2 Results and discussion

5.2.1 Nanopore sequencing

In the previous chapters nanopore sequencing approaches were optimized and here the full-length 16S rRNA gene-targeted nanopore sequencing approach was tested against 16 mice gut faecal samples (nevertheless, we couldn't test the differences between

the Ligation Kit and the 16S sequencing kit due to lack of time). Nine of those samples were collected from infant mice exposed to ethanol *in utero* and the remaining seven were collected from non-exposed infant mice. Since the full-length 16S rRNA gene-targeted sequencing approach gave the most consistent results (see Chapter 4), it was chosen as the bacterial genomic target for the nanopore sequencing approach performed in this part of the work. Additionally, based on the nanopore sequencing approach optimization performed in Chapter 3, the profiling of the gut mycobiome in the same samples was also attempted.

5.2.1.1 Profiling the gut mycobiome

PCR was performed to generate fungal ITS amplicons from the infant mice samples. However, no amplification was detected by gel electrophoresis from any sample (Figure 14). This observation suggests that fungal members were absent or in very low abundance in the infant mice gut.

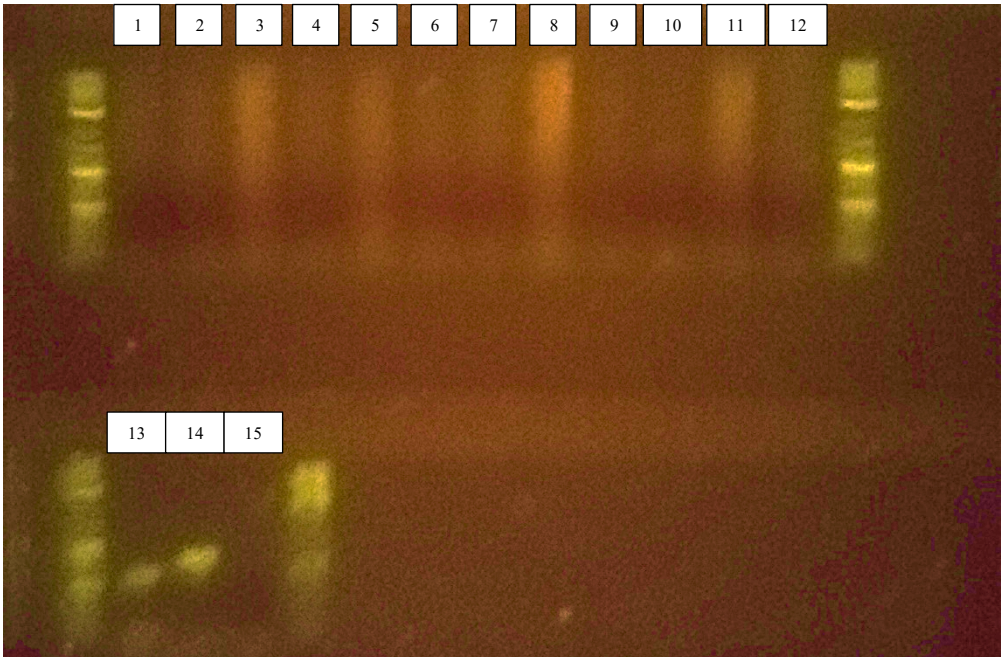


Figure 14 – Gel electrophoresis image showing the amplification ITS band profile of the 12 infant mice DNA extracts. Each line (1-15) represents one sample. Numbers 1-12 represent the amplification bands from the infant mice DNA templates. Number 13 represents the amplification band from DNA templates of the first positive control, *Meyerozyma guilliermondii*. Number 14 represents amplification band from DNA templates of the second positive control, *Candida glabrata*. Number 15 represents the negative control. At the edges is observed the 1 kb Plus DNA Ladder (New England Biolabs, UK).

In a recent study, Aykut and colleagues (2019) analysed the fungal microbiota of a mouse model of pancreatic ductal adenocarcinoma. The authors tested 14 experimental mouse gut and 12 control mice gut samples through Illumina sequencing (Aykut *et al*., 2019). Gut mycobiomes of 30-week-old disease model mice were analysed by 18S ITS targeted sequencing focusing on the ITS1 region of the 18S gene from the genomic DNA using the primer set ITS1F and ITS2 (Walters *et al*., 2016). The authors found significant differences in the relative abundances of several taxa, such as genera *Saccharomyces*, *Phallus*, *Thermomyces*, *Malassezia*, *Ascochyta*, and *Aspergillus* (Aykut *et al*., 2019).

Therefore, the absent fungal PCR amplification here reported is unexpected. Although a different primer set was used in other reports, it should not explain the lack of PCR amplification found. One explanation might be the suboptimal PCR composition and cycle conditions which were optimised for a set of fungal species that may not be found in the mouse gut samples. In Chapter 3, a nanopore sequencing approach was optimized to analyse fungi communities. Three DNA extraction protocols were tested to maximize the DNA quality and quantity obtained, several PCR cycle conditions were tried, and two PCR products purification protocols were compared (data not shown). Indeed, these optimizations attempts were not enough, and more work must be done in order to surpass technical and computational limitations (*i.e.*, reducing potential biases) when analysing mouse gut mycobiomes (Mims *et al*., 2021, Neff, 2021). Moreover, these optimizations were performed in pure fungal cultures. When mice guts were analysed, total genomic DNA were extracted directly from stool samples, which differ greatly from the DNA extraction conditions when the mock community was analysed, for instance regarding the host material that might contaminate and mask the gut microbial content. Thus, all downstream protocols needed to be re-optimized, including tests for different primer sets, as performed by Aykut and colleagues (2019), with the suboptimal fungal amplicons' generation implying a major limitation of this study. However, the hypothesis of the fully absent mice gut mycobiome cannot be discarded, due to the highly transient feature of the gut fungi members in the mouse model under investigation in this dissertation (Fiers *et al*., 2021). This later hypothesis is also supported by the shotgun/metagenomic nanopore sequencing approach, which did not retrieve any fungal assignments in the taxonomical analysis (see Chapter 4). In this work, infant mice were kept under a controlled environment, which perhaps did not provide the opportunity for development of a gut mycobiome. We infer that fungi were likely absent from the samples, or at least below the detection levels of both the ITS-targeted PCR and of the metagenomic sequencing assay.

5.2.1.2 Quality assessment of nanopore sequencing targeted approaches

Following the apparent absence of a gut mycobiome in our sampled infant mice, this chapter focuses on the bacterial gut microbiota of the same infant mice. The bacterial gut microbiome was thus assessed using the full-length 16S rRNA gene-targeted nanopore sequencing approach described in Chapter 4. In order to gather the highest number of sequenced reads possible, four independent sequencing runs were performed (Table 12). The minimum basecall mean Qscore obtained during the four nanopore sequencing runs was 10.1, which corresponds to an error-rate of ~10%, and the maximum was 10.5, corresponding to a mean error-rate of ~9%. During the four nanopore sequencing runs, more than 50% of the reads showed less than 10% error rate (Q10), and between 3-9% of reads showed less than 6.3% error rate (Q12). Despite differences in the number of reads and bases sequenced between the four runs, the average read length and average basecall quality scores were similar, indicating a good sequencing precision as previously observed and discussed. In Table 13 the quality parameters are shown for each barcoded sample within the respective nanopore sequencing run. It can be observed that the number of reads differed between the sequencing runs though similar sequencing performance parameters were obtained in terms of basecall quality, sequenced read lengths, and N50 values.

Table 12 – Nanopore sequencing performance parameters of 4 independent sequencing runs of infant mice faecal samples exposed to ethanol in utero and control samples.

Barcoded samples ID[1]	**No. reads**	**No. bases (bp)**	**Mean read length (bp)**	**Read length N50 (bp)**	**Mean read quality (Qscore)**	**Percentage of reads above quality cutoff (%)**	
						Q10	**Q12**
Run A	42,073	> 67 million	1,606	1,616	10.3	63.4	3.7
Run B	625,975	> 1,0 billion	1,615	1,621	10.4	62.2	9.7
Run C	557,179	> 899 million	1,614	1,623	10.1	52.7	3.9
Run D	989,109	> 1,5 billion	1,603	1,614	10.5	67.9	9.6

[1] – In Run A, two control samples were sequenced; In Run B, three control samples were sequenced; In Run C, four test and two control samples (one of which was a repetated sample) were sequenced; In Run D, five test and 1 control samples were sequenced.

Table 13 – Nanopore sequencing performance parameters of all experimental samples.

Sample ID[1]	No. reads	No. bases (Mbases)	Mean read length (bp)	Mean read quality (Qscore)
Control 9	14,036	24.3	1,589	10.86
Control 11	277,196	458.9	1,587	11.02
Control 13a[2]	100,310	167.0	1,613	10.67
Control 13b	117,108	195.3	1,614	10.65
Control 14	7,597	12.5	1,597	11.08
Control 16	317,219	523.8	1,605	11.05
Control 17	21,839	37.5	1,584	10.93
Control 18	95,029	156.2	1,569	11.18
Test 1	68,573	113.5	1,583	10.70
Test 2	95,347	159.6	1,606	10.64
Test 3	57,064	94.02	1,593	10.66
Test 4	98,644	162.2	1,576	10.63
Test 8	165,896	272.3	1,559	11.15
Test 9	202,391	337.1	1,600	11.15
Test 11	123,931	207.9	1,624	11.20
Test 13	125,991	208.9	1,533	11.13
Test 14	185,468	304.2	1,564	11.17

[1] – 'Test' corresponds to samples gathered from infant mice exposed to ethanol *in utero*, and 'Control' correspond to samples gathered from infant mice not exposed

[2] – 'a' and 'b' correspond to the same sample sequenced twice by being attributed a different sequencing barcode

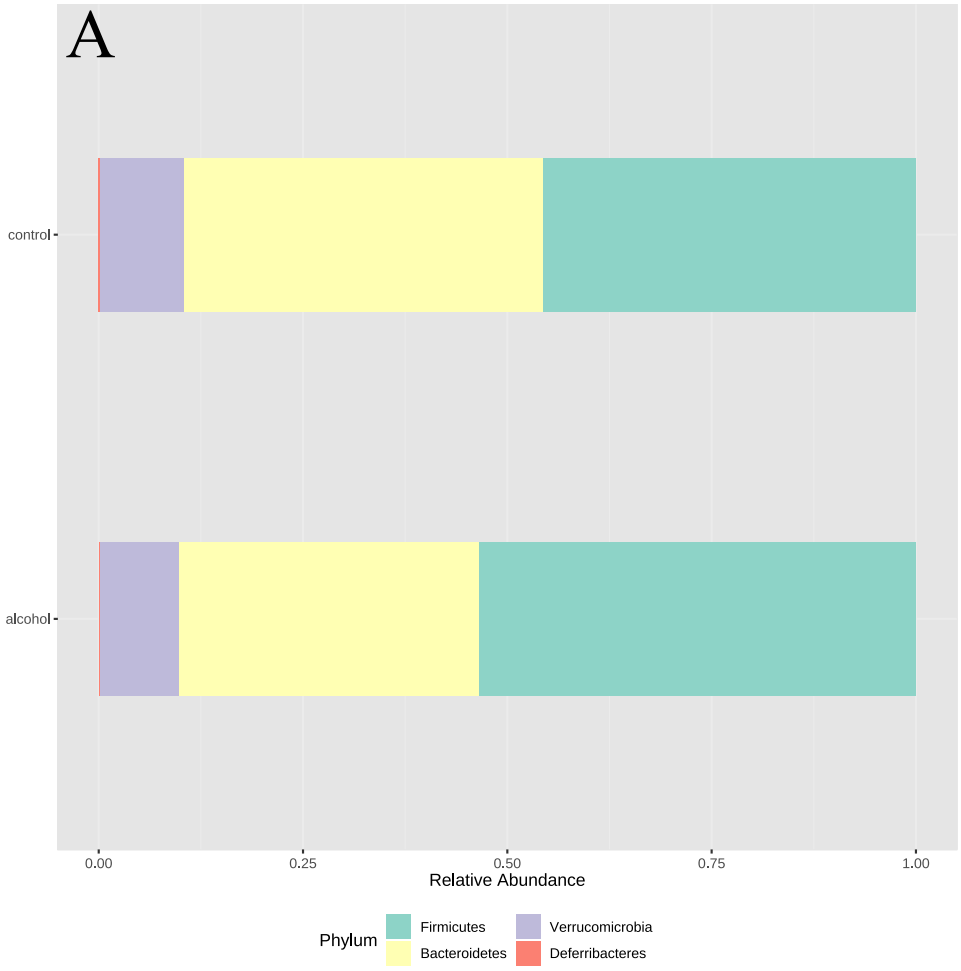
5.2.1.3 Microbial profiles of infant mice gut samples

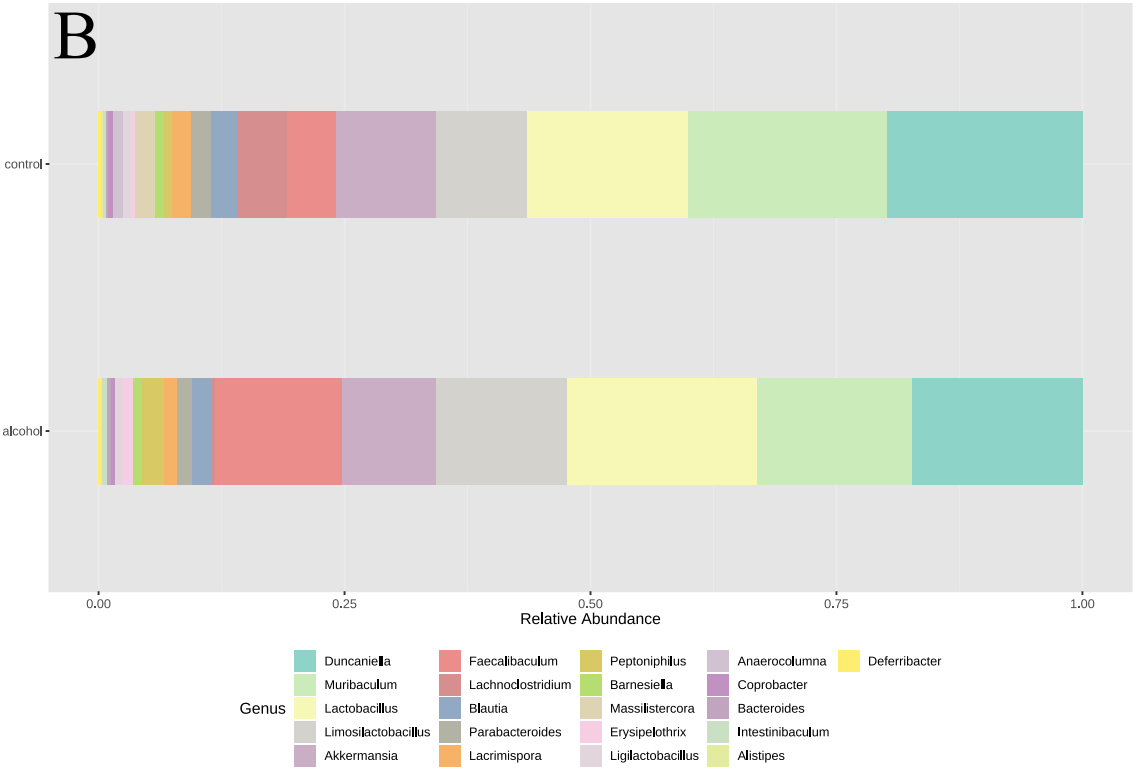
In the ethanol-exposed group, *Firmicutes* was the most represented phylum (53%), followed by *Bacteroidetes* (37%), *Verrucomicrobia* (10%), and *Deferribacteres* which did not exceed 1% of the assigned reads. At the genus level, the most prevalent genera were *Lactobacillus* (19%), followed by *Duncaniella* (17%), *Muribaculum* (16%), *Limosilactobacillus* (13%), *Faecalibaculum* (13%), and *Akkermansia* (10%) with the remaining genera had less than 3% of relative abundance. In the ethanol-exposed group, a total of 21 species were assigned. The most represented species were *Duncaniella muris* (17%), *Lactobacillus johnsonii* (17%), *Muribaculum intestinale* (16%). *Limosilactobacillus reuteri* (13%), *Faecalibaculum rodentium* (13%), and *Akkermansia muciniphila* (10%) with the remaining 14 species had less than 3% of percentual relative abundance.

In the control group, *Firmicutes* was the most represented phylum (46%), followed by *Bacteroidetes* (44%), *Verrucomicrobia* (10%), and *Deferribacteres* that accounted for less than 1% of the reads. At the genus level, the most represented genera were *Muribaculum* (20%), followed by *Duncaniella* (20%), *Lactobacillus* (16%), *Akkermansia* (10%), *Limosilactobacillus* (9%), *Faecalibaculum* (5%), and *Lachnoclostridium* (5%). The remaining genera had less than 3% of percentual relative abundance. In the control group, a total of 27 species was observed. The most represented was *M. intestinale* (20%), followed by *D. muris* (20%), *L. johnsonii* (14%), *A. muciniphila* (10%), *L. reuteri* (9%), and *F. rodentium* (5%). The remaining 21 species had less than 3% of percentual relative abundance.

5.2.1.4 Ethanol effect on microbial profiles of infant mice gut samples exposed *in utero*

In Figure 15 it is depicted the taxonomical relative abundance comparison between ethanol-exposed and control groups. Despite the similar microbial profiles obtained between the ethanol-exposed and the control group, some differences were observed. The genera *Massilistercora* (2%) and *Anaerocolumma* (1%) were only observed in the control group, though at a low level. At the species level, 5 species were only detected in the control group – *Massilisterocora timonensis* (2%), *Clostridium hylemonae* (2%), *Clostridium scindens* (2%), *Anaerocolumna* sp. (1%), and *Lachnoclostridium phocaense* (1%). The low prevalence of these taxa poses questions regarding their correct taxonomical assignment due to the high error-rate from nanopore sequencing basecalling. Moreover, the biological relevance of low abundant taxa should be carefully addressed. The rare taxa in each ecosystem are termed the “rare biosphere” and their importance towards the host or associated niche cannot be fully neglected however challenging to assess (Reid *et al*., 2011, Lynch and Neufield, 2015, Lauder *et al*., 2016, Jousset *et al*., 2017, Lim *et al*., 2018, de Goffau *et al*., 2019, Sanidad and Zeng, 2020). These microbial profiles observed show differences to previously reported data (Ferrere *et al*., 2017, Wang *et al*., 2018, Jiang *et al*., 2020, Li *et a*., 2021). The changes are not easily comparable due mostly to different experimental conditions performed, such as mice model used, diet, and ethanol administration regimen (Bull-Otterson *et al*., 2013, Samuelson *et al*., 2017). Specifically, and in accordance with other reports, neither *Prevotella*, *Pediococcus*, *Faecalibacterium*, *Bifidobacterium*, *Escherichia*, nor *Staphylococcus* were detected in this study (Jiang *et al*., 2020).


A
control
alcohol
0.00
0.25
0.50
0.75
1.00
Relative Abundance
Phylum
Firmicutes
Bacteroidetes
Verrucomicrobia
Deferribacteres


B
control
alcohol
0.00
0.25
0.50
0.75
1.00
Relative Abundance
Genus
Duncaniella
Muribaculum
Lactobacillus
Limosilactobacillus
Akkermansia
Faecalibaculum
Lachnoclostridium
Blautia
Parabacteroides
Lacrimispora
Peptoniphilus
Barnesiella
Massilistercora
Erysipelothrix
Ligilactobacillus
Anaerocolumna
Coprobacter
Bacteroides
Intestinibaculum
Alistipes
Deferribacter

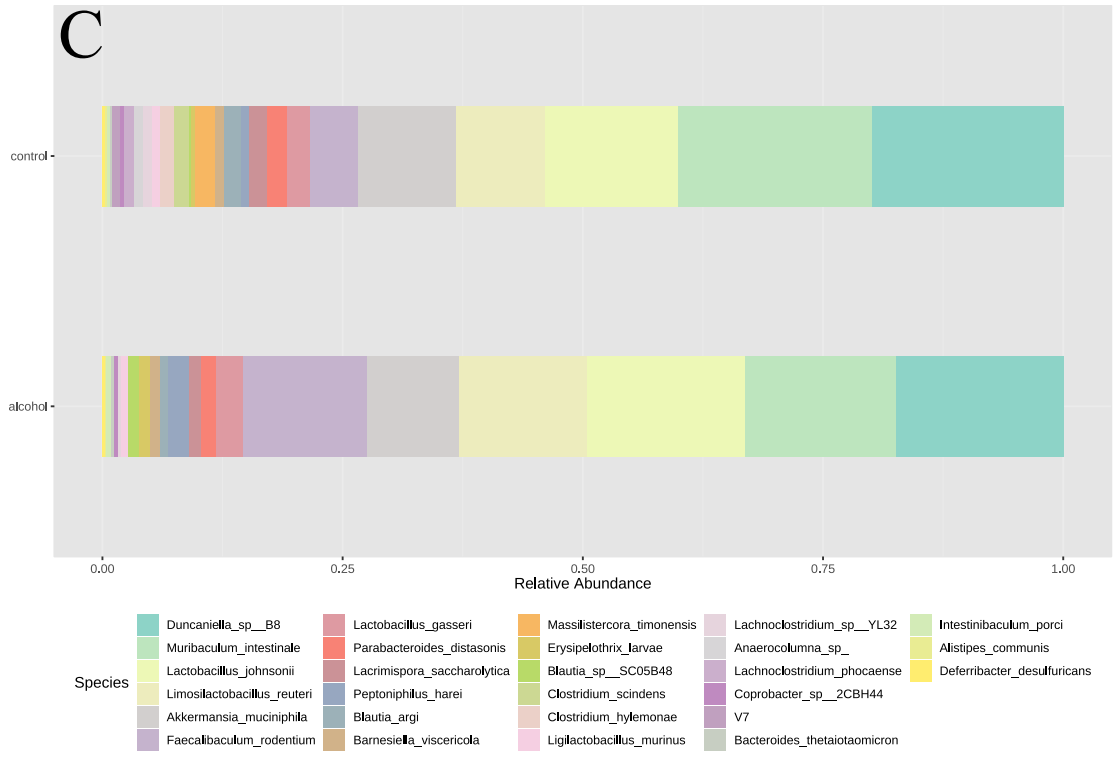


Figure 15 – Bar chart showing the relative abundance of taxa at the phylum (A), genus (B), and species level (C) detected in the ethanol-exposed and control groups by targeted nanopore sequencing. 'V7' corresponds to the relative abundance attributed to the non-assigned reads at the mentioned taxonomical rank level.

The mean F/B ratio was 2.41 in the ethanol-exposed group and 1.66 in the control group. However, discrepancies were found in the individual samples – between 0.2-3 points (Table 14). Taxa detected by nanopore sequencing may be due to other factors than the true microbial composition of the samples, like high basecalling error-rate and misrepresentation in the reference databases, and so the F/B ratio may represent a skewed result and mislead biological interpretations based on it.

Table 14 – Firmicutes/Bacteroidetes ratio distribution in the infant mice experimental samples.

Sample group	Sample ID	F/B ratio	Difference to Mean
Control	Control 9	0.11	-1.54
	Control 11	1.00	-0.66
	Control 13a[1]	2.23	0.58
	Control 13b	2.33	0.58
	Control 14	4.71	3.06
	Control 16	0.89	-0.77
	Control 17	0.90	-0.75

Table 14 – (continued).

Sample group	Sample ID	F/B ratio	Difference to Mean
Control	Control 18	1.06	-0.60
Ethanol-exposed	Test 1	2.13	-0.29
	Test 2	3.30	0.89
	Test 3	1.72	-0.69
	Test 4	0.19	-2.22
	Test 8	0.76	-1.65
	Test 9	2.13	-0.28
	Test 11	10.33	7.92
	Test 13	0.40	-2.01
	Test 14	0.73	-1.68

[1] – The sample control sample (Nr 13) was sequenced in the same sequencing run to assess the precision and accuracy of nanopore sequencing.

The control group presented a slightly higher alpha-diversity in comparison with ethanol-exposed group at the phylum, genus, and species level (Shannon distance measure alpha-diversity index at phylum level = 0.76 *vs* 0.70, genus level = 1.78 *vs* 1.71, species level = 1.93 *vs* 1.79). However, the differences were not significant (Mann-Whitney U Test, p-value > 0.05). Additionally, no differences were found in the diversity detected between the control and ethanol-exposed groups' samples for phylum, genus, and species ranks (PERMDISP, PERMANOVA, ANOSIM, p-value > 0.05). These observations are contrary to previous reports that observed higher alpha-diversity in ethanol-withdrawal mice models after ethanol consumption (Wang *et al*., 2018). Therefore, the identical alpha- and beta-diversity found between ethanol-exposed and control groups reveal that ethanol does not pose a strong effect on the microbial diversity composition found in the infant mice guts.

The microbial profiles retrieved by nanopore sequencing were almost identical between the experimental groups. One species showed different relative abundance between the experimental groups. In accordance with other reports, *F. rodentium* was detected in 13% of the reads in the ethanol-exposed group, and in the control group about 5% of the reads were assigned to the same taxon (Bjørkhaug *et al*, 2019). The clinical relevance of this taxon is being increasingly studied due to its importance in triggering mice depression (Wang *et al*., 2021), protection of intestinal tumour growth (alongside with its human homologue *Holdemanella biformis*) (Zagato *et al*., 2020), and further associations within the gut-bran axis (Wang *et al*., 2021).

Some taxa were exclusively detected in one of the experimental groups, however in a relatively low abundance. In the control group, *C. hylemonae* (number of samples where this taxon was detected = 4, relative abundance < 2.2%) *M. timonensis* (n = 3, relative abundance < 3.9%), *Anaerocolumna* sp. CTTW (n = 3, relative abundance < 2.1%), *C. scindens* (n = 2, relative abundance < 4.8%), *L. phocaeense* (n = 2, relative abundance < 1.9%), *Anaerostipes caccae* (n = 1, relative abundance < 1.2%), and *Anaerostipes hadrus* (n = 1, relative abundance < 1.5%). *Anaerocolumna* sp. CTTW was recently isolated from anoxic soil samples subjected to biological or reductive soil disinfestation and renamed as *Anaerocolumna chitinilytica* (Ueki *et al*., 2021). Its presence is controversial and can be explained either by husbandry or food contamination, or misclassification during bioinformatic classification steps. *M. timonensis* is a newly described species isolated from a stool sample of an 85 years old hospitalized woman (Tall *et al*., 2020). However, further research is necessary to elucidate the relevance of this species to humans. *L. phocaense* was recently isolated from a urine sample collected from a patient after kidney transplantation (Dandachi *et al*., 2021). *Lachnoclostridium* spp. are known members of the human gut microbiota (Lopetuso *et al*., 2013) and their relevance to the host includes anti-inflammatory effects, and butyrate production, thus contributing to the host intestinal homeostasis (Guo *et al*., 2020). *C. hylemonae*, *C. scindens*, and *A. caccae* are members of the *Clostridium coccoides* group, commonly detected in the human intestine, constituting 25-60% of the total microbiota (Kurakawa *et al*., 2015). *C. coccoides* group members may affect the host intestine in many ways. For instance, *A. caccae* is a butyrate producer (Barcenilla *et al*., 2000, Loius *et al*., 2009), and *C. scindens* and *C. hylemonae* have high levels of bile acid 7α-dehydroxylating activity (Kitahara *et al*., 2000). In the mouse colon, taxa from this group impact host immune homeostasis by regulatory T-cell production induction (Atarashi *et al*., 2011, Atarashi *et al*., 2013). Although being predominant in the human gut and showing relevant phenotypes towards the human and mouse host in terms of immunology, nutrition, and pathological processes, the taxonomical composition of the *C. coccoides* group is not fully characterized (Kurakawa *et al*., 2015).

Despite their low prevalence, the relevance of rare taxa cannot be neglected, and its study should take a justifiable significance in gut microbiome functional investigations (Reid *et al*., 2011). Commonly, these rara taxa, represent the majority of species found in environmental studies, displaying unique ecology and biogeography (Lynch and Neufield, 2015). Additionally, the rare biosphere represents the broad diversity and encompasses a wide ecological function arsenal and resiliency, providing the habitat a

functional redundancy and flexibility (Lynch and Neufield, 2015, Jousset *et al*., 2017). However, sequencing noise and biological artifacts may also contribute to the incorrect taxonomical assignment of part of the rare biosphere composition, which is explained by the high error-rate associated with nanopore sequencing, amongst other factors (Lynch and Neufield, 2015). Therefore, the clinical or homeostatic relevance of such rare taxa should be further addressed independently to unveil their true relevance, while deciphering if their presence is not a sequencing artifact (Lynch and Neufield, 2015, Lauder *et al*., 2016, Lim *et al*., 2018, de Goffau *et al*., 2019, Sanidad and Zeng, 2020). The F/B ratios showed a dispersed pattern between samples and experimental groups, and therefore is difficult to evaluate its meaning (Table 14). Indeed, technological biases may hinder the correct identification of microbial taxa, neglecting biologically relevant differences in composition mostly when analysing small sample sizes (Walters *et al*., 2014). Moreover, F/B ratios' usefulness have been discussed in other reports, as there is no consensual data to clear support its relevance as a robust biomarker for microbiome dysbiosis in clinical settings (Magne *et al*., 2020).

5.2.1.5 Microbial profiling correlation analyses and biomarker discovery

At the phylum level, *Firmicutes* was negatively correlated with *Bacteroidetes* (*rho* = -0.76; p-value < 0.05; T-stat = 1438; FDR < 0.05). In Table 15 is shown the correlations between genera found in the ethanol-exposed and control groups. These results show the complex network of associations between the mentioned taxa found in the infant mice gut. The correlation pairs *Erysipelothrix-Peptoniphilus*, *Erysipelothrix-Intestinibaculum*, *Faecalibaculum-Peptinophilus*, *Intestinibaculum-Peptoniphilus*, *Erysipelothrix-Faecalibaculum*, and *Faecalibaculum-Intestinaculum* represented the genera interconnected in the ethanol-exposed group (*rho* > 0.70; p-value < 0.05; FDR < 0.05). On the other hand, the correlation network found in the control group was supported by the *Anaerocolumna-Lachnoclostridium* and *Lachnoclostridium-Massilistercora* correlation pairs (*rho* = 0.8; p-value < 0.05; FDR < 0.05). *Blautia*, *Muribaculum*, *Duncaniella*, *Lactobacillus*, and *Limosilactobacillus* were not associated with any of the experimental groups and therefore they may be present in the infant mice gut core microbiota (Gu *et al*., 2013, Xiao *et al*., 2015, Wang *et al*., 2019, Yang *et al*., 2021). However, care must be taken to not directly extrapolate biological meaning from the statistically significant correlations, since some commonly found taxa may not possess a

true causative effect, nor represent any relevant host-impacting phenotype (Hanage, 2014).

Table 15 – Correlation table showing the Spearman ranks between genera detected through all experimental infant mice samples.

Taxon 1	Taxon 2	Spearman's rank correlation (*rho*)	p-value	T-stat
Erysipelothrix	*Peptoniphilus*	0.92	0	> 65
Erysipelothrix	*Intestinibaculum*	0.89	0	> 87
Faecalibaculum	*Peptoniphilus*	0.85	0	> 121
Intestinibaculum	*Peptoniphilus*	0.82	1.00E-4	> 145
Anaerocolumna	*Lachnoclostridium*	0.81	1.00E-4	> 157
Lachnoclostridium	*Massilistercora*	0.81	1.00E-4	> 157
Erysipelothrix	*Faecalibaculum*	0.79	1.00E-4	> 175
Blautia	*Lacrimispora*	0.76	4.00E-4	> 192
Barnesiella	*Muribaculum*	0.75	5.00E-4	> 204
Duncaniella	*Muribaculum*	0.72	0.001	> 225
Blautia	*Lachnoclostridium*	0.70	0.0017	> 243
Faecalibaculum	*Intestinibaculum*	0.70	0.0017	> 244
Coprobacter	*Muribaculum*	0.70	0.0017	> 244
Barnesiella	*Coprobacter*	0.64	0.0057	> 294
Anaerocolumna	*Blautia*	0.62	0.0086	> 314
Blautia	*Massilistercora*	0.62	0.0086	> 314
Lactobacillus	***Limosilactobacillus***	0.58	**0.014**[1]	> 340
Lachnoclostridium	***Lacrimispora***	0.53	**0.0298**	> 386
Coprobacter	***Duncaniella***	0.51	**0.0355**	> 397
Lactobacillus	***Muribaculum***	-0.52	**0.0339**	> 1237
Lactobacillus	***Parabacteroides***	-0.52	**0.032**	> 1241
Duncaniella	***Limosilactobacillus***	-0.56	**0.0209**	> 1268
Blautia	***Peptoniphilus***	-0.56	**0.0185**	> 1275
Barnesiella	***Lachnoclostridium***	-0.58	**0.0138**	> 1292
Limosilactobacillus	*Muribaculum*	-0.65	0.0047	> 1346
Duncaniella	*Lactobacillus*	-0.67	0.0033	> 1362
Barnesiella	*Limosilactobacillus*	-0.75	5.00E-4	> 1428

[1] – bold, FDR > 0.05

In Figure 16 it is observed that genera *Anaerocolumna*, *Massilistercora*, *Lachnoclostridium*, *Lacrimispora*, and *Blautia*, which are only or mainly present in the control group, are positively correlated between each other, indicating that ethanol may pose a stress to these taxa. Indeed, *Lachnoclostridium* and *Blautia* are negatively correlated with *Barnesiella*, and *Peptinophilus* genera, which are mostly found in the

ethanol-exposed group. Therefore, ethanol might have a selective effect on certain taxa. However, such hypothesis needs further experimentation.

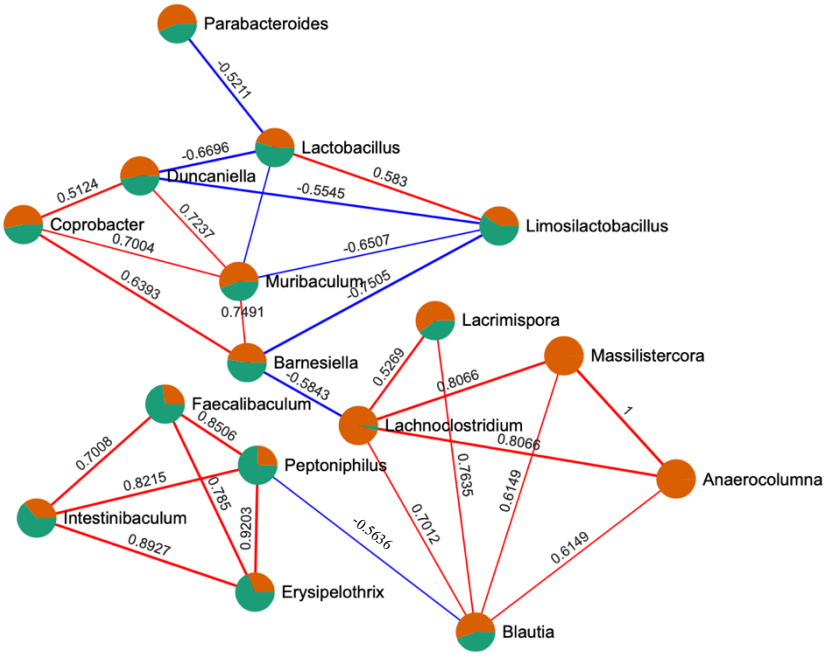


Figure 16 – Correlation network showing connected genera. Green sections within each circular node corresponds to the number of reads attributed to the target taxon in the ethanol-exposed group, and in orange are the number of reads attributed to the same taxa in the control group. Red lines represent positive Spearman correlations and blue lines represents negative Spearman correlations. Numbers above lines show the Spearman correlation coefficient between the taxa connected by the line.

Since the genera *Anaerocolumna*, *Massilistercora* and *Lachnoclostridium* are only or mainly present in the control group, it was expected that they were positively correlated with the control group, *i.e.*, reinforcing the negative effect of ethanol exposure to these taxa. Both genera showed moderate correlation with the control group (*rho* = 0.48-0.49, p-value < 0.05) (Figure 17). However, these observed correlations were not statistically significant (FDR > 0.05; FDR = 0.36, meaning that there is chance of 36% of finding false-positives), but they may be discussed focusing their absence in the ethanol-exposed group. The same observations were made at the species level, where *C. hylemonae* (*rho* = 0.58; p-value < 0.05), *Anaerocolumna* sp. (*rho* = 0.49; p-value < 0.05), and *M. timonensis* (*rho* = 0.49; p-value < 0.05) appeared to be positively correlated with the control group but such correlations were not confirmed after the FDR statistic (FDR = 0.31; *i.e.*, ~31% chance of being considered false-positive).

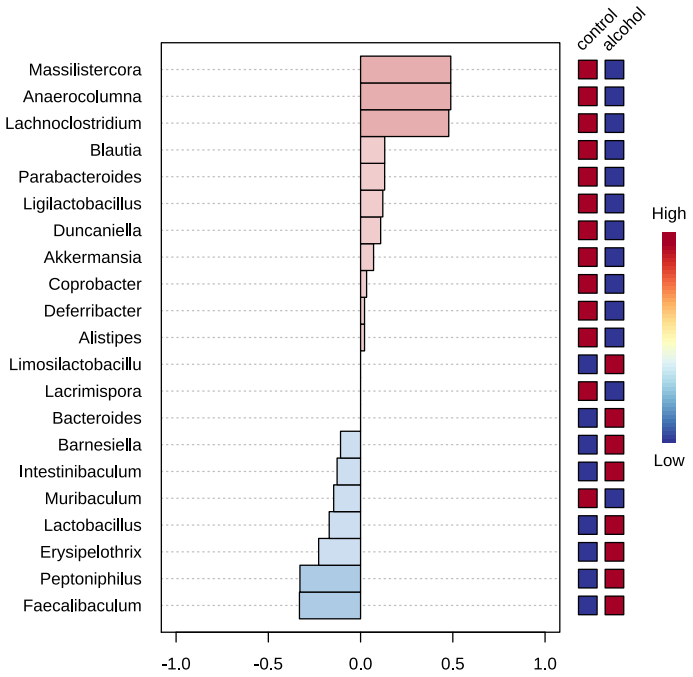


Figure 17 – Nanopore sequencing: Pattern search bar plot showing Spearman correlation coefficients between target genera and experimental conditions. Red bars correspond to positive correlations and blue bars correspond to negative correlations – deep bar colours mean that the correlation is stronger. Bars growing to the right indicate that the taxon is positively correlated with the experimental group coloured in red square; left-growing bars indicate that the taxa is negatively correlated with the experimental group in a blue coloured square. Red squares means that the specific taxon is highly abundant at the respective experimental condition, and blue squares means that the taxon is not abundant.

No correlation was found between any taxa and the experimental groups, so there are no taxon that could be considered a biomarker for condition tested (LEfSe test, Appendix V), and by univariate statistical comparisons one species was found to have a different abundance profile between the experimental groups – *C. hylemonae* (p-value < 0.05; Mann-Whitney U Test = 54). Despite most probably meaning a false-positive (FDR = 0.40, *i.e.*, 40% chance of being considered false-positive), the difference must be discussed in terms of the potential biological meaning, as previously stated (Hanage, 2014).

Unsupervised Random Forest classification showed that no significant predictor could affect the classification found in the control and ethanol-exposed group (OOB error in the control group at the phylum level = 0.63, genus level = 0.38, species level = 0.44; OOB error in the ethanol-exposed group at the phylum level = 0.67, genus level = 0.44, species level = 0.33). OOB is an indicator of how good the prediction is at the taxonomical rank level under evaluation, *i.e.*, lower OOB values mean a good predictability of the

associated taxon. Thus, through OOB rates, it can be observed that species level rank can predict the group classification by less than 67% (overall OOB error at the phylum level = 0.65, genus level = 0.4, species level = 0.35). Additionally, both groups showed a perfect equality if any particular taxa were absent from the classification (Mean Decrease Coefficient < 0.03) (Figure 18). These results confirm the previously similarity observed between the microbial profiles of ethanol-exposed and control group.

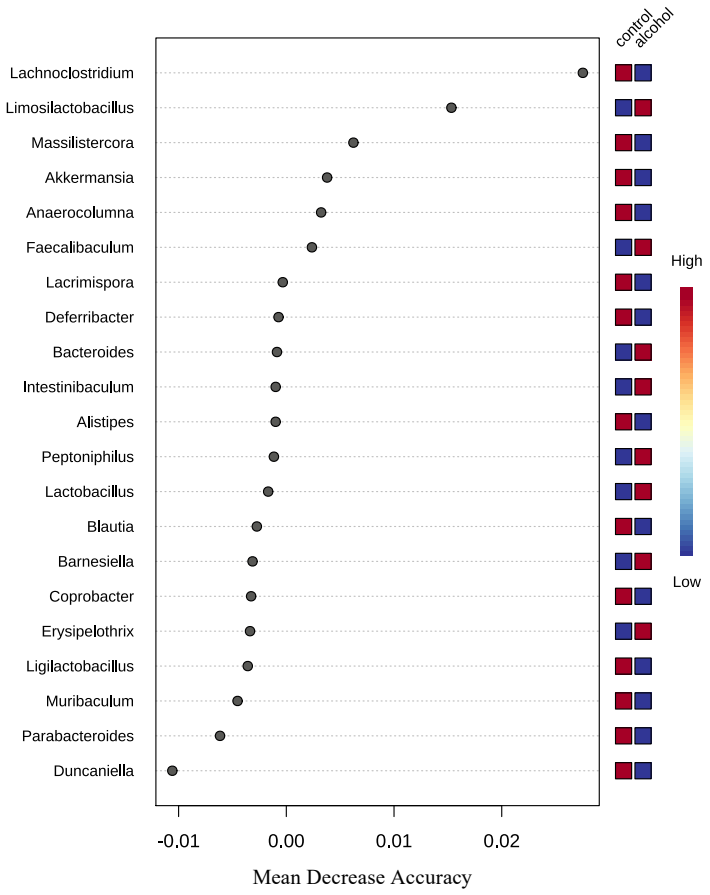


Figure 18 – Nanopore sequencing: Unsupervised RF classification using ethanol-exposure and controls as classes and taxa relative abundances as classifiers along a decision tree. In the y-axis is depicted the most relevant classifying taxa, and in the x-axis is shown the Mean Decrease Accuracy, which is a measure of loss of accuracy if the taxon (classifier) on the y-axis is removed from the classification. On the right is a mini heat map showing the abundance in the target group. Red squares means that the specific taxon is highly abundant and blue squares means that the taxon is not abundant.

Since ethanol was available *ad libitum* it cannot be discarded the hypothesis that the differences here detected are explained by the different ethanol consumed by the mothers in the experimental and control groups.

5.2.1.6 16S rRNA gene-targeted microbial profiles comparison between Optimization and Test experiments

To assess the precision of nanopore sequencing approaches, the results for the 12 ethanol-exposed mice gut samples used during the 'optimization stage of the assays' (termed optimization samples) (presented in Chapter 4) were compared with the results obtained from the nine ethanol-exposed mice samples studied in the current chapter (termed 'Test samples'). The faecal samples analysed during the optimization stage and in the current chapter were obtained from a different offspring of infant mice. The 12 'optimization-stage samples' and the nine 'test samples' were compared using MicrobiomeAnalyst as described in Chapter 2.

The two set of samples showed significant differences in their microbial profiles at the genus and species level (Figure 19). The Test samples showed increased abundances of *Lactobacillus*, *Parabacteroidetes*, and *Barnesiella* genera, and *L. johnsonii*, *B. viscericola*, and *P. distasonis*. The Optimization set of samples displayed a higher abundance of *Peptinophilus* and *Amedibacterium* genera, and *Peptinophilus harei*, and *Amedibacterium intestinale* (see Appendix VI for a detailed heat tree of each set of samples). These results highlight the challenging interpretation of nanopore sequencing microbial profiles when infant mice were exposed to ethanol *in utero*. As introduced before, the mother-foetus interplay is complex and it is not totally clear how or when vertical microbial transmission occurs, nor if the womb is a sterile or a colonized environment (Lauder *et al*., 2016, Lim *et al*., 2018, de Goffau *et al*., 2019). Still, since the analysed infant mice belonged to different offsprings, and therefore corresponding to independent husbandry experiences, the hypothesis of either population or individual inner diversity cannot be ruled out (Gu *et al*., 2013, Xiao *et al*., 2015, Wang *et al*., 2019, Yang *et al*., 2021).

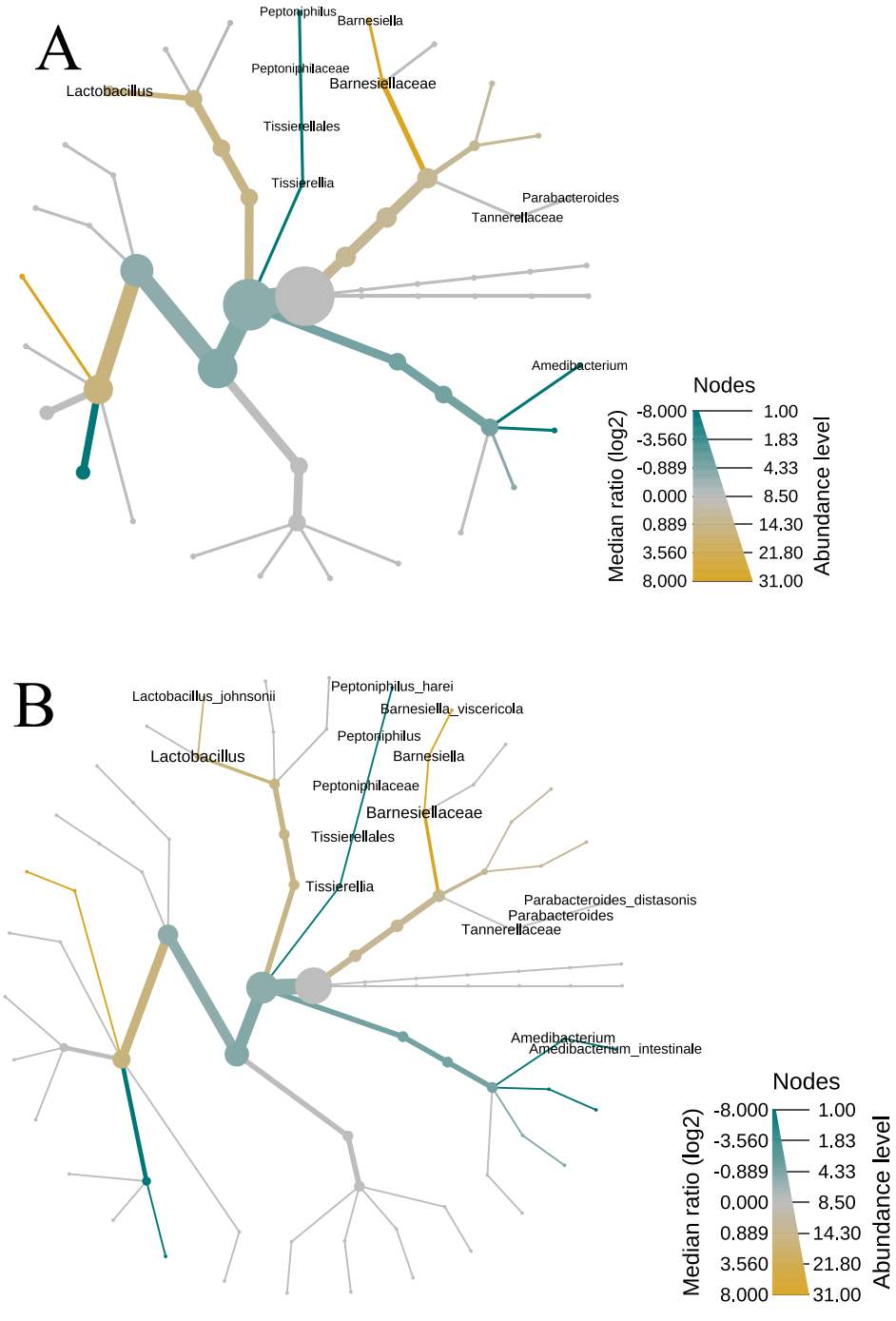


Figure 19 – Heat trees displaying the community structure of the two set of samples. Size and colour of nodes and edges represent correlations with the abundance of organisms in each community. Only significantly different (Wilcoxon p-vale < 0.05) taxa are labelled. 'test' represents the combined Test set of samples, and 'optimization' represents the combined nanopore sequencing Optimization set of samples. A, represents the comparison at the genus level, and B at the species level. Taxa names on top of yellow lines mean taxa significantly overrepresented in Test samples. Taxa names on top of green lines mean taxa significantly overrepresented in Optimization samples.

There were no significant differences detected in the diversity found either within or between optimization and test samples (alpha-diversity and beta-diversity indices). However, sample 4, 11, and 13 in the Test set showed a low alpha-diversity index, while samples 1, 2, 3, and 5 in the Optimization set showed a high alpha-diversity index (Figure 20). The discrepancies detected in the microbial diversity within and between the optimization and test samples support the hypothesis of unique microbial signatures at the individual level that show the probable infant mice core microbiota diversity (Gu *et al*., 2013, Xiao *et al*., 2015, Wang *et al*., 2019, Yang *et al*., 2021).

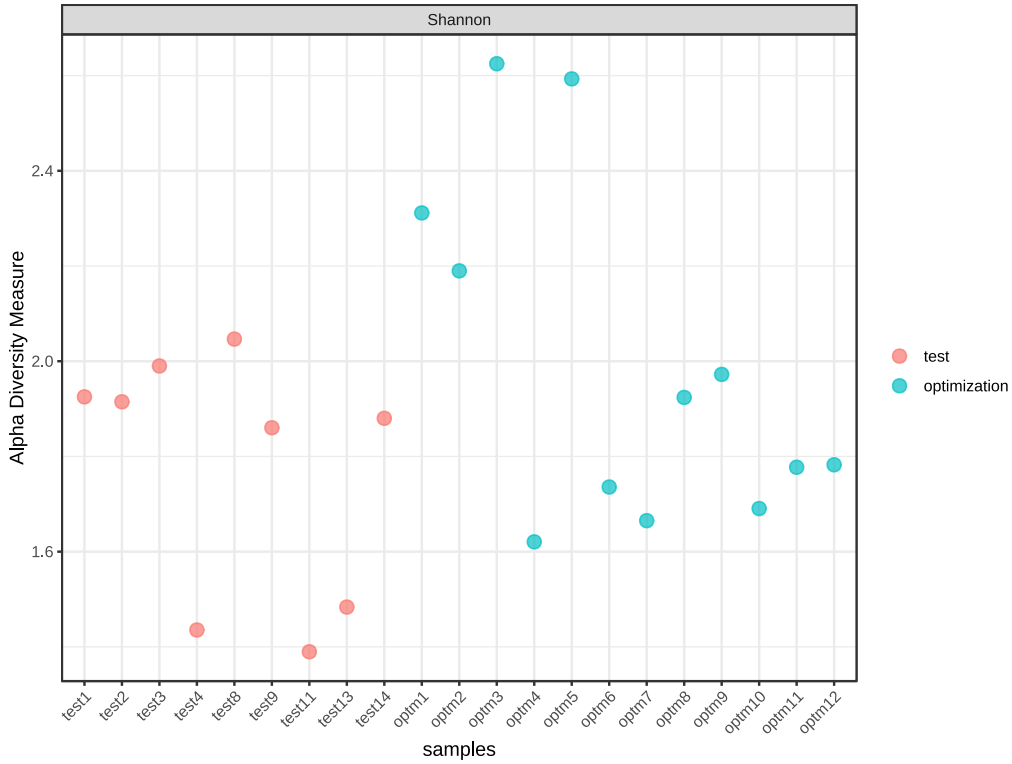


Figure 20 – Sample-level dot plot displaying the alpha-diversity indices based on Shannon distance measure. The x-axis shows the individual infant mice gut samples analysed and the y-axis shows the Shannon alpha-diversity index for each sample.

A single significant correlation was detected with *Bacteroidetes* showing a significant positive correlation with the Test set of samples (*rho* = 0.55; p-vale < 0.05; FDR < 0.05), therefore were detected in greater abundance in the Test samples compared with the Optimization samples (Figure 21). This observation is contrary to what was observed in a previous report (Wang *et al*., 2018), where *Bacteroidetes* members were underrepresented in the ethanol-exposed mice group. The correlation found here with *Bacteroidetes* phylum is not easily interpreted apart from the chance or unique

individual/population-level microbial signatures (Bull-Otterson *et al*., 2013, Gu *et al*., 2013, Xiao *et al*., 2015, Samuelson *et al*., 2017, Wang *et al*., 2019, Yang *et al*., 2021).

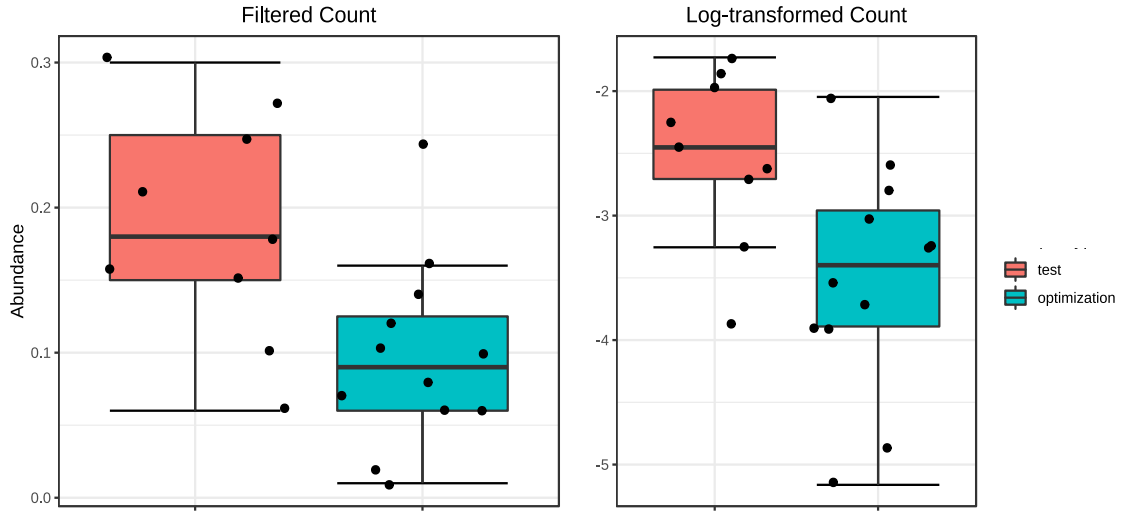


Figure 21 – Boxplot displaying the relative abundances of phylum *Bacteroidetes* (*rho* = 0.5; p-value < 0.05; FDR < 0.05) in the two set of samples. 'test' represents the combined Test set of samples, and 'optimization' represents the combined nanopore sequencing Optimization set of samples. Left: raw relative abundances; Right: log-transformed relative abundances.

In Figure 22 it is the displayed the correlation network of all genera that displayed significant correlations between other taxa (*rho* > 0.75; p-value < 0.05). All the significant correlations found were positive. A correlation network was found at the species level (see Appendix VII), however, the respective biological meaning of such correlations must be assessed in further studies in order to avoid false-positive or irrelevant correlations (Hanage, 2014). As previously discussed in this Chapter, the correlations observed are mainly observed within the same experimental group and so they might be explained by associated experimental conditions or unique microbial profile patterns.

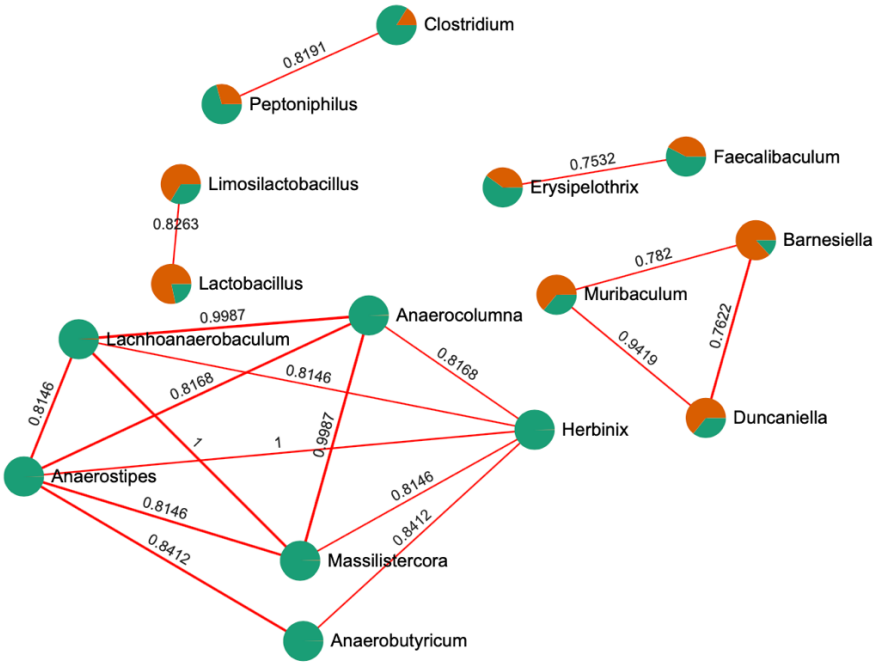


Figure 22 – Correlation network showing connected genera (*rho* > 0.75; p-value < 0.05). Green sections within each circular node corresponds to the number of reads attributed to taxa present in the Optimization set of samples, and in orange are the number of reads attributed to the same taxa in the Test set of samples. Red lines represent positive Spearman correlations and blue lines represents negative Spearman correlations. Numbers above lines show the Spearman correlation coefficient between the taxa connected by the line.

Supporting the infant mice core microbiota existence hypothesis, *M. intestinale*, *D. muris*, *F. rodentium*, *L. reuteri*, *L. johnsonii*, and *P. harei* were detected in more than 50% of the samples with a relative abundance above 24%, thus they can be considered part of the infant mice gut core microbiota when these mice are exposed to ethanol *in utero* (Gu *et al*., 2013, Xiao *et al*., 2015, Wang *et al*., 2019, Yang *et al*., 2021).

The correlation network identified reveals that a complex net of interconnectedness is present in the infant mice gut (Gonzalez-Perez and Lamousé-Smith, 2017, Soderborg *et al*., 2018, Lo *et al*., 2021). Interestingly, three type of correlations seems to be identified: i) ethanol-related correlations (Figure 16 and 17), ii) control-group related correlations (Figure 16 and 17), iii) core gut microbiota correlations (Table 15, Figure 22). Mostly correlations found were positive, although three negative correlation pairs were identified, *Limosilactobacillus-Muribaculum* (*rho* = -0.65; p-value < 0.05; FDR < 0.05), *Duncaniella-Lactobacillus* (*rho* = -0.67; p-value < 0.05; FDR < 0.05), and *Barnesiella-Limosilactobacillus* (*rho* = -0.75; p-value < 0.05; FDR < 0.05). Apart from *Barnesiella* which is mostly present in the control group, all the taxa present in the correlation pairs were present in more than 5 samples in both experimental groups. This

may indicate the complex relationship found in the core microbiota. The correlation pairs *Erysipelothrix-Peptoniphilus*, *Erysipelothrix-Intestinibaculum*, *Faecalibaculum-Peptinophilus*, *Intestinibaculum-Peptoniphilus*, *Erysipelothrix-Faecalibaculum*, and *Faecalibaculum-Intestinaculum* are more prevalent in the control group (*rho* > 0.70; p-value < 0.05; FDR < 0.05). Except for *Faecalibaculum*, these species have not been extensively studied, in spite of their pathogenic relevance in many animals, including mice and humans (Brown *et al*., 2014, Kim *et al*., 2019, Opriessnig *et al*., 2020, Aujoulat *et al*., 2021). *Anaerocolumna-Lachnoclostridium* and *Lachnoclostridium-Massilistercora* correlation pairs represent taxa more prevalent in the ethanol-exposed group (*rho* = 0.8; p-value < 0.05; FDR < 0.05). There is not much data regarding genera *Anaerocolumna*, and *Massilistercora*, which were recently described and discovered, respectively (Ueki *et al*., 2016, Tall *et al*., 2020). However, *Lachnosclostrium* were recently associated with colorectal adenoma and cancer (Liang *et al*., 2020), atherosclerosis (Cai *et al*., 2022), albeit being commonly found in the human gut microbiota (Dandachi *et al*., 2021). Therefore, further research must be performed in order to clarify the relevance of these taxa to host homeostasis. However, the detection of the abovementioned taxa in ethanol-exposed and control groups were not abundant and only occurred in a limited number of samples. This observation strengthens the hypotheses that these taxa are members of the rare gut microbiota taxa and should be interpreted with that notion in mind, as previously discussed (Lynch and Neufield 2015, Lauder *et al*., 2016, Lim *et al*., 2018, de Goffau *et al*., 2019, Sanidad and Zeng, 2020). The remaining correlation pairs were not attributed to any of the experimental groups and the correlated taxa, specifically the genera *Blautia*, *Muribaculum*, *Duncaniella*, *Lactobacillus*, and *Limosilactobacillus,* are most probably members of the infant mice gut core microbiota (Gu *et al*., 2013, Xiao *et al*., 2015, Wang *et al*., 2019, Yang *et al*., 2021).

In conclusion, ethanol exposure *in utero* may exert a selective negative effect on certain taxa. However, this hypothesis needs further experimentation and validation. As stated before, since ethanol was available *ad libitum* it cannot be discarded the hypothesis that the differences here detected are explained by the different ethanol consumed by the mothers in the two set of samples. Additionally, since infant mice were studied, it may happen that their gut microbiota needs more time to develop and its maturation only occurs later in life (Van Best *et al*., 2020), and so no differences can be found at early ages despite the experimental conditions. Still, no significant correlation was found between any taxa and the experimental groups, which indicates that further investigation must be performed in order to clarify if the abovementioned correlations between taxa

pairs were only found by chance. No taxa were identified as a good taxonomical classification predictor for any of the experimental groups under investigation, corroborating the hypothesis of the identical microbial profiles when ethanol-exposure experimental conditions were tested (Figure 18).

5.2.2 Illumina sequencing

5.2.2.1 Illumina sequencing targeted approach quality assessment

A total of 1,988,055 reads were obtained through Illumina sequencing, with a total yield of 1,132,178 kilobases. All reads achieved the quality filter, this means that they have passed the default Illumina chastity filter procedure. More than 60% of the reads had passed Q30 and the mean Q obtained was 29.2 (Table 16).

Table 16 – General performance of Illumina sequencing of the 16 DNA templates obtained from infant mice gut samples.

Sample name	No. read pairs	Yield (Kbp)	No. reads with > Q30 (%)	Mean Qscore[1]
Control 9	114,051	64,953	64.16	29.27
Control 11	105,147	59,845	63.48	29.09
Control 13	116,990	66,578	63.61	29.14
Control 14	124,211	70,721	64.24	29.35
Control 16	160,921	91,697	64.16	29.30
Control 17	127,194	72,441	64.61	29.36
Control 18	113,532	64,652	63.88	29.19
Test 1	139,378	79,390	63.46	29.10
Test 2	121,694	69,285	65.29	29.63
Test 3	130,043	74,013	62.51	28.85
Test 4	149,689	85,223	63.82	29.19
Test 8	136,214	77,592	63.21	29.06
Test 9	107,287	61,137	63.71	29.14
Test 11	90,803	51,708	63.93	29.29
Test 13	124,669	71,027	64.19	29.25
Test 14	126,232	71,916	62.67	28.87

[1] – The Qscore is a prediction of the probability of a wrong base call and Q30 represents the percentage of bases with a quality score of at least 30 (inferred base call accuracy of 99.99%).

A total of 388 of OTUs were assigned to taxa (62-69% of the reads). The number of OTUs correlates with the diversity of the data set. According to literature, a copy-number correction was performed only for species taxonomical rank (Angly *et al*., 2014).

Specifically, the number of reads assigned to a species was divided by the known or assumed copy-number of marker gene region. Therefore, the resulting corrected total count may be significantly lower than the raw total number of assigned reads (see Appendix I).

5.2.2.2 Microbial profiling of infant mice gut samples exposed to ethanol *in utero*

In the ethanol-exposed group *Bacteroidetes* was the most abundant phylum represented (70%) followed by *Firmicutes* (14%), *Actinobacteria* (12%), and *Proteobacteria* (3%). *Bacillota* and *Deferribacteres* did not exceed 1% relative abundance and were only observed in three and four samples, respectively. Among the data, 0.5% of the detected reads was not assigned to any phylum. At the genus level, *Muribaculum* accounted for most of the reads observed (47%), followed by *Bifidobacterium* (7%), *Lactobacillus* (7%), *Paramuribaculum* (6%), *Olsenella* (5%), and *Duncaniella* (4%). Around 6% of the reads did not have any genus attributed. The remaining 37 genera were represented by less than 3% of the reads each. Amongst all reads, 39 species were observed and 26% of the reads were not resolved to the species level. The top species detected were *Muribaculum* sp. (42%), *Paramuribaculum intestinale* (6%), and *Duncaniella dubosi* (4%). The remaining 36 species observed represented less than 3% each.

In the control group *Bacteroidetes* was the most abundant phylum represented (68%) followed by *Firmicutes* (18%), *Actinobacteria* (8%), and *Proteobacteria* (3%). *Deferribacteres* and *Bacillota* did not exceed 1% relative abundance and were only observed in five and one sample, respectively. Among the data, 0.6% of the detected reads was not assigned to any phylum. At the genus level, *Muribaculum* accounted for most of the reads observed (46%), followed by *Paramuribaculum* (6%), *Olsenella* (5%), *Lactobacillus* (5%), *Kineothrix* (4%), *Bacteroides* (4%), and *Duncaniella* (4%). Around 7% of the reads did not have any genus attributed. The remaining 33 genera were represented by less than 3% of the reads each. Amongst all reads, 42 species were observed and 23% of the reads were not resolved to the species level. The top species detected were *Muribaculum* sp. (41%), *P. intestinale* (6%), *Kineothrix alysoides* (4%), and *D. dubosi* (4%). The remaining 37 species observed represented less than 3% each.

5.2.2.3 Ethanol effect on the infant mice gut microbiota when mice were exposed *in utero*

There was observed a great level of similarity between the microbial profiles where the taxa present in both groups did not differ more than 2 percentual relative abundance points (data not shown). As stated before, the mice gut microbiota development and function are age-dependent and it may take time to achieve maturation (Van Best *et al*., 2020). This may be sufficient to explain the similarities found in this thesis. Despite the similar microbial profiles obtained between the ethanol-exposed and the control group, some differences were observed (Figure 23). The genera *Rhizobium* and *Gemella* although low in abundance (< 0.1%) were only present in the ethanol-exposed group, and the genera *Mordavella*, *Millionella*, and *Marvinbryantia* were also present at a very low level (< 0.3%) and only in the control group. Some species were only present in the ethanol-exposed group, such as *Bifidobacterium pseudolongum*, *Gemella* sp., and *Rhizobium* sp. UB 12. On the other hand, *Clostridium* sp. C5 48, *Clostridium* sp. Clone 49, *Clostridium* sp. Marseille P3122, *Marvinbryantia* sp., *Millionella massiliensis*, and *Mordavella* sp. were only observed in the control group. Therefore, these low abundant taxa may belong to the rare biosphere in the infant mice gut, as previously discussed (Reid *et al*., 2011, Lynch and Neufield, 2015, Lauder *et al*., 2016, Jousset *et al*., 2017, Lim *et al*., 2018, de Goffau *et al*., 2019, Sanidad and Zeng, 2020). Additionally, and as previously discussed for nanopore sequencing approaches' results, the microbial profiles detected by Illumina sequencing do not match with previous reports and the disparities may be explained by the different experimental conditions performed in other reports (Bull-Otterson *et al*., 2013, Samuelson *et al*., 2017). Specifically, and in accordance with other reports, neither *Pediococcus*, *Faecalibacterium*, *Escherichia*, nor *Staphylococcus* were detected in this study by Illumina sequencing (Jiang *et al*., 2020).

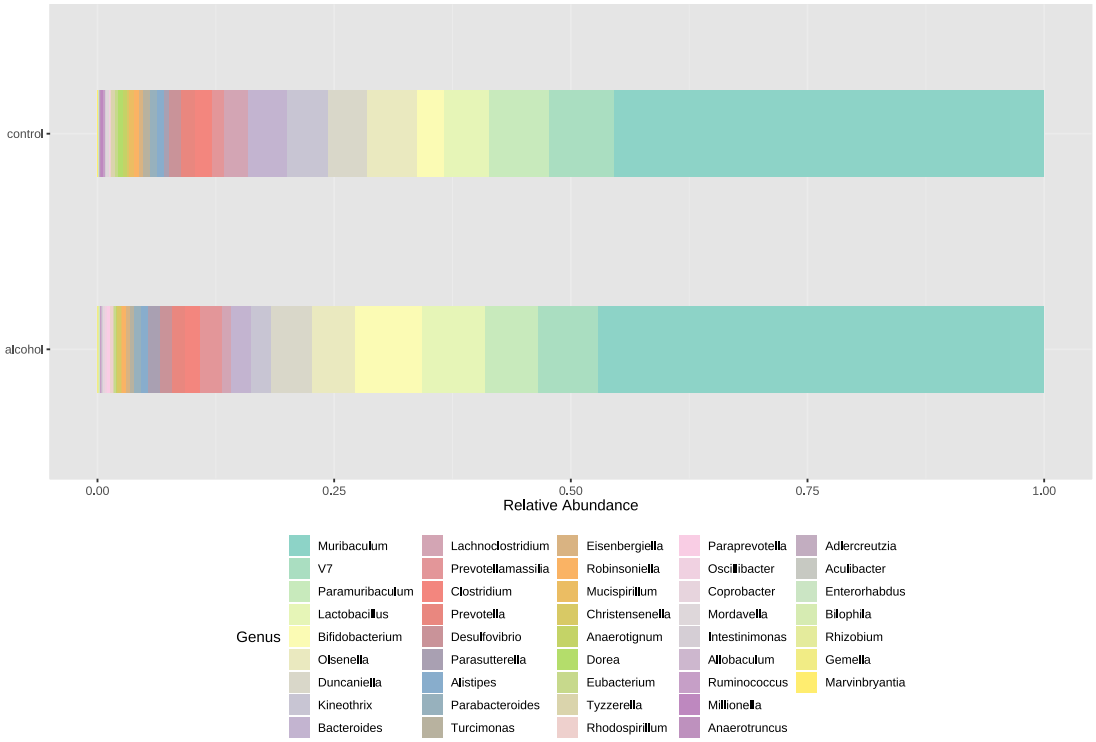


Figure 23 – Bar chart showing the relative abundance of all combined samples at the genus level detected in the ethanol exposed and control groups. The x-acis shows the combined relative abundance of each taxa, and the y-axis shows the groups analysed. 'V7' corresponds to the relative abundance attributed to the non-assigned reads at the mentioned taxonomical rank level.

The F/B ratio was 0.22 in the ethanol-exposed group and 0.27 in the control group. However, in the ethanol-exposed group, three samples showed ratios below 0.1 (Test 3 = 0.07; Test 8 = 0.08; Test 14 = 0.06), and two samples showed elevated ratios in comparison with the group average (Test 2 = 0.6; Test 11 = 0.56).

The control group presented a higher alpha-diversity in comparison with ethanol-exposed group at the phylum (Shannon distance measure alpha-diversity index = 0.90 *vs* 0.85), genus (2.21 *vs* 2.03), and species level (2.46 *vs* 2.30). However, the differences were not significant (p-value > 0.05). Additionally, no differences were found in the diversity detected between samples in the control and ethanol-exposed groups at the phylum, genus, and species level (NMDS ordination, Bray-Curtis index, PERMDSIP, PERMANOVA, ANOSIM, p-value > 0.05). These results are in accordance with what was observed by targeted nanopore sequencing microbial profiling, and they contradict previous reports (Wang *et al*., 2018), thus supporting the hypothesis that ethanol does not pose a strong selection on the microbial diversity composition found in the infant mice guts. As stated before, since ethanol was available *ad libitum* it cannot be discarded the

hypothesis that the differences here detected are explained by the different ethanol consumed by the mothers in the experimental and control groups.

5.2.2.4 Microbial profiling correlation analyses and biomarker discovery

None of the detected taxa were significantly correlated with the ethanol stressor. No significant correlation was found within the control group as well. However, the genera *Millionella*, *Dorea*, *Kineothrix*, and *Mordavella* appeared to be reasonable correlated with the control group (*rho* > 0.5; p-value < 0.05) (Figure 24). Although not truly significant (FDR > 0.28, *i.e.*, around 28% chance of being considered false-positive), such correlations must be discussed in the sense of their biological meaning to clarify the true function of these taxa to the host (Hanage, 2014). The same observation was made for the species *M. massiliencis*, *Clostridium* sp. Marseille P3122, *K. alysoides*, *Clostridium* sp. Clone 49, and *Mordavella* sp. (*rho* > 0.5; p-value < 0.05; FDR > 0.15, *i.e.*, around 15% chance of being considered false-positive). Although only *K. alysoides* was assigned to more than 1% of the reads (control group average relative abundance = 4%; ethanol-exposed group average relative abundance = 2%), it can still be considered a member of the rare gut biosphere (Reid *et al*., 2011, Lynch and Neufield, 2015, Lauder *et al*., 2016, Jousset *et al*., 2017, Lim *et al*., 2018, de Goffau *et al*., 2019, Sanidad and Zeng, 2020). Additionally, no statistically significant observations were obtained by univariate statistical comparisons nor by LEfSe which supports the hypothesis of microbial gut profiles' similarity despite the ethanol exposure *in utero*.

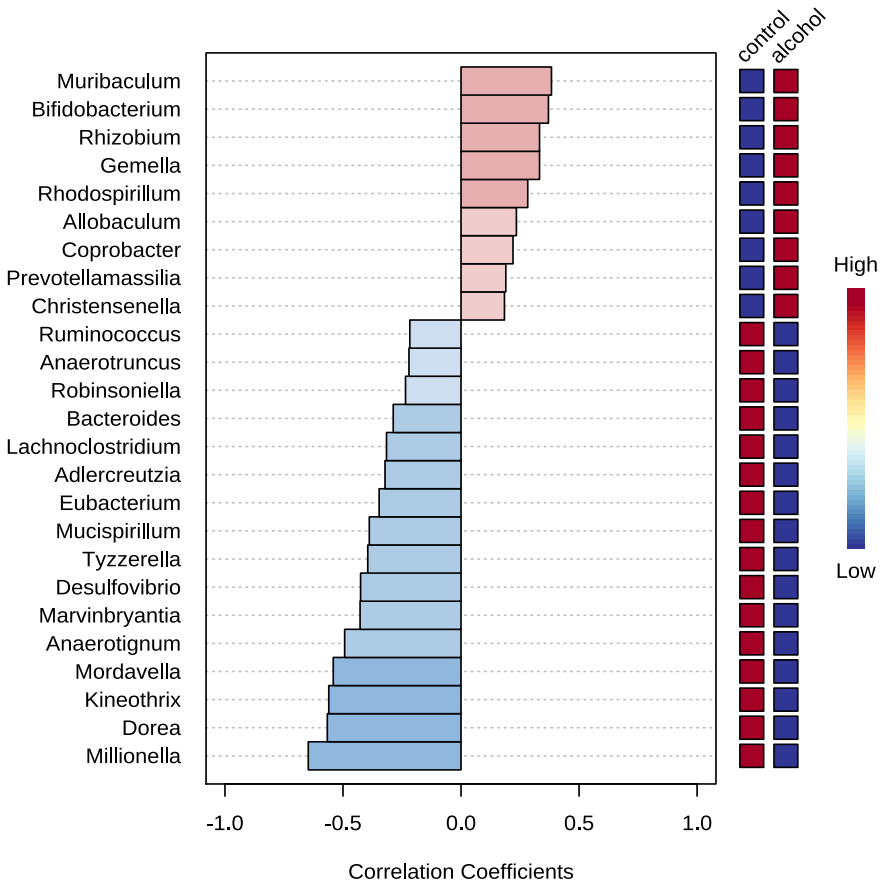


Figure 24 – Illumina sequencing: Pattern search bar plot showing Spearman correlation coefficients between target genera and experimental conditions. Red bars correspond to positive correlations and blue bars correspond to negative correlations – deep bar colours mean that the correlation is stronger. Bars growing to the right indicate that the taxon is positive correlated with the experimental group coloured in red square; left-growing bars indicate that the taxa is negatively correlated with the experimental group in a blue coloured square. Red squares means that the specific taxon is highly abundant at the respective experimental condition, and blue squares means that the taxon is not abundant.

Unsupervised Random Forest analysis showed that no significant predictor could affect the classification found in the control (OOB error = 0.6) and ethanol-exposed group (OOB error = 0.2-0.4). However, in the ethanol-exposed group, OOB showed that genus and species ranks can predict its classification by 78%. These results show that the genera and species level profiles are somehow dependent on the experimental factor under investigation (ethanol-exposure *vs* non-exposure). However, the remaining results here presented do not support this evidence, due to the high similarity found at the compositional profiles and diversity indices between the two experimental groups. Additionally, both groups showed a perfect equality if any particular taxa were absent

from the classification (Mean Decrease Coefficient < 0.012) (Figure 25), therefore no specific taxon seems to be a feature of none of the experimental groups.

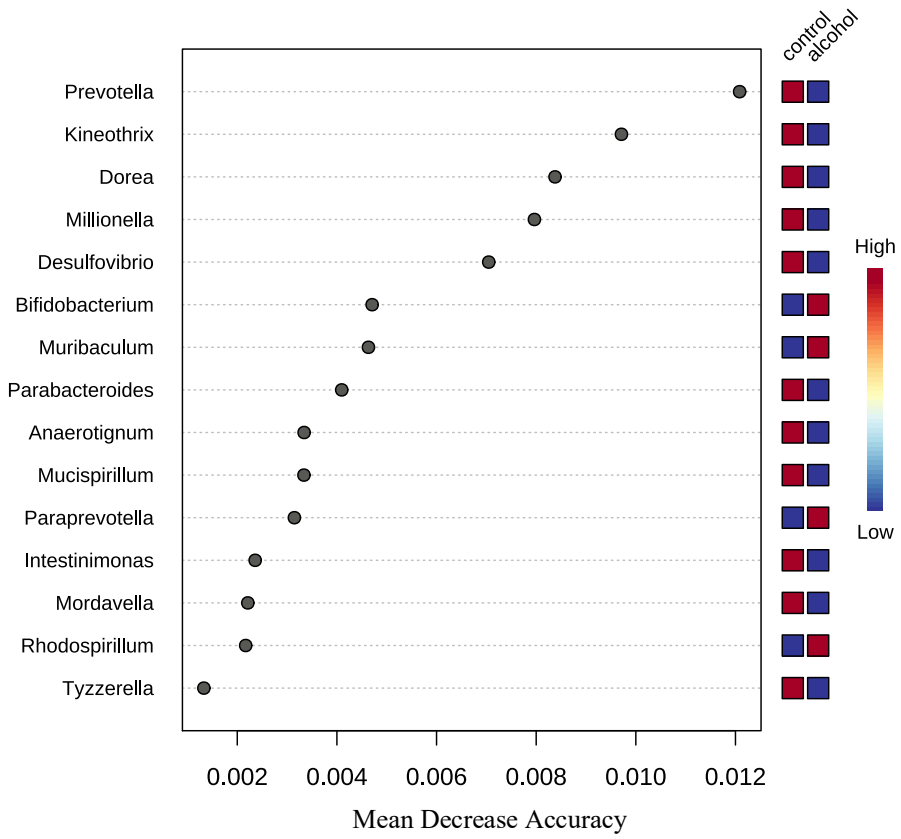


Figure 25 – Illumina sequencing: Unsupervised RF classification using ethanol-exposure and controls as classes and taxa relative abundances as classifiers along a decision tree. In the y-axis is depicted the most relevant classifying taxa, and in the x-axis is shown the Mean Decrease Accuracy, which is a measure of loss of accuracy if the taxon (classifier) on the y-axis is removed from the classification. On the right is a mini heat map showing the abundance in the target group. Red squares means that the specific taxon is highly abundant and blue squares means that the taxon is not abundant.

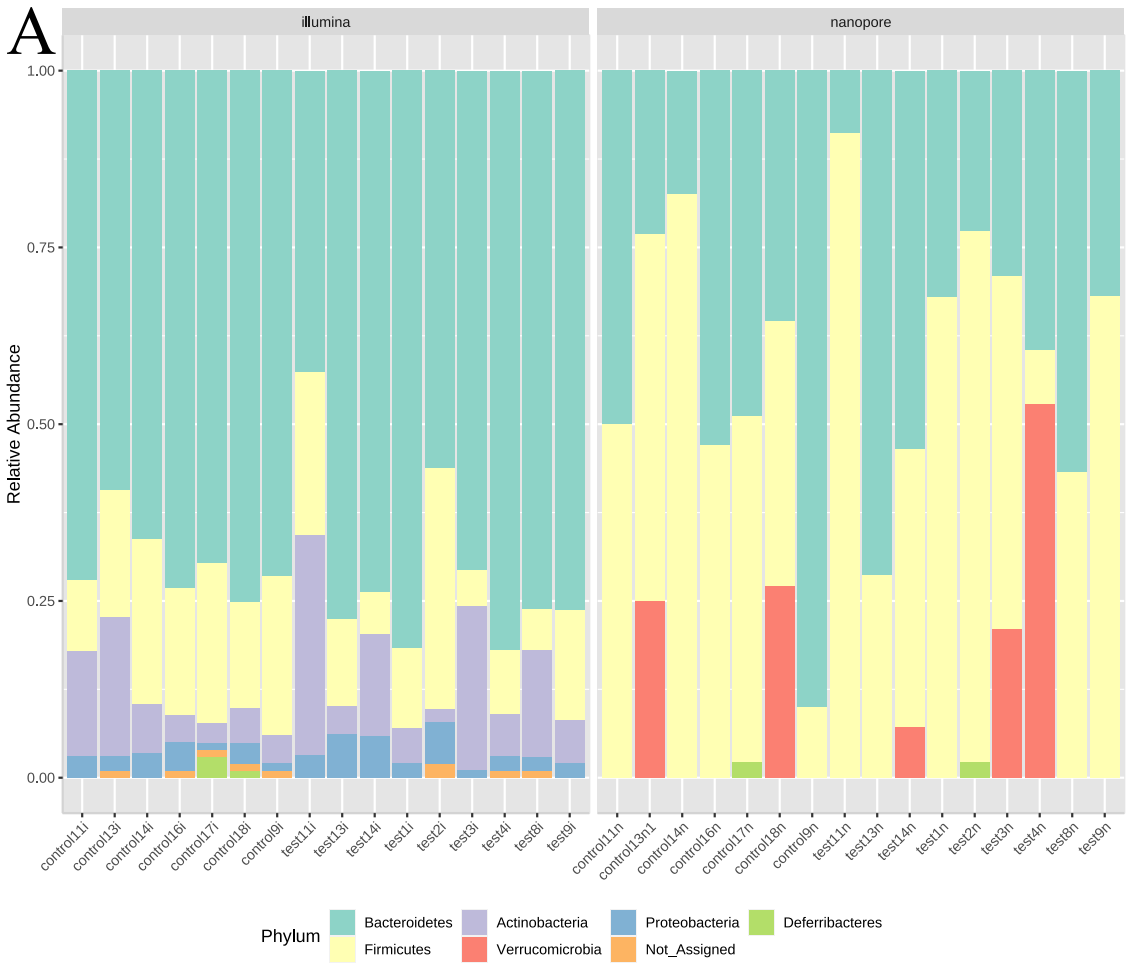
5.2.3 Comparison between microbial profiles of infant mice guts exposed to ethanol *in utero* obtained by nanopore and Illumina sequencing approaches

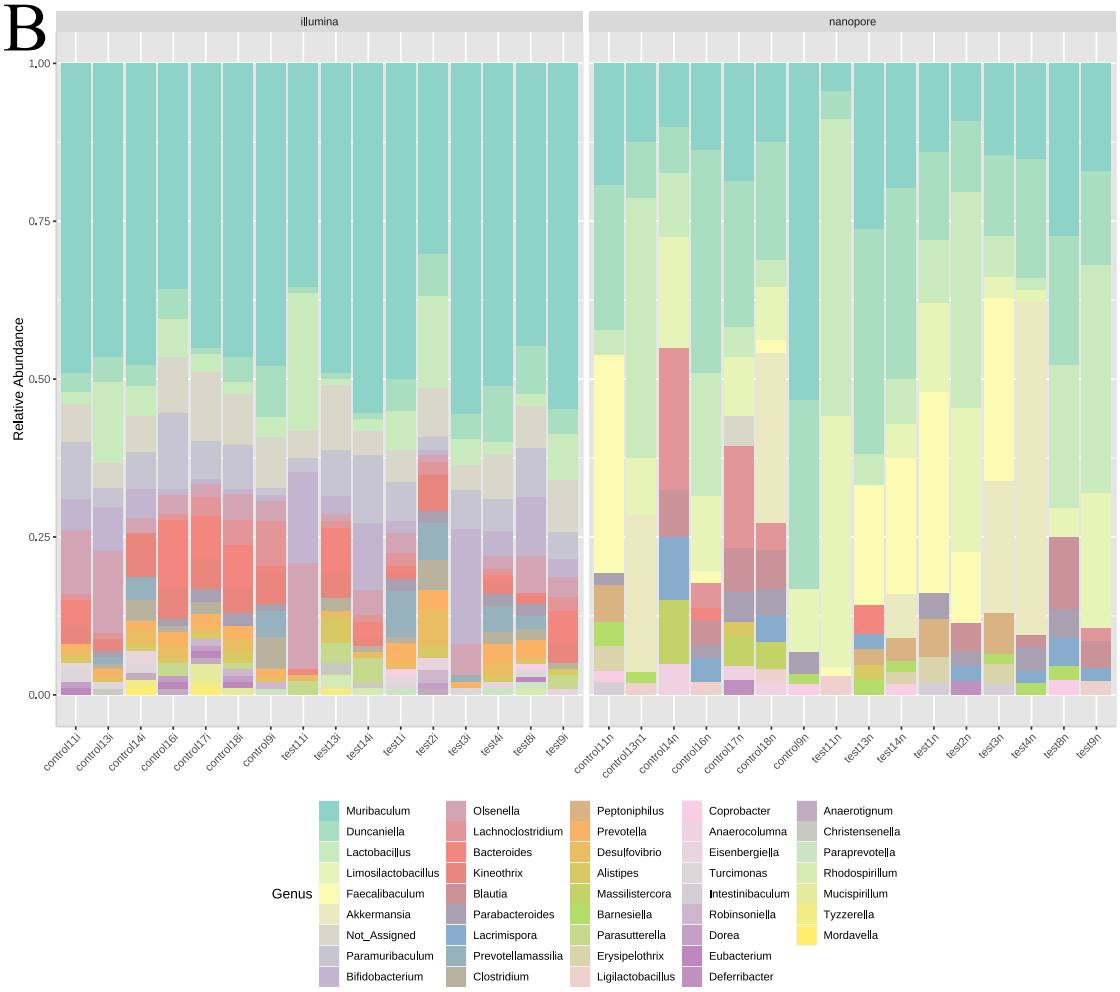
Not many reports have explored nanopore sequencing to microbial profile mice gut microbiomes. One study was performed by Shin and colleagues (2016), who evaluated the performance of nanopore and Illumina sequencing applied to the same samples. The results here showed significant differences with the microbial profile obtained by Shin and colleagues (Shin *et al*., 2016). Only genera *Lactobacillus* and *Akkermansia* were matched between this study and the report published by Shin and colleagues (2016). Moreover, *Faecalibaculum* was also detected by nanopore sequencing

but not abundantly (Shin *et al*., 2016). Since Shin and colleagues' report date to 2016, advances in the sequencing chemistry embedded in the flowcell apparatus (R7.3 *vs* current R9.4.1) and the overall improvement of the nanopore sequencing-related bioinformatics pipelines occurred, namely in the basecallers and that can probably explain the differences here detected (Shin *et al*., 2016).

As expected, Illumina sequencing retrieved a decreased average error-rate in comparison with nanopore sequencing (~0.01% *vs* ~9%) (nanopore sequencing: Table 12 and 13; Illumina sequencing: Table 16). Note that in this experiment, the Illumina sequencing bioinformatic pipeline was much more curated than the nanopore sequencing pipeline's, which was performed in order to achieve the fastest and easiest microbial profile data (Nilsson *et al*., 2019, Santos *et al*., 2020). As observed in Table 12, the targeted nanopore sequencing approach was divided in four sequencing runs to obtain the maximum output possible in terms of the number of sequenced reads (and to avoid carryover of DNA from previous sequencing runs). During these four nanopore sequencing runs, it was included one sample repetition to evaluate the accuracy within the same sequencing run (Table 13). All the nanopore sequencing quality parameters were identical between the four runs (Table 12), which represent evidence of the precision of the technique.

Nanopore and Illumina sequencing approaches retrieved different microbial profiles from the same infant mice gut faecal samples collected from mice exposed and non-exposed to ethanol *in utero* (Figure 26 and Appendix VIII for the sequencing approaches comparison between same experimental group).


A
illumina
nanopore
Relative Abundance
1.00
0.75
0.50
0.25
0.00
control11i
control13i
control14i
control16i
control17i
control18i
control9i
test11i
test13i
test14i
test1i
test2i
test3i
test4i
test8i
test9i
control11n
control13n1
control14n
control16n
control17n
control18n
control9n
test11n
test13n
test14n
test1n
test2n
test3n
test4n
test8n
test9n
Phylum
Bacteroidetes
Firmicutes
Actinobacteria
Verrucomicrobia
Proteobacteria
Not_Assigned
Deferribacteres


B
illumina
nanopore
Relative Abundance
1.00
0.75
0.50
0.25
0.00
control1i
control3i
control4i
control6i
control7i
control8i
control9i
test1i
test3i
test4i
test2i
test5i
test6i
test8i
test9i
control1n
control3n1
control4n
control6n
control7n
control8n
control9n
test1n
test3n
test4n
test2n
test5n
test6n
test8n
test9n
Genus
Muribaculum
Duncaniella
Lactobacillus
Limosilactobacillus
Faecalibaculum
Akkermansia
Not_Assigned
Paramuribaculum
Bifidobacterium
Olsenella
Lachnoclostridium
Bacteroides
Kineothrix
Blautia
Parabacteroides
Lacrimispora
Prevotellamassilia
Clostridium
Peptoniphilus
Prevotella
Desulfovibrio
Alistipes
Massilistercora
Barnesiella
Parasutterella
Erysipelothrix
Ligilactobacillus
Coprobacter
Anaerocolumna
Eisenbergiella
Turcimonas
Intestinibaculum
Robinsoniella
Dorea
Eubacterium
Deferribacter
Anaerotignum
Christensenella
Paraprevotella
Rhodospirillum
Mucispirillum
Tyzzerella
Mordavella

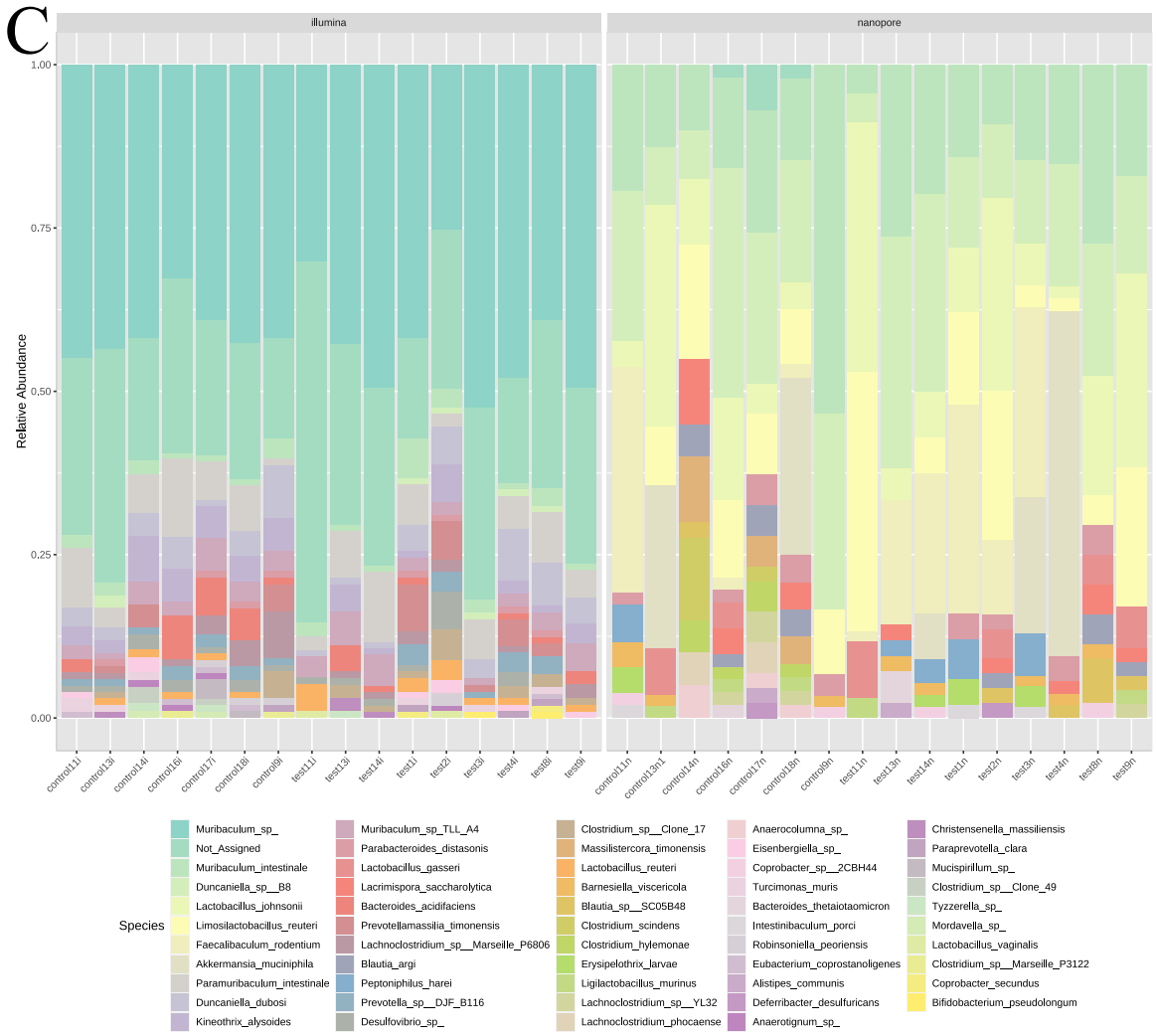


Figure 26 – Stacked bar plot displaying percentage abundances comparison between nanopore and Illumina sequencing approaches of assigned taxa at the phylum (A), genus (B), and species level (C). Right: log-transformed relative abundances, and species level (C).

Major differences were detected between nanopore and Illumina sequencing approaches. Illumina sequencing detected a higher number of reads assigned to *Bacteroidetes* in comparison to nanopore sequencing (70% *vs* 51%), and less reads assigned to *Firmicutes* (16% *vs* 50%). Additionally, no reads were assigned to *Verrucomicrobia* by Illumina sequencing (comparted to 9% detected by nanopore sequencing), and 10% of the reads were assigned to *Actinobacteria* where in the nanopore sequencing approach this phylum was not detected. An average of 24% (control group) and 27% (ethanol-exposed group) of the reads were not assigned to the species level in Illumina sequencing. In the microbial profile retrieved by the Illumina sequencing, the only genera found that matched the ones detected by Shin and colleagues (2016) were *Lactobacillus* and *Bacteroides*. One of the major differences was the detection of genera *Muribaculum* and *Paramuribaculum*, which were abundantly detected in this study and

absent from Shin and colleagues' report (Shin *et al*., 2016) (Figures 26B). Specifically, these genera were recently described and attributed new names (Lagkouvardos *et al*., 2016, Oren *et al*., 2016, Lagkouvardos *et* al., 2019, Park *et al*., 2021). Additionally, the *Muribaculaceae* family have been under scrutiny recently and more taxa were attributed to this family, such as *Duncaniella* gen., *D. muris*, *Paramuribaculum* gen., and *P. intestinale* (Langkouvardos *et al*., 2019). The newly described *Muribaculaceae* family was previously known as S24-7 (or simply as MIB (Salzman *et al*., 2002)), or 'Homeothermaceae' (Ormerod *et al*., 2016) and belongs to the *Bacteroidetes* phylum and *Bacteroidales* order, and despite their lack of phenotypical and functional description, they are commonly detected in animal guts (Seedorf *et al*., 2014, Ormerod *et al*., 2016, Thompson *et al*., 2017). Moreover, the representativeness of bacterial genomes in the reference databases has been exponentially increasing recently and its effect might as well be felt when comparing microbial profiling studies performed in different points in time (Uritskiy *et al*., 2018, Almeida *et al*., 2019, Breitwieser *et al*., 2019, Santos *et al*., 2020, Segerman, 2020). However, similarly to what was observed with nanopore sequencing some taxa were exclusively detected in one of the experimental groups, although showing relatively low abundance. The exceptions were the genus *Bifidobacterium* that were only abundant in the control group (Figure 23). Contrary to some previous independent reports performed in humans, the control group showed a higher prevalence of *Bifidobacterium* (Tsuruya *et al*., 2016, Dubinka *et al*., 2017, Le Roy *et al*., 2020). However, other reports observed similar *Bifidobacterium* abundances in the control/healthy subjects as reported here (Kirpich *et al*., 2008, Yan *et al*., 2011, Feng *et al*., 2015, Meroni *et al*., 2019, Jung *et al*., 2021). Since *Bifidobacterium* has a known potential to accumulate acetaldehyde under ethanol presence (Tsuruya *et al*., 2016) by encoding an acetaldehyde dehydrogenase (Jung *et al*., 2021), the results presented in this study reveal that this taxon may have a more complex phenotypic regulation under ethanol presence than previous thought (Tsuruya *et al*., 2016, Dubinkina *et al*., 2017, Tian *et al*., 2020). Moreover, the infant mice gut microbiota may also be unaffected by the ethanol exposure *in utero*, which can support, in part, the hypothesis of the 'sterile womb' (Lauder *et al*., 2016, Lim *et al*., 2018, de Goffau *et al*., 2019, Sanidad and Zeng, 2020). Other explanation is that the gut microbiota needs more time to developed and mature (Van Best *et al*., 2020) and so analysing infant mice gives a similar microbial profile that is actually a signal of a premature gut microbiota.

In terms of microbial composition retrieved by Illumina sequencing, both ethanol-exposed and control groups showed similar profiles at the phylum, genus, and species

level. Nonetheless, there were detected some minor differences in taxa exclusively found in each of the experimental groups. However, those taxa were only present in a limited number of samples and in a very low abundance. In the ethanol-exposed group, the exclusive taxa were *Rhizobium* sp. UB12 (n = 2, relative abundance < 1%), *Gemella* sp. (n = 2, relative abundance < 1%), and *B. pseudolongum* (n = 2, relative abundance < 1%); and in the control group, *Mordavella* sp (n = 3, relative abundance < 1%)., *M. massiliensis* (n = 4, relative abundance < 1%), *Marvinbryantia* sp. (n = 2, relative abundance < 1%), *Clostridium* sp. C5 48 (n = 2, relative abundance < 1%), *Clostridium* sp. Clone 49 (n = 3, relative abundance < 2%), and *Clostridium* sp. Marseille P3122 (n = 4, relative abundance < 1%). The presence of *Rhizobium* sp. may constitute a contamination or a misassignment as previously observed with other taxa, since it was observed with relatively low abundance. *Gemella* spp. have been previously collected from the human microbiome and potential representing pathogenic risk (Ulger-Toprak *et al*., 2010, Hung *et al*., 2014, Hikone *et al*., 2017). *Mordavella* sp. is a species that was previously isolated from a faecal sample obtained from a healthy volunteer (Ndongo *et al*., 2017) and its presence in laboratory mice gut samples has been linked to food contamination (Gurbanov *et al*., 2021). *M. massiliensis* is a species that have been detected in the human colon right side, however, its characterization is still insufficient to draw any conclusion about its importance to the human or mouse host (Mailhe *et al*., 2020). Recently *M. massiliensis* was associated with a dysbiotic state in an obesity mouse model (Zhang *et al*., 2020). *Marvinbryantia* spp. have been associated with brain lesions in mice, increasing in abundance after traumatic brain injury (Treangen *et al*., 2018). In humans, these taxa were observed with decreased abundance in faecal samples collected from hypertension patients and showed to be correlated with intestinal vitamin D, which is a risk factor for hypertension (Zuo *et al*., 2019).

Interestingly, only the nanopore sequencing detected *A. muciniphila* (in 3 ethanol-exposed and 2 control samples, with an average relative abundance of 5%). This observation is contrary with the observation of Shin and colleagues (2016) where they detected *A. muciniphila* using Illumina and nanopore sequencing. Nevertheless, *A. muciniphila* prevalence differed in the ethanol-exposed group (4-28%), which might implicate that ethanol imposes stress on this species. Thus, *A. muciniphila* may have a relevant role in ethanol addiction. This hypothesis is supported by other reports that stated that increasing ethanol habits decrease the abundance of *A. muciniphila* in humans (Ciocan *et al*., 2018, Gurwara *et al*., 2020). Additionally, the functional capacity of this species to produce/metabolize ethanol (Leclerq *et al*., 2020), and its role in the

inflammation of the gut barrier (Grander *et al*., 2018) makes the abovementioned hypothesis essential to be investigated. However, caution must be made when using mice models to study the host-microbe interactions of *A. muciniphila*, because despite being considered a commensal member of the human gastrointestinal tract, recent reports have supported the hypothesis that *A. muciniphila* beneficial properties and specific responses are context dependent and differ between human and mice (Ansaldo *et al*., 2019). Still, the nanopore sequencing platform basecalling error-rate should be taken into consideration because it may under or overrepresent the true prevalence of this and other affiliated taxa. These observations constitute a major difference between the nanopore and Illumina approaches and, thus, the capacity of each platform to truly assign the correct taxonomy of reads obtained from certain taxa, such as *A. muciniphila*.

Illumina sequencing assigned the majority of the reads to *Muribaculum* sp., more than twice of the reads that nanopore sequencing approach assigned to *M. intestinale* alone (42% *vs* 18%). In detail, Illumina sequencing assigned 3% of the total reads to *Muribaculaceae* family, 2% to *M. intestinale*, and 3% to *Muribaculum* sp. TLL-A4 (recently classified as *Muribaculum gordoncarteri* (Miyake *et al*., 2020b). The distinctive experimental and bioinformatic pipelines performed by Illumina and nanopore sequencing platforms can cause major discrepancies in terms of the microbial profile obtained while analysing the same samples. Here, Illumina sequencing overrepresented *Muribaculum* sp. while nanopore sequencing retrieved less than half of that number of reads to *M. intestinale*. Additionally, nanopore sequencing attributed much more reads to *M. intestinale* than Illumina sequencing (18% *vs* 4%). In the family *Muribaculaceae* there was observed disparities that reveals the taxonomical discrimination capacity of either sequencing platform. Indeed, nanopore sequencing was able to discriminate more species than Illumina sequencing, which concentrated the assignments as *Muribaculum* sp.. This observation supports the hypothesis that nanopore sequencing is more discriminatory than Illumina sequencing in assigning taxonomy identifiers to lower taxonomical ranks (Matsuo *et al*., 2021).

Only three species were detected in both sequencing platforms, *M. intestinale*, *P. distasonis*, and *D. muris* (Table 17). And two of these taxa, *M. intestinale* and *D. muris*, were assigned to more than 50% of the reads in both sequencing approaches irrespectively of the experimental group. Therefore, these species may be extrapolated as members of the murine gastrointestinal tract (Gu *et al*., 2013, Xiao *et al*., 2015, Wang *et al*., 2019, Yang *et al*., 2021). However, they were more abundantly detected in the ethanol-exposed group analysed by nanopore sequencing approach. Indeed, these two taxa were positively

correlated with the nanopore sequencing approach (Figure 27). Therefore, further investigation must be performed to study the relevance of these taxa to the host physiology. Since both species are members of the C57BL/6J mice gut (Miyake *et al*., 2020a, Miyake *et al*., 2020b) their increased presence in the ethanol-exposed group may be interpreted as a response to the ethanol stress. Despite the *M. intestinale* role in the mouse gut physiology is not totally understood, this anaerobic species can degrade galactose (Lagkouvardos *et al*., 2016) and it is associated with homoserine and serine metabolism (Snijders *et al*., 2016), thus showing potential to impact the mouse gut homeostasis (Dowden *et al*., 2020). *D. muris* have been negatively correlated with fat-diet in mice (Nakamura *et al*., 2021), and it was reported as a protection effector against colitis and gut epithelium injury in a mouse model of dextran sulfate sodium-induced injury (Chang *et al*., 2021).

Table 17 – Relative abundances of the common species detected by nanopore and Illumina sequencing approaches.

Species	**Illumina sequencing**		**Nanopore sequencing**	
	Control group	**Ethanol-exposed group**	**Control group**	**Ethanol-exposed group**
Duncaniella sp. B8 (*D. muris*)	0.3%	0.4%	9.5%	8.9%
Muribaculum intestinale	1.7%	0.4%	9.8%	8.1%
Parabacteroides distasonis	0.4%	0.4%	0.8%	0.7%

In Table 18 it is observed the most abundant species only detected by one of the sequencing platforms. A total of 36 species were exclusively detected by nanopore sequencing and 74 taxa by Illumina sequencing. Besides *D. muris*, *M. intestinale*, and *P. distasonis*, the remaining taxa exclusively detected by each sequencing approach may be considered members of the rare gut microbiota due to their low relative abundance (Reid *et al*., 2011, Lynch and Neufield, 2015, Lauder *et al*., 2016, Jousset *et al*., 2017, Lim *et al*., 2018, de Goffau *et al*., 2019, Sanidad and Zeng, 2020). However, the hypothesis of erroneous taxonomical assignment of those sequence reads cannot be ruled out, specifically in the nanopore sequencing approach. The differences detected in the reads assigned to *Muribaculum* spp. are most probably explained by the distinctive sequencing experimental and/or bioinformatic procedures. When both datasets were analysed together, no correlations were found between any taxa and ethanol-exposure or between the control group. All taxa displayed in Table 18 were correlated with their associated sequencing approach (*rho* > 0.6; p-value < 0.05; FDR < 0.05). The exception was *A.*

muciniphila that showed a moderate but significant correlation with the nanopore sequencing approach (*rho* = 0.4; p-value < 0.05; FDR < 0.05).

Table 18 – Most abundant taxa exclusively detected (> 2%) by one of the sequencing platforms, separated by the correspondent experimental group.

Sequencing platform	Control group (No. samples - mean percentual relative abundance)	Ethanol-exposed group (No. samples – mean relative abundance)
Illumina	*Muribaculum* sp. (7 – 40%)	*Muribaculum* sp. (9 – 42%)
	Paramuribaculum intestinale (7 – 6%)	*Paramuribaculum intestinale* (9 – 6%)
	Olsenella gen. (7 – 5%)	*Olsenella* gen. (9 – 5%)
	Kineothrix alysoides (7 – 4%)	*Duncaniella dubosi* (9 – 4%)
	Duncaniella dubosi (7 – 4%)	*Bifidobacterium* gen. (9 – 7%)
	Bacteroides acidifaciens (7 – 3%)	*Muribaculum* sp. TLL-A4 (9 – 3%)
	Bifidobacterium gen. (7 – 3%)	*Prevotellamassilia timonensis* (5 – 2%)
	Muribaculum sp. TLL-A4 (7 – 3%)	
	Lachnoclostridium sp. Marseille-P6806 (6 – 2%)	
Nanopore	*Lactobacillus johnsonnii* (6 – 8%)	*Lactobacillus johnsonnii* (9 – 9%)
	Akkermansia muciniphila (2 – 6%)	*Limosilactobacillus reuteri* (8 – 7%)
	Limosilactobacillus reuteri (6 – 5%)	*Faecalibaculum rodentium* (6 – 7%)
	Faecalibaculum rodentium (3 – 2%)	*Akkermansia muciniphila* (3 – 5%)

Prevotella and *Bifidobacterium* are commonly found in the mice gut microbiota (Jiang *et al*., 2020). Thus, the incapability of nanopore sequencing to detect these taxa represent a substantial handicap. Although present in a more than half of samples, taxa displayed in Table 18 were not abundantly represented in the assigned reads (relative abundance < 9%) – with the exception of *Muribaculum* spp. Hence, they may be considered as members of the rare gut biosphere, and the relevance of the rare gut biosphere for the host homeostasis must be studied taking into account every species' individual and combined effect (Reid *et al*., 2011, Lynch and Neufield, 2015, Lauder *et al*., 2016, Jousset *et al*., 2017, Lim *et al*., 2018, de Goffau *et al*., 2019, Sanidad and Zeng, 2020).

In terms of alpha-diversity, it was observed that both Shannon and Simpson coefficients accepted the null hypothesis of no significant differences at the phylum, genus, and species level within both experimental groups in the two sequencing platforms (Mann-Whitney U Test, p-value > 0.05). These observations support the evidence that ethanol exposure *in utero* does not affect the infant mice gut microbiota despite the abovementioned minor differences in the rare and exclusive taxa (Lauder *et al*., 2016, Lim *et al*., 2018, de Goffau *et al*., 2019, Sanidad and Zeng, 2020).

In terms of beta-diversity, the two sequencing platforms differed at every taxonomical rank and in both experimental groups (NMDS ordination, Bray-Curtis index, PERMANOVA R^2 = 0.5-0.7, ANOSIM R = 0.6-1.0, p-value < 0.05). However, the differences found at the phylum level in control and ethanol-exposed groups may be a false-positive (PERMDISP: p-vale > 0.05; ANOSIM: p-vale < 0.05), partially explained by the different abundance detected in the members of the *Bacteroidetes* phylum (*Muribaculum* spp.) previously discussed in this chapter.

In the control and ethanol-exposed groups two taxa were detected in more than 50% of the samples analysed by the two sequencing platforms: *M. intestinale* and *D. muris*. These two species showed a positive correlation with the nanopore sequencing approach (*rho* > 0.6: p-value < 0.05; FDR < 0.05), which may be considered a sequence bias associated with the nanopore sequencing approach (Figure 27). Moreover, there are more taxa that were correlated with one of the sequencing platforms (Appendix XIX). However, such correlations are in accordance with the previously diversity dissimilarities found within and between the microbial compositions obtained by Illumina and nanopore sequencing.

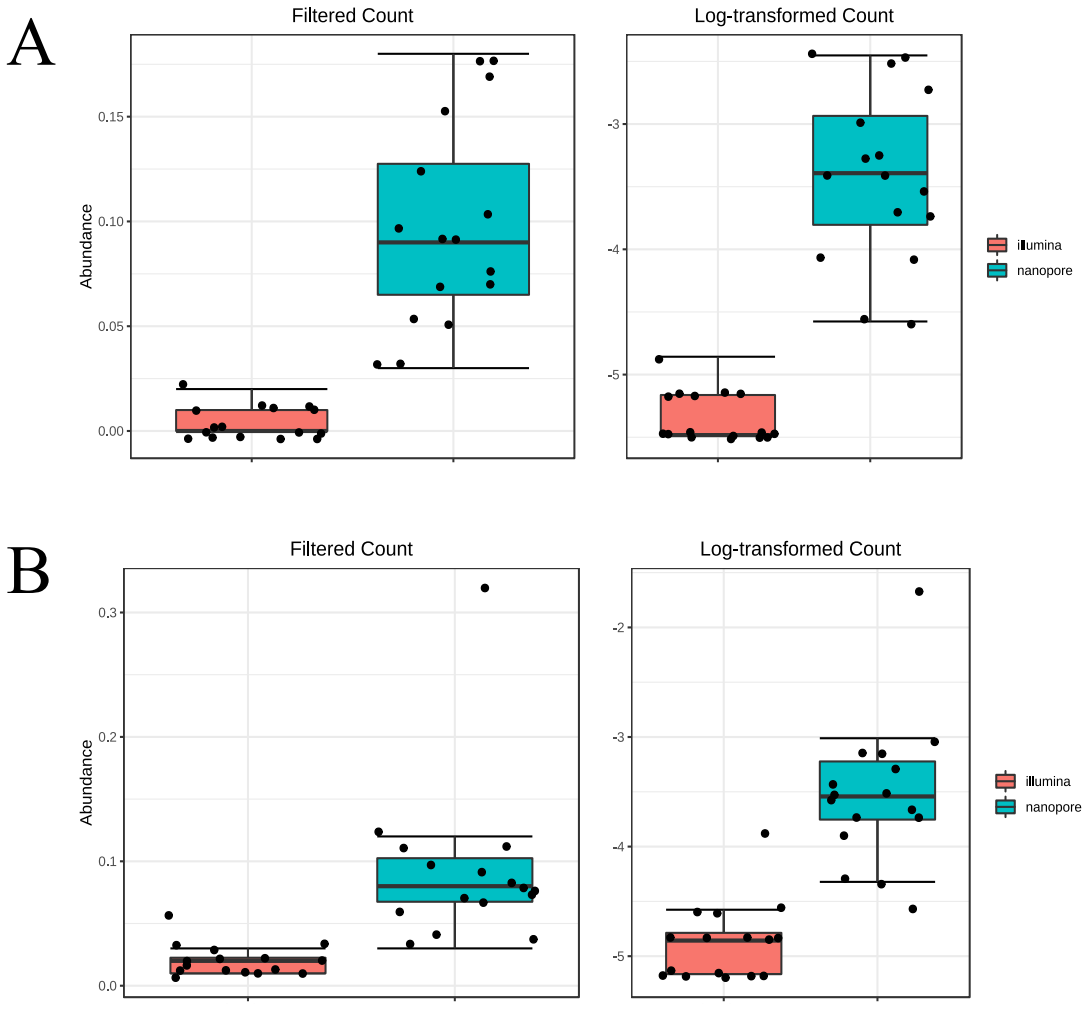
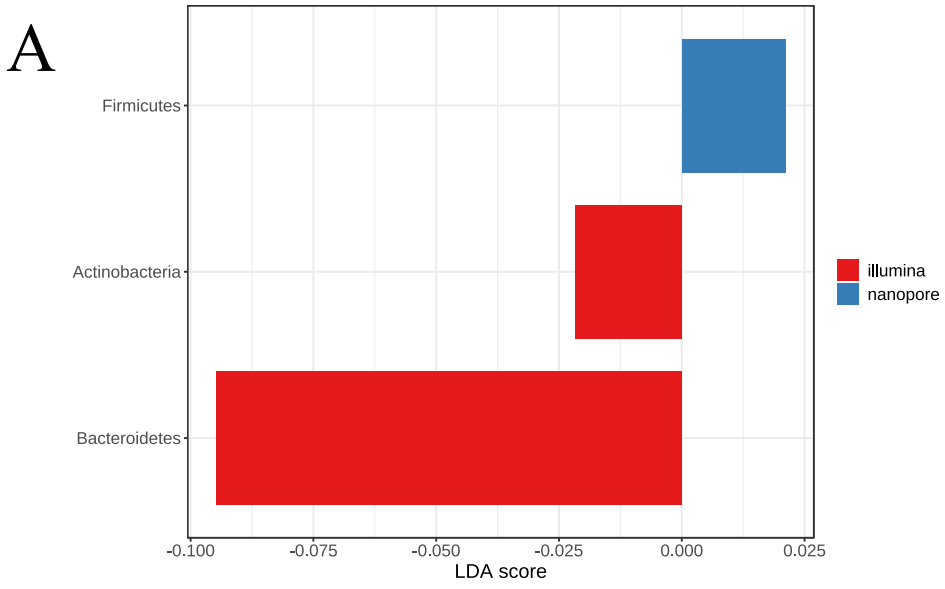


Figure 27 – Boxplot displaying the relative abundances of *Duncaniella muris* (A) and *M. intestinale* (B) (Spearman correlation coefficient > 0.8; p-value < 0.05; FDR < 0.05) in the two sequencing approaches. Left: raw relative abundances; Right: log-transformed relative abundances.

Additionally, and in accordance with the already mentioned differences found between experimental groups and sequencing platforms, there were also differences detected in the abundance profiles of several taxa at the phylum, genus, and species level (integer Log LDA score > 2; p-value < 0.05: FDR < 0.05) (Figure 28). Due to the overrepresentation of *Bacteroidetes* phylum in Illumina sequencing, this taxon may work as a marker (sequencing bias) when Illumina sequencing is performed. This observation can be explained by high number of reads assigned to *Muribaculum* spp. However, its biological meaning needs to be further investigated to clarify the hypothesis of sequencing bias since Illumina sequencing has handicaps when assigning taxa to lower taxonomical ranks.



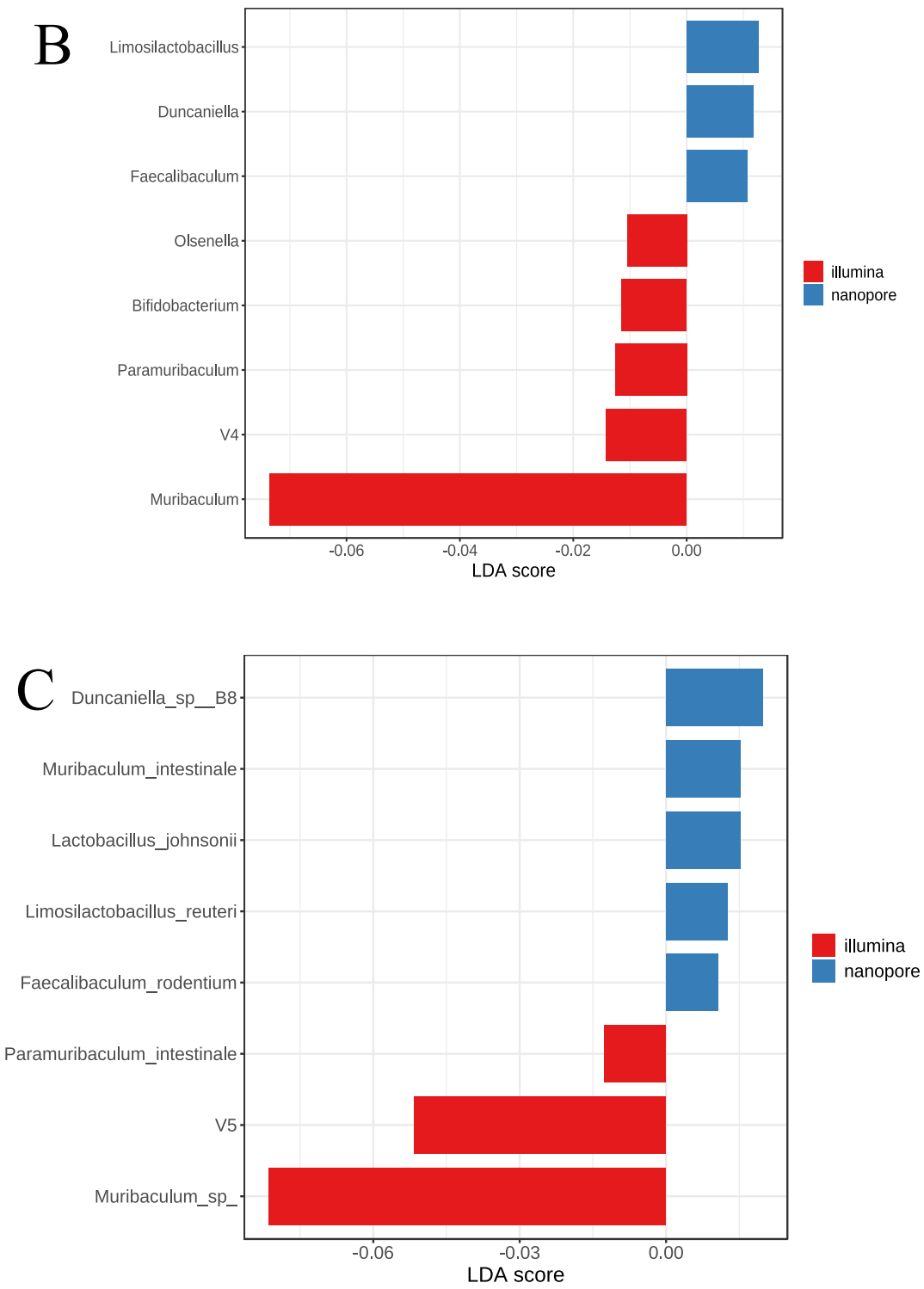


Figure 28 – Linear Discriminant Analysis (LDA) Effect Size (LEfSe) showing the relevant taxa significantly abundant and associated with nanopore or Illumina sequencing approaches. The length of each bar corresponds to the LDA score cutoff of 2 (A, phylum level), 1 (B, genus and C, species level); (p-value < 0.05). 'V4' and 'V5' correspond to the relative abundance attributed to the non-assigned reads at the mentioned taxonomical rank level.

In conclusion, there were a number of taxa correlated with the sequencing approaches, but such correlations may lack biological meaning due to their exclusiveness detection in one the sequencing approaches. A high number of taxa were exclusively obtained by each sequencing platform (Table 18). However, these sequencing platform-exclusive taxa were not abundantly observed (Table 18). Therefore, these taxa may be considered rare taxa, and their presence can potentially be explained due to either lab-related or bioinformatic-related experimental errors or limitations (Lynch and Neufield, 2015, Lauder *et al*., 2016, Lim *et al*., 2018, de Goffau *et al*., 2019, Sanidad and Zeng, 2020). Despite the differences found in the microbial profiles and relative abundances of both experimental groups by Illumina and nanopore sequencing, there were no significant differences detected in the diversity indices within samples. The beta-diversity differences were more robust at lower taxonomical ranks, which mean that the microbial profiles obtained by Illumina or nanopore sequencing approaches greatly differ, which was a typical discussed topic throughout this Chapter. Despite the differences between sequencing methodologies, *M. intestinale*, and *D. muris* were detected by both approaches, therefore validating their presence in the mice gut since early age. Both sequencing platforms validated the same conclusion in terms of the effect of the ethanol-exposure *in utero*. Ethanol exposure did not affect the microbial profiles although imposing a selective pressure to some species such as *L. jonhsonii* and *L. reuteri*, both detected by nanopore sequencing. Nevertheless, further studies must be performed to detail the effect of ethanol exposure to these taxa since no clear pattern was detected throughout this study.

Finally, the differences here detected between nanopore and Illumina sequencing can be explained by several factors, for instance the bioinformatic pipeline. The Illumina bioinformatic pipeline is more curated, including a copy number normalization approach, and a chimeric reads removal step, which were absent in the nanopore sequencing methodology. Additionally, the reference datasets used were not the same (NCBI database in WIMP *vs* DC-MEGABLAST) and therefore incongruences might appear. Still, since the goal is to decipher the gut microbiota from mice stool samples of unknown composition, it is challenging to assess the ground truth about the microbial composition and so to decide whether Illumina or nanopore sequencing is the appropriate approach.

Chapter 6 – Conclusions, novelty, and future work

6.1 Conclusions

The work presented here focused on the development, evaluation, and optimization of novel nanopore sequencing approaches to profile the gut microbiota of infant mice gut samples. The initial aim was to profile both the bacterial and fungal mice gut microbiota originating from infant mice exposed to ethanol *in utero*. Firstly, a fungal mock community comprised of five fungi species was used to optimize a targeted nanopore sequencing approach using the ITS genomic region as a genomic barcode. Then, twelve faecal samples were collected from infant mice exposed to ethanol *in utero* and targeted and untargeted nanopore sequencing approaches were tested to microbial profile such samples. In the targeted approaches, full-length 16S rRNA gene and 16S-ITS-23S region of the *rrn* operon were used to generate bacterial amplicons for sequencing. The untargeted metagenomic approach aimed at the sequencing of both fungal and bacterial nucleic acids. Finally, nanopore sequencing was performed to microbially profile nine faecal samples collected from guts recovered from infant mice exposed to ethanol *in utero* and of seven non-exposed control samples. The optimized 16S rRNA gene-targeted nanopore sequencing approach was compared with the sequencing output retrieved from Illumina sequencing targeting the conventional V3-V4 region of the 16S rRNA gene (using the same samples as the template). PCR ITS-targeted amplification failed to detect the fungal mice gut microbiota and consequently, only the nanopore sequencing bacterial profiling was performed using the full-length 16S rRNA gene as taxonomical genomic barcode. No gender-based differences were found throughout the statistical analyses of the results.

6.1.1 ITS-targeted nanopore sequencing

A pure culture of *C. glabrata* was used to generate ITS amplicons by PCR amplification and then a library was prepared for nanopore sequencing. The technology behind nanopore sequencing is quite sensitive and the performance of the nanopores rely on both good flowcell chemistry – depending on shelf life and storage conditions - and good library preparation – depending on the original DNA quantity and quality parameters during library preparation steps. Despite the low performance of the nanopores and an average basecalling error-rate of ~10%, the classification *centrifuge* score of the sequenced reads was above 300 in 80% of the reads and more than 90% of the reads were correctly assigned to *C. glabrata* showing the expected fragment size. Due to the high basecalling error-rate a large number of misclassified reads were obtained that

corresponded to sequencing 'taxonomical noise'. However, contamination inserted during sample handling or during library preparation steps cannot be excluded, as well as internal contamination present in the DNA extraction kits used – 'DNA kitome' (Olomu *et al*., 2020). Still, the low abundance of such misclassified reads (assigned taxa below 3% of relative abundance) is a good indicator of the precision of targeted-nanopore sequencing approaches used.

A fungal mock community comprising *C. lusitaniae*, *C. glabrata*, *C. parapsilosis*, *M. guilliermondii*, and *P. kudriavzevii* was then used to assess the precision and accuracy of targeted nanopore sequencing protocols based on the use of the fungal ITS genomic taxonomical marker. The goal was to check the capacity of nanopore sequencing to correctly assign the community composition while detecting the correct abundances of each species. Nanopore sequencing was able to assign the correct fungal mock community taxonomy, although overrepresenting species with shorter ITS genomic fragments (*C. lusitaniae*) and underrepresenting those with longer ITS genomic fragments (*C. glabrata*). Therefore, nanopore sequencing showed some yet undetermined bias that could be attributed to the nanopores preference for sequencing shorter fragments. In the fungal mock community analyses, the BugSeq pipeline was used since EPI2ME's WIMP generated unexpected erroneous taxonomical classifications. Additionally, BugSeq's *recentrifuge* classification score as low as 54 could still perform the correct taxonomical assignments and reasonably differentiate the different species' relative abundances. Therefore, it is absolutely crucial to carefully evaluate the bioinformatic pipeline choice, since they can profoundly impact the taxonomic and microbial abundances analyses in the samples under study. In conclusion, these preliminary results suggest that either upstream experimental steps, downstream bioinformatics analyses, or both, can introduce biases that are confounding the fungal relative abundances from biological samples.

6.1.2 Targeted-, and untargeted nanopore sequencing approaches reveal taxonomical sequencing biases

Twelve infant mice faecal samples were tested for microbial profiling. These samples originated from infant mice exposed to ethanol *in utero*. The aim was to develop and optimize a targeted nanopore sequencing approach to profile the gut microbiota of the infant mice gut. PCR was unable to amplify fungal ITS amplicons from the infant mice DNA templates, which means that fungal members were absent or below the PCR detection limit in the DNA templates we were able to prepare. Other reports have been

able to detect some fungi from mice gut samples (Aykut *et al*., 2019). However, since infant mice were studied in this dissertation, it is possible that the gut microbiota of the mice infants is not yet fully matured, and fungi members could eventual colonize later in life with the exposure to environmental microbes.

In terms of bacterial gut diversity, similarities were found between the microbial profiles detected between targeted and untargeted nanopore sequencing approaches (*i.e.*, targeting the 16S rRNA gene or the 16S-23S-ITS region *vs* shotgun/metagenomic). These approaches showed an expected sequencing performance in terms of the average basecalling error-rate, and genomic barcode fragment sizes detected. While the nanopore sequencing metagenomics approach yielded the longest reads, it was unable to assign taxonomical identifiers to more than 76% of the sequenced reads. Since it was an untargeted approach, the metagenomic nanopore sequencing approach also retrieved a high number of non-microbial taxonomical assignments, attributed either to the host or to an eventual contamination introduced during samples' handling, DNA extraction, and/or library preparation.

Different microbial profiles were detected depending on the 16S rRNA gene-, 16S-ITS-23S region-targeted, and metagenomic nanopore sequencing approaches attempted. Such differences may be probably explained by the variable representativeness of genomic sequences and whole genomes in the reference databases. Additionally, an unknown number of taxa may possess non-linked 16S-ITS-23 rRNA sequences which increases the taxonomical assignment challenge when this genomic target is used. Therefore, the diversity detected within samples were unique to each nanopore sequencing approach, and every nanopore sequencing approach retrieved a characteristic microbial profile. The discrepant microbial profiles also revealed correlations of certain taxa towards a particular nanopore sequencing approach at the phylum, genus, and species level. *Actinobacteria*, *Proteobacteria*, and *Bacteroidetes* were correlated with the metagenomic sequencing approach. At the species level, *C. flaccumfaciens*, *P. distasonis*, *B caecimuris*, *B. animalis*, *H. typhlonius*, *M. intestinale*, and *Lachnoclostridium* sp. were also correlated with the metagenomic approach. Additionally, *P. harei*, and *L. reuteri*, were correlated with the full-length 16S rRNA gene-targeted nanopore sequencing approach. Nevertheless, such correlations may be explained by the different classification capacity of the referred approaches, which can under and overrepresent several taxa. Additionally, some differences were observed with specific taxa. *A. muciniphila* was only detected by the 16S rRNA gene-targeted approach, and no *Firmicutes* members were assigned by the metagenomic approach, while being the only approach that detected

Proteobacteria members (represented by the 21% reads assigned to *C. flaccumfaciens*, which is a plant-pathogen (Osdaghi *et al*., 2022). The differences may be explained by the underrepresentation of 16S-ITS-23S sequences and whole genomes in the reference databases, therefore conferring increased reliability for the 16S rRNA gene genomic sequencing target. Thus, it is crucial that more good quality sequences are added to these databases to allow a more precise and trustful microbial profiling using different targets. Only *M. intestinale*, *Duncaniella muris*, *B. caecimuris*, and *B. animalis* were matched between WIMP and BugSeq classification algorithms when the 10 longest reads retrieved by the metagenomic approach were analysed. Such differences can be attributed to the superior curation of the BugSeq bioinformatic pipeline, which includes additional steps for contaminants' removal. As previously observed, BugSeq's *recentrifuge* classification algorithm could attribute a taxonomical identifier even with a score as low as 29.

These observations support the conclusion that it is necessary be very cautious when selecting the most appropriate bioinformatic pipeline for an analysis. Nevertheless, the longer reads that the metagenomic nanopore sequencing can retrieve represent a distinct advantage for answering genomic questions otherwise impossible to assess through sort-read Illumina sequencing, as exemplified by the Human genome assembly using ultra-long reads (Jain *et al*., 2018). One way to assess the limitations of the nanopore experimental approach and bioinformatic pipeline here used is by the use of a bacterial mock community (Fujiyoshi *et al*., 2020). This represents a handicap of this thesis since the analyses of a bacterial mock community comprised of the main colonizers of the mice gut would reveal the limitations of nanopore sequencing to detect the presence and the relative abundance of certain taxa (Fujiyoshi *et al*., 2020). Additionally, there are other experimental approaches available to perform nanopore sequencing like the 16S barcoding kit (Kai *et al*., 2019, Fujiyoshi *et al*., 2020). Accordingly to recent reports, there are several limitations in the experimental steps for the 16S rRNA gene analysis using nanopore sequencing, like DNA purification and primer sequence selection (Kai *et al*., 2019), and PCR cycle conditions (Fujiyoshi *et al*., 2020). Even though using the 16S kit from Oxford Nanopore Technologies might have helped, nanopore sequencing poses several challenges in microbial profiling such as DNA handle and treatment, high error-rate, different classification assignments depending on the classification algorithm used, and genomic target-dependent biases. In conclusion, our nanopore sequencing optimization attempts showed that biases may occur which can potentially under and overrepresent certain species in the microbial profiles detected. These nanopore

sequencing biases seem to be dependent on the sequencing approach performed, either targeted or untargeted.

6.1.3 Differences detected between the microbial profiles of infant mice gut obtained by nanopore and Illumina sequencing

After the nanopore sequencing microbial profiling optimization attempts based on the use of different genomic targets, the full-length 16S rRNA gene was chosen to microbially profile the infant mice gut microbiota when exposed to ethanol *in utero* in comparison with non-exposed controls. Between the three nanopore sequencing approaches tested, the one based on the full-length 16S rRNA gene seemed to be the one less affected by reference databases skewed representation, thus showing the most reliability. Next-generation Illumina sequencing was also employed to assess the sequencing depth, accuracy, and precision of the targeted nanopore sequencing results previously obtained from the infant mice gut microbial profiles from animals exposed and non-exposed to ethanol *in utero*. As expected, the Illumina short-read sequencing approach (targeting the V3-V4 region of the 16S rRNA gene) showed a lower error-rate than nanopore sequencing, although it presented an elevated taxonomical noise. Illumina sequencing was unable to assign species taxonomical identifiers to ~25% of the sequenced reads. In accordance with the microbial profiles detected by the targeted nanopore sequencing approach, Illumina sequencing showed identical microbial profiles between the ethanol-exposed and non-exposed samples. However, some compositional differences were observed. Using nanopore sequencing, *F. rodentium* was more abundantly detected in the control groups compared to the group exposed to ethanol, while not detected by Illumina sequencing. Using Illumina sequencing, *Bifidobacterium* gen. was only abundant in the control group, while it was absent in nanopore sequencing analysis. Other taxa were also more prone to be observed in the control group, but their relative abundance was low, and they were not evenly present in all samples. Only *P. distasonis*, *M. intestinale*, and *D. muris* were detected by both sequencing platforms. Despite the *Muribaculum* spp. incongruences detected (overrepresentation of *Bacteroidetes* members using the Illumina sequencing approach), that are most probably explained by the specific experimental and bioinformatic pipeline setup employed by each sequencing platform, these exclusive taxa discrepancies were not of abundantly prevalent organisms and should be considered members of the rare taxa. These observations support the fact that they are part of the rare biosphere, whose relevance needs to be further studied considering the mice core microbiota as well. The relevance

of the rare gut biosphere members has been unfairly overlooked since their diversity likely enhances the resiliency of the ecosystem (Lynch and Neufield, 2015, Jousset *et al*., 2017). Therefore, a transcriptomics analysis should be performed in order to uncover the potential functionality of these underrepresented species and to define the abundance and phenotypical thresholds upon which their impact is significant (Hausmann *et al*., 2019, Sato *et al*., 2019), as exemplified in studies with asthma patients (Sbihi *et al*., 2019). Additionally, the relevance of these taxa was evidenced here by the interconnected correlation network observed, thus showing that the importance of rare taxa may also be seen at the population level. Nevertheless, *Blautia*, *Muribaculum*, *Duncaniella*, *Lactobacillus*, and *Limosilactobacillus*, also demonstrated a correlation network between each other, and since they were present evenly in the samples, they may comprise part of the infant mice gut core microbiota. These species need be studied individually to fully characterize their function in the mice and human gastrointestinal tract, as exemplified by Ju and colleagues (2019), who performed *in vitro* and *in vivo* studies to describe *Parasutturella* colonization pattern, mediation of gut bile acids, and related immunological effects. Although no major differences were reported between the experimental groups, a complex correlation network was observed. Such a network may represent evidence of an ethanol selective effect on particular genera *Anaerocolumna*, *Massilistercora*, *Lachnoclostridium*, *Lacrimispora*, and *Blautia* in this instance. Nonetheless, further investigation must be performed to describe the biological meaning of these observed correlations (Hanage, 2014).

In conclusion, nanopore sequencing and Illumina sequencing did not identify the same microbial composition from the same infant mice gut samples. Nanopore sequencing overrepresented *Firmicutes* members while Illumina sequencing overrepresented *Bacteroides* members. Additionally, *A. muciniphila* was only detected by nanopore sequencing; while *Bifidobacterium*, *Prevotella*, and *Olsenella* were only detected by nanopore and Illumina sequencing, respectively. Due to the particular relevance of these taxa to the host, the inability of the sequencing approaches used to reliably detect them constitutes a clear handicap of these technologies. This poses questions in the use of these sequencing technologies, either Illumina or nanopore sequencing, to analyse the microbial profiles of gut environments. The recent field of systems biology may help better define and consequently improve these methodological limitations, taking advantage of multi-omics big datasets which rely on the cross-validation of results obtained by each approach (Qian *et al*., 2020). Nanopore and Illumina sequencing showed different infant mice gut microbial profiles when analysing

the same samples. These differences can be explained by several factors, for instance the bioinformatic pipeline. The Illumina bioinformatic pipeline applied here is more complete, including a copy number normalization approach, and a chimeric reads removal step, which were absent in the nanopore sequencing methodology. For instance, nanopore basecalling was performed with the *fast basecalling* mode using *guppy* due to technical constraints, however the *high accuracy* mode could have improved the error-rate and retrieved better results overall. Recently, Egeter and colleagues (2022) applied a polished pipeline to analyse nanopore sequence data. Their bioinformatics pipeline could represent a way to increase the reliability of our approach allowing a better comparison with the Illumina results. Briefly, it included a i) clustering step with *isonclust* and *racon*; ii) clustered reads were then clustered at 99% sequence identity with *cd-hit*, and a centroid (representative of each read) was selected; iii) primer trimming and length filtering of the polished centroids; iv) alignment of the polished reads using BLAST; and v) taxonomy assignment using a lowest common ancestor approach (Egeter *et al*., 2022). Specifically, the use of other alignment tools rather than *centrifuge* used by WIMP could have improved the taxonomical assignment of our data, for instance applying a pipeline performing alignments with LAST followed by MEGAN (Bağcı *et al*., 2019). Using LAST has advantages when analysing longer reads data because these data require frame-shift aware alignment tools due to frame-shift and break translation-based alignments caused by insertions or deletions (Bağcı *et al*., 2019).

Contrary to other reports our results differed between platforms (Fujiyoshi *et al*., 2020). So, the PCR conditions, library reparation (e.g. by using the 16S barcoding kit) and the nanopore bioinformatic pipeline still need further optimization to give more confidence into the results obtained (Bağcı *et al*., 2019, Kai *et al*., 2019, Fujiyoshi *et al*., 2020). Additionally, the reference datasets used were not the same (WIMP *vs* DC-MEGABLAST) and therefore incongruences might appear. Still, since the goal is to decipher the gut microbiota from mice stool samples of unknown composition, it is challenging to assess the ground truth about and so to decide whether Illumina or nanopore sequencing is the appropriate approach.

However, based on both sequencing strategies, and despite differences attributed to experimental and bioinformatical protocols and parameters, ethanol does not seem to induce any relevant effect on the diversity of infant mice gut microbiota. Although some differences were detected between ethanol-exposed and non-exposed mice, these differences were not significant and may not be relevant to the infant mice physiology.

When the two independent sets of faecal samples collected from infant mice exposed to ethanol *in utero* were compared by nanopore sequencing, different microbial profiles were detected at the individual level. Therefore, it can be concluded that individual mice possess unique microbial profiles while maintaining the same overall diversity. *L. johnsonni*, *Barnesiella viscericola*, *P. distasonis*, *P. harei*, and *A. intestinale*, showed different representativeness between the two set of samples. Additionally, *M. intestinale*, *Duncaniella* spp. *F. rodentium*, *L. reuteri*, and *L. johnsonii*, may be members of the infant mice core microbiota, particularly in mice exposed to ethanol *in utero*.

To summarize, the results here presented and discussed pose a number of questions regarding the use of nanopore and Illumina sequencing for the analysis of the gut microbiota of infant mice. Each platform showed limitations, either in detecting taxa that are members of the mice core gut microbiota or by showing sequencing biases towards the genomic barcode used for sequencing. Moreover, the choice of the appropriate bioinformatic pipeline is not trivial and should be careful evaluated to avoid inaccurate microbial composition and abundance results. Therefore, the standardization of experimental methodologies, during mice husbandry (mice model, diet, and ethanol consumption regimen, *etc*), library preparation (sequencing equipment/technology, specific library preparation protocols, and third-party reagents, *etc*), and in the choice of bioinformatic analysis performed is of paramount importance. Our results also show that ethanol exposure in *utero* does not appear to affect the diversity of the infant mice microbiota in the model used, although it may potentially be exerting a selective effect on particular taxa. Due to the correlation network found between members of the infant mice gut microbiota, these taxa must be individually studied to assess their relevance to the host and their adaptive phenotypes when ethanol is present.

6.2 Novelty

Nanopore sequencing is a novel sequencing methodology that claims to revolutionize the genomics field (Leggett *et al*., 2017). However, current nanopore approaches have not been standardized which may confound the biological interpretations of high-throughput sequencing datasets. Additionally, nanopore sequencing still present a high error-rate compared to other more mature sequencing technologies, such as Illumina sequencing. These technological handicaps create the need to study and optimize nanopore sequencing approaches to answer biological questions, such as interrogations of the microbial composition and abundance of clinical and environmental samples. In this dissertation, four nanopore sequencing approaches were designed and attempted to

optimize fungal and bacterial profiling sequencing methodologies. Three targeted methods based on the fungal ITS gene, the bacterial 16S rRNA gene, and 16S-ITS-23S *rrn* operon region, and one untargeted shotgun/metagenomic approach were tested. Lastly, 16S rRNA gene-targeted nanopore sequencing was compared with the current 'gold standard' Illumina sequencing approach to interrogate the gut microbial profile of infant mice exposed to ethanol *in utero*.

Detailed knowledge about the interactions between gut microbes and the developing nervous system is still scarce. FASD represent a clinically relevant set of conditions with cumbersome diagnostic and treatment. In this dissertation, the microbial profiles of infant mice gut exposed to ethanol *in utero* were analysed through third-generation Illumina and optimized next-generation nanopore sequencing technologies. The fungal (although not detected) and bacterial microbial profiles here obtained through nanopore and Illumina sequencing represent a technological and biological advancement towards a better comprehension of the microbial landscape in FASD at early post-natal periods.

6.3 Future work and perspectives

Despite much recent work focusing on improving our understanding of the human and mice gut microbiome, their fungal components are still not well understood. Therefore, further optimizations need to be performed to develop better detection and sequencing methodologies to help decipher the composition and relevance of the gut mycobiome.

In order to decrease the error-rate and other handicaps related to nanopore sequencing, a better curated bioinformatic pipeline must be developed and employed a greater number of biological samples. Additionally, nanopore sequencing should be also complemented with other sequencing technologies, such as Illumina and/or PacBio, to aid completion of potential sequencing gaps and reduce errors.

A complex network of microbial correlations was detected, and this needs to be assessed to characterize and fully describe their individual and collective relevance to the host. These host-microbe interactions' studies would benefit from a standardized mice *in vivo* experimental setup, thus eliminating any unwanted experimental bias. Moreover, microbe-microbe association's studies may also be performed to delineate, highlight, and guide further *in vivo* studies at the population-level inside the gastrointestinal tract. Since *F. rodentium* and *A. muciniphila* were observed with different relative abundances between the experimental groups, it would be interesting to use Germ-free mice models

to analyse the host-microbe interactions occurring in gut when mice are colonized with these species. Additionally, alcohol-use disorder mice models can be studied with a diet supplemented with *F. rodentium* and/or *A. muciniphila* to assess both the microbe-microbe and host-microbe interactions towards disease amelioration.

Finally, a fully catalogued and characterized fungal and bacterial microbial profile in FASD since early birth may be explored, challenged, and serve as guide, in pre-clinical studies.

One relevant question is the effect of particular species in gut homeostasis in FASD cases. Due to the limitations associated with the study of host gut microbial colonization and its effect on host development during gestation and at early ages, a new avenue of research could exploit *in vitro gut* models (gut-on-a-chip). Recently, Grassart and colleagues (2019), observed that *Shigella* exploits the architecture and mechanical forces of the gut to maximize infectivity, using the crypt-like structure of the intestine for enhanced invasion. These observations were only possible due to the use of an intestine-chip technology (Grassart *et al*., 2019). This type of technology could be exploited to expand beyond more traditional *in vitro* studies (Ingber, 2020) and may represent a novel avenue to validate results obtained by conventional approaches before starting clinical trials or *in vivo* studies (Ingber, 2022). For instance, it would be interesting to evaluate the phenotypic arsenal of the rare taxa (namely *A. muciniphila*) found in this thesis when ethanol is present in the gut lumen.

The detailed genetic and metabolic characterization of these microbial adaptive responses to ethanol could highlight potential routes of research in terms of development of preventive measures, such as drugs, pre or probiotics, to help mitigate FASD impairments. The rational is that ethanol presence in the gut will impact the gut barrier, creating conditions for leakage. This leaky gut condition will hamper the host homeostasis and change the balance of gut metabolites (e.g. LPS and cytokines) in circulation which can reach the brain through the vagus nerve. Therefore, it is necessary to assess the gut "metabolome" during ethanol exposure to test the effect on neurological conditions and to validate the endotoxin hypothesis of neurodegeneration (Brown, 2019). The identification of patterns related to target metabolites will pave the way for the predictive relevance of such biomarkers of alcohol-related diseases.

Bibliography

Aagaard, Kjersti, et al. "The placenta harbors a unique microbiome." *Science translational medicine* 6.237 (2014): 237ra65-237ra65.

Abdel-Haq, Reem, et al. "Microbiome–microglia connections via the gut–brain axis." *Journal of Experimental Medicine* 216.1 (2019): 41-59.

Abele-Horn, Marianne, et al. "High-density vaginal *Ureaplasma urealyticum* colonization as a risk factor for chorioamnionitis and preterm delivery." *Acta obstetricia et gynecologica Scandinavica* 79.11 (2000): 973-978.

Adachi, Yukito, et al. "Antibiotics prevent liver injury in rats following long-term exposure to ethanol." *Gastroenterology*108.1 (1995): 218-224.

Agostoni, E., et al. "Functional and histological studies of the vagus nerve and its branches to the heart, lungs and abdominal viscera in the cat." *The Journal of physiology* 135.1 (1957): 182.

Ahn, Hyeonju, et al. "Enhanced symbiotic characteristics in bacterial genomes with the disruption of rRNA operon." *Biology*9.12 (2020): 440.

Ahn, Jiyoung, et al. "Human gut microbiome and risk for colorectal cancer." *Journal of the National Cancer Institute*105.24 (2013): 1907-1911.

Akira, Shizuo, and Kiyoshi Takeda. "Toll-like receptor signalling." *Nature reviews immunology* 4.7 (2004): 499-511.

Alfonso-Loeches, Silvia, et al. "Ethanol-induced TLR4/NLRP3 neuroinflammatory response in microglial cells promotes leukocyte infiltration across the BBB." *Neurochemical research* 41.1-2 (2016): 193-209.

Alfonso-Loeches, Silvia, et al. "Pivotal role of TLR4 receptors in ethanol-induced neuroinflammation and brain damage." *Journal of Neuroscience* 30.24 (2010): 8285-8295.

Almeida, Alexandre, et al. "A new genomic blueprint of the human gut microbiota." *Nature* 568.7753 (2019): 499-504.

Almeida, Alexandre, et al. "A unified catalog of 204,938 reference genomes from the human gut microbiome." *Nature biotechnology* 39.1 (2021): 105-114.

Almeida, Laura, et al. "Murine models for the study of fetal ethanol spectrum disorders: an overview." *Frontiers in Pediatrics* 8 (2020).

Alneberg, Johannes, et al. "Binning metagenomic contigs by coverage and composition." *Nature methods* 11.11 (2014): 1144-1146.

Ames, Sasha K., et al. "Using populations of human and microbial genomes for organism detection in metagenomes." *Genome research* 25.7 (2015): 1056-1067.

Ammer-Herrmenau, Christoph, et al. "Comprehensive wet-bench and bioinformatics workflow for complex microbiota using Oxford Nanopore technologies." *Msystems* 6.4 (2021): e00750-21.

Anand, Rahul J., et al. "The role of the intestinal barrier in the pathogenesis of necrotizing enterocolitis." *Shock* 27.2 (2007): 124-133.

Anderson, Marti J. "A new method for non-parametric multivariate analysis of variance." *Austral ecology* 26.1 (2001): 32-46.

Anderson, Marti J. "Distance-based tests for homogeneity of multivariate dispersions." *Biometrics* 62.1 (2006): 245-253.

Angly, Florent E., et al. "CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction." *Microbiome* 2.1 (2014): 1-13.

Ansaldo, Eduard, et al. "*Akkermansia muciniphila* induces intestinal adaptive immune responses during homeostasis." *Science* 364.6446 (2019): 1179-1184.

Arrieta, Marie-Claire, et al. "Associations between infant fungal and bacterial dysbiosis and childhood atopic wheeze in a nonindustrialized setting." *Journal of Allergy and Clinical Immunology* 142.2 (2018): 424-434.

Arthur, Janelle C., et al. "Intestinal inflammation targets cancer-inducing activity of the microbiota." *science* 338.6103 (2012): 120-123.

Ashikawa, Sae, et al. "Rapid identification of pathogens from positive blood culture bottles with the MinION nanopore sequencer." *Journal of Medical Microbiology* 67.11 (2018): 1589-1595.

Atarashi, Koji, et al. "Induction of colonic regulatory T cells by indigenous *Clostridium* species." *Science* 331.6015 (2011): 337-341.

Atarashi, Koji, et al. "Treg induction by a rationally selected mixture of *Clostridia* strains from the human microbiota." *Nature*500.7461 (2013): 232-236.

Auchtung, Thomas A., et al. "Investigating colonization of the healthy adult gastrointestinal tract by fungi." *MSphere* 3.2 (2018): e00092-18.

Aujoulat, Fabien, et al. "*Peptoniphilus nemausensis* sp. nov. A new Gram-positive anaerobic coccus isolated from human clinical samples, an emendated description of the genus *Peptoniphilus* and an evaluation of the taxonomic status of *Peptoniphilus* species with not validly published names." *Systematic and applied microbiology* 44.5 (2021): 126235.

Aykut, Berk, et al. "The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL." *Nature* 574.7777 (2019): 264-267.

Bağcı, Caner, et al. "Introduction to the analysis of environmental sequences: metagenomics with MEGAN." *Evolutionary Genomics*. Humana, New York, NY, 2019. 591-604.

Bahram, Mohammad, et al. "Newly designed 16S rRNA metabarcoding primers amplify diverse and novel archaeal taxa from the environment." *Environmental microbiology reports* 11.4 (2019): 487-494.

Bajaj, Jasmohan S., et al. "Altered profile of human gut microbiome is associated with cirrhosis and its complications." *Journal of hepatology* 60.5 (2014): 940-947.

Bajaj, Jasmohan S., et al. "Continued ethanol misuse in human cirrhosis is associated with an impaired gut–liver axis." *Ethanolism: Clinical and Experimental Research* 41.11 (2017): 1857-1865.

Bajaj, Jasmohan S., et al. "Fungal dysbiosis in cirrhosis." *Gut*67.6 (2018): 1146-1154.

Bala, Shashi, et al. "Acute binge drinking increases serum endotoxin and bacterial DNA levels in healthy individuals." *PloS one* 9.5 (2014): e96864.

Balvočiūtė, Monika, and Daniel H. Huson. "SILVA, RDP, Greengenes, NCBI and OTT—how do these taxonomies compare?." *BMC genomics* 18.2 (2017): 1-8.

Barcenilla, Adela, et al. "Phylogenetic relationships of butyrate-producing bacteria from the human gut." *Applied and environmental microbiology* 66.4 (2000): 1654-1661.

Barr, Tasha, et al. "Concurrent gut transcriptome and microbiota profiling following chronic ethanol consumption in nonhuman primates." *Gut microbes* 9.4 (2018): 338-356.

Baxter, Nielson T., et al. "Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions." *Genome medicine* 8.1 (2016): 1-10.

Bell, J. S., et al. "Invited Review: From nose to gut–the role of the microbiome in neurological disease." *Neuropathology and applied neurobiology* 45.3 (2019): 195-215.

Bercik, Premysl, et al. "The intestinal microbiota affect central levels of brain-derived neurotropic factor and behavior in mice." *Gastroenterology* 141.2 (2011): 599-609.

Berer, Kerstin, et al. "Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination." *Nature* 479.7374 (2011): 538-541.

Beresford-Jones, Benjamin S., et al. "The Mouse Gastrointestinal Bacteria Catalogue enables translation between the mouse and human gut microbiotas via functional mapping." *Cell host & microbe* (2021).

Biagi, Elena, et al. "Gut microbiota and extreme longevity." *Current Biology* 26.11 (2016): 1480-1485.

Biagi, Elena, et al. "The gut microbiota of centenarians: signatures of longevity in the gut microbiota profile." *Mechanisms of ageing and development* 165 (2017): 180-184.

Bik, Holly M. "Phinch: an interactive, exploratory data visualization framework for–Omic datasets." *BioRxiv* (2014): 009944.

Bishehsari, Faraz, et al. "Ethanol and gut-derived inflammation." *Ethanol research: current reviews* 38.2 (2017): 163.

Bjørkhaug, Steinar Traae, et al. "Characterization of gut microbiota composition and functions in patients with chronic ethanol overconsumption." *Gut microbes* 10.6 (2019): 663-675.

Blaser, Martin J., et al. "Lessons learned from the prenatal microbiome controversy." *Microbiome* 9.1 (2021): 1-7.

Bockhorst, K. H., et al. "Early postnatal development of rat brain: *in vivo* diffusion tensor imaging." *Journal of neuroscience research* 86.7 (2008): 1520-1528.

Boer, Wietse de, et al. "Living in a fungal world: impact of fungi on soil bacterial niche development." *FEMS microbiology reviews* 29.4 (2005): 795-811.

Boix-Amorós, Alba, et al. "Mycobiome profiles in breast milk from healthy women depend on mode of delivery, geographic location, and interaction with bacteria." *Applied and environmental microbiology* 85.9 (2019): e02994-18.

Borges, Francis M., et al. "Fungal diversity of human gut microbiota among eutrophic, overweight, and obese individuals based on aerobic culture-dependent approach." *Current microbiology* 75.6 (2018): 726-735.

Bowers, Robert M., et al. "Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea." *Nature biotechnology* 35.8 (2017): 725-731.

Boža, Vladimír, Broňa Brejová, and Tomáš Vinař. "DeepNano: deep recurrent neural networks for base calling in MinION nanopore reads." *PloS one* 12.6 (2017): e0178751.

Breitwieser, Florian P., and Steven L. Salzberg. "Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification." *Bioinformatics* 36.4 (2020): 1303-1304.

Breitwieser, Florian P., Jennifer Lu, and Steven L. Salzberg. "A review of methods and databases for metagenomic classification and assembly." *Briefings in bioinformatics* 20.4 (2019): 1125-1136.

Brennan, Caitlin A., and Wendy S. Garrett. "*Fusobacterium nucleatum*—symbiont, opportunist and oncobacterium." *Nature Reviews Microbiology* 17.3 (2019): 156-166.

Brewer, Tess E., et al. "Unlinked rRNA genes are widespread among bacteria and archaea." *The ISME journal* 14.2 (2020): 597-608.

Brown, Bonnie L., et al. "MinION™ nanopore sequencing of environmental metagenomes: a synthetic approach." *Gigascience* 6.3 (2017): gix007.

Brown, Christopher T., et al. "Measurement of bacterial replication rates in microbial communities." *Nature biotechnology* 34.12 (2016): 1256-1263.

Brown, Clive G., and James Clarke. "Nanopore development at Oxford nanopore." *Nature biotechnology* 34.8 (2016): 810-811.

Brown, Guy C. "The endotoxin hypothesis of neurodegeneration." *Journal of neuroinflammation* 16.1 (2019): 1-10.

Brown, K., et al. "Bloodstream infections due to *Peptoniphilus* spp.: report of 15 cases." *Clinical Microbiology and Infection*20.11 (2014): O857-O860.

Browne, Hilary P., et al. "Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation." *Nature* 533.7604 (2016): 543-546.

Browne, Patrick Denis, et al. "GC bias affects genomic and metagenomic reconstructions, underrepresenting GC-poor organisms." *GigaScience* 9.2 (2020): giaa008.

Buffington, Shelly A., et al. "Microbial reconstitution reverses maternal diet-induced social and synaptic deficits in offspring." *Cell* 165.7 (2016): 1762-1775.

Bukin, Yu S., et al. "The effect of 16S rRNA region choice on bacterial community metabarcoding results." *Scientific Data* 6.1 (2019): 1-14.

Bull-Otterson, Lara, et al. "Metagenomic analyses of ethanol induced pathogenic alterations in the intestinal microbiome and the effect of Lactobacillus rhamnosus GG treatment." *PloS one*8.1 (2013): e53028.

Burokas, Aurelijus, et al. "Targeting the microbiota-gut-brain axis: prebiotics have anxiolytic and antidepressant-like effects and reverse the impact of chronic stress in mice." *Biological psychiatry* 82.7 (2017): 472-487.

Cai, Shaofang, et al. "Ethanol drinking and the risk of colorectal cancer death." *European Journal of Cancer Prevention* 23.6 (2014): 532-539.

Cai, Yuan-Yuan, et al. "Integrated metagenomics identifies a crucial role for trimethylamine-producing *Lachnoclostridium* in promoting atherosclerosis." *NPJ biofilms and microbiomes* 8.1 (2022): 1-12.

Calus, Szymon T., Umer Z. Ijaz, and Ameet J. Pinto. "NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform." *Gigascience* 7.12 (2018): giy140.

Camacho, Christiam, et al. "BLAST+: architecture and applications." *BMC bioinformatics* 10.1 (2009): 1-9.

Cao, Zhifang, et al. "Ubiquitin ligase TRIM62 regulates CARD9-mediated anti-fungal immunity and intestinal inflammation." *Immunity* 43.4 (2015): 715-726.

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

Castaño, Carles, et al. "Optimized metabarcoding with Pacific biosciences enables semi-quantitative analysis of fungal communities." *New Phytologist* 228.3 (2020).

Castellarin, Mauro, et al. "*Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma." *Genome research*22.2 (2012): 299-306.

Cavaillon, Jean-Marc, and Sandra Legout. "Centenary of the death of Elie Metchnikoff: a visionary and an outstanding team leader." *Microbes and Infection* 18.10 (2016): 577-594.

Cavalieri, Duccio, et al. "Genomic and phenotypic variation in morphogenetic networks of two Candida albicans isolates subtends their different pathogenic potential." *Frontiers in immunology* 8 (2018): 1997.

Cericco, M., et al. "*Corynebacterium xerosis* endocarditis associated with alcoholic cirrhosis." *Gastroenterologie clinique et biologique* 20.5 (1996): 514-514.

Ceylani, Taha, et al. "The effects of repeated antibiotic administration to juvenile BALB/c mice on the microbiota status and animal behavior at the adult age." *Heliyon* 4.6 (2018): e00644.

Chadwick, Derek J., and Jamie A. Goode, eds. *Acetaldehyde-related Pathology: Bridging the Trans-disciplinary Divide*. John Wiley & Sons, 2007.

Chalupowicz, L., et al. "Diagnosis of plant diseases using the Nanopore sequencing platform." *Plant Pathology* 68.2 (2019): 229-238.

Chang, Cherng-Shyang, et al. "Identification of a gut microbiota member that ameliorates DSS-induced colitis in intestinal barrier enhanced Dusp6-deficient mice." *Cell reports* 37.8 (2021): 110016.

Charalampous, Themoula, et al. "Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection." *Nature biotechnology* 37.7 (2019): 783-792.

Charlet, Rogatien, et al. "Remodeling of the *Candida glabrata* cell wall in the gastrointestinal tract affects the gut microbiota and the immune response." *Scientific reports* 8.1 (2018): 1-12.

Chen, Chen, et al. "The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases." *Nature communications* 8.1 (2017): 1-11.

Chen, Peng, et al. "Microbiota protects mice against acute ethanol-induced liver injury." *Ethanolism: Clinical and Experimental Research* 39.12 (2015): 2313-2323.

Chen, Peng, et al. "Microbiota protects mice against acute ethanol-induced liver injury." *Ethanolism: Clinical and Experimental Research* 39.12 (2015): 2313-2323.

Chen, Shifu, et al. "fastp: an ultra-fast all-in-one FASTQ preprocessor." *Bioinformatics* 34.17 (2018): i884-i890.

Chen, Weiguang, et al. "Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer." *PloS one* 7.6 (2012): e39743.

Chen, Zigui, et al. "Impact of preservation method and 16S rRNA hypervariable region on gut microbiota profiling." *Msystems* 4.1 (2019): e00271-18.

Chikina, Aleksandra S., et al. "Macrophages maintain epithelium integrity by limiting fungal product absorption." *Cell* 183.2 (2020): 411-428.

Chong, Jasmine, et al. "Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data." *Nature Protocols* 15.3 (2020): 799-821.

Chu, Derrick M., et al. "Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery." *Nature medicine* 23.3 (2017): 314-326.

Chu, Huikuan, et al. "The *Candida albicans* exotoxin candidalysin promotes ethanol-associated liver disease." *Journal of hepatology* 72.3 (2020): 391-400.

Chung, Hachung, et al. "Gut immune maturation depends on colonization with a host-specific microbiota." *Cell* 149.7 (2012): 1578-1593.

Chung, Liam, et al. "*Bacteroides fragilis* toxin coordinates a pro-carcinogenic inflammatory cascade via targeting of colonic epithelial cells." *Cell host & microbe* 23.2 (2018): 203-214.

Ciocan, Dragos, et al. "Characterization of intestinal microbiota in alcoholic patients with and without alcoholic hepatitis or chronic alcoholic pancreatitis." *Scientific reports* 8.1 (2018): 1-12.

Clarke, Gerard, et al. "The microbiome-gut-brain axis during early life regulates the hippocampal serotonergic system in a sex-dependent manner." *Molecular psychiatry* 18.6 (2013): 666-673.

Cole, James R., et al. "Ribosomal Database Project: data and tools for high throughput rRNA analysis." *Nucleic acids research* 42.D1 (2014): D633-D642.

Coleman Jr, Leon G., et al. "Postnatal day 7 ethanol treatment causes persistent reductions in adult mouse brain volume and cortical neurons with sex specific effects on neurogenesis." *Ethanol* 46.6 (2012): 603-612.

Collado, Maria Carmen, et al. "Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid." *Scientific reports* 6.1 (2016): 1-13.

Collins, Stephen M., and Premsyl Bercik. "Intestinal bacteria influence brain activity in healthy humans." *Nature reviews Gastroenterology & hepatology* 10.6 (2013): 326-327.

Corvilain, Emilie, Jean-Laurent Casanova, and Anne Puel. "Inherited CARD9 deficiency: invasive disease caused by ascomycete fungi in previously healthy children and adults." *Journal of clinical immunology* 38.6 (2018): 656-693.

Cox, Laura M., and Martin J. Blaser. "Antibiotics in early life and obesity." *Nature Reviews Endocrinology* 11.3 (2015): 182-190.

Cryan, John F., and Timothy G. Dinan. "More than a gut feeling: the microbiota regulates neurodevelopment and behavior." *Neuropsychopharmacology* 40.1 (2015): 241.

Cryan, John F., et al. "The gut microbiome in neurological disorders." *The Lancet Neurology* 19.2 (2020): 179-194.

Cryan, John F., et al. "The microbiota-gut-brain axis." *Physiological reviews* (2019).

Cuscó, Anna, et al. "Microbiota profiling with long amplicons using Nanopore sequencing: full-length 16S rRNA gene and the 16S-ITS-23S of the rrn operon." *F1000Research* 7 (2018).

Cuskin, Fiona, et al. "Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism." *Nature* 517.7533 (2015): 165-169.

D'Andreano, Sara, Anna Cuscó, and Olga Francino. "Rapid and real-time identification of fungi up to species level with long amplicon nanopore sequencing from clinical samples." *Biology Methods and Protocols* 6.1 (2020): bpaa026.

Dallas, David C., et al. "Digestion of protein in premature and term infants." *Journal of nutritional disorders & therapy* 2.3 (2012): 112.

Dandachi, Iman, et al. "Genome analysis of *Lachnoclostridium phocaeense* isolated from a patient after kidney transplantation in Marseille." *New Microbes and New Infections* 41 (2021): 100863.

Dandachi, Iman, et al. "Genome analysis of *Lachnoclostridium phocaeense* isolated from a patient after kidney transplantation in Marseille." *New Microbes and New Infections* 41 (2021): 100863.

De Coster, Wouter, et al. "NanoPack: visualizing and processing long-read sequencing data." *Bioinformatics* 34.15 (2018): 2666-2669.

de Goffau, Marcus C., et al. "Human placenta has no microbiome but can contain potential pathogens." *Nature* 572.7769 (2019): 329-334.

de Oliveira Martins, Leonardo, et al. "Taxonomic resolution of the ribosomal RNA operon in bacteria: implications for its use with long-read sequencing." *NAR genomics and bioinformatics* 2.1 (2020): lqz016.

de Timary, Philippe, et al. "A dysbiotic subpopulation of ethanol-dependent subjects." *Gut microbes* 6.6 (2015): 388-391.

Deamer, David, Mark Akeson, and Daniel Branton. "Three decades of nanopore sequencing." *Nature biotechnology* 34.5 (2016): 518-524.

Delahaye, Clara, and Jacques Nicolas. "Sequencing DNA with nanopores: Troubles and biases." *PloS one* 16.10 (2021): e0257521.

Desbonnet, Lieve, et al. "Gut microbiota depletion from early adolescence in mice: implications for brain and behaviour." *Brain, behavior, and immunity* 48 (2015): 165-173.

Desbonnet, Lieve, et al. "Microbiota is essential for social development in the mouse." *Molecular psychiatry* 19.2 (2014): 146-148.

Deveau, Aurelie, et al. "Bacterial–fungal interactions: ecology, mechanisms and challenges." *FEMS microbiology reviews* 42.3 (2018): 335-352.

Dhariwal, Achal, et al. "MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data." *Nucleic acids research* 45.W1 (2017): W180-W188.

Di Pierro, Francesco. "Gut Microbiota Parameters Potentially Useful in Clinical Perspective." *Microorganisms* 9.11 (2021): 2402.

Dobbing, John, and Jean Sands. "Comparative aspects of the brain growth spurt." *Early human development* 3.1 (1979): 79-83.

Donohoe, Dallas R., et al. "A gnotobiotic mouse model demonstrates that dietary fiber protects against colorectal tumorigenesis in a microbiota-and butyrate-dependent manner." *Cancer discovery* 4.12 (2014): 1387-1397.

Dowden, R. A., et al. "Host genotype and exercise exhibit species-level selection for members of the gut bacterial communities in the mouse digestive system." *Scientific reports*10.1 (2020): 1-12.

Drew, Paul D., et al. "Pioglitazone blocks ethanol induction of microglial activation and immune responses in the hippocampus, cerebellum, and cerebral cortex in a mouse model of fetal ethanol spectrum disorders." *Ethanolism: Clinical and Experimental Research* 39.3 (2015): 445-454.

Dubinkina, Veronika B., et al. "Links of gut microbiota composition with ethanol dependence syndrome and alcoholic liver disease." *Microbiome* 5.1 (2017): 1-14.

Durack, Juliana, et al. "Delayed gut microbiota development in high-risk for asthma infants is temporarily modifiable by Lactobacillus supplementation." *Nature communications* 9.1 (2018): 1-9.

Dutra, Jared, et al. "The effects of alcohol consumption and increased body mass on the gut microbiota of Parkinson's Disease patients." *Undergraduate Journal of Experimental Microbiology and Immunology* 26 (2021).

Edgar, Robert C., et al. "UCHIME improves sensitivity and speed of chimera detection." *Bioinformatics* 27.16 (2011): 2194-2200.

Edwards, Arwyn, et al. "In-field metagenome and 16S rRNA gene amplicon nanopore sequencing robustly characterize glacier microbiota." *BioRxiv* (2019): 073965.

Egeter, Bastian, et al. "Speeding up the detection of invasive bivalve species using environmental DNA: A Nanopore and Illumina sequencing comparison." *Molecular Ecology Resources*(2022).

Eisenhofer, Raphael, et al. "Contamination in low microbial biomass microbiome studies: issues and recommendations." *Trends in microbiology* 27.2 (2019): 105-117.

Elamin, Elhaseen E., et al. "Ethanol metabolism and its effects on the intestinal epithelial barrier." *Nutrition reviews* 71.7 (2013): 483-499.

Eren, A. Murat, et al. "Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences." *The ISME journal* 9.4 (2015): 968-979.

Eren, A. Murat, et al. "Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data." *Methods in ecology and evolution* 4.12 (2013): 1111-1119.

Eric, S. Donkor, TKD Dayie Nicholas, and K. Adiku Theophilus. "Bioinformatics with basic local alignment search tool (BLAST) and fast alignment (FASTA)." *Journal of Bioinformatics and sequence analysis* 6.1 (2014): 1-6.

Ericsson, Aaron C., et al. "Differential susceptibility to colorectal cancer due to naturally occurring gut microbiota." *Oncotarget*6.32 (2015): 33689.

Erny, Daniel, et al. "Host microbiota constantly control maturation and function of microglia in the CNS." *Nature neuroscience* 18.7 (2015): 965-977.

Ervin, Samantha M., Siddharth Venkat Ramanan, and Aadra P. Bhatt. "Relationship between the gut microbiome and systemic chemotherapy." *Digestive diseases and sciences* 65.3 (2020): 874-884.

Escobar-Zepeda, Alejandra, et al. "Analysis of sequencing strategies and tools for taxonomic annotation: defining standards for progressive metagenomics." *Scientific reports* 8.1 (2018): 1-13.

Evseev, Peter, et al. "*Curtobacterium* spp. and *Curtobacterium flaccumfaciens*: Phylogeny, Genomics-Based Taxonomy, Pathogenicity, and Diagnostics." *Current Issues in Molecular Biology* 44.2 (2022): 889-927.

Ewels, Philip, et al. "MultiQC: summarize analysis results for multiple tools and samples in a single report." *Bioinformatics*32.19 (2016): 3047-3048.

Fan, Jeremy, Steven Huang, and Samuel D. Chorlton. "BugSeq: a highly accurate cloud platform for long-read metagenomic analyses." *BMC bioinformatics* 22.1 (2021): 1-12.

Fang, Cheng, et al. "Solid-state fermented Chinese alcoholic beverage (baijiu) and ethanol resulted in distinct metabolic and microbiome responses." *The FASEB Journal* 33.6 (2019): 7274-7288.

Feng, Qiang, et al. "Gut microbiome development along the colorectal adenoma–carcinoma sequence." *Nature communications* 6.1 (2015): 1-13.

Fernandez-Lizarbe, Sara, Jorge Montesinos, and Consuelo Guerri. "Ethanol induces TLR 4/TLR 2 association, triggering an inflammatory response in microglial cells." *Journal of neurochemistry* 126.2 (2013): 261-273.

Ferrere, Gladys, et al. "Fecal microbiota manipulation prevents dysbiosis and ethanol-induced liver injury in mice." *Journal of hepatology* 66.4 (2017): 806-815.

Fiers, William D., Iris H. Gao, and Iliyan D. Iliev. "Gut mycobiota under scrutiny: fungal symbionts or environmental transients?." *Current opinion in microbiology* 50 (2019): 79-86.

Fletcher, E. V., and G. Z. Mentis. "Motor circuit dysfunction in spinal muscular atrophy." *Spinal Muscular Atrophy*. Academic Press, 2017. 153-165.

Fodor, Anthony A., et al. "The "most wanted" taxa from the human microbiome for whole genome sequencing." (2012): e41294.

Foster, Zachary SL, Thomas J. Sharpton, and Niklaus J. Grünwald. "Metacoder: An R package for visualization and manipulation of community taxonomic diversity data." *PLoS computational biology* 13.2 (2017): e1005404.

Francis, Owen E., et al. "Pathoscope: species identification and strain attribution with unassembled sequencing data." *Genome research* 23.10 (2013): 1721-1729.

Frau, Alessandra, et al. "DNA extraction and amplicon production strategies deeply inf luence the outcome of gut mycobiome studies." *Scientific reports* 9.1 (2019): 1-17.

Frey-Klett, Pascale, et al. "Bacterial-fungal interactions: hyphens between agricultural, clinical, environmental, and food microbiologists." *Microbiology and molecular biology reviews* 75.4 (2011): 583-609.

Fricke, W. Florian, and Jacques Ravel. "Microbiome or no microbiome: are we looking at the prenatal environment through the right lens?." *Microbiome* 9.1 (2021): 1-2.

Fröhlich, Esther E., et al. "Cognitive impairment by antibiotic-induced gut dysbiosis: analysis of gut microbiota-brain communication." *Brain, behavior, and immunity* 56 (2016): 140-155.

Fujii, Yusuke, et al. "Fecal metabolite of a gnotobiotic mouse transplanted with gut microbiota from a patient with Alzheimer's disease." *Bioscience, biotechnology, and biochemistry* 83.11 (2019): 2144-2152.

Fujimura, Kei E., et al. "Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation." *Nature medicine* 22.10 (2016): 1187-1191.

Fujiyoshi, So, Ai Muto-Fujita, and Fumito Maruyama. "Evaluation of PCR conditions for characterizing bacterial communities with full-length 16S rRNA genes using a portable nanopore sequencer." *Scientific reports* 10.1 (2020): 1-10.

Fukui, Hirokazu, et al. "Effect of probiotic *Bifidobacterium bifidum* G9-1 on the relationship between gut microbiota profile and stress sensitivity in maternally separated rats." *Scientific reports*8.1 (2018): 1-10.

Fung, Thomas C., Christine A. Olson, and Elaine Y. Hsiao. "Interactions between the microbiota, immune and nervous systems in health and disease." *Nature neuroscience* 20.2 (2017): 145-155.

Furness, John B. "The enteric nervous system and neurogastroenterology." *Nature reviews Gastroenterology & hepatology* 9.5 (2012): 286-294.

Gacias, Mar, et al. "Microbiota-driven transcriptional changes in prefrontal cortex override genetic differences in social behavior." *elife* 5 (2016): e13442.

Gao, Bin, et al. "Innate immunity in alcoholic liver disease." *American Journal of Physiology-Gastrointestinal and Liver Physiology* 300.4 (2011): G516-G525.

Garcia-Solache, Monica A., and Arturo Casadevall. "Global warming will bring new fungal diseases for mammals." *MBio* 1.1 (2010): e00061-10.

Garud, N. R., et al. "Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. bioRxiv." *Preprint], doi* 10 (2017): 210955.

Giatsis, Christos, et al. "The colonization dynamics of the gut microbiota in tilapia larvae." *PloS one* 9.7 (2014): e103641.

Giménez-Gómez, Pablo, et al. "Changes in brain kynurenine levels via gut microbiota and gut-barrier disruption induced by chronic ethanol exposure in mice." *The FASEB Journal* 33.11 (2019): 12900-12914.

Glavas, Maria M., et al. "Effects of prenatal ethanol exposure on basal limbic–hypothalamic–pituitary–adrenal regulation: role of corticosterone." *Ethanolism: Clinical and Experimental Research* 31.9 (2007): 1598-1610.

Gonzalez-Perez, Gabriela, and Esi SN Lamousé-Smith. "Gastrointestinal microbiome dysbiosis in infant mice alters peripheral CD8+ T cell receptor signaling." *Frontiers in immunology* 8 (2017): 265.

Gouba, Nina, Didier Raoult, and Michel Drancourt. "Plant and fungal diversity in gut microbiota as revealed by molecular and culture investigations." *PloS one* 8.3 (2013): e59474.

Grander, Christoph, et al. "Recovery of ethanol-induced *Akkermansia muciniphila* depletion ameliorates alcoholic liver disease." *Gut* 67.5 (2018): 891-901.

Grassart, Alexandre, et al. "Bioengineered human organ-on-chip reveals intestinal microenvironment and mechanical forces impacting Shigella infection." *Cell host & microbe* 26.3 (2019): 435-444.

Greenblum, Sharon, Rogan Carr, and Elhanan Borenstein. "Extensive strain-level copy-number variation across human gut microbiome species." *Cell* 160.4 (2015): 583-594.

Greninger, Alexander L., et al. "Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis." *Genome medicine* 7.1 (2015): 1-13.

Grigor'eva, Irina N. "Gallstone disease, obesity and the *Firmicutes*/*Bacteroidetes* ratio as a possible biomarker of gut dysbiosis." *Journal of personalized medicine* 11.1 (2020): 13.

Grimaldi, Roberta, et al. "A prebiotic intervention study in children with autism spectrum disorders (ASDs)." *Microbiome*6.1 (2018): 1-13.

Gruber, Karl. "Here, there, and everywhere: From PCR s to next-generation sequencing technologies and sequence databases, DNA contaminants creep in from the most unlikely places." *EMBO reports* 16.8 (2015): 898-901.

Gu, Shenghua, et al. "Bacterial community mapping of the mouse gastrointestinal tract." *PloS one* 8.10 (2013): e74957.

Guerri, Consuelo. "Neuroanatomical and neurophysiological mechanisms involved in central nervous system dysfunctions induced by prenatal ethanol exposure." *Ethanolism: Clinical and Experimental Research* 22.2 (1998): 304-312.

Guida, Francesca, et al. "Antibiotic-induced microbiota perturbation causes gut endocannabinoidome changes, hippocampal neuroglial reorganization and depression in mice." *Brain, behavior, and immunity* 67 (2018): 230-245.

Guo, Pingting, et al. "*Clostridium* species as probiotics: potentials and challenges." *Journal of Animal Science and Biotechnology* 11.1 (2020): 1-10.

Guo, Rui, and Jun Ren. "Ethanol and acetaldehyde in public health: from marvel to menace." *International journal of environmental research and public health* 7.4 (2010): 1285-1301.

Gurbanov, Rafig, Uygar Kabaoğlu, and Tuba Yağci. "Metagenomic profiling of gut microbiota in urban and rural rats: A comparative study." (2021).

Gurwara, Shawn, et al. "Ethanol use alters the colonic mucosa–associated gut microbiota in humans." *Nutrition Research* 83 (2020): 119-128.

Gutierrez, Mackenzie W., and Marie-Claire Arrieta. "The intestinal mycobiome as a determinant of host immune and metabolic health." *Current Opinion in Microbiology* 62 (2021): 8-13.

Ha, Connie WY, et al. "Translocation of viable gut microbiota to mesenteric adipose drives formation of creeping fat in humans." *Cell* 183.3 (2020): 666-683.

Hale, Vanessa L., et al. "Shifts in the fecal microbiota associated with adenomatous polyps." *Cancer Epidemiology and Prevention Biomarkers* 26.1 (2017): 85-94.

Halkias, Joanna, et al. "CD161 contributes to prenatal immune suppression of IFN-γ–producing PLZF+ T cells." *The Journal of clinical investigation* 129.9 (2019): 3562-3577.

Hallen-Adams, Heather E., and Mallory J. Suhr. "Fungi in the healthy human gastrointestinal tract." *Virulence* 8.3 (2017): 352-358.

Hallen-Adams, Heather E., et al. "Fungi inhabiting the healthy human gastrointestinal tract: a diverse and dynamic community." *Fungal Ecology* 15 (2015): 9-17.

Hanage, William P. "Microbiology: microbiome science needs a healthy dose of scepticism." *Nature* 512.7514 (2014): 247-248.

Harnisch, James P., et al. "Diphtheria among alcoholic urban adults: a decade of experience in Seattle." *Annals of internal medicine* 111.1 (1989): 71-82.

Harris, Vanessa C., et al. "The intestinal microbiome in infectious diseases: the clinical relevance of a rapidly emerging field." *Open forum infectious diseases*. Vol. 4. No. 3. Oxford University Press, 2017.

Hartmann, Phillipp, Caroline T. Seebauer, and Bernd Schnabl. "Alcoholic liver disease: the gut microbiome and liver cross talk." *Ethanolism: Clinical and Experimental Research* 39.5 (2015): 763-775.

Hartmann, Phillipp, et al. "Deficiency of intestinal mucin-2 ameliorates experimental alcoholic liver disease in mice." *Hepatology* 58.1 (2013): 108-119.

Hartmann, Phillipp, et al. "Modulation of the intestinal bile acid/farnesoid X receptor/fibroblast growth factor 15 axis improves alcoholic liver disease in mice." *Hepatology* 67.6 (2018): 2150-2166.

Hausmann, Bela, et al. "Long-term transcriptional activity at zero growth of a cosmopolitan rare biosphere member." *MBio* 10.1 (2019): e02189-18.

Hedayati, M. T., et al. "*Aspergillus flavus*: human pathogen, allergen and mycotoxin producer." *Microbiology* 153.6 (2007): 1677-1692.

Heijtz, Rochellys Diaz, et al. "Normal gut microbiota modulates brain development and behavior." *Proceedings of the National Academy of Sciences* 108.7 (2011): 3047-3052.

Heisel, Timothy, et al. "Breastmilk and NICU surfaces are potential sources of fungi for infant mycobiomes." *Fungal Genetics and Biology* 128 (2019): 29-35.

Heisel, Timothy, et al. "High-fat diet changes fungal microbiomes and interkingdom relationships in the murine gut." *Msphere* 2.5 (2017): e00351-17.

Hendrikx, Tim, et al. "Bacteria engineered to produce IL-22 in intestine induce expression of REG3G to reduce ethanol-induced liver disease in mice." *Gut* 68.8 (2019): 1504-1515.

Hikone, Mayu, et al. "The first case report of infective endocarditis caused by *Gemella taiwanensis*." *Journal of infection and chemotherapy* 23.8 (2017): 567-571.

Hildebrand, Falk, et al. "Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice." *Genome biology* 14.1 (2013): 1-15.

Hillemacher, Thomas, et al. "Ethanol, microbiome, and their effect on psychiatric disorders." *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 85 (2018): 105-115.

Hoban, Alan E., et al. "Regulation of prefrontal cortex myelination by the microbiota." *Translational psychiatry* 6.4 (2016): e774-e774.

Hoffmann, Christian, et al. "Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents." *PloS one* 8.6 (2013): e66019.

Hold, Georgina L., et al. "Role of the gut microbiota in inflammatory bowel disease pathogenesis: what have we learnt in the past 10 years?." *World journal of gastroenterology: WJG* 20.5 (2014): 1192.

Honkanen, Jarno, et al. "Fungal dysbiosis and intestinal inflammation in children with beta-cell autoimmunity." *Frontiers in immunology* 11 (2020): 468.

Hsiao, Elaine Y., et al. "Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders." *Cell* 155.7 (2013): 1451-1463.

Hu, Yiheng, et al. "Inferring species compositions of complex fungal communities from long-and short-read sequence data." *bioRxiv* (2021).

Human Microbiome Jumpstart Reference Strains Consortium, et al. "A catalog of reference genomes from the human microbiome." *Science* 328.5981 (2010): 994-999.

Human Microbiome Project Consortium. "Structure, function and diversity of the healthy human microbiome." *nature* 486.7402 (2012): 207.

Hung, Wei-Chun, et al. "*Gemella parahaemolysans* sp. nov. and *Gemella taiwanensis* sp. nov., isolated from human clinical specimens." *International journal of systematic and evolutionary microbiology* 64.Pt_6 (2014): 2060-2065.

Huseyin, Chloe E., et al. "Forgotten fungi—the gut mycobiome in human health and disease." *FEMS microbiology reviews* 41.4 (2017): 479-511.

Huson, Daniel H., et al. "MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data." *PLoS computational biology* 12.6 (2016): e1004957.

IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. "COAL-TAR PITCH." *Chemical Agents and Related Occupations*. International Agency for Research on Cancer, 2012.

Iliev, Iliyan D., et al. "Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis." *Science* 336.6086 (2012): 1314-1317.

Imai, Kazuo, et al. "Whole genome sequencing of influenza A and B viruses with the MinION sequencer in the clinical setting: a pilot study." *Frontiers in Microbiology* (2018): 2748.

Ingber, Donald E. "Human organs-on-chips for disease modelling, drug development and personalized medicine." *Nature Reviews Genetics* (2022): 1-25.

Ingber, Donald E. "Is it time for reviewer 3 to request human organ chip experiments instead of animal validation studies?." *Advanced Science* 7.22 (2020): 2002030.

Inkelis, Sarah M., et al. "Neurodevelopment in adolescents and adults with fetal ethanol spectrum disorders (FASD): a magnetic resonance region of interest analysis." *Brain research* 1732 (2020): 146654.

Ip, Camilla LC, et al. "MinION Analysis and Reference Consortium: Phase 1 data release and analysis." *F1000Research* 4 (2015).

Jain, Miten, et al. "Nanopore sequencing and assembly of a human genome with ultra-long reads." *Nature biotechnology*36.4 (2018): 338-345.

Jain, Miten, et al. "The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community." *Genome biology* 17.1 (2016): 1-11.

Jamy, Mahwash, et al. "Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity." *Molecular Ecology Resources*20.2 (2020): 429-443.

Jeon, Soomin, et al. "Positive effect of *Lactobacillus acidophilus* EG004 on cognitive ability of healthy mice by fecal microbiome analysis using full-length 16S-23S rRNA metagenome sequencing." *Microbiology spectrum* 10.1 (2022): e01815-21.

Jiang, Xian-Wan, et al. "New strain of *Pediococcus pentosaceus* alleviates ethanol-induced liver injury by modulating the gut microbiota and short-chain fatty acid metabolism." *World Journal of Gastroenterology* 26.40 (2020): 6224.

Jiao, Li, et al. "Ethanol use and risk of pancreatic cancer: the NIH-AARP Diet and Health Study." *American journal of epidemiology* 169.9 (2009): 1043-1051.

Johnson, Jethro S., et al. "Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis." *Nature communications* 10.1 (2019): 1-11.

Jousset, Alexandre, et al. "Where less may be more: how the rare biosphere pulls ecosystems strings." *The ISME journal* 11.4 (2017): 853-862.

Ju, Tingting, et al. "Defining the role of *Parasutterella*, a previously uncharacterized member of the core gut microbiota." *The ISME journal* 13.6 (2019): 1520-1534.

Jung, Su-Jin, et al. "Regulation of ethanol and acetaldehyde metabolism by a mixture of *Lactobacillus* and *Bifidobacterium* species in human." *Nutrients* 13.6 (2021): 1875.

Juul, Sissel, et al. "What's in my pot? Real-time species identification on the MinION™." *BioRxiv* (2015): 030742.

Kai, Shinichi, et al. "Rapid bacterial identification by direct PCR amplification of 16S rRNA genes using the MinION™ nanopore sequencer." *FEBS open bio* 9.3 (2019): 548-557.

Kaminen-Ahola, Nina, et al. "Maternal ethanol consumption alters the epigenotype and the phenotype of offspring in a mouse model." *PLoS genetics* 6.1 (2010): e1000811.

Kane, Cynthia JM, and Paul D. Drew. "Neuroinflammatory contribution of microglia and astrocytes in fetal ethanol spectrum disorders." *Journal of neuroscience research* 99.8 (2021): 1973-1985.

Kane, Cynthia JM, et al. "Protection of neurons and microglia against ethanol in a mouse model of fetal ethanol spectrum disorders by peroxisome proliferator-activated receptor-γ agonists." *Brain, behavior, and immunity* 25 (2011): S137-S145.

Kang, Dae-Wook, et al. "Microbiota transfer therapy alters gut ecosystem and improves gastrointestinal and autism symptoms: an open-label study." *Microbiome* 5.1 (2017): 1-16.

Kao, Amy Chia-Ching, et al. "Prebiotic attenuation of olanzapine-induced weight gain in rats: analysis of central and peripheral biomarkers and gut microbiota." *Translational psychiatry* 8.1 (2018): 1-12.

Karst, Søren M., et al. "Enabling high-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing." *BioRxiv* (2020): 645903.

Karst, Søren M., et al. "Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias." *Nature biotechnology* 36.2 (2018): 190-195.

Kazemi, Asma, et al. "Effect of probiotic and prebiotic vs placebo on psychological outcomes in patients with major depressive disorder: A randomized clinical trial." *Clinical Nutrition* 38.2 (2019): 522-528.

Kim, Daehwan, et al. "Centrifuge: rapid and sensitive classification of metagenomic sequences." *Genome research*26.12 (2016): 1721-1729.

Kim, Ji-Sun, et al. "*Intestinibaculum porci* gen. nov., sp. nov., a new member of the family *Erysipelotrichaceae* isolated from the small intestine of a swine." *Journal of Microbiology* 57.5 (2019): 381-387.

Kim, Yun-Gi, et al. "Gut dysbiosis promotes M2 macrophage polarization and allergic airway inflammation via fungi-induced PGE2." *Cell host & microbe* 15.1 (2014): 95-102.

Kiraly, Drew D., et al. "Alterations of the host microbiome affect behavioral responses to cocaine." *Scientific reports* 6.1 (2016): 1-12.

Kirpich, Irina A., et al. "Probiotics restore bowel flora and improve liver enzymes in human ethanol-induced liver injury: a pilot study." *Ethanol* 42.8 (2008): 675-682.

Kitahara, Maki, et al. "Assignment of *Eubacterium* sp. VPI 12708 and related strains with high bile acid 7alpha-dehydroxylating activity to *Clostridium scindens* and proposal of *Clostridium hylemonae* sp. nov., isolated from human faeces." *International journal of systematic and evolutionary microbiology* 50.3 (2000): 971-978.

Kleiber, Morgan L., Elise Wright, and Shiva M. Singh. "Maternal voluntary drinking in C57BL/6J mice: advancing a model for fetal ethanol spectrum disorders." *Behavioural brain research* 223.2 (2011): 376-387.

Kleiber, Morgan L., et al. "Long-term genomic and epigenomic dysregulation as a consequence of prenatal ethanol exposure: a model for fetal ethanol spectrum disorders." *Frontiers in genetics* 5 (2014): 161.

Klug, Lisa, and Günther Daum. "Yeast lipid metabolism at a glance." *FEMS yeast research* 14.3 (2014): 369-388.

Koh, Hyunwook, et al. "A distance-based kernel association test based on the generalized linear mixed model for correlated microbiome studies." *Frontiers in genetics* 10 (2019): 458.

Kollerov, V. V., et al. "Deoxycholic acid transformations catalyzed by selected filamentous fungi." *Steroids* 107 (2016): 20-29.

Kolligs, Frank T. "Diagnostics and epidemiology of colorectal cancer." *Visceral medicine* 32.3 (2016): 158-164.

Kosnicki, Kassi L., et al. "Effects of moderate, voluntary ethanol consumption on the rat and human gut microbiome." *Addiction biology* 24.4 (2019): 617-630.

Kostic, Aleksandar D., et al. "Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma." *Genome research* 22.2 (2012): 292-298.

Kuczynski, Justin, et al. "Experimental and analytical tools for studying the human microbiome." *Nature Reviews Genetics* 13.1 (2012): 47-58.

Kulakov, Leonid A., et al. "Analysis of bacteria contaminating ultrapure water in industrial systems." *Applied and environmental microbiology* 68.4 (2002): 1548-1555.

Kurakawa, Takashi, et al. "Diversity of intestinal *Clostridium coccoides* group in the Japanese population, as demonstrated by reverse transcription-quantitative PCR." *PLoS One* 10.5 (2015): e0126226.

La Rosa, Patricio S., et al. "Patterned progression of bacterial populations in the premature infant gut." *Proceedings of the National Academy of Sciences* 111.34 (2014): 12522-12527.

Laforest-Lapointe, Isabelle, and Marie-Claire Arrieta. "Patterns of early-life gut microbial colonization during human immune development: an ecological perspective." *Frontiers in immunology* 8 (2017): 788.

Lagier, Jean-Christophe, et al. "Culture of previously uncultured members of the human gut microbiota by culturomics." *Nature microbiology* 1.12 (2016): 1-8.

Lagier, Jean-Christophe, et al. "The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota." *Clinical microbiology reviews* 28.1 (2015): 237-264.

Lagkouvardos, Ilias, et al. "Sequence and cultivation study of *Muribaculaceae* reveals novel species, host preference, and functional potential of this yet undescribed family." *Microbiome*7.1 (2019): 1-15.

Lagkouvardos, Ilias, et al. "The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota." *Nature microbiology* 1.10 (2016): 1-15.

Lang, Sonja, et al. "Intestinal fungal dysbiosis and systemic immune response to fungi in patients with alcoholic hepatitis." *Hepatology* 71.2 (2020): 522-538.

Langille, Morgan GI, et al. "Microbial shifts in the aging mouse gut." *Microbiome* 2.1 (2014): 1-12.

Lauder, Abigail P., et al. "Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota." *Microbiome* 4.1 (2016): 1-11.

Le Roy, Tiphaine, et al. "Intestinal microbiota determines development of non-alcoholic fatty liver disease in mice." *Gut*62.12 (2013): 1787-1794.

Leclercq, Sophie, et al. "Gut microbiota-induced changes in β-hydroxybutyrate metabolism are linked to altered sociability and depression in ethanol use disorder." *Cell reports* 33.2 (2020): 108238.

Leclercq, Sophie, et al. "Intestinal permeability, gut-bacterial dysbiosis, and behavioral markers of ethanol-dependence severity." *Proceedings of the National Academy of Sciences*111.42 (2014): E4485-E4493.

Leclercq, Sophie, et al. "Low-dose penicillin in early life induces long-term changes in murine gut microbiota, brain cytokines and behavior." *Nature communications* 8.1 (2017): 1-12.

Lee, Eunjung, and Jang-Eun Lee. "Impact of drinking ethanol on gut microbiota: Recent perspectives on ethanol and alcoholic beverage." *Current Opinion in Food Science* 37 (2021): 91-97.

Lee, Jang-Eun, et al. "Alteration of gut microbiota composition by short-term low-dose ethanol intake is restored by fermented rice liquor in mice." *Food Research International* 128 (2020): 108800.

Leggett, Richard M., and Matthew D. Clark. "A world of opportunities with nanopore sequencing." *Journal of experimental botany* 68.20 (2017): 5419-5429.

Leggett, Richard M., et al. "Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens." *Nature microbiology* 5.3 (2020): 430-442.

Lehnardt, Seija. "Innate immunity and neuroinflammation in the CNS: The role of microglia in Toll-like receptor-mediated neuronal injury." *Glia* 58.3 (2010): 253-263.

Levan, Sophia R., et al. "Elevated faecal 12, 13-diHOME concentration in neonates at high risk for asthma is produced by gut bacteria and impedes immune tolerance." *Nature microbiology* 4.11 (2019): 1851-1861.

Li, Heng. "Minimap2: pairwise alignment for nucleotide sequences." *Bioinformatics* 34.18 (2018): 3094-3100.

Li, Huizhen, et al. "*Lactobacillus plantarum* KLDS1. 0344 and *Lactobacillus acidophilus* KLDS1. 0901 mixture prevents chronic alcoholic liver injury in mice by protecting the intestinal barrier and regulating gut microbiota and liver-related pathways." *Journal of Agricultural and Food Chemistry* 69.1 (2020): 183-197.

Li, Na, et al. "Memory CD4+ T cells are generated in the human fetal intestine." *Nature immunology* 20.3 (2019): 301-312.

Liang, Jessie Qiaoyi, et al. "A novel faecal *Lachnoclostridium* marker for the non-invasive diagnosis of colorectal adenoma and cancer." *Gut* 69.7 (2020): 1248-1257.

Liang, Qiaoyi, et al. "Fecal bacteria act as novel biomarkers for noninvasive diagnosis of colorectal cancer." *Clinical Cancer Research* 23.8 (2017): 2061-2070.

Liang, Siyuan, et al. "Gut microbiome associated with APC gene mutation in patients with intestinal adenomatous polyps." *International journal of biological sciences* 16.1 (2020): 135.

Liaw, Andy, and Matthew Wiener. "Classification and regression by random forest." *R news* 2.3 (2002): 18-22.

Lim, Efrem S., Cynthia Rodriguez, and Lori R. Holtz. "Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community." *Microbiome* 6.1 (2018): 1-8.

Limon, Jose J., et al. "Malassezia is associated with Crohn's disease and exacerbates colitis in mouse models." *Cell host & microbe* 25.3 (2019): 377-388.

Lippai, Dora, et al. "Chronic ethanol-induced microRNA-155 contributes to neuroinflammation in a TLR4-dependent manner in mice." *PloS one* 8.8 (2013): e70945.

Litvak, Yael, et al. "Dysbiotic *Proteobacteria* expansion: a microbial signature of epithelial dysfunction." *Current opinion in microbiology* 39 (2017): 1-6.

Liu, Chang, et al. "The Mouse Gut Microbial Biobank expands the coverage of cultured bacteria." *Nature communications* 11.1 (2020): 1-12.

Liu, Yong-Xin, et al. "A practical guide to amplicon and metagenomic analysis of microbiome data." *Protein & cell* 12.5 (2021): 315-330.

Llopis, Marta, et al. "Intestinal microbiota contributes to individual susceptibility to alcoholic liver disease." *Gut* 65.5 (2016): 830-839.

Lo, Yu-Chun, et al. "Neonatal hyperoxia induces gut dysbiosis and behavioral changes in adolescent mice." *Journal of the Chinese Medical Association* 84.3 (2021): 290-298.

Loman, Nicholas J., and Aaron R. Quinlan. "Poretools: a toolkit for analyzing nanopore sequence data." *Bioinformatics* 30.23 (2014): 3399-3401.

Lopetuso, Loris R., et al. "Commensal *Clostridia*: leading players in the maintenance of gut homeostasis." *Gut pathogens* 5.1 (2013): 1-8.

Louis, Petra, and Harry J. Flint. "Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine." *FEMS microbiology letters* 294.1 (2009): 1-8.

Lowe, Patrick P., et al. "Ethanol-related changes in the intestinal microbiome influence neutrophil infiltration, inflammation and steatosis in early alcoholic hepatitis in mice." *PloS one* 12.3 (2017): e0174544.

Lowe, Patrick P., et al. "Reduced gut microbiome protects from ethanol-induced neuroinflammation and alters intestinal and brain inflammasome expression." *Journal of neuroinflammation*15.1 (2018): 1-12.

Lu, Jennifer, and Steven L. Salzberg. "Removing contaminants from databases of draft genomes." *PLoS computational biology*14.6 (2018): e1006277.

Luczynski, Pauline, et al. "Growing up in a bubble: using germ-free animals to assess the influence of the gut microbiota on brain and behavior." *International Journal of Neuropsychopharmacology* 19.8 (2016).

Luk, Berkley, et al. "Postnatal colonization with human “infant-type" *Bifidobacterium* species alters behavior of adult gnotobiotic mice." *PLoS One* 13.5 (2018): e0196510.

Lukić, Iva, et al. "Role of tryptophan in microbiota-induced depressive-like behavior: evidence from tryptophan depletion study." *Frontiers in behavioral neuroscience* 13 (2019): 123.

Lundberg, Randi, et al. "Human microbiota-transplanted C57BL/6 mice and offspring display reduced establishment of key bacteria and reduced immune stimulation compared to mouse microbiota-transplantation." *Scientific reports* 10.1 (2020): 1-16.

Luo, Yuanyuan, et al. "Gut microbiota regulates mouse behaviors through glucocorticoid receptor pathway genes in the hippocampus." *Translational psychiatry* 8.1 (2018): 1-10.

Lusk, Richard W. "Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data." *PloS one* 9.10 (2014): e110808.

Lynch, Michael DJ, and Josh D. Neufeld. "Ecology and exploration of the rare biosphere." *Nature Reviews Microbiology*13.4 (2015): 217-229.

Lyte, Mark, et al. "Induction of anxiety-like behavior in mice during the initial stages of infection with the agent of murine colonic hyperplasia Citrobacter rodentium." *Physiology & behavior* 89.3 (2006): 350-357.

Macfarlane, S., and J. F. Dillon. "Microbial biofilms in the human gastrointestinal tract." *Journal of applied microbiology* 102.5 (2007): 1187-1196.

Magne, Fabien, et al. "The *Firmicutes*/*Bacteroidetes* ratio: a relevant marker of gut dysbiosis in obese patients?." *Nutrients*12.5 (2020): 1474.

Magoč, Tanja, and Steven L. Salzberg. "FLASH: fast length adjustment of short reads to improve genome assemblies." *Bioinformatics* 27.21 (2011): 2957-2963.

Mailhe, M., et al. "'*Millionella massiliensis'* gen. nov., sp. nov., a new bacterial species isolated from human right colon." *New Microbes and New Infections* 17 (2017): 11-12.

Makki, Kassem, et al. "The impact of dietary fiber on gut microbiota in host health and disease." *Cell host & microbe* 23.6 (2018): 705-715.

Malaguarnera, Giulia, et al. "Gut microbiota in alcoholic liver disease: pathogenetic role and therapeutic perspectives." *World Journal of Gastroenterology: WJG* 20.44 (2014): 16639.

Marchet, Camille, et al. "De novo clustering of long reads by gene from transcriptomics data." *Nucleic Acids Research* 47.1 (2019): e2-e2.

Marcobal, Angela, et al. "*Bacteroides* in the infant gut consume milk oligosaccharides via mucus-utilization pathways." *Cell host & microbe* 10.5 (2011): 507-514.

Markey, Laura, et al. "Colonization with the commensal fungus *Candida albicans* perturbs the gut-brain axis through dysregulation of endocannabinoid signaling." *Psychoneuroendocrinology* 121 (2020): 104808.

Martí, Jose Manuel. "Recentrifuge: Robust comparative analysis and contamination removal for metagenomics." *PLoS computational biology* 15.4 (2019): e1006967.

Martijn, Joran, et al. "Confident phylogenetic identification of uncultured prokaryotes through long read amplicon sequencing of the 16S-ITS-23S rRNA operon." *Environmental microbiology*21.7 (2019): 2485-2498.

Martín, Rebeca, et al. "Characterization of indigenous vaginal lactobacilli from healthy women as probiotic candidates." *International Microbiology* 11.4 (2009): 261-266.

Martín, Rocío, et al. "Human milk is a source of lactic acid bacteria for the infant gut." *The Journal of pediatrics* 143.6 (2003): 754-758.

Martinez-Taboada, Fernando, and Jose Ignacio Redondo. "The SIESTA (SEAAV Integrated evaluation sedation tool for anaesthesia) project: Initial development of a multifactorial sedation assessment tool for dogs." *PloS one* 15.4 (2020): e0230799.

Matsuo, Yoshiyuki, et al. "Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution." *BMC microbiology* 21.1 (2021): 1-13.

Mattson, Sarah N., et al. "A decrease in the size of the basal ganglia in children with fetal ethanol syndrome." *Ethanolism: Clinical and Experimental Research* 20.6 (1996): 1088-1093.

Mayer, Emeran A., Kirsten Tillisch, and Arpana Gupta. "Gut-brain axis and the microbiota." *The Journal of clinical investigation*125.3 (2015): 926-938.

Mayer, François L., Duncan Wilson, and Bernhard Hube. "*Candida albicans* pathogenicity mechanisms." *Virulence* 4.2 (2013): 119-128.

MCC, Canesso, et al. "Comparing the effects of acute ethanol consumption in germ-free and conventional mice: the role of the gut microbiota." *BMC microbiology* 14.1 (2014): 1-10.

McDonald, Daniel, et al. "An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea." *The ISME journal* 6.3 (2012): 610-618.

McGovern, Naomi, et al. "Human fetal dendritic cells promote prenatal T-cell immune suppression through arginase-2." *Nature* 546.7660 (2017): 662-666.

McKnight, Donald T., et al. "Methods for normalizing microbiome data: an ecological perspective." *Methods in Ecology and Evolution* 10.3 (2019): 389-400.

Meckel, Katherine R., and Drew D. Kiraly. "A potential role for the gut microbiome in substance use disorders." *Psychopharmacology* 236.5 (2019): 1513-1530.

Mehta, Raaj S., et al. "Association of dietary patterns with risk of colorectal cancer subtypes classified by *Fusobacterium nucleatum* in tumor tissue." *JAMA oncology* 3.7 (2017): 921-927.

Meroni, Marica, Miriam Longo, and Paola Dongiovanni. "Ethanol or gut microbiota: who is the guilty?." *International journal of molecular sciences* 20.18 (2019): 4568.

Merseguel, Karina Bellinghausen, et al. "Genetic diversity of medically important and emerging Candida species causing invasive infection." *BMC infectious diseases* 15.1 (2015): 1-11.

Mims, Tahliyah S., et al. "The gut mycobiome of healthy mice is shaped by the environment and correlates with metabolic outcomes in response to diet." *Communications biology* 4.1 (2021): 1-11.

Mims, Tahliyah S., et al. "The gut mycobiome of healthy mice is shaped by the environment and correlates with metabolic outcomes in response to diet." *Communications biology* 4.1 (2021): 1-11.

Mir, Hina, et al. "Occludin deficiency promotes ethanol-induced disruption of colonic epithelial junctions, gut barrier dysfunction and liver damage in mice." *Biochimica et Biophysica Acta (BBA)-General Subjects* 1860.4 (2016): 765-774.

Mirzayi, Chloe, et al. "Reporting guidelines for human microbiome research: the STORMS checklist." *Nature medicine* 27.11 (2021): 1885-1892.

Mitsuhashi, Satomi, et al. "A portable system for rapid bacterial composition analysis using a nanopore-based sequencer and laptop computer." *Scientific reports* 7.1 (2017): 1-9.

Miyake, Sou, et al. "Cultivation and description of *Duncaniella dubosii* sp. nov., *Duncaniella freteri* sp. nov. and emended description of the species *Duncaniella muris*." *International Journal of Systematic and Evolutionary Microbiology* 70.5 (2020a): 3105-3110.

Miyake, Sou, et al. "Muribaculum gordoncarteri sp. nov., an anaerobic bacterium from the faeces of C57BL/6J mice." *International Journal of Systematic and Evolutionary Microbiology* 70.8 (2020b): 4725-4729.

Mizoue, Tetsuya, et al. "Ethanol drinking and colorectal cancer in Japanese: a pooled analysis of results from five cohort studies." *American journal of epidemiology* 167.12 (2008): 1397-1406.

Monteggia, Lisa M., et al. "Essential role of brain-derived neurotrophic factor in adult hippocampal function." *Proceedings of the National Academy of Sciences* 101.29 (2004): 10827-10832.

Moossavi, Shirin, et al. "Human milk fungi: environmental determinants and inter-kingdom associations with milk bacteria in the CHILD Cohort Study." *BMC microbiology* 20 (2020): 1-13.

Morgan, Timothy R., Sarathy Mandayam, and M. Mazen Jamal. "Ethanol and hepatocellular carcinoma." *Gastroenterology* 127.5 (2004): S87-S96.

Morgan, Xochitl C., and Curtis Huttenhower. "Chapter 12: Human microbiome analysis." *PLoS computational biology* 8.12 (2012): e1002808.

Muadcheingka, Thaniya, and Pornpen Tantivitayakul. "Distribution of *Candida albicans* and non-albicans *Candida* species in oral candidiasis patients: correlation between cell surface hydrophobicity and biofilm forming activities." *Archives of oral biology* 60.6 (2015): 894-901.

Mukamolova, Galina V., et al. "The *rpf* gene of *Micrococcus* luteus encodes an essential secreted growth factor." *Molecular microbiology* 46.3 (2002): 611-621.

Mutlu, Ece A., et al. "Colonic microbiome is altered in ethanolism." *American Journal of Physiology-Gastrointestinal and Liver Physiology* 302.9 (2012): G966-G978.

Nakamura, Saori, et al. "Detection and isolation of β-conglycinin-susceptible gut indigenous bacteria from ICR mice fed high-sucrose diet." *Food Bioscience* 41 (2021): 100994.

Nash, Andrea K., et al. "The gut mycobiome of the Human Microbiome Project healthy cohort." *Microbiome* 5.1 (2017): 1-13.

Nasko, Daniel J., et al. "RefSeq database growth influences the accuracy of k-mer-based lowest common ancestor species identification." *Genome biology* 19.1 (2018): 1-10.

Nayfach, Stephen, and Katherine S. Pollard. "Toward accurate and quantitative comparative metagenomics." *Cell* 166.5 (2016): 1103-1116.

Nayfach, Stephen, et al. "An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography." *Genome research* 26.11 (2016): 1612-1625.

Nayfach, Stephen, et al. "New insights from uncultivated genomes of the global human gut microbiome." *Nature* 568.7753 (2019): 505-510.

Ndongo, S., et al. "'*Collinsella phocaeensis*' sp. nov., '*Clostridium merdae*' sp. nov., '*Sutterella massiliensis*' sp. nov., '*Sutturella timonensis*' sp. nov., '*Enorma phocaeensis*' sp. nov., '*Mailhella massiliensis*' gen. nov., sp. nov., '*Mordavella massiliensis*' gen. nov., sp. nov. and '*Massiliprevotella massiliensis*' gen. nov., sp. nov., 9 new species isolated from fresh stool samples of healthy French patients." *New Microbes and New Infections* 17 (2017): 89-95.

Nearing, Jacob T., André M. Comeau, and Morgan GI Langille. "Identifying biases and their potential solutions in human microbiome studies." *Microbiome* 9.1 (2021): 1-22.

Needham, Brittany D., Rima Kaddurah-Daouk, and Sarkis K. Mazmanian. "Gut microbial molecules in behavioural and neurodegenerative conditions." *Nature Reviews Neuroscience* 21.12 (2020): 717-731.

Neff, Ellen. "A focus on fungi in the mouse gut." *Lab animal* 50.5 (2021): 124-124.

Neufeld, K. M., et al. "Reduced anxiety-like behavior and central neurochemical change in germ-free mice." *Neurogastroenterology & Motility* 23.3 (2011): 255-e119.

Nielsen, H. Bjørn, et al. "Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes." *Nature biotechnology* 32.8 (2014): 822-828.

Nilsson, R. Henrik, et al. "Mycobiome diversity: high-throughput sequencing and identification of fungi." *Nature Reviews Microbiology* 17.2 (2019): 95-109.

Nosova, Tatiana, et al. "Aldehyde dehydrogenase activity and acetate production by aerobic bacteria representing the normal flora of human large intestine." *Ethanol and ethanolism* 31.6 (1996): 555-564.

Nosova, Tatiana, et al. "Characteristics of aldehyde dehydrogenases of certain aerobic bacteria representing human colonic flora." *Ethanol and Ethanolism* 33.3 (1998): 273-280.

Nosova, Tatjana, et al. "Acetaldehyde production and metabolism by human indigenous and probiotic *Lactobacillus* and *Bifidobacterium* strains." *Ethanol and Ethanolism* 35.6 (2000): 561-568.

Nougayrède, Jean-Philippe, et al. "*Escherichia coli* induces DNA double-strand breaks in eukaryotic cells." *Science* 313.5788 (2006): 848-851.

Noverr, Mairi C., et al. "Development of allergic airway disease in mice following antibiotic therapy and fungal microbiota increase: role of host genetics, antigen, and interleukin-13." *Infection and immunity* 73.1 (2005): 30-38.

Ogbonnaya, Ebere S., et al. "Adult hippocampal neurogenesis is regulated by the microbiome." *Biological psychiatry* 78.4 (2015): e7-e9.

Ohkusa, Toshifumi, et al. "Commensal bacteria can enter colonic epithelial cells and induce proinflammatory cytokine secretion: a possible pathogenic mechanism of ulcerative colitis." *Journal of medical microbiology* 58.Pt 5 (2009): 535.

Olm, Matthew R., et al. "The source and evolutionary history of a microbial contaminant identified through soil metagenomic analysis." *MBio* 8.1 (2017): e01969-16.

Olomu, Isoken Nicholas, et al. "Elimination of “kitome” and “splashome” contamination results in lack of detection of a unique placental microbiome." *BMC microbiology* 20.1 (2020): 1-19.

Ondov, Brian D., Nicholas H. Bergman, and Adam M. Phillippy. "Interactive metagenomic visualization in a Web browser." *BMC bioinformatics* 12.1 (2011): 1-10.

Opriessnig, Tanja, Taya Forde, and Yoshihiro Shimoji. "*Erysipelothrix* spp.: past, present, and future directions in vaccine research." *Frontiers in Veterinary Science* 7 (2020): 174.

Oren, Aharon, and George M. Garrity. "List of new names and new combinations previously effectively, but not validly, published." *International journal of systematic and evolutionary microbiology* 66.11 (2016): 4299-4305.

Ormerod, Kate L., et al. "Genomic characterization of the uncultured *Bacteroidales* family S24-7 inhabiting the guts of homeothermic animals." *Microbiome* 4.1 (2016): 1-17.

Op De Beeck, Michiel, et al. "Comparison and validation of some ITS primer pairs useful for fungal metabarcoding studies." *PloS one* 9.6 (2014): e97629.

Osdaghi, Ebrahim, Anthony J. Young, and Robert M. Harveson. "Bacterial wilt of dry beans caused by *Curtobacterium flaccumfaciens* pv. *flaccumfaciens*: A new threat from an old enemy." *Molecular Plant Pathology* 21.5 (2020): 605-621.

Osdaghi, Ebrahim, et al. "Whole genome resources of 17 *Curtobacterium flaccumfaciens* strains including pathotypes of *C. flaccumfaciens* pv. *betae*, *C. flaccumfaciens* pv. *oortii* and *C. flaccumfaciens* pv. *poinsettiae*." *Molecular Plant-Microbe Interactions* ja (2022).

Pal, M. "Morbidity and mortality due to fungal infections." *J. Appl. Microbiol. Biochem* 1.1 (2017): 2.

Paolicelli, Rosa C., et al. "Synaptic pruning by microglia is necessary for normal brain development." *science* 333.6048 (2011): 1456-1458.

Park, Beom Seok, and Jie-Oh Lee. "Recognition of lipopolysaccharide pattern by TLR4 complexes." *Experimental & molecular medicine* 45.12 (2013): e66-e66.

Park, John Chulhoon, and Sin-Hyeog Im. "Of men in mice: the development and application of a humanized gnotobiotic mouse model for microbiome therapeutics." *Experimental & molecular medicine* 52.9 (2020): 1383-1396.

Park, Jun Kyu, et al. "*Heminiphilus faecis* gen. nov., sp. nov., a member of the family *Muribaculaceae*, isolated from mouse faeces and emended description of the genus *Muribaculum*." *Antonie Van Leeuwenhoek* 114.3 (2021): 275-286.

Parks, Donovan H., et al. "A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life." *Nature biotechnology* 36.10 (2018): 996-1004.

Parks, Donovan H., et al. "CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes." *Genome research* 25.7 (2015): 1043-1055.

Pascual, María, et al. "Cytokines and chemokines as biomarkers of ethanol-induced neuroinflammation and anxiety-related behavior: role of TLR4 and TLR2." *Neuropharmacology* 89 (2015): 352-359.

Pascual, María, et al. "Neuroimmune activation and myelin changes in adolescent rats exposed to high-dose ethanol and associated cognitive dysfunction: a review with reference to human adolescent drinking." *Ethanol and ethanolism* 49.2 (2014): 187-192.

Pascual, María, et al. "TLR4 response mediates ethanol-induced neurodevelopment alterations in a model of fetal ethanol spectrum disorders." *Journal of neuroinflammation* 14.1 (2017): 1-15.

Pasolli, Edoardo, et al. "Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle." *Cell* 176.3 (2019): 649-662.

Patel, Sheena, et al. "Ethanol and the Intestine." *Biomolecules*5.4 (2015): 2573-2588.

Patlevič, Peter, et al. "Reactive oxygen species and antioxidant defense in human gastrointestinal diseases." *Integrative medicine research* 5.4 (2016): 250-258.

Patterson, Angela M., et al. "Human gut symbiont *Roseburia hominis* promotes and regulates innate immunity." *Frontiers in immunology* 8 (2017): 1166.

Pelvig, Dorte P., et al. "Neocortical glial cell numbers in human brains." *Neurobiology of aging* 29.11 (2008): 1754-1762.

Pearman, William S., Nikki E. Freed, and Olin K. Silander. "Testing the advantages and disadvantages of short-and long-read eukaryotic metagenomics using simulated reads." *BMC bioinformatics* 21.1 (2020): 1-15.

Perez-Lopez, Araceli, et al. "Mucosal immunity to pathogenic intestinal bacteria." *Nature Reviews Immunology* 16.3 (2016): 135-148.

Perrone, Giancarlo, and Antonia Gallo. "*Aspergillus* species and their associated mycotoxins." *Mycotoxigenic Fungi* (2017): 33-49.

Perrone, Giancarlo, and Antonia Susca. "*Penicillium* species and their associated mycotoxins." *Mycotoxigenic fungi* (2017): 107-119.

Petrelli, Berardino, Joanne Weinberg, and Geoffrey G. Hicks. "Effects of prenatal ethanol exposure (PAE): insights into FASD using mouse models of PAE." *Biochemistry and Cell Biology*96.2 (2018): 131-147.

Poyet, M., et al. "A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research." *Nature medicine* 25.9 (2019): 1442-1452.

Pradier, Bruno, et al. "Microglial IL-1β progressively increases with duration of ethanol consumption." *Naunyn-Schmiedeberg's archives of pharmacology* 391.4 (2018): 455-461.

Prechtl, James C., and Terry L. Powley. "B-afferents: a fundamental division of the nervous system mediating homeostasis?." *Behavioral and Brain Sciences* 13.2 (1990): 289-300.

Qian, Gordon, and Joshua WK Ho. "Challenges and emerging systems biology approaches to discover how the human gut microbiome impact host physiology." *Biophysical Reviews* 12.4 (2020): 851-863.

Qin, Junjie, et al. "A human gut microbial gene catalogue established by metagenomic sequencing." *nature* 464.7285 (2010): 59-65.

Qin, Junjie, et al. "A metagenome-wide association study of gut microbiota in type 2 diabetes." *Nature* 490.7418 (2012): 55-60.

Qin, Nan, et al. "Alterations of the human gut microbiome in liver cirrhosis." *Nature* 513.7516 (2014): 59-64.

Quast, Christian, et al. "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools." *Nucleic acids research* 41.D1 (2012): D590-D596.

Queipo-Ortuño, María Isabel, et al. "Influence of red wine polyphenols and ethanol on the gut microbiota ecology and biochemical biomarkers." *The American journal of clinical nutrition* 95.6 (2012): 1323-1334.

Quinlan, Aaron R., and Ira M. Hall. "BEDTools: a flexible suite of utilities for comparing genomic features." *Bioinformatics* 26.6 (2010): 841-842.

Rackaityte, Elze, et al. "Viable bacterial colonization is highly limited in the human intestine in utero." *Nature medicine* 26.4 (2020): 599-607.

Raimondi, Stefano, et al. "Longitudinal survey of fungi in the human gut: ITS profiling, phenotyping, and colonization." *Frontiers in microbiology* 10 (2019): 1575.

Raineki, Charlis, et al. "Neurocircuitry underlying stress and emotional regulation in animals prenatally exposed to ethanol and subjected to chronic mild stress in adulthood." *Frontiers in endocrinology* 5 (2014): 5.

Ramchandani, V. A., W. F. Bosron, and T. K. Li. "Research advances in ethanol metabolism." *Pathologie Biologie* 49.9 (2001): 676-682.

Ramette, Alban. "Multivariate analyses in microbial ecology." FEMS microbiology ecology 62.2 (2007): 142-160.

Rao, R. K. "Acetaldehyde-induced barrier disruption and paracellular permeability in Caco-2 cell monolayer." *Ethanol*. Humana Press, 2008. 171-183.

Ravussin, Yann, et al. "Responses of gut microbiota to diet composition and weight loss in lean and obese mice." *Obesity*20.4 (2012): 738-747.

Reid, Ann, Merry Buckley, and M. Mcfall. "The rare biosphere." *Washington DC* (2011).

Rezasoltani, Sama, et al. "The association between fecal microbiota and different types of colorectal polyp as precursors of colorectal cancer." *Microbial pathogenesis* 124 (2018): 244-249.

Rhodes, Johanna, et al. "Genomic epidemiology of the UK outbreak of the emerging human fungal pathogen *Candida auris*." *Emerging microbes & infections* 7.1 (2018): 1-12.

Riley, Edward P., and Christie L. McGee. "Fetal ethanol spectrum disorders: an overview with emphasis on changes in brain and behavior." *Experimental biology and medicine* 230.6 (2005): 357-365.

Riley, Edward P., M. Alejandra Infante, and Kenneth R. Warren. "Fetal ethanol spectrum disorders: an overview." *Neuropsychology review* 21.2 (2011): 73.

Rintala, Anniina, et al. "Gut microbiota analysis results are highly dependent on the 16S rRNA gene target region, whereas the impact of DNA extraction is minor." *Journal of biomolecular techniques: JBT* 28.1 (2017): 19.

Robertson, Ruairi C., et al. "The human microbiome and child growth–first 1000 days and beyond." *Trends in microbiology* 27.2 (2019): 131-147.

Rodríguez, M. Mar, et al. "Obesity changes the human gut mycobiome." *Scientific reports* 5.1 (2015): 1-15.

Rognes, Torbjørn, et al. "VSEARCH: a versatile open source tool for metagenomics." *PeerJ* 4 (2016): e2584.

Rosero, Jaime A., et al. "Reclassification of *Eubacterium rectale* (Hauduroy et al. 1937) Prévot 1938 in a new genus *Agathobacter* gen. nov. as *Agathobacter rectalis* comb. nov., and description of *Agathobacter ruminis* sp. nov., isolated from the rumen contents of sheep and cows." *International Journal of Systematic and Evolutionary Microbiology* 66.2 (2016): 768-773.

Rossi, Marco, et al. "Colorectal cancer and ethanol consumption—populations to molecules." *Cancers* 10.2 (2018): 38.

Roswall, Nina, and Elisabete Weiderpass. "Ethanol as a risk factor for cancer: existing evidence in a global perspective." *Journal of Preventive Medicine and Public Health* 48.1 (2015): 1.

Salaspuro, Mikko. "Acetaldehyde as a common denominator and cumulative carcinogen in digestive tract cancers." *Scandinavian journal of gastroenterology* 44.8 (2009): 912-925.

Salaspuro, Mikko. "Key role of local acetaldehyde in upper GI tract carcinogenesis." *Best Practice & Research Clinical Gastroenterology* 31.5 (2017): 491-499.

Salaspuro, Mikko. "Microbial metabolism of ethanol and acetaldehyde and clinical consequences." *Addiction Biology* 2.1 (1997): 35-46.

Salaspuro, Ville, et al. "Ethanol oxidation and acetaldehyde production in vitro by human intestinal strains of *Escherichia coli* under aerobic, microaerobic, and anaerobic conditions." *Scandinavian journal of gastroenterology* 34.10 (1999): 967-973.

Salter, Michael W., and Beth Stevens. "Microglia emerge as central players in brain disease." *Nature medicine* 23.9 (2017): 1018-1027.

Salter, Susannah J., et al. "Reagent and laboratory contamination can critically impact sequence-based microbiome analyses." *BMC biology* 12.1 (2014): 1-12.

Salzman, Nita H., et al. "Analysis of 16S libraries of mouse gastrointestinal microflora reveals a large new group of mouse intestinal." *Microbiology* 148.11 (2002): 3651-3660.

Sampson, Timothy R., et al. "Gut microbiota regulate motor deficits and neuroinflammation in a model of Parkinson's disease." *Cell* 167.6 (2016): 1469-1480.

Samuelson, Kristin W., et al. "Neuropsychological functioning in posttraumatic stress disorder and ethanol abuse." *Neuropsychology* 20.6 (2006): 716.

Sanapareddy, Nina, et al. "Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans." *The ISME journal* 6.10 (2012): 1858-1868.

Sanidad, Katherine Z., and Melody Y. Zeng. "Neonatal gut microbiome and immunity." *Current opinion in microbiology* 56 (2020): 30-37.

Santos, Andres, et al. "Computational methods for 16S metabarcoding studies using Nanopore sequencing data." *Computational and Structural Biotechnology Journal* 18 (2020): 296-305.

Sato, Yuya, et al. "Transcriptome analysis of activated sludge microbiomes reveals an unexpected role of minority nitrifiers in carbon metabolism." *Communications biology* 2.1 (2019): 1-8.

Sbihi, Hind, et al. "Thinking bigger: How early-life environmental exposures shape the gut microbiome and influence the development of asthma and allergic disease." *Allergy* 74.11 (2019): 2103-2115.

Scanlan, Pauline D., et al. "Culture-independent analysis of the gut microbiota in colorectal cancer and polyposis." *Environmental microbiology* 10.3 (2008): 789-798.

Schloss, Patrick D. "The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies." *PLoS computational biology* 6.7 (2010): e1000844.

Schreurs, Renée RCE, et al. "Human fetal TNF-α-cytokine-producing CD4+ effector memory T cells promote intestinal development and mediate inflammation early in life." *Immunity* 50.2 (2019): 462-476.

Sczyrba, Alexander, et al. "Critical assessment of metagenome interpretation—a benchmark of metagenomics software." *Nature methods* 14.11 (2017): 1063-1071.

Sears, Cynthia L., and Drew M. Pardoll. "Perspective: alpha-bugs, their microbial partners, and the link to colon cancer." *Journal of Infectious Diseases* 203.3 (2011): 306-311.

Seed, Patrick C. "The human mycobiome." *Cold Spring Harbor perspectives in medicine* 5.5 (2015): a019810.

Seedorf, Henning, et al. "Bacteria from diverse habitats colonize and compete in the mouse gut." *Cell* 159.2 (2014): 253-266.

Seekatz, Anna M., et al. "Recovery of the gut microbiome following fecal microbiota transplantation." *MBio* 5.3 (2014): e00893-14.

Segata, Nicola, et al. "Metagenomic biomarker discovery and explanation." *Genome biology* 12.6 (2011): 1-18.

Segerman, Bo. "The most frequently used sequencing technologies and assembly methods in different time segments of the bacterial surveillance and RefSeq genome databases." *Frontiers in cellular and infection microbiology* (2020): 571.

Seitz, Helmut K., and Felix Stickel. "Acetaldehyde as an underestimated risk factor for cancer development: role of genetics in ethanol metabolism." *Genes & nutrition* 5.2 (2010): 121-128.

Seitz, Helmut K., and Peter Becker. "Ethanol metabolism and cancer risk." *Ethanol Research & Health* 30.1 (2007): 38.

Sekizuka, Tsuyoshi, et al. "Characterization of *Fusobacterium varium* Fv113-g1 isolated from a patient with ulcerative colitis based on complete genome sequence and transcriptome analysis." *PloS one* 12.12 (2017): e0189319.

Sender, Ron, Shai Fuchs, and Ron Milo. "Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans." *Cell* 164.3 (2016): 337-340.

Seo, Boram, et al. "*Roseburia* spp. abundance associates with ethanol consumption in humans and its administration ameliorates alcoholic fatty liver in mice." *Cell Host & Microbe*27.1 (2020): 25-40.

Seol, Donghyeok, et al. "Microbial identification using rRNA operon region: database and tool for metataxonomics with long-read sequence." *Microbiology Spectrum* (2022): e02017-21.

Severance, Emily G., et al. "*Candida albicans* exposures, sex specificity and cognitive deficits in schizophrenia and bipolar disorder." *NPJ schizophrenia* 2.1 (2016): 1-7.

Severance, Emily G., et al. "Gastrointestinal inflammation and associated immune activation in schizophrenia." *Schizophrenia research* 138.1 (2012): 48-53.

Shafin, Kishwar, et al. "Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes." *Nature biotechnology* 38.9 (2020): 1044-1053.

Shaiber, Alon, and A. Murat Eren. "Composite metagenome-assembled genomes reduce the quality of public genome repositories." *MBio* 10.3 (2019): e00725-19.

Sharon, Gil, et al. "Human gut microbiota from autism spectrum disorder promote behavioral symptoms in mice." *Cell* 177.6 (2019): 1600-1618.

Shen, Qian, Nan Shang, and Pinglan Li. "*In vitro* and *in vivo* antioxidant activity of *Bifidobacterium animalis* 01 isolated from centenarians." *Current microbiology* 62.4 (2011): 1097-1103.

Sherwin, Eoin, et al. "A gut (microbiome) feeling about the brain." *Current opinion in gastroenterology* 32.2 (2016): 96-102.

Sherwin, Eoin, et al. "Microbiota and the social brain." *Science* 366.6465 (2019): eaar2016.

Sherwin, Eoin, Timothy G. Dinan, and John F. Cryan. "Recent developments in understanding the role of the gut microbiota in brain health and disease." *Annals of the New York Academy of Sciences* 1420.1 (2018): 5-25.

Shin, Jongoh, et al. "Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing." *Scientific reports* 6.1 (2016): 1-10.

Shin, Na-Ri, Tae Woong Whon, and Jin-Woo Bae. "Proteobacteria: microbial signature of dysbiosis in gut microbiota." *Trends in biotechnology* 33.9 (2015): 496-503.

Silverman, Justin D., et al. "Measuring and mitigating PCR bias in microbiota datasets." *PLoS computational biology* 17.7 (2021): e1009113.

Simon, Gary L., and Sherwood L. Gorbach. "Intestinal flora in health and disease." *Gastroenterology* 86.1 (1984): 174-193.

Simon, H. Ye, et al. "Benchmarking metagenomics tools for taxonomic classification." *Cell* 178.4 (2019): 779-794.

Simoneau, Joël, et al. "Current RNA-seq methodology reporting limits reproducibility." *Briefings in Bioinformatics* 22.1 (2021): 140-145.

Singer, Jeffrey R., et al. "Preventing dysbiosis of the neonatal mouse intestinal microbiome protects against late-onset sepsis." *Nature medicine* 25.11 (2019): 1772-1782.

Singh, Ramandeep, et al. "Fecal microbiota transplantation against intestinal colonization by extended spectrum beta-lactamase producing Enterobacteriaceae: a proof of principle study." *BMC research notes* 11.1 (2018): 1-6.

Slattery, Martha L., et al. "Energy balance and colon cancer—beyond physical activity." *Cancer research* 57.1 (1997): 75-80.

Smiley, John F., et al. "Selective reduction of cerebral cortex GABA neurons in a late gestation model of fetal ethanol spectrum disorder." *Ethanol* 49.6 (2015): 571-580.

Snijders, Antoine M., et al. "Influence of early life exposure, host genetics and diet on the mouse gut microbiome and metabolome." *Nature microbiology* 2.2 (2016): 1-8.

So, Daniel, et al. "Dietary fiber intervention on gut microbiota composition in healthy adults: a systematic review and meta-analysis." *The American journal of clinical nutrition* 107.6 (2018): 965-983.

Sobhani, Iradj, et al. "Microbial dysbiosis in colorectal cancer (CRC) patients." *PloS one* 6.1 (2011): e16393.

Soderborg, Taylor K., et al. "The gut microbiota in infants of obese mothers increases inflammation and susceptibility to NAFLD." *Nature communications* 9.1 (2018): 1-12.

Sokol, Harry, et al. "Fungal microbiota dysbiosis in IBD." *Gut* 66.6 (2017): 1039-1048.

Song, Mingyang, Andrew T. Chan, and Jun Sun. "Influence of the gut microbiome, diet, and environment on risk of colorectal cancer." *Gastroenterology* 158.2 (2020): 322-340.

Song, Wooseok, et al. "Divergent rRNAs as regulators of gene expression at the ribosome level." *Nature microbiology* 4.3 (2019): 515-526.

Staley, Christopher, et al. "Stable engraftment of human microbiota into mice with a single oral gavage following antibiotic conditioning." *Microbiome* 5.1 (2017): 1-13.

Stärkel, Peter, et al. "Intestinal dysbiosis and permeability: the yin and yang in ethanol dependence and alcoholic liver disease." *Clinical Science* 132.2 (2018): 199-212.

Steel, Jennifer H., et al. "Bacteria and inflammatory cells in fetal membranes do not always cause preterm labor." *Pediatric research* 57.3 (2005): 404-411.

Stewart, Robert D., et al. "Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen." *Nature communications* 9.1 (2018): 1-11.

Stojanov, Spase, Aleš Berlec, and Borut Štrukelj. "The influence of probiotics on the *Firmicutes*/*Bacteroidetes* ratio in the treatment of obesity and inflammatory bowel disease." *Microorganisms*8.11 (2020): 1715.

Strati, Francesco, et al. "Age and gender affect the composition of fungal population of the human gastrointestinal tract." *Frontiers in microbiology* 7 (2016): 1227.

Strati, Francesco, et al. "Altered gut microbiota in Rett syndrome." *Microbiome* 4.1 (2016): 1-15.

Strati, Francesco, et al. "New evidences on the altered gut microbiota in autism spectrum disorders." *Microbiome* 5.1 (2017): 1-11.

Sudo, Nobuyuki, et al. "Postnatal microbial colonization programs the hypothalamic–pituitary–adrenal system for stress response in mice." *The Journal of physiology* 558.1 (2004): 263-275.

Suhr, Mallory J., and Heather E. Hallen-Adams. "The human gut mycobiome: pitfalls and potentials—a mycologist's perspective." *Mycologia* 107.6 (2015): 1057-1073.

Sunagawa, Shinichi, et al. "Metagenomic species profiling using universal phylogenetic marker genes." *Nature methods* 10.12 (2013): 1196-1199.

Surana, Neeraj K., and Dennis L. Kasper. "Moving beyond microbiome-wide associations to causal microbe identification." *Nature* 552.7684 (2017): 244-247.

Swidsinski, Alexander, et al. "Mucosal flora in inflammatory bowel disease." *Gastroenterology* 122.1 (2002): 44-54.

Swidsinski, Alexander, et al. "Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease." *Journal of clinical microbiology* 43.7 (2005): 3380-3389.

Szabo, Gyongyi, et al. "TLR4, ethanol, and lipid rafts: a new mechanism of ethanol action with implications for other receptor-mediated effects." *The Journal of Immunology* 178.3 (2007): 1243-1249.

Szabo, Gyongyi. "Gut–liver axis in alcoholic liver disease." *Gastroenterology* 148.1 (2015): 30-36.

Tabouy, Laure, et al. "Dysbiosis of microbiome and probiotic treatment in a genetic model of autism spectrum disorders." *Brain, Behavior, and Immunity* 73 (2018): 310-319.

Tall, Mamadou Lamine, et al. "*Massilistercora timonensis* gen. nov., sp. nov., a new bacterium isolated from the human microbiota." *New Microbes and New Infections* 35 (2020): 100664.

Tang, Yueming, et al. "The role of miR-212 and iNOS in ethanol-induced intestinal barrier dysfunction and steatohepatitis." *Ethanolism: Clinical and Experimental Research* 39.9 (2015): 1632-1641.

Taş, Neslihan, et al. "Metagenomic tools in microbial ecology research." *Current Opinion in Biotechnology* 67 (2021): 184-191.

Tavanti, Arianna, et al. "*Candida orthopsilosis* and *Candida metapsilosis* spp. nov. to replace *Candida parapsilosis* groups II and III." *Journal of clinical microbiology* 43.1 (2005): 284-292.

Taylor, Louise H., Sophia M. Latham, and Mark EJ Woolhouse. "Risk factors for human disease emergence." *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 356.1411 (2001): 983-989.

Tedersoo, Leho, et al. "Global diversity and geography of soil fungi." *science* 346.6213 (2014): 1256688.

Teng, Haotian, et al. "Chiron: translating nanopore raw signal directly into nucleotide sequence using deep learning." *GigaScience* 7.5 (2018): giy037.

Thion, Morgane Sonia, et al. "Microbiome influences prenatal and adult microglia in a sex-specific manner." *Cell* 172.3 (2018): 500-516.

Thompson, Luke R., et al. "A communal catalogue reveals Earth's multiscale microbial diversity." *Nature* 551.7681 (2017): 457-463.

Tian, Xiaozhu, et al. "*Bifidobacterium breve* ATCC15700 pretreatment prevents alcoholic liver disease through modulating gut microbiota in mice exposed to chronic ethanol intake." *Journal of Functional Foods* 72 (2020): 104045.

Tiew, Pei Yee, et al. "The mycobiome in health and disease: Emerging concepts, methodologies and challenges." *Mycopathologia* 185.2 (2020): 207-231.

Tomas, Julie, Philippe Langella, and Claire Cherbuy. "The intestinal microbiota in the rat model: major breakthroughs from new technologies." *Animal health research reviews* 13.1 (2012): 54-63.

Treangen, Todd J., et al. "Traumatic brain injury in mice induces acute bacterial dysbiosis within the fecal microbiome." *Frontiers in immunology* 9 (2018): 2757.

Tsuruya, Atsuki, et al. "Major anaerobic bacteria responsible for the production of carcinogenic acetaldehyde from ethanol in the colon and rectum." *Ethanol and Ethanolism* 51.4 (2016): 395-401.

Turner, Andrew D., et al. "New invasive nemertean species (*Cephalothrix simula*) in England with high levels of tetrodotoxin and a microbiome linked to toxin metabolism." *Marine drugs* 16.11 (2018): 452.

Uban, Kristina A., et al. "Basal regulation of HPA and dopamine systems is altered differentially in males and females by prenatal ethanol exposure and chronic variable stress." *Psychoneuroendocrinology* 38.10 (2013): 1953-1966.

Ueki, Atsuko, et al. "*Anaerocolumna chitinilytica* sp. nov., a chitin-decomposing anaerobic bacterium isolated from anoxic soil subjected to biological soil disinfestation." *International Journal of Systematic and Evolutionary Microbiology* 71.9 (2021): 004999.

Ueki, Atsuko, et al. "Descriptions of *Anaerotaenia torta* gen. nov., sp. nov. and *Anaerocolumna cellulosilytica* gen. nov., sp. nov. isolated from a methanogenic reactor of cattle waste and reclassification of *Clostridium aminovalericum*, *Clostridium jejuense* and *Clostridium xylanovorans* as *Anaerocolumna* species." *International Journal of Systematic and Evolutionary Microbiology* 66.8 (2016): 2936-2943.

Uesugi, Takehiko, et al. "Toll-like receptor 4 is involved in the mechanism of early ethanol-induced liver injury in mice." *Hepatology* 34.1 (2001): 101-108.

Ulger-Toprak, Nurver, et al. "*Gemella asaccharolytica* sp. nov., isolated from human clinical specimens." *International Journal of Systematic and Evolutionary Microbiology* 60.5 (2010): 1023-1026.

UniProt Consortium. "UniProt: the universal protein knowledgebase." *Nucleic acids research* 46.5 (2018): 2699.

Uritskiy, Gherman V., Jocelyne DiRuggiero, and James Taylor. "MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis." *Microbiome* 6.1 (2018): 1-13.

Ussar, Siegfried, Shiho Fujisaka, and C. Ronald Kahn. "Interactions between host genetics and gut microbiome in diabetes and metabolic syndrome." *Molecular metabolism* 5.9 (2016): 795-803.

Van Best, N., et al. "Bile acids drive the newborn's gut microbiota maturation." Nature communications 11.1 (2020): 1-13.

van der Merwe, Marie, et al. "Time of feeding alters obesity-associated parameters and gut bacterial communities, but not fungal populations, in C57BL/6 male mice." *Current developments in nutrition* 4.2 (2020): nzz145.

van Overbeek, Leonard S., and Kari Saikkonen. "Impact of bacterial–fungal interactions on the colonization of the endosphere." *Trends in plant science* 21.3 (2016): 230-242.

van Tilburg Bernardes, Erik, et al. "Intestinal fungi are causally implicated in microbiome assembly and immune development in mice." *Nature communications* 11.1 (2020): 1-16.

Vartoukian, Sonia R., Richard M. Palmer, and William G. Wade. "Strategies for culture of 'unculturable' bacteria." *FEMS microbiology letters* 309.1 (2010): 1-7.

Vesterlund, Satu, et al. "Adhesion of bacteria to resected human colonic tissue: quantitative analysis of bacterial adhesion and viability." *Research in microbiology* 156.2 (2005): 238-244.

Vieira-Silva, Sara, and Eduardo PC Rocha. "The systemic imprint of growth and its uses in ecological (meta) genomics." *PLoS genetics* 6.1 (2010): e1000808.

Vogtmann, Emily, et al. "Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing." *PloS one* 11.5 (2016): e0155362.

Volgin, Denys V. "Perinatal ethanol exposure leads to prolonged upregulation of hypothalamic GABAA receptors and increases behavioral sensitivity to gaboxadol." *Neuroscience letters* 439.2 (2008): 182-186.

Walters, William A., Zech Xu, and Rob Knight. "Meta-analyses of human gut microbes associated with obesity and IBD." *FEBS letters* 588.22 (2014): 4223-4233.

Walters, William, et al. "Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys." *Msystems* 1.1 (2016): e00009-15.

Wang, Guanhao, et al. "Gut microbiota and relevant metabolites analysis in ethanol dependent mice." *Frontiers in microbiology* (2018): 1874.

Wang, Jingjing, et al. "Core gut bacteria analysis of healthy mice." *Frontiers in microbiology* 10 (2019): 887.

Wang, Mian, Amit K. Roy, and Thomas J. Webster. "Development of chitosan/poly (vinyl ethanol) electrospun nanofibers for infection related wound healing." *Frontiers in physiology* 7 (2017): 683.

Wang, Shaogui, et al. "A mechanistic review of cell death in ethanol-induced liver injury." *Ethanolism: Clinical and Experimental Research* 40.6 (2016): 1215-1223.

Wang, Siming, et al. "Ingestion of *Faecalibaculum rodentium* causes depression-like phenotypes in resilient Ephx2 knock-out mice: A role of brain–gut–microbiota axis via the subdiaphragmatic vagus nerve." *Journal of Affective Disorders*292 (2021): 565-573.

Wang, Xin-Cun, et al. "ITS 1: a DNA barcode better than ITS 2 in eukaryotes?." *Molecular ecology resources* 15.3 (2015): 573-586.

Wang, Yunhao, et al. "Nanopore sequencing technology, bioinformatics and applications." *Nature biotechnology* 39.11 (2021): 1348-1365.

Ward, Tonya L., et al. "Development of the human mycobiome over the first month of life and across body sites." *MSystems* 3.3 (2018): e00140-17.

Warner, Barbara B. "The contribution of the gut microbiome to neurodevelopment and neuropsychiatric disorders." *Pediatric research* 85.2 (2019): 216-224.

Watson, Mick, et al. "poRe: an R package for the visualization and analysis of nanopore sequencing data." *Bioinformatics* 31.1 (2015): 114-115.

Wawina-Bokalanga, Tony, et al. "Complete genome sequence of a new Ebola virus strain isolated during the 2017 Likati outbreak in the Democratic Republic of the Congo." *Microbiology Resource Announcements* 8.20 (2019): e00360-19.

Wei, Po-Li, et al. "Classification of changes in the fecal microbiota associated with colonic adenomatous polyps using a long-read sequencing platform." *Genes* 11.11 (2020): 1374.

West, James R., Wei-Jung A. Chen, and Nicholas J. Pantazis. "Fetal ethanol syndrome: the vulnerability of the developing brain and possible mechanisms of damage." *Metabolic brain disease* 9.4 (1994): 291-322.

Wheeler, Matthew L., et al. "Immunological consequences of intestinal fungal dysbiosis." *Cell host & microbe* 19.6 (2016): 865-873.

Wick, Ryan R., Louise M. Judd, and Kathryn E. Holt. "Performance of neural network basecalling tools for Oxford Nanopore sequencing." *Genome biology* 20.1 (2019): 1-10.

Wilson, D. A., et al. "Developmental ethanol exposure-induced sleep fragmentation predicts adult cognitive impairment." *Neuroscience* 322 (2016): 18-27.

Wilson, Matthew R., et al. "The human gut bacterial genotoxin colibactin alkylates DNA." *Science* 363.6428 (2019): eaar7785.

Wong, Ma-Li, et al. "Inflammasome signaling affects anxiety-and depressive-like behavior and gut microbiome composition." *Molecular psychiatry* 21.6 (2016): 797-805.
World Health Organization. *Global status report on ethanol and health 2018*. World Health Organization, 2019.

Wu, Lu, et al. "Age-related variation of bacterial and fungal communities in different body habitats across the young, elderly, and centenarians in Sardinia." *Msphere* 5.1 (2020): e00558-19.

Wu, Ping-Feng, et al. "Epidemiology and antifungal susceptibility of candidemia isolates of non-albicans *Candida* species from cancer patients: Non-albicans candidemia in cancer patients." *Emerging microbes & infections* 6.1 (2017): 1-7.

Xia, Yinglin, and Jun Sun. "Hypothesis testing and statistical analysis of microbiome." *Genes & diseases* 4.3 (2017): 138-148.

Xia, Yinglin, Jun Sun, and Ding-Geng Chen. "Community diversity measures and calculations." *Statistical analysis of microbiome data with R*. Springer, Singapore, 2018. 167-190.

Xiao, Hui-wen, et al. "Gut microbiota modulates ethanol withdrawal-induced anxiety in mice." *Toxicology letters* 287 (2018): 23-30.

Xiao, Liang, et al. "A catalog of the mouse gut metagenome." *Nature biotechnology* 33.10 (2015): 1103-1108.

Yachida, Shinichi, et al. "Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer." *Nature medicine* 25.6 (2019): 968-976.

Yan, Arthur W., et al. "Enteric dysbiosis associated with a mouse model of alcoholic liver disease." *Hepatology* 53.1 (2011): 96-105.

Yang, An-Ming, et al. "Intestinal fungi contribute to development of alcoholic liver disease." *The Journal of clinical investigation* 127.7 (2017): 2829-2841.

Yang, Huan, et al. "A critical cysteine is required for HMGB1 binding to Toll-like receptor 4 and activation of macrophage cytokine release." *Proceedings of the National Academy of Sciences* 107.26 (2010): 11942-11947.

Yang, Junwon, and Jongsik Chun. "Taxonomic composition and variation in the gut microbiota of laboratory mice." *Mammalian Genome* 32.4 (2021): 297-310.

Yang, Rui-Heng, et al. "Evaluation of the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the analysis of fungal diversity by deep sequencing." *PloS one* 13.10 (2018): e0206428.

Yeung, Frank, et al. "Altered immunity of laboratory mice in the natural environment is associated with fungal colonization." *Cell host & microbe* 27.5 (2020): 809-822.

Yokoyama, Akira, et al. "Salivary acetaldehyde concentration according to alcoholic beverage consumed and aldehyde dehydrogenase-2 genotype." *Ethanolism: clinical and experimental research* 32.9 (2008): 1607-1614.

Yoshioka, Hajime, Ken-ichi Iseki, and Kozo Fujita. "Development and differences of intestinal flora in the neonatal period in breast-fed and bottle-fed infants." *Pediatrics* 72.3 (1983): 317-321.

Yu, Chaoheng, et al. "*Prevotella copri* is associated with carboplatin-induced gut toxicity." *Cell death & disease* 10.10 (2019): 1-15.

Yu, Jinyan, et al. "Maternal exposure to farming environment protects offspring against allergic diseases by modulating the neonatal TLR-Tregs-Th axis." *Clinical and translational allergy* 8.1 (2018): 1-13.

Yu, TaChung, et al. "*Fusobacterium nucleatum* promotes chemoresistance to colorectal cancer by modulating autophagy." *Cell* 170.3 (2017): 548-563.

Zackular, Joseph P., et al. "The human gut microbiome as a screening tool for colorectal cancer." *Cancer prevention research*7.11 (2014): 1112-1121.

Zagato, Elena, et al. "Endogenous murine microbiota member *Faecalibaculum rodentium* and its human homologue protect from intestinal tumour growth." *Nature microbiology* 5.3 (2020): 511-524.

Zakhari, Samir. "Overview: how is ethanol metabolized by the body?." *Ethanol research & health* 29.4 (2006): 245.

Zeller, Georg, et al. "Potential of fecal microbiota for early-stage detection of colorectal cancer." *Molecular systems biology* 10.11 (2014): 766.

Zeng, Y. H., et al. "Long PCR-RFLP of 16 S-ITS-23 S r RNA genes: a high-resolution molecular tool for bacterial genotyping." *Journal of applied microbiology* 114.2 (2013): 433-447.

Zhai, Baohui, et al. "Rapamycin relieves anxious emotion and synaptic plasticity deficits induced by hindlimb unloading in mice." *Neuroscience letters* 677 (2018): 44-48.

Zhang, Chenhong, et al. "Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice." *The ISME journal* 4.2 (2010): 232-241.

Zhang, Jing, et al. "Novel high-docosahexaenoic-acid tuna oil supplementation modulates gut microbiota and alleviates obesity in high-fat diet mice." *Food Science & Nutrition* 8.12 (2020): 6513-6527.

Zhang, Lin, et al. "The role of gut mycobiome in health and diseases." *Therapeutic advances in gastroenterology* 14 (2021): 17562848211047130.

Zhang, Yanan, et al. "Caecal infusion of the short-chain fatty acid propionate affects the microbiota and expression of inflammatory cytokines in the colon in a fistula pig model." *Microbial biotechnology* 11.5 (2018): 859-868.

Zhao, Shijie, et al. "Adaptive evolution within gut microbiomes of healthy people." *Cell host & microbe* 25.5 (2019): 656-667.

Zheng, Peng, et al. "Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism." *Molecular psychiatry* 21.6 (2016): 786-796.

Zheng, Peng, et al. "The gut microbiome from patients with schizophrenia modulates the glutamate-glutamine-GABA cycle and schizophrenia-relevant behaviors in mice." *Science advances* 5.2 (2019): eaau8317.

Zhi, Xiao-Yang, et al. "Prokaryotic systematics in the genomics era." *Antonie Van Leeuwenhoek* 101.1 (2012): 21-34.

Zhou, Da, et al. "Total fecal microbiota transplantation alleviates high-fat diet-induced steatohepatitis in mice via beneficial regulation of gut microbiota." *Scientific reports* 7.1 (2017): 1-11.

Zuo, Kun, et al. "Dysbiotic gut microbes may contribute to hypertension by limiting vitamin D production." *Clinical cardiology* 42.8 (2019): 710-719.

Zuo, Tao, et al. "Bacteriophage transfer during faecal microbiota transplantation in Clostridium difficile infection is associated with treatment outcome." *Gut* 67.4 (2018): 634-643.

Appendices

Appendix I – Illumina Sequencing reads' quality control detail

All sequenced reads passing the standard Illumina chastity filter (PF reads) were demultiplexed according to their index sequences, followed by primer clipping where the target region specific for forward and reverse primer sequences were identified and clipped from the starts of the raw forward and reverse reads. No mismatches were allowed and if primer sequences could not be perfectly matched, read pairs were removed and only high-quality reads were retained. Removed read pairs were not considered for microbial profiling. Next, reads were merged if their forward and reverse ends overlap to obtain a single read covering the entire target region. Reads shorter or longer than 2 times the target region were considered artifacts and only the forward read was retained for further steps. Paired-end reads were merged using the software FLASH (2.2.00, Magoc and Salzberg, 2011). In brief, the FLASH algorithm considers all possible overlaps at or above a minimum length between the reads in a pair and chooses the overlap that results in the lowest proportion of mismatched bases in the overlapped region. Thus, FLASH computes a consensus sequence in the overlapped region by selecting at each overlapped position the base with the higher quality value. If both bases have an identical quality value, one is selected randomly. Pairs were merged with a minimum overlap size of 10 bp to reduce false-positive merges. The next step involved quality filter where merged reads were length filtered according to the expected length variations of the V3-V4 region, *i.e.* ca 445 bps. Merged reads that were significantly shorter or longer than the expected minimal or maximal target region length, respectively, and merged reads containing ambiguous bases ("N") were discarded.

Table A-1 – Illumina sequencing summarized statistics of the 16 DNA templates obtained from the infant mice gut samples.

Variable nominator	**No. reads**	**% reads**
Total number of input sequences	1,987,886	100.0
Remaining sequences after preprocessing and quality filter	1,987,646	100.0
Remaining sequences after chimera detection and filtering	1,981,707	99.7
Total number of sequences assigned to OTUs	1,289,435	64.9
Total number of sequences assigned to taxa	1,289,435	64.9
Total number of OTUs	388	100.0
Number of OTUs assigned to taxa	388	100.0

Table A-2 – Illumina sequencing summarized statistics separated by sample.

Sample name	No. input sequences	% Sequences after preprocessing and chimera removal	% Sequences assigned to OTUs/taxa	No. sequences after lineage-specific copy-number correction	Median sequence length after preprocessing
Control 9	114,046	99.5	64.1	25,290	399
Control 11	105,139	99.7	66.7	27,292	399
Control 13	116,983	99.7	66.9	30,114	399
Control 14	124,204	99.7	63.0	27,106	399
Control 16	160,911	99.7	65.4	35,949	406
Control 17	127,172	99.7	64.9	28,839	399
Control 18	113,525	99.6	65.1	26,540	399
Test 1	139,368	99.8	64.9	31,094	399
Test 2	121,680	99.9	68.7	27,097	417
Test 3	130,033	99.8	64.8	30,814	286
Test 4	149,684	99.6	65.6	33,874	399
Test 8	136,208	99.5	62.2	30,873	406
Test 9	107,284	99.6	63.2	23,214	299
Test 11	90,784	99.9	63.6	22,668	417
Test 13	124,635	99.8	64.3	27,423	399
Test 14	126,230	99.5	64.3	28,865	286

Appendix II – Detailed description of the statistical test rationale performed for beta-diversity analysis using Bray-Curtis and Jaccard dissimilarity indices

Two of the most used tests for group difference analysis are PERMANOVA and ANOSIM. PERMANOVA is a widely used non-parametric method of multivariate analysis of variance based on pairwise distances (Anderson, 2001), which aim is to fit multivariate models to microbiome data (Xia *et al*., 2017). PERMANOVA calculates the differences between the groups centroids or means, as well how much of the variance can be explained by the specified group (through R^2). ANOSIM is a non-parametric procedure that tests for significant differences between two or more groups obtained by any distance measure (Chong *et al*., 2020). The rationale is that the mean of ranks of distances between groups are compared with the mean of ranks of distances within groups. The obtained *R* test statistic measures how separate the community structure is (0 – no separation; 1 – separation). *R* values > 0.75 means that there is dissimilarity between the groups, *R* values < 0.25 means that no significant dissimilarity was found, and *R* values > 0.5 means that groups are reasonable dissimilar. In summary, PERMANOVA tests whether distance differ between groups, and ANOSIM calculates whether distances between groups are greater than within groups. Additionally, all experimental groups should have comparable within-group dispersion to avoid false significant results (type II errors) (Ramette *et al*., 2007). Therefore, ANOSIM should be performed in addition to PERMANOVA to give suggestions about the degree of separation between the experimental groups (Giatsis *et al*., 2014). As an example, ANOSIM's *R* near 1 indicates that rank similarity between samples within a group is different/lower than the average rank similarity between samples belonging to different groups, and 0 means that such rank similarity between groups is greater than within groups. This means that *R* values near to 1 represent differences in the experimental groups. Since neither PERMANOVA nor ANOSIM clearly suppose common variance among experimental groups, both approaches are sensitive to differences in multivariate dispersion (however, PERMANOVA more sensitive than ANOSIM) (Giatsis *et al*., 2014). Consequently, PERMDISP should be performed in the first place to test the hypothesis of equality within group dispersion (*i.e*, functional similarity). PERMDSIP analysis is a multivariate non-Euclidean equivalent to traditional Levene's test (Anderson, 2006). PERMDSIP's use is necessary to avoid false-significant differences from PERMANOVA and ANOSIM tests, and to give indications about within and between group variation (Giatsis *et al*., 2014). The rationale of

PERMDISP is to calculate homogeneity of dispersion among groups as an average distance of experimental group samples from the group's centroid. Indeed, PERMDISP tests the null hypothesis of no difference between groups dispersion. As PERMANOVA assumes homogeneity of multivariate dispersion between groups, its use is only recommended if PERMDISP previously accepted the null hypothesis.

Appendix III – Univariate, LEfSe, and Random Forest analyses' details

One of the most common methods used for exploratory data analysis is univariate analysis methods. They serve to identify differentially abundant features in microbiome data through statistical comparisons based on single grouping variable or experimental factor (Chong *et al*., 2020). Linear Discriminant analysis (LDA) effect size (LEfSe) is also employed for biomarker discovery and explanation in high-dimensional metagenomic data (Segata *et al*., 2011). This method incorporates statistical significance with biological consistency (effect size) estimation. Specifically, it performs non-parametric factorial Kruskal-Wallis sum-rank test to identify features with significant differential abundance respective to an experimental factor or class of interest, followed by LDA to calculate the effect size of each differentially abundant features (Chong *et al*., 2020). The LDA score is interpreted as the effect size of all features against all groups or classes. Another method used for metagenomic analysis is Random Forests (FR). RF is a supervised learning algorithm suitable for high dimensional data analysis. It uses an ensemble of classification trees, each of which is grown by random feature selection from a bootstrap sample at each branch. Class prediction is based on the majority vote of the ensemble. Out-of-the bag (OOB) error and variable importance measure is also provided. When compiling tree construction, about a third of the instances are neglected for the bootstrap sample. Such data is considered OOB and it used thereafter as test sample to obtain an unbiased estimate of the classification error (OOB error) (Chong *et al*., 2020OOB is an indicator of how good the feature under examination can predict the results, *i.e.*, the lower OOB is, better the prediction capability of the feature. In MicrobiomeAnalyst, RF analysis is performed using *randomforest* package (Liaw and Wiener, 2002).

$$P = (1\text{-}OOB) \times 100$$

In this work, the features are taxa and through RF, they are ranked by the mean decrease in classification accuracy (or mean Gini score) when they are permuted. Mean decrease in accuracy is defined by the normalized difference in the classification accuracy when the target feature is included *vs* when data is randomly permuted. Thus, mean decrease accuracy shows how classification errors are reduced by the inclusion of each taxon. Model accuracy is calculated by the OOB error estimate, which represents an approximation of how frequently an individual is misclassified. The resulting plot – called

Mean Decrease Accuracy – denotes how much accuracy the model losses by excluding each taxon. The mean decrease in Gini coefficient measures the capacity of each taxon to contribute to the homogeneity of the nodes and leaves in the resulting random forest. The taxa are presented from descending order, and the higher the value of mean decrease accuracy, more relevant is the taxon. This outcome is fundamental because it explains how important each feature is in classifying the data in the RF and how the accuracy suffers by removing each taxon (Martinez-Taboada *et al*., 2020). Higher Mean Decrease Accuracy coefficient values mean perfect inequality (without feature *A*) and values closer to 0 correspond to perfect equality (without feature *A*). If a negative value is observed, the interpretation is that the feature(s) do not show a reasonably ability to predict the classification whatsoever.

Appendix IV – Microbial composition retrieved by WIMP displaying an inaccurate taxonomical assignment

WIMP pipeline in EPI2ME software retrieves a taxonomic tree for the basecalled nanopore reads. Interestingly, the *centrifuge* classification algorithm was unable to correctly assign the composition and abundance of the fungal mock community under study.

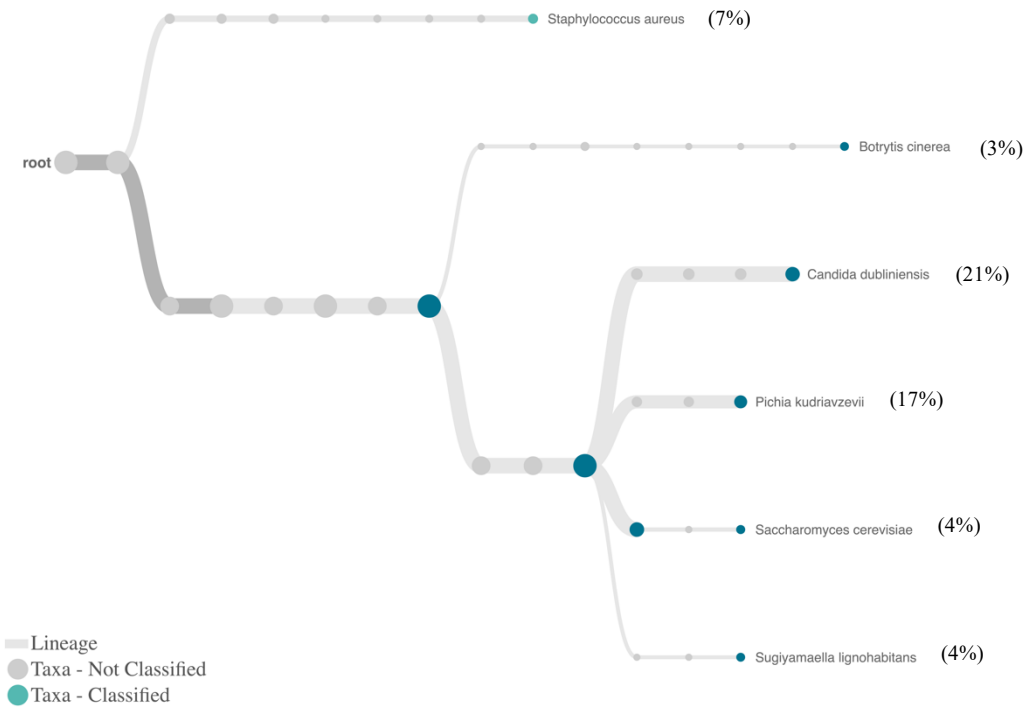


Figure A-1 – Heat trees displaying the community structure of the Mock community A. A total of 1,314,572 reads were analysed with WIMP, from which 1,084,471 were classified. The average quality score was 9.43 and the average sequence length was 550 bps. Only taxa with relative abundances above 3% are showed (in parenthesis are depicted their relative abundances).

Appendix V – No statistically significant taxa were associated with the experimental groups by Linear Discriminant Analysis (LDA) Effect Size (LEfSe) scores

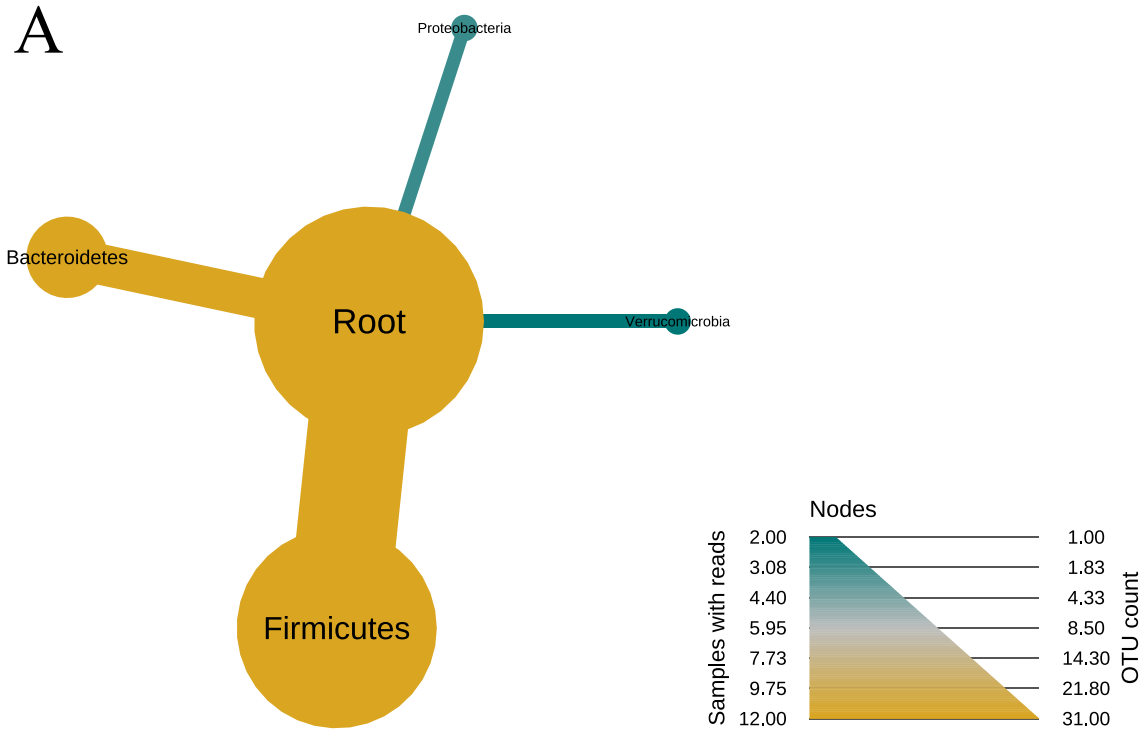
No taxon was associated with any of the experimental approaches. This is an indication of the null effect of ethanol exposure on the microbial infant mice gut composition.

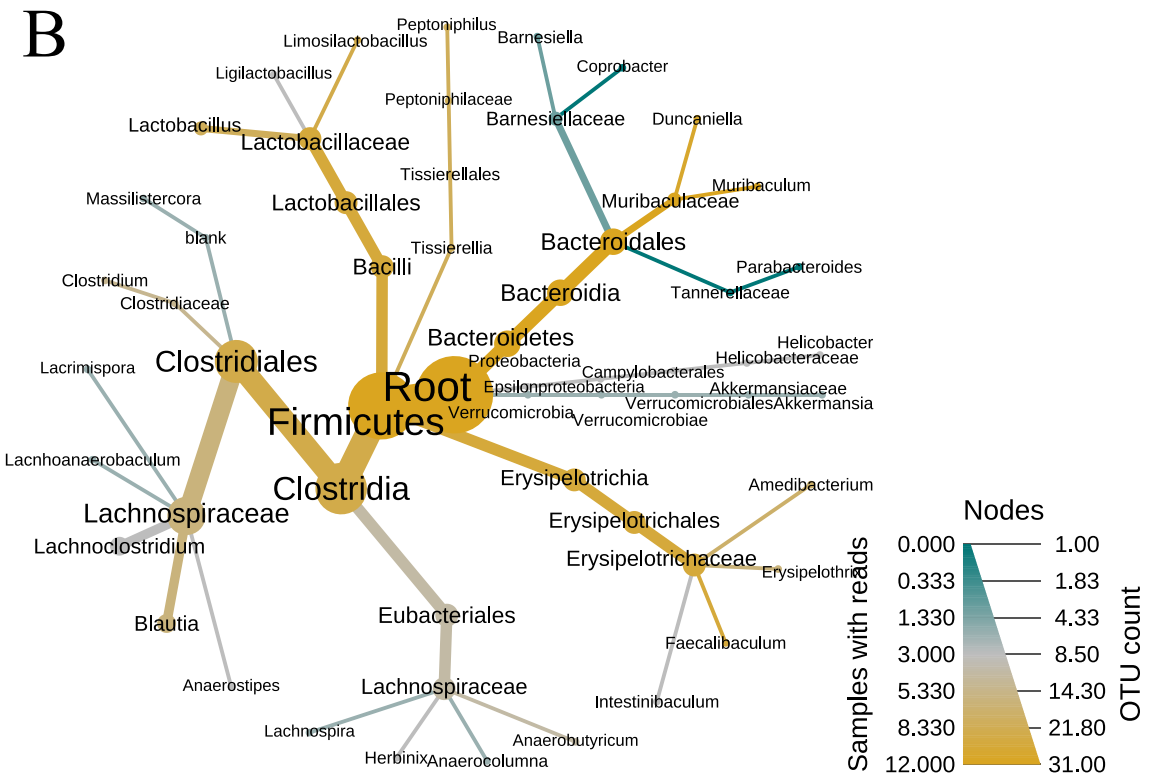
Table A-3 – Linear Discriminant Analysis (LDA) Effect Size (LEfSe) showing the most relevant taxa associated ethanol-exposed and control groups.

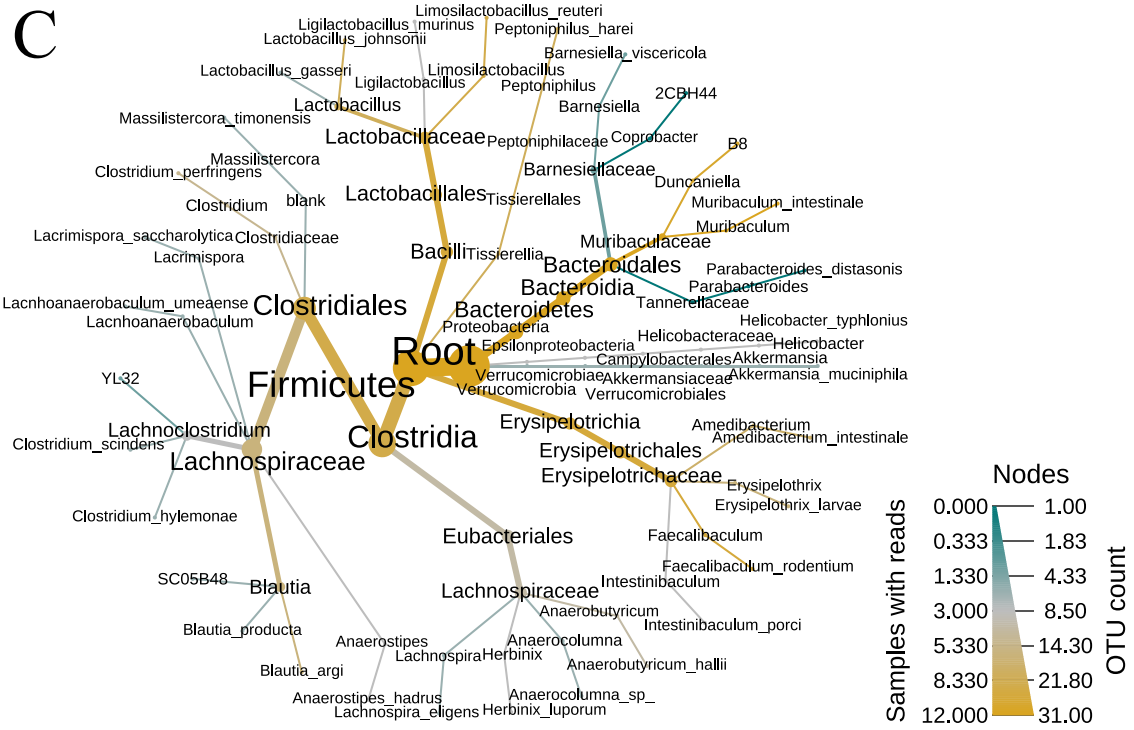
Taxon	**p-value**	**FDR**	**Control Group**	**Ethanol-exposed group**	**log LDA score**
Clostridium hylemonae	0.0197	0.34	0.0075	0	-2.8
Blautia sp.	0.0495	0.34	0.0038	0	-3.1
Massilistercora timonensis	0.0505	0.34	0.0100	0	-2.6
Anaerocolumna sp.	0.0505	0.34	0.0050	0	-3.0
Lachnoclostridium phocaeense	0.1213	0.52	0.0050	0	-3.0

Appendix VI – Heat trees depicting the microbial profiles of the Test and Optimization set of samples obtained by nanopore sequencing.

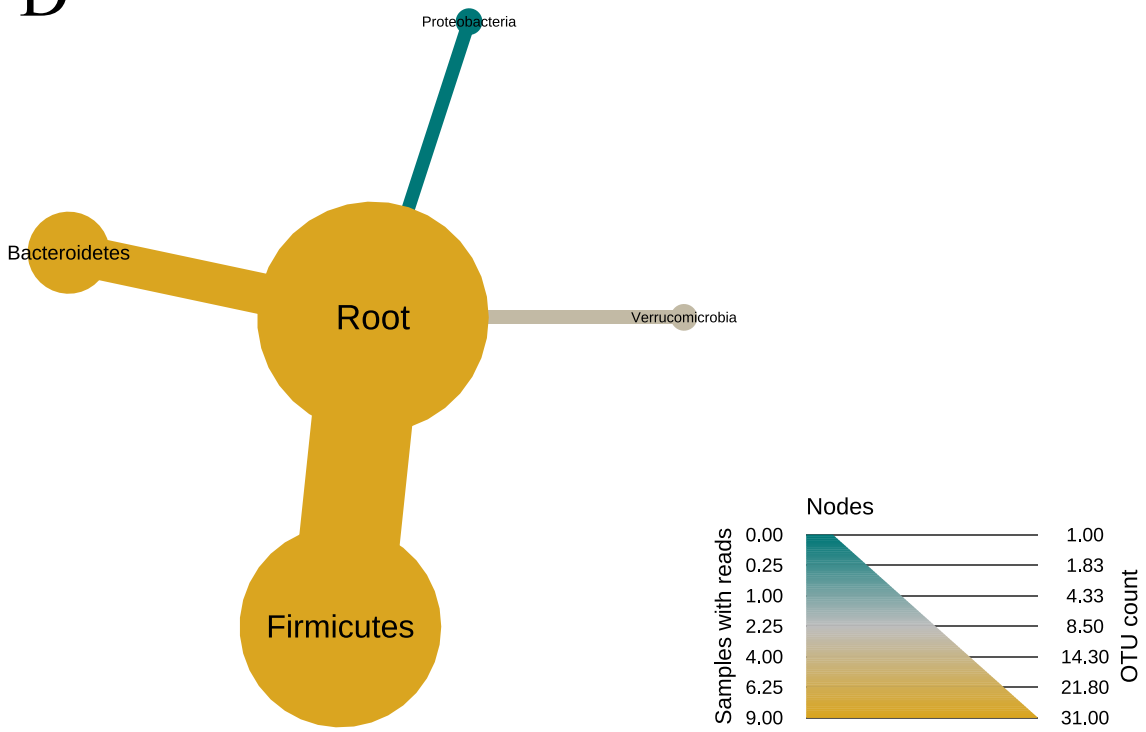
In this study, two set of infant mice gut samples were sequenced by nanopore sequencing using a targeted approach based on the full-length 16S rRNA gene. Interestingly, significant differences were observed albeit the same ethanol exposure experimental procedure.

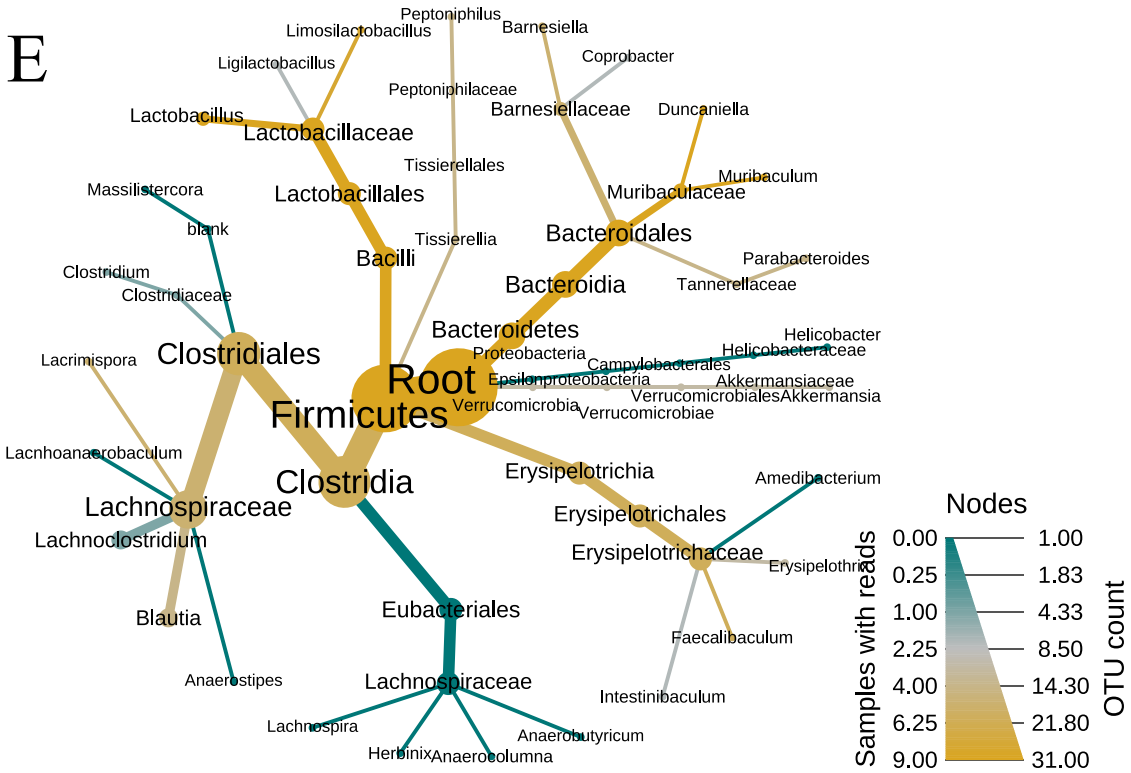



B
Peptoniphilus
Limosilactobacillus
Barnesiella
Ligilactobacillus
Coprobacter
Peptoniphilaceae
Barnesiellaceae
Duncaniella
Lactobacillus
Lactobacillaceae
Tissierellales
Muribaculum
Massilistercora
Lactobacillales
Muribaculaceae
blank
Bacteroidales
Tissierellia
Bacilli
Parabacteroides
Clostridium
Bacteroidia
Tannerellaceae
Clostridiaceae
Bacteroidetes
Helicobacter
Lacrimispora
Clostridiales
Proteobacteria
Helicobacteraceae
Campylobacterales
Root
Epsilonproteobacteria
Akkermansiaceae
Verrucomicrobia
VerrucomicrobialesAkkermansia
Firmicutes
Verrucomicrobiae
Lacnhoanaerobaculum
Erysipelotrichia
Amedibacterium
Clostridia
Lachnospiraceae
Erysipelotrichales
Lachnoclostridium
Erysipelotrichaceae
Erysipelothrix
Eubacteriales
Blautia
Faecalibaculum
Anaerostipes
Lachnospiraceae
Intestinibaculum
Lachnospira
Anaerobutyricum
Herbinix
Anaerocolumna
Nodes
Samples with reads
0.000
0.333
1.330
3.000
5.330
8.330
12.000
1.00
1.83
4.33
8.50
14.30
21.80
31.00
OTU count


C
Limosilactobacillus_reuteri
Ligilactobacillus_murinus
Peptoniphilus_harei
Lactobacillus_johnsonii
Barnesiella_viscericola
Lactobacillus_gasseri
Limosilactobacillus
Ligilactobacillus
Peptoniphilus
2CBH44
Lactobacillus
Barnesiella
Massilistercora_timonensis
Lactobacillaceae
Peptoniphilaceae
Coprobacter
B8
Massilistercora
Barnesiellaceae
Clostridium_perfringens
Duncaniella
Clostridium
blank
Lactobacillales
Tissierellales
Muribaculum_intestinale
Muribaculum
Lacrimispora_saccharolytica
Clostridiaceae
Muribaculaceae
Lacrimispora
Bacilli
Tissierellia
Bacteroidales
Parabacteroides_distasonis
Bacteroidia
Parabacteroides
Lachnoanaerobaculum_umeaense
Clostridiales
Tannerellaceae
Bacteroidetes
Lachnoanaerobaculum
Helicobacter_typhlonius
Proteobacteria
Root
Helicobacteraceae
Epsilonproteobacteria
Helicobacter
Campylobacterales
Akkermansia
YL32
Verrucomicrobiae
Akkermansiaceae
Akkermansia_muciniphila
Firmicutes
Verrucomicrobia
Verrucomicrobiales
Erysipelotrichia
Amedibacterium
Lachnoclostridium
Clostridia
Amedibacterium_intestinale
Clostridium_scindens
Erysipelotrichales
Lachnospiraceae
Erysipelotrichaceae
Erysipelothrix
Erysipelothrix_larvae
Clostridium_hylemonae
Eubacteriales
Faecalibaculum
Faecalibaculum_rodentium
SC05B48
Blautia
Intestinibaculum
Lachnospiraceae
Anaerobutyricum
Blautia_producta
Intestinibaculum_porci
Anaerostipes
Anaerocolumna
Lachnospira
Herbinix
Blautia_argi
Anaerobutyricum_hallii
Anaerostipes_hadrus
Anaerocolumna_sp_
Lachnospira_eligens
Herbinix_luporum
Nodes
Samples with reads
0.000
0.333
1.330
3.000
5.330
8.330
12.000
1.00
1.83
4.33
8.50
14.30
21.80
31.00
OTU count

D


Proteobacteria
Bacteroidetes
Root
Verrucomicrobia
Firmicutes
Nodes
Samples with reads
0.00
0.25
1.00
2.25
4.00
6.25
9.00
1.00
1.83
4.33
8.50
14.30
21.80
31.00
OTU count


E
Peptoniphilus
Limosilactobacillus
Barnesiella
Ligilactobacillus
Coprobacter
Peptoniphilaceae
Barnesiellaceae
Duncaniella
Lactobacillus
Lactobacillaceae
Tissierellales
Muribaculum
Massilistercora
Lactobacillales
Muribaculaceae
blank
Tissierellia
Bacteroidales
Bacilli
Parabacteroides
Clostridium
Bacteroidia
Clostridiaceae
Tannerellaceae
Bacteroidetes
Helicobacter
Clostridiales
Proteobacteria
Helicobacteraceae
Lacrimispora
Campylobacterales
Root
Epsilonproteobacteria
Akkermansiaceae
Firmicutes
Verrucomicrobia
VerrucomicrobialesAkkermansia
Verrucomicrobiae
Lacnhoanaerobaculum
Erysipelotrichia
Clostridia
Amedibacterium
Lachnospiraceae
Erysipelotrichales
Lachnoclostridium
Erysipelotrichaceae
Erysipelothrix
Eubacteriales
Blautia
Faecalibaculum
Anaerostipes
Lachnospiraceae
Intestinibaculum
Lachnospira
Anaerobutyricum
Herbinix
Anaerocolumna
Nodes
Samples with reads
0.00
0.25
1.00
2.25
4.00
6.25
9.00
1.00
1.83
4.33
8.50
14.30
21.80
31.00
OTU count

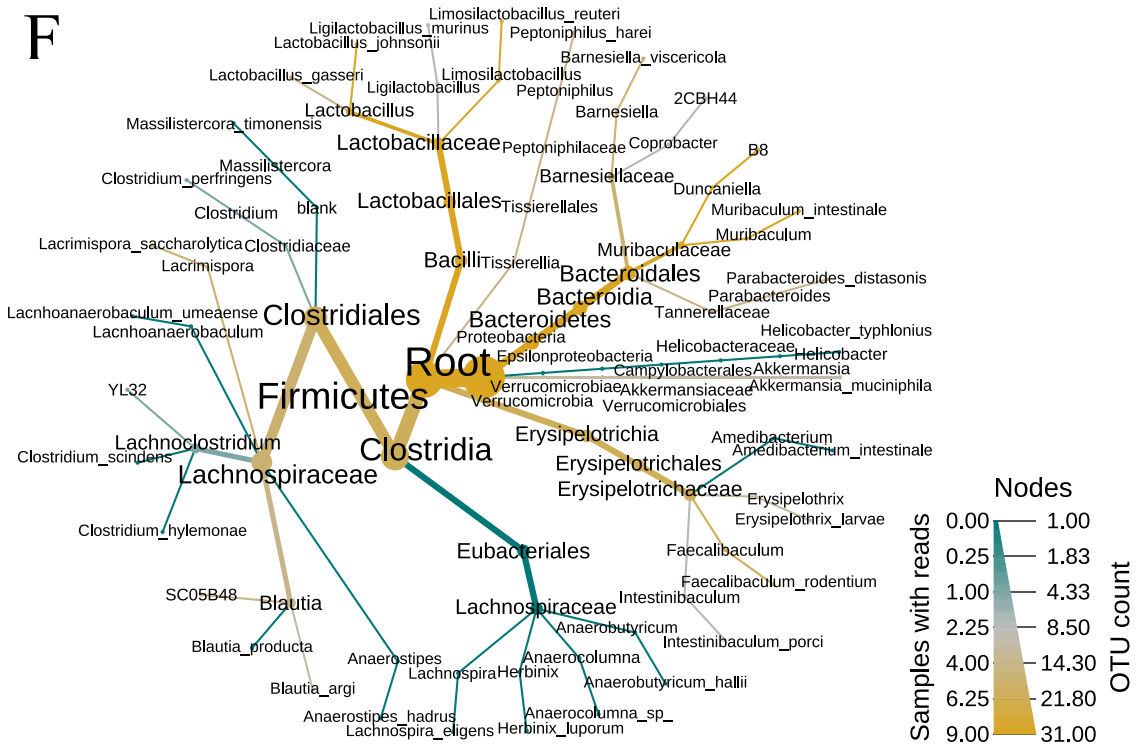


Figure A-2 – Heat trees displaying the community structure of the two set of samples. Size and colour of nodes and edges represent correlations with the abundance of organisms in each community. Only significantly different (Wilcoxon p-vale < 0.05) taxa are labelled. 'test' represents the combined test set of samples, and 'optimization' represents the combined nanopore sequencing optimization set of samples. A, B, and C display data from the Optimization samples. D, E, and F display data from the Test samples. A and D, represents the comparison at the phylum level, B and E at the genus level, and C and F at the species level.

Appendix VII – Correlation network of combined species detected in the Test and Optimized set of samples

The correlation network shows how the infant mice gut microbes may be associated with each other, showing a dense community effect that may be relevant to the host.

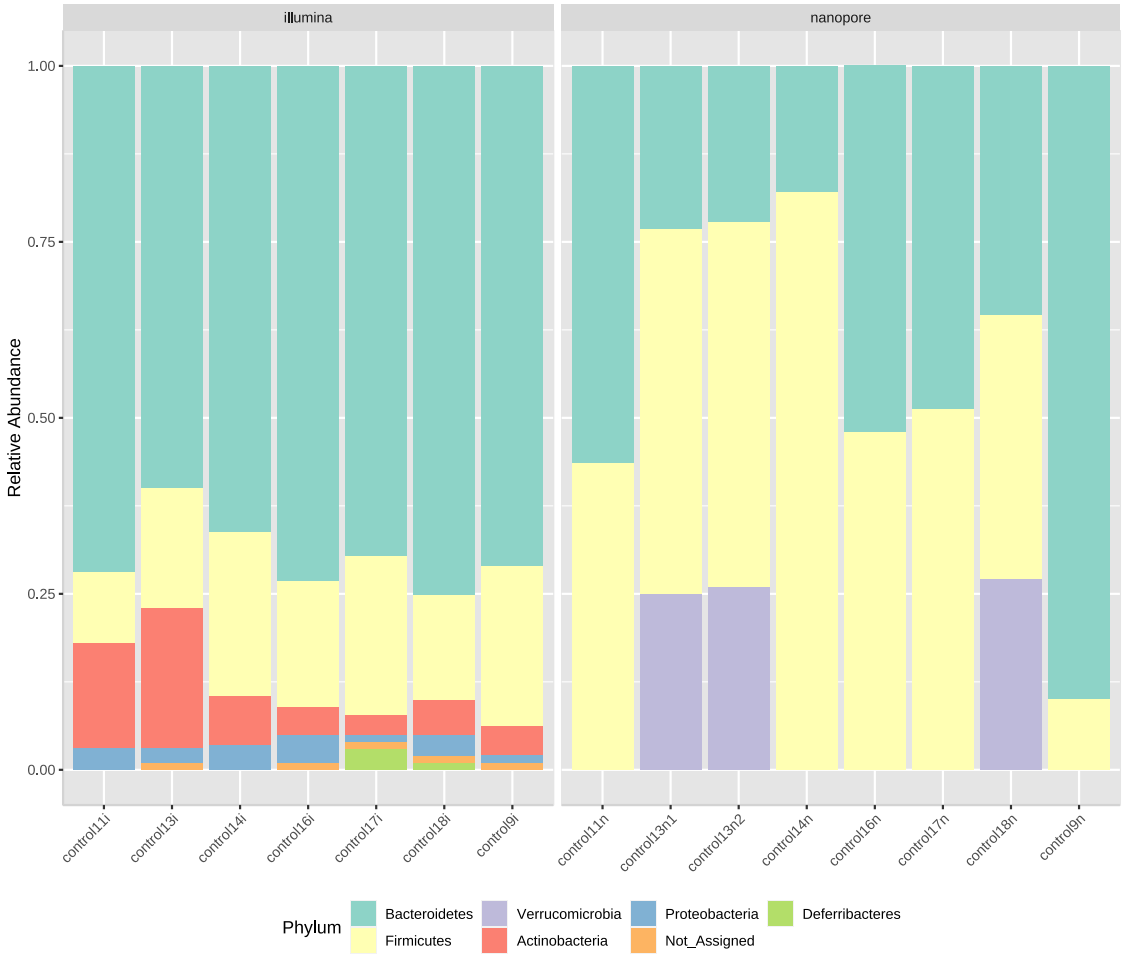
Table A-4 – Correlation table displaying all the species with Spearman correlation coefficient (*rho*) above 0.8.

Taxon 1	**Taxon 2**	***rho***	**p-value**	**T-stat**
Anaerobutyricum hallii	*Anaerostipes hadrus*	0.8	0	> 244
Anaerobutyricum hallii	*Herbinix luporum*	0.8	0	> 244
Anaerocolumna sp.	*Anaerostipes hadrus*	0.8	0	> 282
Anaerocolumna sp.	*Clostridium hylemonae*	1.0	0	> 1
Anaerocolumna sp.	*Clostridium scindens*	1.0	0	0
Anaerocolumna sp.	*Herbinix luporum*	0.8	0	> 282
Anaerocolumna sp.	*Lacnhoanaerobaculum umeaense*	1.0	0	> 1
Anaerocolumna sp.	*Massilistercora timonensis*	1.0	0	> 1
Anaerostipes hadrus	*Clostridium hylemonae*	0.8	0	> 285
Anaerostipes hadrus	*Clostridium scindens*	0.8	0	> 282
Anaerostipes hadrus	*Herbinix luporum*	1.0	0	0
Anaerostipes hadrus	*Lacnhoanaerobaculum umeaense*	0.8	0	> 285
Anaerostipes hadrus	*Massilistercora timonensis*	0.8	0	> 285
Barnesiella viscericola	*Duncaniella* sp. B8	0.8	1.0E-04	> 366
Barnesiella viscericola	*Muribaculum intestinale*	0.8	0	> 335
Clostridium hylemonae	*Clostridium scindens*	01.0	0	> 1
Clostridium hylemonae	*Herbinix luporum*	0.8	0	> 285
Clostridium hylemonae	*Lacnhoanaerobaculum umeaense*	1.0	0	0
Clostridium hylemonae	*Massilistercora timonensis*	1.0	0	0
Clostridium perfringens	*Peptoniphilus harei*	0.8	0	> 278
Clostridium scindens	*Herbinix luporum*	0.8	0	> 282
Clostridium scindens	*Lacnhoanaerobaculum umeaense*	1.0	0	> 1
Clostridium scindens	*Massilistercora timonensis*	1.0	0	> 1
Duncaniella sp. B8	*Muribaculum intestinale*	1.0	0	> 89
Erysipelothrix larvae	*Faecalibaculum rodentium*	0.8	1.0E-04	> 380
Herbinix luporum	*Lacnhoanaerobaculum umeaense*	0.8	0	> 285
Herbinix luporum	*Massilistercora timonensis*	0.8	0	> 285
Lacnhoanaerobaculum umeaense	*Massilistercora timonensis*	1.0	0	0
Lactobacillus gasseri	*Lactobacillus johnsonii*	0.8	0	> 308
Lactobacillus johnsonii	*Limosilactobacillus reuteri*	0.8	0	> 247

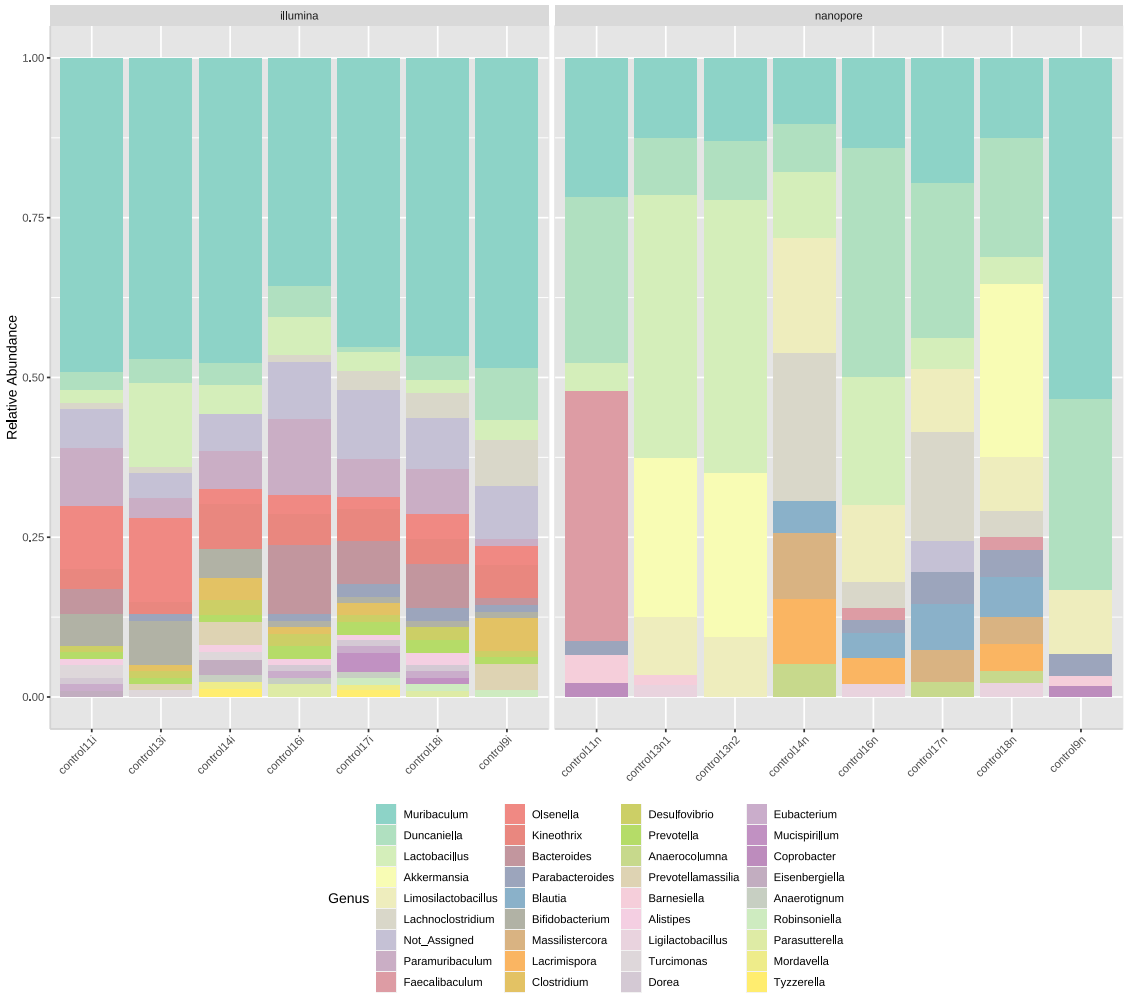
Appendix VIII – Microbial profile comparison between nanopore and Illumina sequencing in the same experimental mice group

A different microbial profile was observed when comparing the same infant mice gut samples analysed by nanopore and Illumina sequencing. These differences pose questions in terms of the reliability of using nanopore sequencing approach alone.

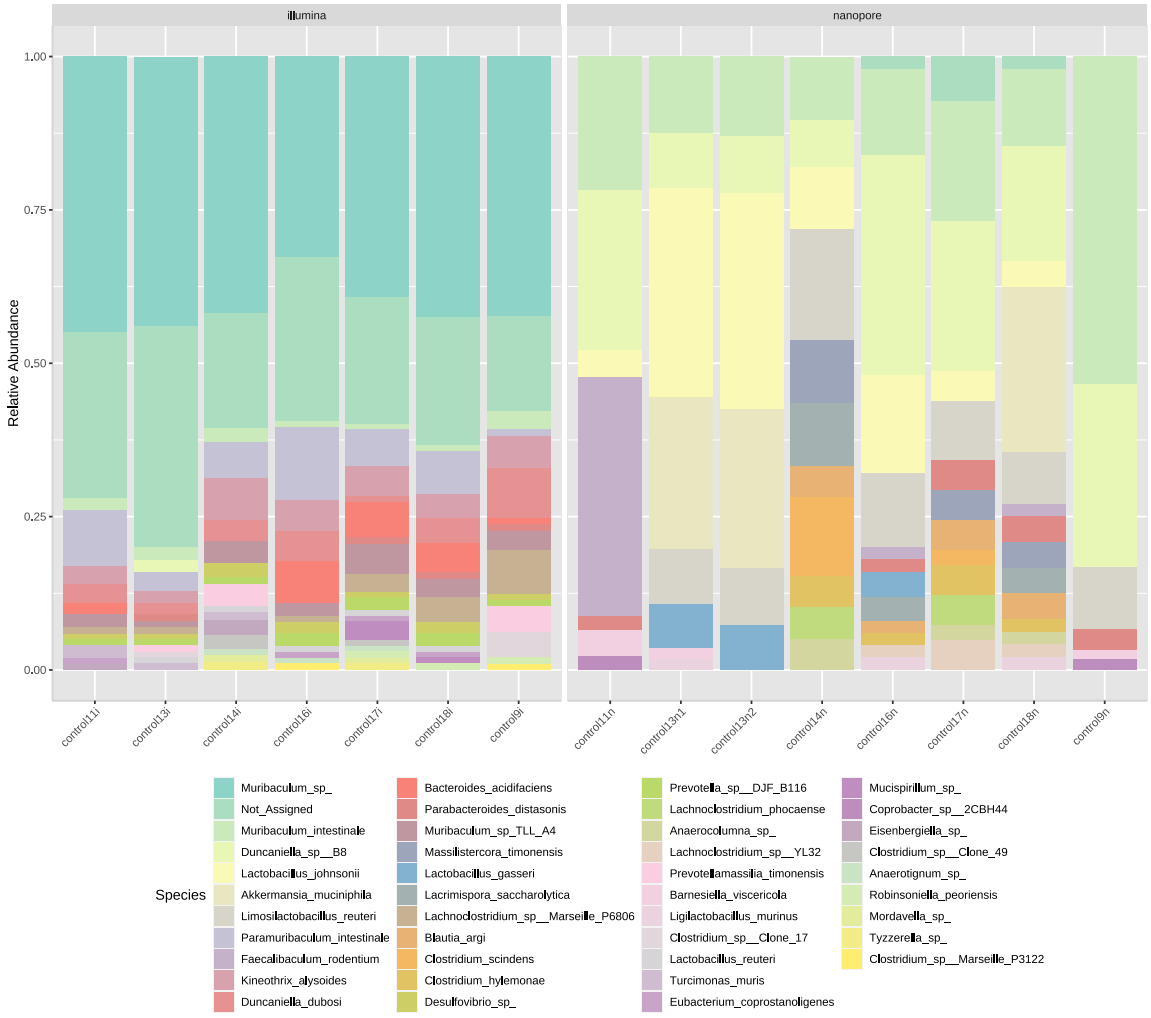
A1



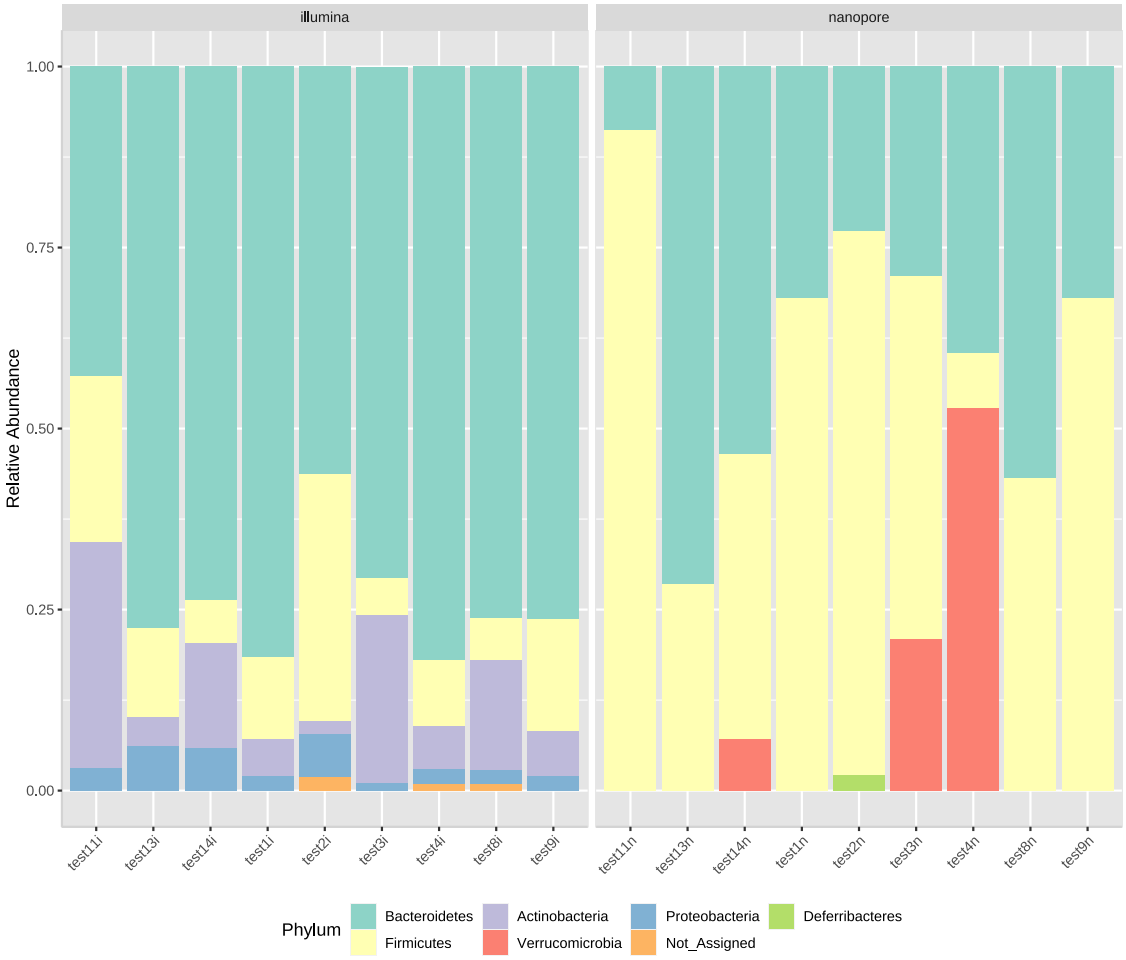
A2


illumina
nanopore
Relative Abundance
1.00
0.75
0.50
0.25
0.00
control11i
control13i
control14i
control16i
control17i
control18i
control9i
control11n
control13n1
control13n2
control14n
control16n
control17n
control18n
control9n
Genus
Muribaculum
Duncaniella
Lactobacillus
Akkermansia
Limosilactobacillus
Lachnoclostridium
Not_Assigned
Paramuribaculum
Faecalibaculum
Olsenella
Kineothrix
Bacteroides
Parabacteroides
Blautia
Bifidobacterium
Massilistercora
Lacrimispora
Clostridium
Desulfovibrio
Prevotella
Anaerocolumna
Prevotellamassilia
Barnesiella
Alistipes
Ligilactobacillus
Turcimonas
Dorea
Eubacterium
Mucispirillum
Coprobacter
Eisenbergiella
Anaerotignum
Robinsoniella
Parasutterella
Mordavella
Tyzzerella

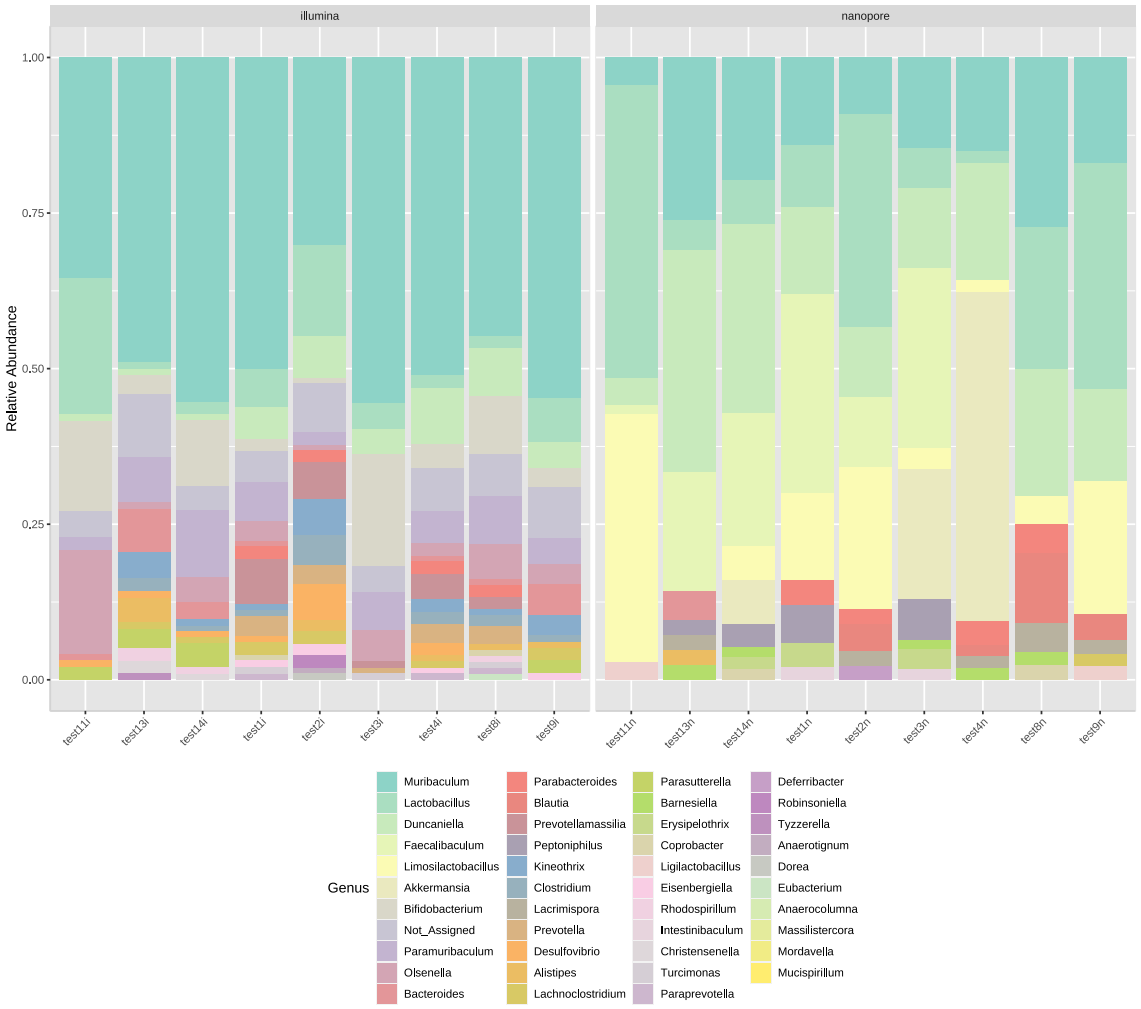
A3


illumina
nanopore
Relative Abundance
1.00
0.75
0.50
0.25
0.00
control11i
control13i
control14i
control16i
control17i
control18i
control9i
control11n
control13n1
control13n2
control14n
control16n
control17n
control18n
control9n
Species
Muribaculum_sp_
Not_Assigned
Muribaculum_intestinale
Duncaniella_sp__B8
Lactobacillus_johnsonii
Akkermansia_muciniphila
Limosilactobacillus_reuteri
Paramuribaculum_intestinale
Faecalibaculum_rodentium
Kineothrix_alysoides
Duncaniella_dubosi
Bacteroides_acidifaciens
Parabacteroides_distasonis
Muribaculum_sp_TLL_A4
Massilistercora_timonensis
Lactobacillus_gasseri
Lacrimispora_saccharolytica
Lachnoclostridium_sp__Marseille_P6806
Blautia_argi
Clostridium_scindens
Clostridium_hylemonae
Desulfovibrio_sp_
Prevotella_sp__DJF_B116
Lachnoclostridium_phocaense
Anaerocolumna_sp_
Lachnoclostridium_sp__YL32
Prevotellamassilia_timonensis
Barnesiella_viscericola
Ligilactobacillus_murinus
Clostridium_sp__Clone_17
Lactobacillus_reuteri
Turcimonas_muris
Eubacterium_coprostanoligenes
Mucispirillum_sp_
Coprobacter_sp__2CBH44
Eisenbergiella_sp_
Clostridium_sp__Clone_49
Anaerotignum_sp_
Robinsoniella_peoriensis
Mordavella_sp_
Tyzzerella_sp_
Clostridium_sp__Marseille_P3122

B1


illumina
nanopore
Relative Abundance
1.00
0.75
0.50
0.25
0.00
test11i
test13i
test14i
test1i
test2i
test3i
test4i
test8i
test9i
test11n
test13n
test14n
test1n
test2n
test3n
test4n
test8n
test9n
Phylum
Bacteroidetes
Firmicutes
Actinobacteria
Verrucomicrobia
Proteobacteria
Not_Assigned
Deferribacteres

B2


illumina
nanopore
Relative Abundance
1.00
0.75
0.50
0.25
0.00
test11i
test13i
test14i
test1i
test2i
test3i
test4i
test8i
test9i
test11n
test13n
test14n
test1n
test2n
test3n
test4n
test8n
test9n
Genus
Muribaculum
Lactobacillus
Duncaniella
Faecalibaculum
Limosilactobacillus
Akkermansia
Bifidobacterium
Not_Assigned
Paramuribaculum
Olsenella
Bacteroides
Parabacteroides
Blautia
Prevotellamassilia
Peptoniphilus
Kineothrix
Clostridium
Lacrimispora
Prevotella
Desulfovibrio
Alistipes
Lachnoclostridium
Parasutterella
Barnesiella
Erysipelothrix
Coprobacter
Ligilactobacillus
Eisenbergiella
Rhodospirillum
Intestinibaculum
Christensenella
Turcimonas
Paraprevotella
Deferribacter
Robinsoniella
Tyzzerella
Anaerotignum
Dorea
Eubacterium
Anaerocolumna
Massilistercora
Mordavella
Mucispirillum

B3

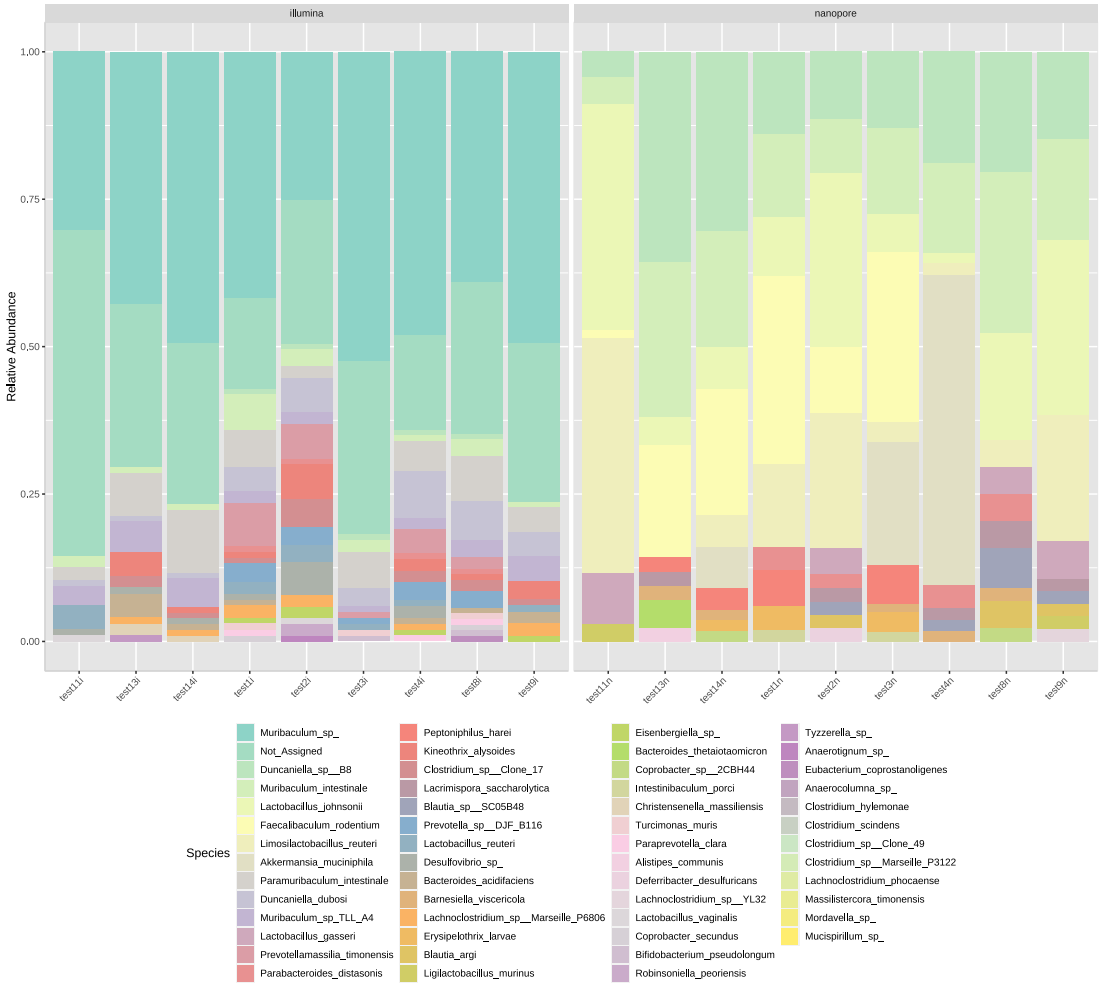


Figure A-3 – Stacked bar plot displaying percentage abundances comparison between nanopore and Illumina sequencing approaches separated by experimental group samples. A, control experimental group; B, ethanol-exposed experimental group; 1, taxa assigned to the phylum level; 2, taxa assigned to the genus level; 3, taxa assigned to the species level.

Appendix IX – Taxa correlated with nanopore or Illumina sequencing

The correlation network here displayed shows the taxa correlated with one of the sequencing approaches. These correlations may be seen as sequencing biases and care must be taken when detecting them.

Table A-5 – Correlation table showing species correlated (Spearman correlation coefficient correlations - *rho*) with one sequencing approach.

Sequencing approach	Taxa[1]	*rho*	p-value	T-stat	FDR
Nanopore	s_*Lactobacillus johnsonii*	0.9	2.5E-11	624	1.6E-10
	s_*Duncaniella* sp. B8	0.9	2.6E-11	644	1.6E-10
	s_*Muribaculum intestinale*	0.8	1.8E-9	872	1.1E-8
	s_*Limosilactobacillus reuteri*	0.8	2.5E-9	892	1.2E-8
	s_*Faecalibaculum rodentium*	0.6	1.8E-4	2,100	5.8E-4
	s_*Barnesiella viscericola*	0.6	5.8E-4	2,320	0.002
	s_*Lacrimispora saccharolytica*	0.6	6.5E-4	2,341	0.002
	s_*Blautia argi*	0.5	0.002	2,590	0.005
	s_*Lactobacillus gasseri*	0.5	0.006	2,851	0.012
	s_*Blautia* sp. SC05B48	0.4	0.014	3,114	0.025
	s_*Ligilactobacillus murinus*	0.4	0.014	3,114	0.025
	s_*Akkermansia muciniphila*	0.4	0.014	3,118	0.025
	s_*Peptinophilus harei*	0.4	0.014	3,118	0.025
	s_*Coprobacter* sp. 2CBH44	0.4	0.033	3,394	0.049
	s_*Erysipelothrix larvae*	0.4	0.033	3,397	0.049
	s_*Lachnoclostridium* sp. YL32	0.4	0.033	3,397	0.049
	s_*Clostridium hylemonae*	0.4	0.033	3,398	0.049
Illumina	f_*Muribaculaceae*	0.9	6.9E-15	10,548	2.2E-13
	o_*Bacteroidales*	0.9	7.6E-15	10,546	2.2E-13
	s_*Muribaculum gordoncarteri*	0.9	1.3E-14	10,533	2.2E-13
	g_*Bifidobacterium*	0.9	2.0E-14	10,520	2.2E-13
	s_*Duncaniella dubosii*	0.9	2.2E-14	10,519	2.2E-13
	g_*Lactobacillus*	0.9	2.2E-14	10,518	2.2E-13
	g_*Olsenella*	0.9	2.3E-14	10,517	2.2E-13
	s_*Paramuribaculum intestinale*	0.9	2.7E-14	10,512	2.2E-13
	s_*Muribaculum* sp.	0.9	2.8E-14	10,511	2.2E-13
	s_*Kineothrix alysoides*	0.8	2.4E-9	10,023	1.2E-8
	s_*Desulfovibrio* sp.	0.8	3.9E-8	9,822	1.8E-7
	s_*Lachnoclostridium* sp. Marseille P6806	0.8	7.4E-7	9,552	3.1E-6
	s_*Prevotella* sp. DJF B116	0.8	7.6E-7	9,550	3.1E-6
	s_*Lactobacillus reuteri*	0.7	5.1E-6	9,334	2.0E-5
	s_*Bacteroides acidifaciens*	0.7	6.6E-6	9,303	2.4E-5

Table A-5 – Correlation table showing species correlated (Spearman correlation coefficient correlations - *rho*) with one sequencing approach.

Sequencing approach	Taxa[1]	*rho*	p-value	T-stat	FDR
	g_*Alistipes*	0.7	3.4E-5	9,079	1.2E-4
	s_*Clostridium* sp. Clone 17	0.6	1.8E-4	8,816	5.7E-4
	k_*Bacteria*	0.6	5.8E-4	8,592	0.002
	g_*Bacteroides*	0.6	6.6E-4	8,567	0.002
	s_*Prevotellamassilia timonensis*	0.6	6.6E-4	8,565	0.002
	g_*Parabacteroides*	0.5	0.002	8,343	0.005
	o_*Lactonacillales*	0.5	0.005	8,077	0.012
	s_*Turicimonas muris*	0.5	0.006	8,069	0.012
	s_*Eisenbergiella* sp.	0.5	0.006	8,064	0.012
	s_*Parasutturella* sp.	0.5	0.006	8,061	0.012
	g_*Dorea*	0.4	0.014	7,804	0.025
	s_*Eubacterium coprostanoligenes*	0.4	0.014	7,804	0.025
	f_*Lachnospiraceae*	0.4	0.021	7,678	0.035
	s_*Anaerotignum* sp.	0.4	0.033	7,518	0.049
	s_*Paraprevotella clara*	0.4	0.033	7,518	0.049
	s_*Robinsoniella peoriensis*	0.4	0.033	7,515	0.049

[1] – k, kingdom; o, order; f, family; g, genus; s, species