

Exploring the impact of the human *FUT2* gene on gut lactobacilli and bifidobacteria to improve gut health through personalised probiotics

Dissertation

in fulfilment of the requirements for the degree

Doctor rerum naturalium (Dr. rer. nat.)

of the Faculty of Mathematics and Natural Sciences

at Christian-Albrechts-Universität zu Kiel

submitted by

Ana Sofia Galrinho Borges

Department of Microbiology and Biotechnology

Max Rubner-Institut, Federal Research Institute of Nutrition and Food

Kiel, 2024

First examiner: Prof. Dr. Charles M. A. P. Franz

Second examiner: Prof. Dr. Tal Dagan

Third examiner: Prof. Dr. Hinrich Graf von der Schulenburg

Date of the oral examination: 22nd April 2024

Declaration

I, Ana Sofia Galrinho Borges, declare that:

Apart from my supervisor's guidance the content and design of the thesis is all my own work; Specific aspects of my thesis were supported by colleagues; their contribution is specified in detail in the section "Contribution of authors";

The thesis has not already been submitted neither partially nor wholly as part of a doctoral degree to another examining body;

Apart from the included published papers no other part of the thesis has been published nor submitted for publishing;

No academic degree has ever been withdrawn;

The thesis has been prepared subject to the Rules of Good Scientific Practice of the German Research Foundation (DFG).

Signature: _____

Table of contents

Declaration	5
Table of contents	7
Acknowledgements.....	9
Summary	11
Zusammenfassung	13
Contribution of authors	15
1. General introduction.....	17
2. Investigation of faecal lactobacilli and bifidobacteria based on host secretor status using 16S rRNA gene amplicon sequencing	39
3. Fast identification method for screening bacteria from faecal samples using Oxford Nanopore Technologies MinION sequencing	55
4. Comparative genomic analysis of lactobacilli and bifidobacteria based on secretor status	78
5. Challenging the diet culture through an evolutionary perspective	119
6. General discussion.....	136
References	147

Acknowledgements

After many hours of staring at the computer and working with the anaerobic chamber, the PhD saga is finally over! It has been a long, long journey and my “evolution” (pun intended) would not have been possible without the unwavering support of those who have cheered me on throughout this wild ride.

First of all, I would like to thank my supervisor Charles for his guidance and insightful feedback throughout the PhD. I would also like to thank my “unofficial” co-supervisor Gyu-Sung for his assistance with the microbiological lab work, and Erik for his unconditional support with the bioinformatics analyses and sequencing runs. Many thanks also to John, Tal, Hinrich and Olivia for their continued support over the years. A special thanks also goes to my tandem partner Meghna for her collaboration, and all the members of the Baines lab for helping me with the bioinformatics analyses during this last year.

My PhD life was always more colourful when I had the chance to work or exchange ideas with fellow students at the MRI. Maria, thank you so much for being the lab mom, the live German translator and the go-to lunch companion. Thank you too, Lu and Benni, for your company and support through the ups and the downs of my academic journey. I would also like to thank all the other technical assistants who have helped me with the seemingly endless lab work: Jessica, Katja, Adrian, Yvonne, Isabell and Jana.

I was fortunate to be part of the Research Training Group (RTG) TransEvo, which allowed me to meet a friendly international community and return to the feeling of “school”, with “classmates” working towards a similar goal. Through TransEvo I was lucky to meet brilliant students who I can now call friends. Thanks to the opinion paper,

I met Freya and now we cannot stop chatting whenever we run into each other. I would also like to thank Aditi and Lina, who welcomed me as if I were part of their lab.

To my friends outside of work, many of whom are also struggling PhD students, a big thank you! Ginevra was my first friend in Kiel and we have been there for each other through thick and thin. It was also a pleasure to meet Irene recently and connect with a like-minded Iberian student. Thanks also to my friends around the world: Eli, Ana Beatriz, Teresa and Marta, I have truly appreciated your encouragement.

I would also like to thank my family for their unwavering support. My dad, mum and sister in particular have been subjected to wild emotions and complicated scientific explanations along the way. A big “obrigada” to you three!

Finally, I would like to dedicate my thesis to my dear grandpa José Galrinho, who would have been very happy to see me graduate. He was a great advocate of education, despite never having had the opportunity himself. As he wished, may more and more of us invest in education and knowledge for the betterment of society!

Sofia

Summary

The human *FUT2* gene encodes an enzyme that mediates the attachment of fucose to glycan chains on mucosal surfaces, such as the intestinal mucosa. Individuals with one or two functional *FUT2* alleles are secretors, while those with two non-functional alleles are non-secretors. Secretor status can influence pathogen adherence, susceptibility to infection and the biodiversity of the gut microbiome.

Probiotics can improve gut health, but their relationship with host secretor status is not fully understood yet. The aim of this project was to investigate whether host secretor status influences gut microbiome taxa with probiotic potential, namely bifidobacteria and lactobacilli, in order to identify potential secretor status-based probiotic candidates.

The relationship between secretor status and the abundance of lactobacilli and bifidobacteria was investigated using 16S rRNA gene sequencing data from stool samples of an adult German cohort. There were no significant differences between secretors and non-secretors in phylum abundance, lactobacilli abundance, bifidobacteria abundance, alpha diversity and beta diversity. However, there was a distinct variety of lactobacilli in the two groups, with secretors presenting mainly two typically autochthonous genera, *Limosilactobacillus* and *Ligilactobacillus*, and non-secretors presenting several different genera, including typically transient genera.

To complement the faecal microbiome analysis, a culturomics approach was used in parallel. Stool samples were plated out and the isolated colonies were identified using a 16S rRNA gene-based bacterial identification method developed in this project to select only bifidobacteria and lactobacilli for further studies. The whole genomes of these strains were then sequenced to investigate genomic variations according to secretor status. Bifidobacteria were isolated significantly more often in secretors than

in non-secretors, while lactobacilli were isolated in similar numbers from both groups. However, there were no significant differences according to secretor status in terms of fucose-specific genes and functional categories in lactobacilli, bifidobacteria and *B. longum* subsp. *longum* in particular. Principal coordinate analysis also revealed no significant association between gene presence/absence patterns of *B. longum* subsp. *longum* isolates and secretor status or *FUT2* genotype.

Finally, an opinion piece was developed to explore the evolutionary mismatch between modern weight loss goals and the survival-oriented human physiology, focusing on the influence of factors such as the set-point weight theory, population genetics, transgenerational inheritance and the gut microbiome.

Overall, the results of this study did not support the possibility of enhancing probiotic efficacy through customization based on secretor status. Additional research is essential to further substantiate these conclusions.

Zusammenfassung

Das menschliche *FUT2*-Gen kodiert ein Enzym, das die Bindung von Fucose an Glykanketten auf Schleimhautoberflächen wie der Darmschleimhaut vermittelt. Personen mit einem oder zwei funktionellen *FUT2*-Allelen sind Sekretoren, während Personen mit nicht funktionellen Allelen Nicht-Sekretoren sind. Der Sekretor-Status kann die Bindung von Krankheitserregern an Darmzellen, die Anfälligkeit für Infektionen und die Artenvielfalt des Darmmikrobioms beeinflussen.

Probiotika können die Darmgesundheit verbessern, aber ihr Zusammenhang mit dem Sekretorstatus des Wirts ist noch nicht vollständig geklärt. Ziel dieses Projekts war es zu untersuchen, ob der Sekretorstatus des Wirts bestimmte Taxa des Darmmikrobioms mit probiotischem Potenzial, nämlich Bifidobakterien und Laktobazillen, beeinflusst, um somit potenzielle probiotische Kandidaten auf der Grundlage des Sekretorstatus zu identifizieren.

Der Zusammenhang zwischen dem Sekretor-Status und der Häufigkeit von Laktobazillen und Bifidobakterien wurde mit Hilfe von 16S rRNA-Gen-Sequenzierungsdaten aus Stuhlproben einer deutschen Erwachsenen-Kohorte untersucht. Es gab keine signifikanten Unterschiede zwischen Sekretoren und Nicht-Sekretoren in Bezug auf die Häufigkeit von Phyla, Laktobazillen und Bifidobakterien sowie auf die Alpha-Diversität und Beta-Diversität. Es gab jedoch einen deutlichen Unterschied in der Vielfalt an Laktobazillen in den beiden Gruppen. Die Sekretoren wiesen hauptsächlich zwei typisch autochthone Gattungen, *Limosilactobacillus* und *Ligilactobacillus*, auf, während die Nicht-Sekretoren mehrere verschiedene Gattungen, einschließlich typischer transienter Gattungen, aufwiesen.

Zur Ergänzung der Mikrobiomanalyse der Stuhlproben wurde parallel dazu ein auf Kultivierung basierender Ansatz verwendet. Die Stuhlproben wurden ausplattiert und die isolierten Kolonien mit einer in diesem Projekt entwickelten bakteriellen Identifizierungsmethode auf der Grundlage des 16S rRNA-Gens identifiziert, um lediglich Bifidobakterien und Laktobazillen für weitere Untersuchungen zu erhalten. Die gesamten Genome dieser Stämme wurden dann sequenziert, um genomische Variationen in Abhängigkeit vom Sekretorstatus zu untersuchen. Bifidobakterien wurden signifikant häufiger bei Sekretoren isoliert als bei Nicht-Sekretoren, während Laktobazillen in ähnlicher Anzahl aus beiden Gruppen isoliert wurden. Es gab jedoch keine signifikanten Unterschiede der Sekretorstatus in Bezug auf Fucose-spezifische Gene und funktionelle Kategorien bei Laktobazillen, Bifidobakterien und insbesondere *B. longum* subsp. *longum*. Die Hauptkoordinatenanalyse ergab auch keinen signifikanten Zusammenhang zwischen dem Vorhandensein/Abwesenheitsmuster von Genen in Isolaten von *B. longum* subsp. *longum* und dem Sekretorstatus oder dem *FUT2*-Genotyp.

Schließlich wurde ein Meinungsbeitrag verfasst, welcher darauf abzielte die evolutionäre Diskrepanz zwischen modernen Gewichtsabnahmezielen und der überlebensorientierten menschlichen Physiologie zu untersuchen, wobei der Schwerpunkt auf dem Einfluss von Faktoren wie der Set-Point-Gewichtstheorie, der Populationsgenetik, der transgenerationalen Vererbung und dem Darmmikrobiom lag. Insgesamt konnten die Ergebnisse dieser Studie die Hypothese einer verbesserten Wirksamkeit von Probiotika durch Anpassung an den Sekretorstatus nicht bestätigen. Um diese Schlussfolgerungen zu untermauern, sind jedoch weitere Forschungsarbeiten erforderlich.

Contribution of authors

The General Introduction (Chapter 1) and the General Discussion (Chapter 6) were written by Ana Sofia Galrinho Borges.

“Investigation of faecal lactobacilli and bifidobacteria based on host secretor status using 16S rRNA gene amplicon sequencing” (Chapter 2) is a manuscript in preparation. The authors are Ana Sofia Galrinho Borges, Erik Brinks (EB), Gyu-Sung Cho (GC) and Charles M.A.P. Franz (CF). Ana Sofia Galrinho Borges is the first author. Ana Sofia Galrinho Borges conceptualized the study, performed the data analysis, prepared the figures and wrote the manuscript. EB assisted with the data analysis. GC contributed partially to the conceptualization of the study. CF conceptualized the study, supervised the work and contributed to the writing. John F. Baines and his group assisted with the data analysis. Corinna Bang and Andre Franke provided the human faecal samples and respective ethical approval.

“Fast identification method for screening bacteria from faecal samples using Oxford Nanopore Technologies MinION sequencing” (Chapter 3) is a paper published in the journal *Current Microbiology*. The authors are Ana Sofia Galrinho Borges, Meghna Basu (MB), Erik Brinks (EB), Corinna Bang (CB), Gyu-Sung Cho (GC), John F. Baines (JB), Andre Franke (AF) and Charles M.A.P. Franz (CF). Ana Sofia Galrinho Borges is the first author. Ana Sofia Galrinho Borges conceptualized the study, performed the laboratory work, performed the data analysis, prepared the figures and wrote the manuscript. MB participated in the *Bacteroides* laboratory work. EB performed the 16S rRNA gene sequencing and whole genome sequencing, performed the raw data analysis and contributed to the reviewing process. CB and AF provided the human faecal samples and respective ethical approval. CB and JB contributed to the reviewing process. GC performed the whole genome sequencing data analysis and participated

in the reviewing process. CF contributed to the writing and reviewing processes. ASB, MB, JB, GC and CF developed the concept and novel idea of this manuscript.

“Comparative genomic analysis of lactobacilli and bifidobacteria based on secretor status” (Chapter 4) is a manuscript in preparation. The authors are Ana Sofia Galrinho Borges, Erik Brinks (EB), Gyu-Sung Cho (GC) and Charles M.A.P. Franz (CF). Ana Sofia Galrinho Borges is the first author. Ana Sofia Galrinho Borges conceptualized the study, performed the laboratory work, performed the data analysis, prepared the figures and wrote the manuscript. EB and John F. Baines’ group performed the whole genome sequencing. EB performed the raw data analysis partially, assisted with the data analysis and reviewed part of the manuscript. John F. Baines and his group assisted with the data analysis. GC contributed partially to the conceptualization of the study and performed the raw data analysis partially. CF supervised the work and reviewed the manuscript. Corinna Bang and Andre Franke provided the human faecal samples and respective ethical approval.

“Challenging the diet culture through an evolutionary perspective” (Chapter 5) is a manuscript in preparation. The authors are Ana Sofia Galrinho. Borges, Freya A. Pappert (FP) and Olivia Roth (OR). Ana Sofia Galrinho Borges and FP are the first authors. Ana Sofia Galrinho Borges and FP conceptualized and wrote the manuscript. OR and Charles M.A.P. Franz revised the manuscript.

1. General introduction

1.1. The gut microbiome

The human microbiome is the set of microorganisms that live in and on the human body. This includes bacteria, fungi, viruses and other microorganisms. Over the past 20 years, the microbiome has become increasingly important in research and is now known to be associated with many human health conditions¹.

This thesis focuses on the bacterial gut microbiome. There are approximately 10-100 trillion bacterial cells in the human body, a number which is of the same order of magnitude as the number of human cells^{2,3}. These bacteria are present in several different parts of the human body, including skin, mouth, vagina and gut. The gut microbiome has the highest bacterial content compared to the other areas of the body². It is also highly diverse and unique to each individual. However, there are some recognisable trends. For example, *Firmicutes* (also known as *Bacillota*⁴) and *Bacteroidetes* (also known as *Bacteroidota*⁴) are two phyla that make up around 90% of the gut microbiota of the average healthy individual².

Intestinal bacteria are very important for our health, as they can help us digest food, metabolise drugs and protect us from pathogens⁵⁻⁸. When there is a state of disease, the gut microbiome is usually disturbed as well, which is termed gut dysbiosis⁹. In several diseases, this dysbiotic state of the gut microbiome is characterised by its low diversity and increased proportion of facultative anaerobic bacteria¹⁰. Although there is a correlation between many diseases and gut dysbiosis, it is not known yet whether the microbiome imbalances promote the development of certain medical conditions or vice-versa, or both⁹.

One example of gut dysbiosis is observed in inflammatory bowel diseases (IBDs), such as Crohn's disease and ulcerative colitis, where there is a decrease in beneficial bacteria and an increase in harmful bacteria in the gut¹¹⁻¹⁴. Modulation of the disrupted

gut microbiota, through therapies such as probiotic administration or faecal microbiota transplantation, is being investigated as a potential therapeutic intervention for diseases linked to gut dysbiosis¹⁵.

1.2. The human *FUT2* gene and the secretor status

The human gut microbiome is influenced by both genetics and environmental factors, which include household sharing, diet and medications. Although there is evidence showing that environmental factors have the strongest influence on the gut microbiome¹⁶, certain genes have been shown to have relevant effects on the microbiota as well¹⁷. One example is the human *FUT2* gene, which encodes an $\alpha(1,2)$ -fucosyltransferase, an enzyme that mediates the attachment of an L-fucose sugar to the terminal galactose residue of glycan chains present on mucosal surfaces, such as the gut¹⁸.

Fucose is a 6-deoxy-hexose in the L-configuration that is present in various glycans of different tissues and fluids of the human body¹⁸. Fucose can be produced either by an endogenous *de novo* pathway or by a salvage pathway using free fucose from external sources^{18,19}. Although this monosaccharide is not a major component of the human diet, there are a few food sources which contain this sugar, such as seaweed and mushrooms²⁰. For infants, a source of fucose can also be the human milk oligosaccharides (HMOs) found in breast milk, as 35-50% of HMOs in human milk are fucosylated²¹. These are extensively used by bifidobacteria in the infant gut, as well as by other bacterial taxa such as lactobacilli and *Bacteroides*²².

The attachment of fucose to mucosal glycans leads to the formation of $\alpha(1,2)$ -fucosylated glycans, also known as the H-antigens, which are important for cell communication, recognition, adhesion and the immune response¹⁸. They can either be

anchored to the surface of the mucosa or secreted from it. These glycans can also be further modified into A and B histo-blood group antigens that are present on mucosal surfaces. These are analogous to the antigens present on erythrocytes, but they are only expressed on mucosal surfaces if there is a functional *FUT2* gene. Therefore, although the *FUT2* gene does not regulate the secretion of these glycans itself, it is because of this background that it is called the “secretor” gene²³.

Individuals with at least one functional allele are called “secretors”, while those with two non-functional alleles are called “non-secretors”. Due to the activity of the $\alpha(1,2)$ -fucosyltransferase, secretors will have fucose sugars attached to the glycan chains present on their mucosa, which can act as adhesion sites or as an energy source for bacterial species that metabolise fucose. In the case of non-secretors, fucose is not present at the end of the glycan chains, resulting in an altered microbiome as this potential adhesion site is absent and fucose as a possible energy source is reduced. For example, *Bacteroides* species have been shown to be more prevalent in secretors than in non-secretors²⁴, and this genus is known to include species that can utilise fucose²⁰. The non-secretor genotype is also associated with an increased risk of IBD, type 1 diabetes, vaginal candidiasis and psoriasis, as well as other conditions that are associated with an altered microbiome^{23,25–28}. One proposed hypothesis for this is that in secretors, fucosylation increases during acute or chronic stress to support the gut microbiome and the host-microbiome interactions, which is not possible in non-secretors^{29,30}.

Several mutations lead to the non-secretor genotype. The most common single-nucleotide polymorphism (SNP) causing loss-of-function of the *FUT2* gene in Caucasian, African and Central Asian populations is a guanine-adenine (G-A) substitution at base pair 428 (rs601338)²³. Despite the aforementioned health

disadvantages of the non-secretor genotype, this SNP has been maintained in the human population for 3 million years through balancing selection, which is the set of adaptive mechanisms that maintain genetic variation beyond what would be expected by random chance^{31,32}. Furthermore, there is a surprisingly high percentage of the world's population that has a non-secretor genotype (around 20%). This suggests that there is a selective advantage for non-secretors³¹.

One possible reason for this selective advantage is that individual differences in the glycans present in mucosal tissues can lead to differences in infection susceptibility. For example, several microorganisms possess specific receptors for $\alpha(1,2)$ -fucosylated glycans (e.g. *Helicobacter pylori*, norovirus, and rotavirus), making secretors more susceptible to these pathogens^{23,33}. The *FUT2* gene may also have an indirect effect on infection susceptibility by influencing the gut microbiome. For instance, commensal microbes can confer colonization resistance to pathogens by using $\alpha(1,2)$ -fucosylated glycans and occupying available niches. Additionally, several clinical conditions correlated with the non-secretor genotype are associated with an altered gut microbiome, as mentioned above. Therefore, it is possible that by affecting the microbiome, the secretor status may also influence susceptibility to microbiome-associated diseases^{23,33}.

Although some associations have been shown between the *FUT2* gene and changes in the gut microbiome^{24,34}, more research is needed to clarify exactly how this gene affects gut microbes. Nevertheless, the potential effect of the *FUT2* gene on the gut microbiome makes it an interesting candidate for research into tailoring therapies that modulate the gut microbiota. One such therapy is the administration of probiotics, which would greatly benefit from a more targeted approach. Understanding how the *FUT2* gene influences probiotic bacterial strains could pave the way for more effective

probiotic therapies, if there is an association between such strains and an individual's secretor status.

1.3. Probiotics

Probiotics are “live microorganisms which when administered in adequate amounts confer a health benefit on the host”³⁵. The World Health Organisation (WHO) and the Food and Agriculture Organisation (FAO) have established the key characteristics that a bacterial strain must have for it to be considered a probiotic: it must be able to reach the site of action, provide benefit to the host, pose no risk to the host, remain stable during manufacture and be well preserved in its matrix³⁶.

Probiotic properties are strain-specific³⁷. The bacterial taxa that contain the most probiotic strains are the lactobacilli genera and the *Bifidobacterium* genus, but other genera can also be used, such as *Enterococcus*³⁸, *Streptococcus*³⁹ and *Leuconostoc*^{37,40}.

Probiotic species can improve gut health through three main modes of action. First, they can modulate the host microbiome, for example through a “colonization resistance” effect. Some of the mechanisms involved in this effect are that probiotics can compete with pathogenic bacteria for functional niches or via antagonism, through production of antimicrobial substances (e.g. bacteriocins) and metabolites such as short-chain fatty acids (SCFAs), which can lower the pH to unfavourable levels for bacterial growth. Finally, probiotics have also been shown to improve the gastrointestinal (GI) barrier function and modulate the immune system⁴¹.

There are already some evidence-based applications of probiotics for *Clostridium difficile*- and antibiotic-associated diarrhoea⁴². However, despite the promising insights on probiotics and their effects on gut health, in most fields there is a striking

heterogeneity of results that limits their clinical use. As with most microbiome research, studies tend to be of low quality, with small numbers of participants and low statistical power. In addition, the variability of patients and the different strain combinations used as probiotics in each study can prevent proper comparisons and meta-analyses. Therefore, it is important to conduct future research with larger cohorts, consistent strain combinations, and adequate quality control before establishing clinical guidelines⁴¹.

It is also important to investigate which human factors may influence the efficacy of probiotics, including genetic factors such as the human *FUT2* gene. Mäkituokko *et al.* (2012) have shown that microbes from pooled faecal samples of secretors perform better than microbes from a non-secretor pool when subjected to transit through an upper intestinal tract model (TIM-1), especially bifidobacteria and anaerobes. Furthermore, bifidobacteria and lactobacilli from these pools had a lower number of RAPD-PCR fingerprinting genotypes in the non-secretor pool than in the secretor pool⁴³. This study supports the idea that these two probiotic-like taxa may be influenced by the secretor status. In the following subsections, these taxa will be briefly described and their association with the human *FUT2* gene will be explained.

1.4. Lactobacilli

Lactobacilli are Gram-positive bacteria that live in nutrient-rich habitats that can satisfy their extensive growth requirements. They are facultative anaerobes or microaerophiles. They are found in humans, animals, plants and food as host-adapted species, but they can also exist as free-living or nomadic species⁴⁴.

Lactobacilli can be autochthonous or allochthonous. Autochthonous lactobacilli have co-evolved with their host over time and form stable populations in it⁴⁴. For example, human intestinal strains of the autochthonous *Limosilactobacillus reuteri* have fewer

adhesin-like surface proteins than other lactobacilli, but can instead utilise glycerol and propanediol, suggesting that they have adapted to the intestinal lumen^{45,46}. Allochthonous lactobacilli have a different origin from their environment, but can also form seemingly stable populations if regularly introduced into a habitat⁴⁴. Both can be used as probiotics, but for different purposes. On the one hand, autochthonous species can be used to outcompete pathogens because they have fitness advantages within a particular host, and can also establish relationships with the immune system towards tolerance. On the other hand, allochthonous species can be used to stimulate the immune system since they have not co-evolved with the host. Therefore, it is important to consider the natural history of each lactobacilli species before using it for probiotic therapy⁴⁴.

Until recently, this taxon was considered to be a single genus, containing 261 species. However, in 2020, a taxonomic note proposed a reclassification of this genus due to its family-like diversity, suggesting its division into 25 different genera. These include 23 new genera, *Paralactobacillus* and the emended genus *Lactobacillus*. In addition, the family *Lactobacillaceae* now includes all genera previously included in the emended family *Lactobacillaceae* and the family *Leuconostocaceae*⁴⁷.

Only around 50 species of lactobacilli are found in the human gut. Some of the most common species found in the gut are *Lacticaseibacillus rhamnosus*, *Ligilactobacillus ruminis*, *Lactobacillus delbrueckii*, *Lactiplantibacillus plantarum*, *Lacticaseibacillus casei* and *Ligilactobacillus acidophilus*⁴⁸. While the majority of lactobacilli in the GI tract are considered to be allochthonous, certain species such as *L. rhamnosus* are rarely found in body sites other than the gut, and are therefore regarded as autochthonous gut microbes⁴⁹.

Although lactobacilli are a minor component of the human gut microbiome (~0.3% in the colon), their levels in the gut are often associated with disease. For example, they are enriched in rheumatoid arthritis but depleted in irritable bowel syndrome (IBS). Interestingly, even in the diseases in which this genus is enriched, its administration usually improves patients' symptoms⁴⁹. This may be due to the changes that lactobacilli cause in the epithelial tight junctions of the gut. By improving the GI barrier, there is greater protection against harmful bacteria. Furthermore, lactobacilli stimulate the restoration of homeostasis in gut conditions, leading to balanced levels of inflammatory signalling molecules and hence protection against inflammatory diseases⁵⁰.

1.4.1. Lactobacilli, the human *FUT2* gene and fucose utilization

The relationship between lactobacilli and secretor status has not been extensively investigated yet. However, one study has shown that there is a higher abundance of lactobacilli in gut mucosa samples from healthy secretors⁵¹. In addition, another study has shown that the composition of lactobacilli differs significantly depending on the host secretor status³⁴.

One possible explanation for this is that some lactobacilli strains are able to cleave fucose from fucosylated glycans through an $\alpha(1,2)$ -fucosidase and/or catabolize it, implying that they might thrive better in the secretor gut with its fucosylated glycans.

The most significant example is the *L. rhamnosus* species. For instance, the *L. rhamnosus* GG strain has a gene cluster for fucose catabolism similar to that of *Escherichia coli*, the *fuc* operon, which includes the following genes, also summarized in Figure 1⁵²:

- *fucP*: encodes an L-fucose/H⁺ symporter/permease, which mediates the uptake of L-fucose into the bacterial cell;

- *fucU*: encodes an L-fucose mutarotase, which mediates the conversion of α -L-fucose to β -L-fucose and vice versa;
- *fucI*: encodes an L-fucose isomerase, which converts L-fucose to L-fuculose;
- *fucK*: encodes an L-fuculokinase, which catalyses the phosphorylation of L-fuculose;
- *fucA*: encodes an L-fuculose phosphate aldolase, which cleaves L-fuculose-1-phosphate to dihydroxyacetone phosphate and L-lactaldehyde;
- *fucR*: encodes a transcriptional regulator of the *fuc* operon.

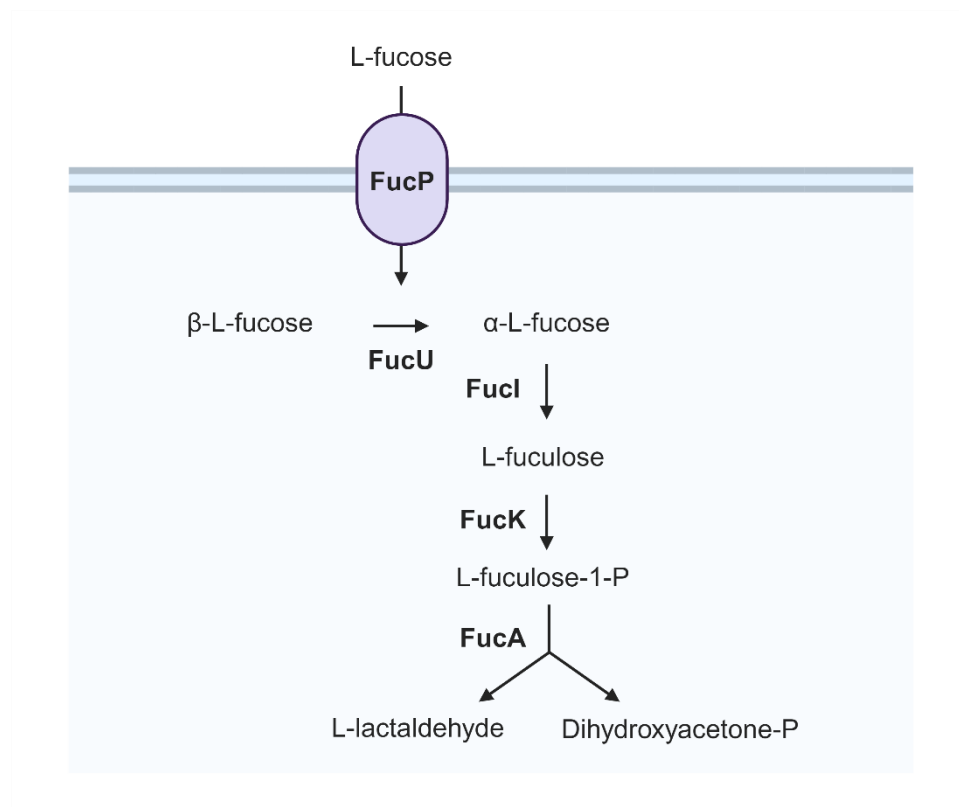


Figure 1 – Schematic representation of the L-fucose catabolic pathway in *L. rhamnosus* GG⁵². Figure created with BioRender.com.

It is likely that after the action of FucA, the generated L-lactaldehyde is directed either to 1,2-propanediol by a lactaldehyde reductase (which in *Escherichia coli* is encoded

by the *fucO* gene) or to lactate by a lactaldehyde dehydrogenase (encoded by the *aldA* gene in *E. coli*), depending on the redox state. The mechanisms for lactaldehyde metabolism in *L. rhamnosus* GG have not yet been identified, but since its fucose catabolic activity only occurs under anaerobic conditions, it is expected that lactaldehyde reductase would be more efficient in this scenario⁵².

Only a few *L. rhamnosus* strains have the ability to catabolize fucose (e.g. the *L. rhamnosus* GG, LRHMDP2, LRHMDP3 and HN001 strains). Therefore, it is a strain-specific mechanism. Furthermore, the *L. rhamnosus* GG strain has fucosidases, but these are intracellular and used on different linkage types, such as (1,3) and (1,6), which can be useful, for example, for the uptake of HMOs (same for *L. casei*)⁵²⁻⁵⁴.

There are also some less well-studied examples. For instance, some *Lacticaseibacillus zeae* isolates have a functional *fuc* operon⁵². Another example is *Limosilactobacillus mucosae*, which can ferment fucose but cannot catabolize fucosylated human milk oligosaccharides such as 2'-fucosyllactose, suggesting that it may have fucose catabolic pathways but no fucosidases⁵⁵. It is also possible that some lactobacilli strains act as scavengers, i.e. if they have a *fuc* operon, they can use fucose that has been cleaved from fucosylated glycans by other species such as, for example, bifidobacteria and *Bacteroides*^{56,57}.

Lactobacilli may also use fucosylated glycans as an adhesion site. However, the mechanisms described so far are not specific for fucosylated glycans, but rather for intestinal glycans in general through several adhesion factors as well as pili in the case of *Lacticaseibacillus paracasei*, *Lactobacillus johnsonii* and *L. rhamnosus*^{58,59}. Therefore, there is no evidence proving that adhesion to fucosylated glycans would confer a particular advantage to lactobacilli in secretors.

Overall, studies on the relationship between the human *FUT2* gene and the abundance and diversity of gut lactobacilli are sparse, probably due to the difficulty of capturing these bacteria through metagenomics and faecal culturing given their low abundance in the human gut. In addition, research on the interaction between fucose and fucosylated glycans and many lactobacilli species is still in its infancy, and several mechanisms remain to be described. Further studies will help in clarifying whether lactobacilli as probiotics can and should be tailored to the human secretor status.

1.4.2. Comparative genomics of lactobacilli

The pan-genome is the total set of genes of a given species, taking into account all its strains^{60,61}. It consists of the core genome, which contains the genes present in all strains of that species, and the accessory genome, which refers to genes that are not present in all strains^{60,62}. The pan-genome can also be calculated at the genus level if a sufficient number of species is considered, but it is not usually calculated for taxa above the genus level. However, as lactobacilli were until recently considered to be a single genus⁴⁷, there have been several comparative genomics studies on this topic, which will be briefly summarised here.

The lactobacilli taxon has an open pan-genome, meaning that it can evolve and acquire an unlimited number of genes⁶³. At a single species-level, lactobacilli usually have a closed pan-genome, i.e. each strain has a limited number of genes that it can acquire. However, some species have a greater genetic diversity, such as *L. plantarum*, which is considered to have an open pan-genome⁶⁴. The genome size of lactobacilli ranges from 1.23 to 4.91 Mb, corresponding to *Fructilactobacillus sanfranciscensis* and *Lentilactobacillus parakefiri* respectively. Lactobacilli are also characterised by a wide range of GC content, from 31.93% to 57.02%⁶⁴.

Core and pan-genome studies of the lactobacilli taxon are highly dependent on the number of strains used for the analysis, because the more strains are used, the smaller the core genome size. Pan-genome analyses also depend on whether draft genomes are included, as their poor quality can lead to an incorrect core genome size. Sun *et al.* (2015) found a core genome of only 73 genes when considering 175 lactobacilli strains plus 26 strains from 8 related genera, with 44,668 gene families contained in the pan-genome. These core genes encoded mainly proteins involved in replication and cell growth⁶⁴. More recently, Inglin *et al.* (2018) identified 266 core genes from 98 fully assembled genomes of 32 lactobacilli species, with a pan-genome containing 20,800 genes⁶³.

The accessory genome of lactobacilli can vary greatly, especially depending on the environment. For example, in a study of *Lactobacillus crispatus*, there were significant differences between isolates from vaginal and faecal samples. For example, isolates from faeces were enriched in genes related to carbohydrate metabolism, which indicates adaptation to the gut since this habitat contains more types of carbohydrates than the vaginal environment and consequently requires a broader diversity of active carbohydrate enzymes⁶⁵.

Just as there are environmental adaptations in the lactobacilli pan-genome, it would be interesting to investigate whether there are genomic adaptations based on the secretor status of the host, as it partially defines the local gut environment by determining whether there are fucosylated glycans attached to the intestinal mucosa.

1.5. Bifidobacteria

Bifidobacterium is a genus of Gram-positive anaerobic bacteria found in the human GI tract. These bacteria have also been isolated from other sources, such as the GI tract

of other mammals, the insect gut, water kefir and sewage⁶⁶. The abundance of bifidobacteria in the human gut varies with age. They are the predominant genus in infants (60-70% of the gut microbiome composition), but after weaning their abundance decreases to 30-40% in children and adolescents. They then continue to decline slowly throughout adulthood to approximately 2-14% of the total microbiota, followed by another dramatic reduction in old age (0-5%), caused not only by the ageing process itself, but also by an increased burden of medical conditions and increased antibiotic use^{67,68}.

By 2017, this genus included 54 species⁶⁹, but only 11 had been isolated from human faeces, namely *B. adolescentis*, *B. angulatum*, *B. animalis* (subspecies *lactis*), *B. bifidum*, *B. breve*, *B. catenulatum*, *B. dentium*, *B. gallicum*, *B. longum* (subspecies *longum* and *infantis*), *B. scardovii* and *B. pseudocatenulatum*⁷⁰. Intestinal bifidobacteria have co-evolved with their host over time⁷¹⁻⁷³. For example, *B. adolescentis* has several glycoside hydrolases necessary to utilise sugars and glycans of plant origin (e.g. starch) that are part of the typical human diet⁷¹.

Bifidobacterium species have been widely used as probiotics because of the health benefits associated with them. These include inhibition of harmful bacteria, promotion of beneficial bacteria through cross-feeding, improvement of the GI barrier, suppression of inflammatory signalling, and improvement of symptoms in medical conditions such as IBS. Although the complex mechanisms underlying these beneficial effects are not yet fully understood, recent studies have shown that *Bifidobacterium* alters the functionality of dendritic cells to enable the immune system to defend against pathogens while adapting its homeostasis to harmless bacteria and antigens⁵⁰.

1.5.1. Bifidobacteria, the human *FUT2* gene and fucose utilization

Certain *Bifidobacterium* species can interact with the $\alpha(1,2)$ -fucosylated glycans present in the mucosa of secretors, either by using them as adhesion sites or by using fucose as an energy source³⁴.

Bifidobacteria possess several non-specific adhesion factors that can use fucosylated glycans as adhesion sites, as well as pili in certain species such as *B. breve*⁷⁴. However, in *B. longum* subsp. *infantis* there are two solute binding proteins, Blon_0343 and Blon_2202, that specifically recognise fucosylated oligosaccharides, which could imply a preferential use of fucosylated glycans for intestinal adhesion and thus an advantage for secretors. Blon_0343 has also been identified in one strain of *B. pseudocatenulatum*⁷⁵.

Regarding fucose metabolism, as described above for lactobacilli, the key elements are $\alpha(1,2)$ -fucosidases and fucose catabolic pathways, and sometimes not all elements are present in the same strain. For example, *B. bifidum* has one extracellular $\alpha(1,2)$ -fucosidase (AfcA) that can cleave the fucose from fucosylated glycans. However, *B. bifidum* does not metabolize fucose itself^{56,76–80}. However, by releasing it, other bacteria that have a *fuc* operon can act as scavengers and use the free fucose molecules. One example is *B. breve*, which can catabolize fucose, although the mechanisms are not fully understood⁸¹. It has also been shown that certain lactobacilli can use fucose released by bifidobacteria through cross-feeding⁵⁷.

Regarding the *B. longum* species, a gene cluster involving the import of fucosylated molecules, fucose metabolism and two intracellular α -fucosidases (one of which was able to cleave $\alpha(1,2)$ fucosyl linkages) was identified in *B. longum* subsp. *infantis*, in one *B. longum* subsp. *suvis* strain and in a few infant *B. longum* subsp. *longum* strains (e.g. M12, BCY01, JDM301, SC596)^{82–84}. This cluster was also found in one *B.*

pseudocatenulatum strain, but in this case only one of the α -fucosidase genes was identified, in particular the $\alpha(1,2)$ -fucosidase gene⁸². The organisation of the genes in this cluster may differ between strains. The most striking difference is that, in *B. infantis*, the fucosidases are separated from the *fuc* cluster⁸².

It is important to note once again that these are intracellular fucosidases. Therefore, they cannot cleave the L-fucose attached to anchored glycan chains, but it would be possible for these bacteria to uptake secreted fucosylated glycans and break them down with these intracellular fucosidases, just as they do with HMOs^{82,85–88}. Having cleaved fucose internally, they can either metabolise it or release it for scavenging bacteria.

The identified *fuc* operon consisted of the following set of genes, also summarized in Figure 2^{83,87,89}:

- *fucP*: encodes an L-fucose/H⁺ symporter/permease, which mediates the uptake of L-fucose into the bacterial cell;
- *fucU*: encodes a fucose mutarotase that mediates the conversion of α -L-fucose to β -L-fucose and vice versa. Strains that only possess the AfcA fucosidase will not have this gene, as this fucosidase performs this configuration inversion on its own. If there is another fucosidase, then the strain will usually have this gene as it is necessary for its mechanism⁸⁹;
- *fucC*: encodes a β -L-fucose dehydrogenase, which oxidises β -L-fucose to L-fucono-1,5-lactone. The latter is unstable and is therefore rapidly converted to L-fucono-1,4-lactone in a non-enzymatic manner⁹⁰;
- *fucD*: encodes an L-fuconolactone hydrolase, which converts L-fucono-1,4-lactone to L-fuconate;

- *fucE*: encodes an L-fuconate dehydratase, which converts L-fuconate to 2-keto-3-deoxy-L-fuconate;
- *fucF*: encodes a 2-keto-3-deoxy-L-fuconate aldolase, which cleaves the 2-keto-3-deoxy-L-fuconate to L-lactaldehyde and pyruvate;
- *fucO*: encodes a lactaldehyde reductase, which reduces L-lactaldehyde to 1,2-propanediol;
- *fucR*: encodes a transcriptional regulator of the *fuc* operon.

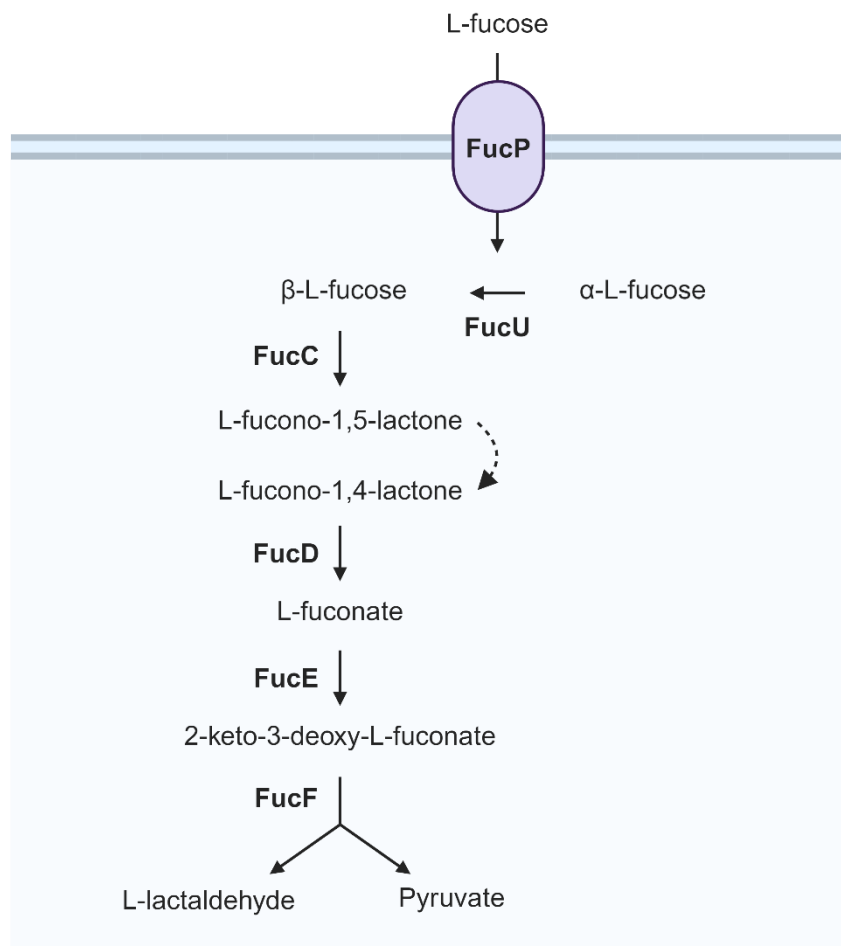


Figure 2 – Schematic representation of the putative L-fucose catabolic pathway in *B. longum* subsp. *infantis*^{83,87,89}. The generated lactaldehyde can be reduced to 1,2-propanediol by the lactaldehyde reductase encoded by *fucO*. It is important to note that

fucose can also be present intracellularly by uptake of fucosylated HMOs, rather than relying on the permease encoded by *fucP*⁸⁷. Figure created with BioRender.com.

The 1,2-propanediol produced by FucO is a precursor of intestinal propionate, which is a SCFA. Therefore, these bacteria may be involved in the balance of SCFAs, which is highly important for human health as SCFAs have anti-inflammatory and immunoregulatory properties⁸³.

It is important to note that many of the above-mentioned species and strains are much more common in the infant gut microbiome than in adults, which is consistent with the fact that bifidobacteria are very abundant in the infant gut and crucial for the metabolism of HMOs present in human milk. Furthermore, these pathways and genes are often a strain-specific mechanism⁹¹.

Given the ability to catabolize fucose present in some *Bifidobacterium* species and strains, it has been hypothesised that this taxon would have a greater colonization capability in secretors when compared to non-secretors⁹². However, it is controversial whether the human *FUT2* gene has an effect on bifidobacterial abundance and diversity. In two studies from 2011 and 2014, a decrease in bifidobacteria diversity, richness and abundance was observed in 14 non-secretors compared to 57 secretors. Interestingly, species that can cleave fucosylated glycans and/or catabolize fucose, such as *B. bifidum* and *B. pseudocatenulatum*, were less abundant in faecal samples from non-secretors than from secretors, as were *B. adolescentis* and *B. catenulatum*⁹². However, two more recent studies involving more than 1000 individuals found no association between secretor status and gut microbiome composition in general and specifically for *Bifidobacterium*^{93,94}.

There are several possible reasons for these conflicting results. Firstly, the origin of the sample (mucosa or faeces) could have a major influence, as *FUT2* expression varies according to tissue and site. It is the highest in the salivary glands, then in the oesophageal mucosa, followed by the terminal ileum. It also influences effects on gut microbiome composition in the colon but with decreasing effects in distal areas. Technical methodological differences may be another source of variation, whether in DNA extraction and/or microbiome profiling. Additionally, diet and other environmental aspects affecting the individuals tested also play a role in the results found⁹⁴. Although there may be no differences in the abundance and diversity of bifidobacteria according to secretor status, there may be local adaptation to the gut at the genomic level, which could be a target for personalised probiotic administration.

1.5.2. Comparative genomics of bifidobacteria

The genus *Bifidobacterium* has an open pan-genome, but at the species level bifidobacteria usually have a closed pan-genome⁹⁵.

This genus is characterised by a high GC content, ranging from 55% to 67%⁹⁶. The genome size of bifidobacteria can range from 1.73 to 3.25 Mbp, corresponding to 1,352 and 2,557 open reading frames (ORFs) in *B. indicum* and *B. biavatii*, respectively⁹⁷. However, the pan-genome of the genus is much larger. A recent study analysed 129 bifidobacteria strains and identified 26,905 gene clusters. It is expected that studies including the existing >400 bifidobacteria strains with publicly available genomes will lead to a much larger pan-genome⁷³.

The core genome of the genus *Bifidobacterium* consists of 400 to 500 genes, which mainly correspond to housekeeping functions^{73,95}. The Clusters of Orthologous Groups (COGs) with proportions greater than 5% include translation, replication, transport and

metabolism of amino acids, carbohydrates and nucleotides⁹⁵. With regard to the accessory genome, it is possible to find truly unique genes (TUGs), i.e. genes that only exist in one strain. The number of TUGs is highly strain-dependent: it can range from 7 in *B. indicum* to 595 in *B. cuniculi* LMG10738⁹⁸. Most TUGs have a distinct GC content, different codon usage, and are flanked by insertion elements, suggesting acquisition by horizontal gene transfer (HGT). Functions of the annotated TUGs include adhesion-mediating proteins and carbohydrate-metabolising enzymes and transporters⁹⁵.

By comparing with other members of the gut microbiome, it was also possible to identify dispensable genes that are not present in all *Bifidobacterium* species, but are shared with other bacteria. Since the colonization of the human gut by bifidobacteria occurs shortly after birth, it is natural that there is HGT between members of the gut microbiome throughout an individual's lifetime. The proteins encoded by the identified dispensable genes were associated with signal transduction, as well as sugar and coenzyme metabolising enzymes and transporters, which can contribute to host adaptation. Importantly, the gene exchange occurs with a wide set of taxa, including not only bacteria from three different phyla (*Chlamydiae*, *Proteobacteria* and *Firmicutes*) but also some Archaea such as *Methanobrevibacterium smithii*⁹⁵.

It is important to note that bifidobacteria have a higher percentage of their genome dedicated to sugar metabolism than other gut bacteria. These genes can either encode proteins necessary for the metabolism of a wide range of host- or diet-derived glycans or for specific carbohydrates. Some of these genes are TUGs, while others are part of the core genome⁹⁷.

Interestingly, a 2018 study showed that, unlike lactobacilli, bifidobacteria from vaginal and faecal sources are not significantly different and may actually be a shared

community⁹⁶. Although there is no appreciable site-specific adaptation, there may be host adaptation instead, for example according to the secretor status and particularly in the gut, given the genomic emphasis on sugar metabolism and the distinct fucose availability in secretors and non-secretors.

1.6. Project aims and thesis overview

The aim of this project was to study the abundance, diversity and genomic differences of human faecal lactobacilli and bifidobacteria depending on the host secretor status.

It was hypothesised that a decrease in the abundance of lactobacilli and bifidobacteria would be observed in non-secretors when compared to secretors, as these taxa include species that can utilise fucose. A second hypothesis was that, in terms of species diversity of lactobacilli and bifidobacteria, it would be possible to identify more species that use fucose (e.g. *L. rhamnosus*) in secretors than in non-secretors. To investigate these two hypotheses, I first characterised the overall microbiome composition of faecal samples from 40 healthy human participants (20 secretors and 20 non-secretors). To do this, I examined the diversity and abundance of different taxa of the microbiome, and in particular lactobacilli and bifidobacteria, using 16S rRNA gene sequencing data obtained from these faecal samples, as described in Chapter 2.

The third hypothesis was that, irrespective of changes in abundance and diversity, there could be species- and/or strain-level variations in the genomes of lactobacilli and bifidobacteria depending on the secretor status due to differences in fucose availability.

The influence of host factors on the gut microbiome is usually only studied in terms of taxonomic diversity and abundance, but it is possible for the same bacterial species to be present at the same abundance in different individuals while being different at the genomic level. In this particular case, it was hypothesised that genes and/or functions

associated with fucose utilisation would have been lost in lactobacilli and bifidobacteria isolated from non-secretors, revealing local adaptation to the secretor status of the host. To test this hypothesis, I first had to develop the necessary methodology to isolate a significant number of isolates of these taxa, which, as described above, are not very abundant in the human gut microbiota. We succeeded in developing a cost-effective pipeline for the identification of bacteria isolated from faecal samples, which was first applied to 180 bacterial colonies and then extended to 833 isolates, as described in Chapter 3⁹⁹. A total of 860 isolates were identified in this project, of which the bifidobacteria and lactobacilli isolates were then sequenced at the whole genome level and comparative genomic analyses were performed to identify any significant differences according to the host's secretor status, as discussed in Chapter 4. The abundance and diversity of the isolates obtained will also be described.

Finally, Chapter 5 consists of the opinion piece written for the Translational Evolutionary Biology course of the RTG TransEvo, which focuses on challenging dieting trends through an evolutionary lens and explores, for example, how the diet industry can profit from hype regarding preliminary gut microbiome research.

I will conclude this thesis with a general discussion focusing on the implications of the results obtained, their contextualisation in the current literature, and possible directions for future research.

2. Investigation of faecal lactobacilli and bifidobacteria based on host secretor status using 16S rRNA gene amplicon sequencing

Ana Sofia G. Borges¹, Erik Brinks¹, Gyu-Sung Cho¹ and Charles M.A.P. Franz^{1*}

Affiliations

¹Department of Microbiology and Biotechnology, Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Hermann-Weigmann-Straße 1, 24103 Kiel, Germany

*Corresponding Author. Contact: charles.franz@mri.bund.de

Manuscript in preparation.

2.1. Introduction and aim

The association between lactobacilli and bifidobacteria abundance and secretor status has been studied over the years, but with variable results. One study found higher lactobacilli abundance in healthy secretors in mucosal samples⁵¹, but there has been a paucity of further studies and data in this regard. As for bifidobacteria, some studies show a reduced abundance of bifidobacteria in non-secretors⁹², but larger studies have not found a general association between secretor status and faecal microbiome composition, nor specifically for bifidobacteria^{93,94}.

In this investigation, we aimed to assess the impact of host secretor status on the relative abundance and diversity of lactobacilli and bifidobacteria in our cohort and to characterise the faecal microbiome composition of the donors of this cohort.

2.2. Materials and methods

2.2.1. Cohort selection

Healthy blood donors from Northern Germany were recruited by mail. From this cohort, 40 donors were selected for our study, of which 20 were secretors and 20 were non-secretors. Of the secretors, 14 were heterozygous (Secretor_Het) and 6 were homozygous (Secretor_Hom). Donor selection aimed to match a secretor with a non-secretor of the same sex, same ABO blood group and similar age (within a 10-year range) when possible. The selection process also aimed to select donors with an ABO blood group distribution that would resemble that of the German population as much as possible¹⁰⁰, and to obtain half female and half male donors. Data on Body Mass Index (BMI) were also collected for some of the donors. Table 1 summarises the metadata of all donors, ordered by match.

Table 1 – Selected donors and respective metadata. YOB, year of birth. NA, not available.

Donor code	ABO blood group	Secretor status	<i>FUT2</i> genotype	Sex	YOB	BMI	Match
G5	B	Secretor	Secretor_Hom	f	1980	32.9	1
G9	B	Non-Secretor	Non-Secretor	f	1970	30.5	1
896	AB	Secretor	Secretor_Hom	m	NA	NA	2
Y2	AB	Non-Secretor	Non-Secretor	m	1961	26.0	2
674	O	Secretor	Secretor_Hom	f	NA	NA	3
Y6	O	Non-Secretor	Non-Secretor	f	1970	29.8	3
G8	A	Secretor	Secretor_Het	f	1953	NA	4
G6	A	Non-Secretor	Non-Secretor	f	1953	NA	4
461	O	Secretor	Secretor_Het	m	NA	NA	5
Y22	O	Non-Secretor	Non-Secretor	m	1985	23.6	5
541	O	Secretor	Secretor_Het	m	1990	NA	6
G12	O	Non-Secretor	Non-Secretor	m	1988	30.7	6
818	A	Secretor	Secretor_Het	m	1988	NA	7
G11	A	Non-Secretor	Non-Secretor	m	1977	28.7	7
936	A	Secretor	Secretor_Hom	f	NA	NA	8
603	A	Non-Secretor	Non-Secretor	f	1989	22.9	8
697	O	Secretor	Secretor_Het	f	1965	NA	9
Y10	O	Non-Secretor	Non-Secretor	f	1962	23.7	9
883	O	Secretor	Secretor_Het	f	1975	NA	10
Y14	O	Non-Secretor	Non-Secretor	f	1971	25.4	10
559	O	Secretor	Secretor_Het	m	1971	32.0	11
Y26	O	Non-Secretor	Non-Secretor	m	1969	24.3	11
907	O	Secretor	Secretor_Het	f	1946	NA	12
Y18	O	Non-Secretor	Non-Secretor	f	1958	NA	12
186	O	Secretor	Secretor_Het	m	1975	NA	13
Y30	O	Non-Secretor	Non-Secretor	m	1971	30.7	13
605	A	Secretor	Secretor_Het	m	1965	29.7	14

G7	A	Non-Secretor	Non-Secretor	m	1967	NA	14
724	AB	Secretor	Secretor_Het	m	1966	NA	15
G2	AB	Non-Secretor	Non-Secretor	m	NA	NA	15
698	A	Secretor	Secretor_Het	m	1957	NA	16
G3	A	Non-Secretor	Non-Secretor	m	1961	24.4	16
744	AB	Secretor	Secretor_Het	f	1969	27.2	17
G1	AB	Non-Secretor	Non-Secretor	f	1965	NA	17
850	O	Secretor	Secretor_Hom	f	1973	NA	18
G4	O	Non-Secretor	Non-Secretor	f	1978	32.4	18
045	A	Secretor	Secretor_Het	m	1973	NA	19
G13	A	Non-Secretor	Non-Secretor	m	1961	33.1	19
690	A	Secretor	Secretor_Hom	f	1969	30.5	20
G10	A	Non-Secretor	Non-Secretor	f	1970	30.1	20

2.2.2. Stool collection and microbial 16S rRNA gene amplicon sequencing

Donors were asked to provide stool samples in buffer solution. Samples were mailed within 24 hours after sample collection. Immediately upon arrival at the Microbiome laboratory of the Institute of Clinical Molecular Biology (IKMB), stool samples were aliquoted and stored at -80°C until further processing. Microbial 16S rRNA gene amplicon sequencing of the faecal samples was performed at the IKMB as previously described in the literature, using primers targeting the V1-V2 region of the 16S rRNA gene¹⁰¹. The study was conducted in accordance with the Declaration of Helsinki, and ethical approval was granted by the ethics committee at Kiel University (AZ A103/14, D590/22).

2.2.3. Microbial 16S rRNA gene amplicon sequencing data analysis

Microbial 16S rRNA gene amplicon sequencing data processing was performed using the DADA2 workflow as previously described^{101–104}, resulting in abundance tables of

amplicon sequence variants (ASVs). The ASVs were taxonomically annotated using the Bayesian classifier from DADA2 and the Ribosomal Database Project (RDP) (v16)¹⁰¹. Both the abundance tables and the taxonomic annotation were then used for analysis using the phyloseq package¹⁰⁵ in RStudio (R 4.2.0)¹⁰⁶. First, chloroplast sequences were removed. For the remaining ASVs, a rarefaction depth of 90% of the minimum sample depth was chosen to ensure a similar sequencing depth for all samples. The data were then transformed into relative abundance data, which were used from then onwards. Afterwards, the data were filtered to remove taxa with a mean abundance of less than 0.001%¹⁰⁷. Statistical tests between groups were done using the ggpubr package¹⁰⁸ in RStudio using the Wilcoxon rank-sum test¹⁰⁹ and p-values were adjusted using the False Discovery Rate (FDR) correction¹¹⁰. The results were plotted using the ggplot2 package¹¹¹ in RStudio. Permutational multivariate analysis of variance (PERMANOVA) was used to test for significance in beta diversity analyses¹¹².

2.3. Results

Removal of chloroplast sequences from the samples resulted in a total of 4,167 ASVs remaining. Subsequent rarefaction removed 56 ASVs by random subsampling, leaving a total of 4,111 ASVs. After transforming the data to relative abundance, filtering out low abundance taxa resulted in a further reduction to a total of 3,221 ASVs remaining for analysis.

Figure 3 shows the abundance of the phyla present in our samples. At the phylum level, the most abundant phyla were *Bacteroidetes* and *Firmicutes*, followed by *Proteobacteria*. Other phyla present at noticeably lower relative abundances were *Actinobacteria* and *Verrucomicrobia*. When the samples were merged according to secretor status, as shown in Figure 4, the abundance of the different phyla was similar regardless of secretor status.

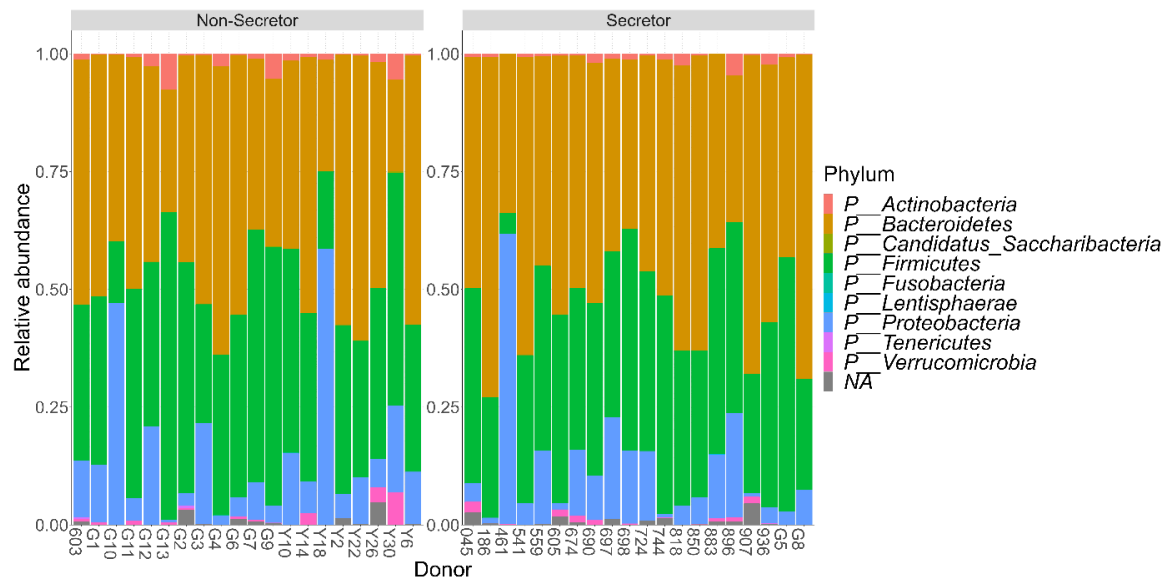


Figure 3 – Phylum-based composition of the faecal microbiota of the cohort, by donor.

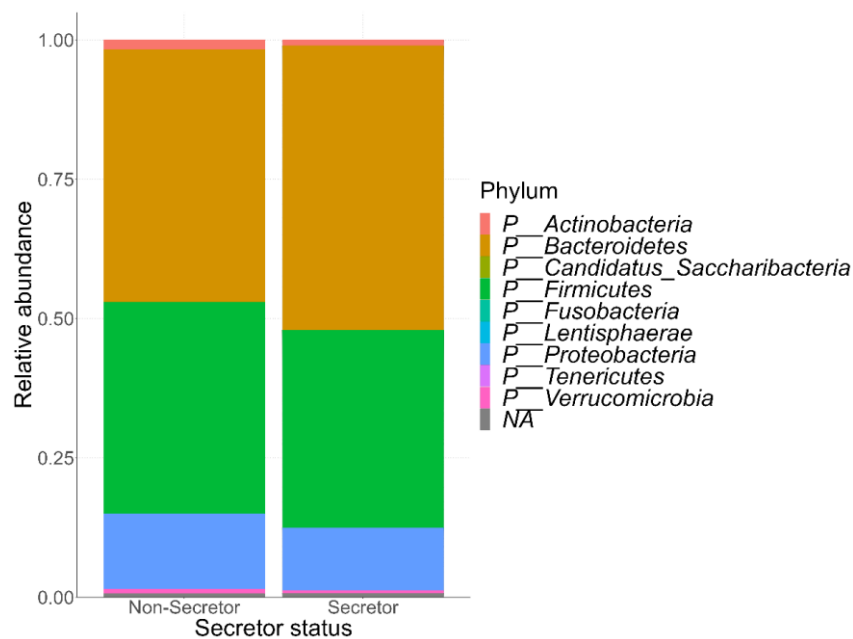


Figure 4 – Phylum-based composition of the faecal microbiota of the cohort, aggregated by secretor status.

Figure 5 shows that there was no significant difference ($p < 0.05$) between secretors and non-secretors across all phyla.

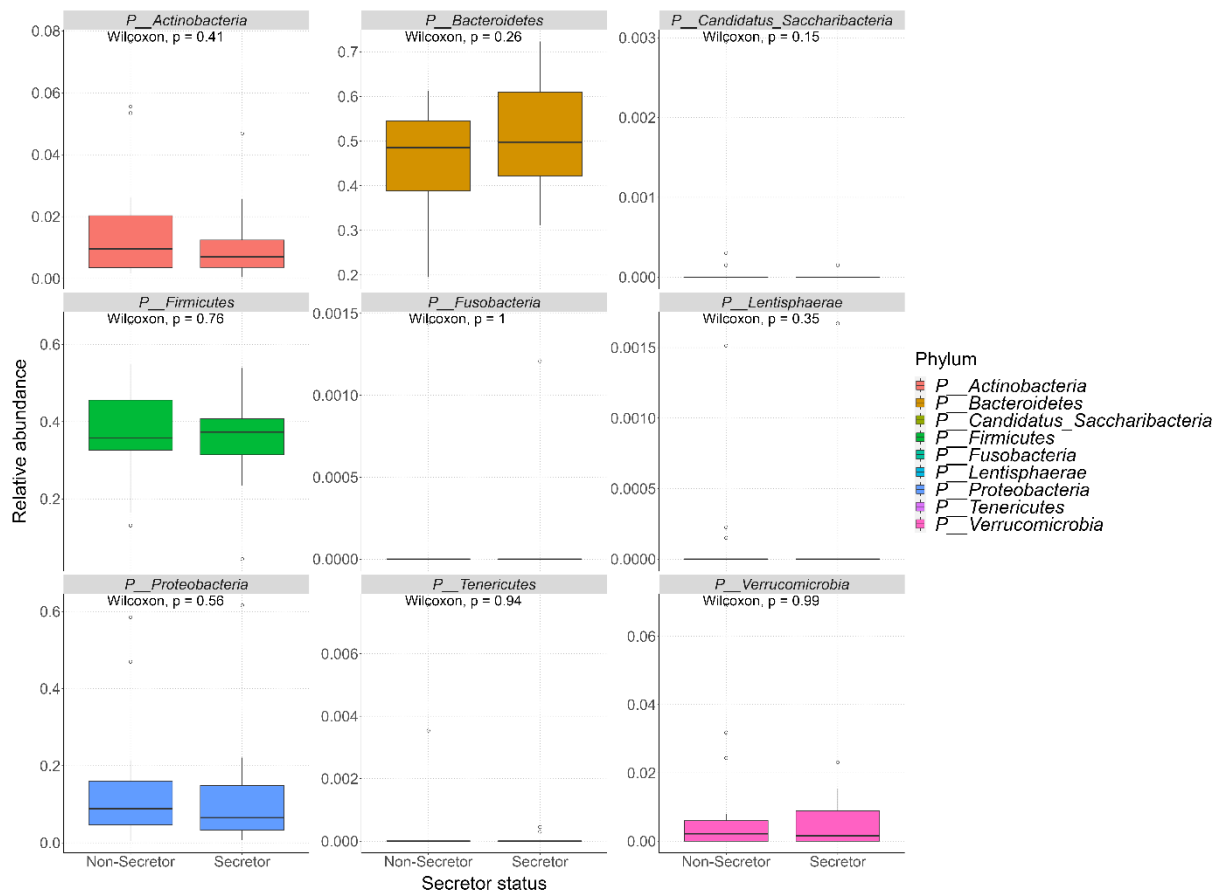


Figure 5 – Phyla relative abundance and corresponding Wilcoxon rank-sum test to assess possible significant differences between secretors and non-secretors.

The taxonomic assignments showed that only 6% of the ASVs could be resolved to the species level (191 out of 3,221 ASVs), while 66% could be classified to the genus level (2,141 out of 3,221 ASVs). To maintain sufficient statistical power without compromising resolution, further analyses were performed at the genus level.

The genus *Lactobacillus* has recently been reclassified into 25 genera⁴⁷. However, this change in taxonomy of this genus has not yet been implemented in all taxonomy-based databases. Our data were initially processed at the IKMB, where taxonomic annotation was performed using the DADA2 classifier and RDP¹⁰¹, resulting in an output based on the previous *Lactobacillus* classification. Therefore, in this study, the previously

assigned genera were manually curated using EZTaxon¹¹³ to obtain the most up-to-date taxonomic classification.

No significant differences were observed between secretors and non-secretors when comparing the relative abundances of bifidobacteria ($p=0.17$; Figure 6) and lactobacilli (all genera combined, $p=0.54$; Figure 7).

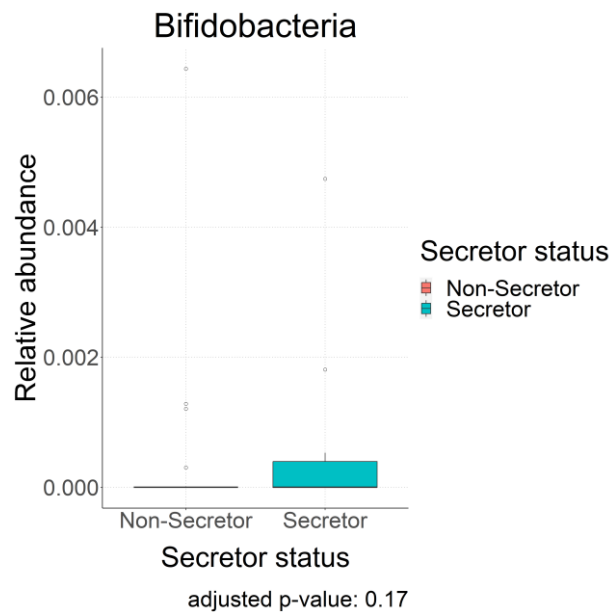


Figure 6 – Relative abundance of bifidobacteria in our cohort.

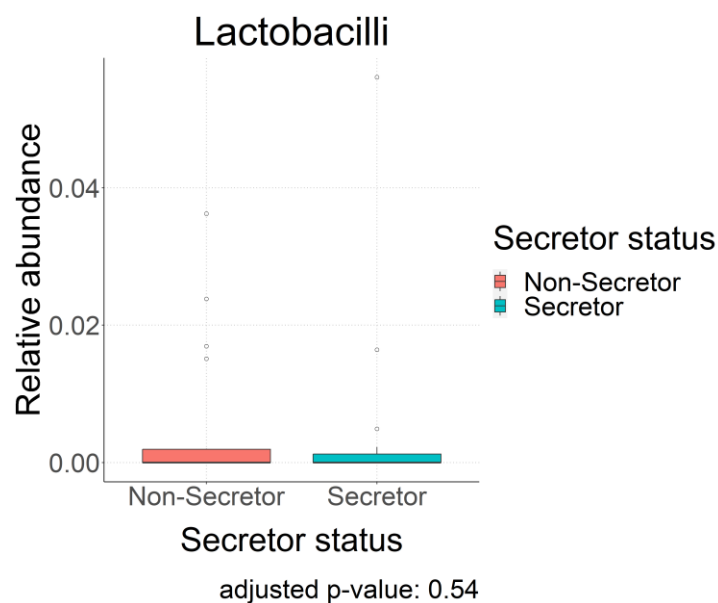


Figure 7 – Relative abundance of lactobacilli (all genera) in our cohort.

We then filtered for ASVs classified as lactobacilli only to investigate the diversity of the lactobacilli genera present in our samples as a function of the secretor status, as can be seen in Figure 8. Bearing in mind that these ASVs had very few counts per donor, it is still interesting to note that different genera are present depending on the secretor status. In fact, it can be observed that secretors present mainly lactobacilli of the genera *Limosilactobacillus* and *Ligilactobacillus*, while non-secretors are characterised by a wider variety of genera, mainly *Limosilactobacillus*, *Levilactobacillus*, *Latilactobacillus*, *Lactiplantibacillus* and *Lacticaseibacillus*.

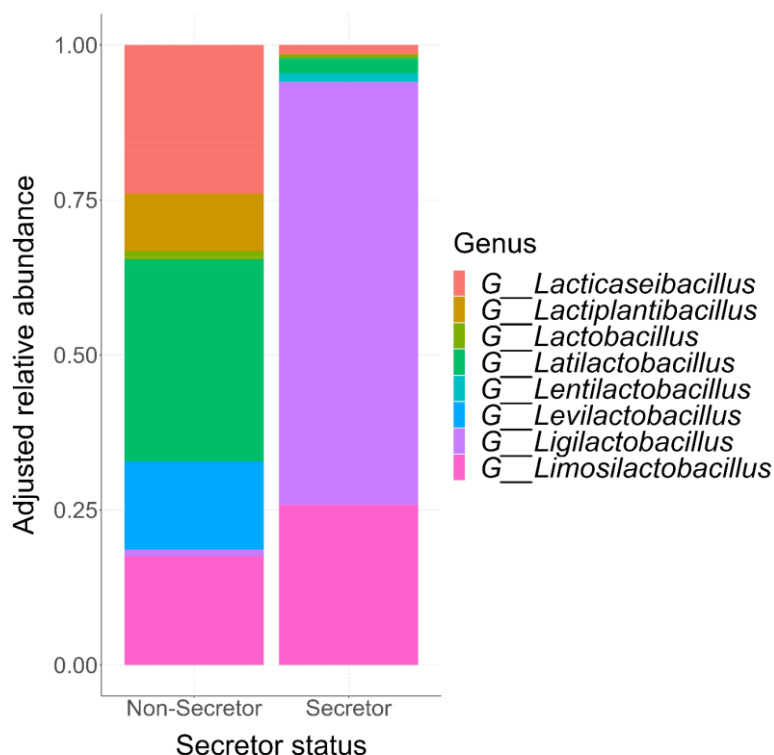


Figure 8 – Relative abundance of each lactobacilli genus according to the new taxonomic classification⁴⁷, adjusted for ASVs classified as lactobacilli only.

Diversity analyses performed on the whole microbiome data showed that there was no significant difference in alpha diversity between secretors and non-secretors. Figure 9 and Figure 10 show that there was no significant difference in the richness of the

microbiome depending on secretor status in our cohort, as measured by the number of observed ASVs and the Chao1 indicator. In Figure 11, there are also no significant differences between secretors and non-secretors in terms of species diversity, as measured by the Shannon index. However, when looking at beta diversity, the homozygous *FUT2*^{+/+} genotype samples appear to be more homogeneous/less diverse than the heterozygous *FUT2*^{+/-} or the recessive *FUT2*^{-/-} genotypes, as shown in Figure 12. However, there was no significant association ($p < 0.05$) between beta diversity and either secretor status or *FUT2* genotype, as tested by PERMANOVA.

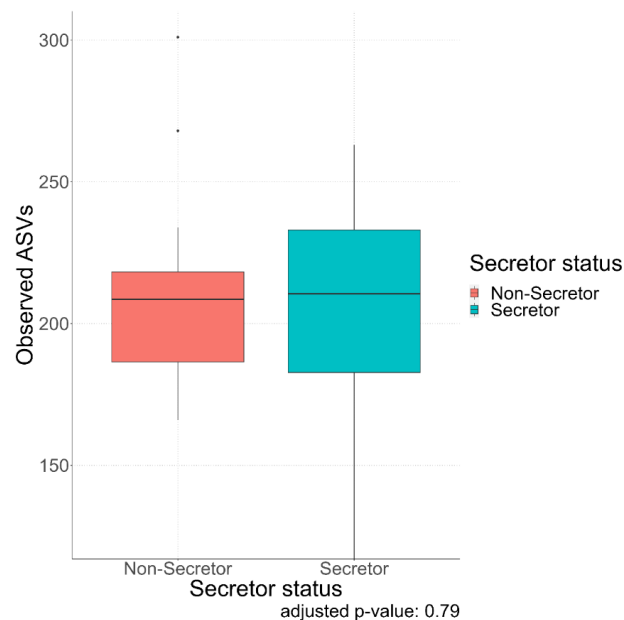


Figure 9 – Number of observed ASVs as a measure of alpha diversity depending on secretor status.

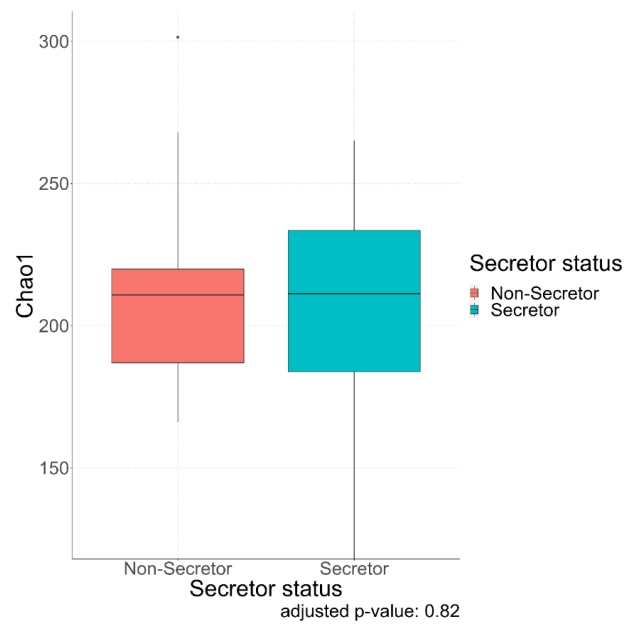


Figure 10 – Chao1 indicator depending on secretor status.

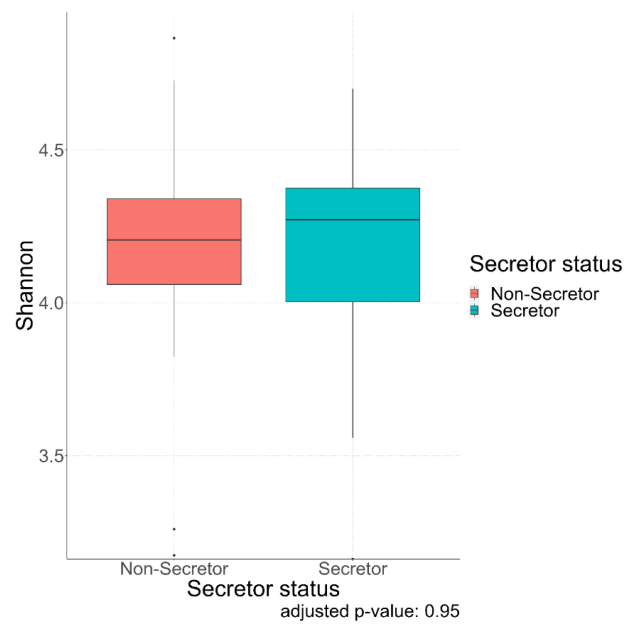


Figure 11 – Shannon index depending on secretor status.

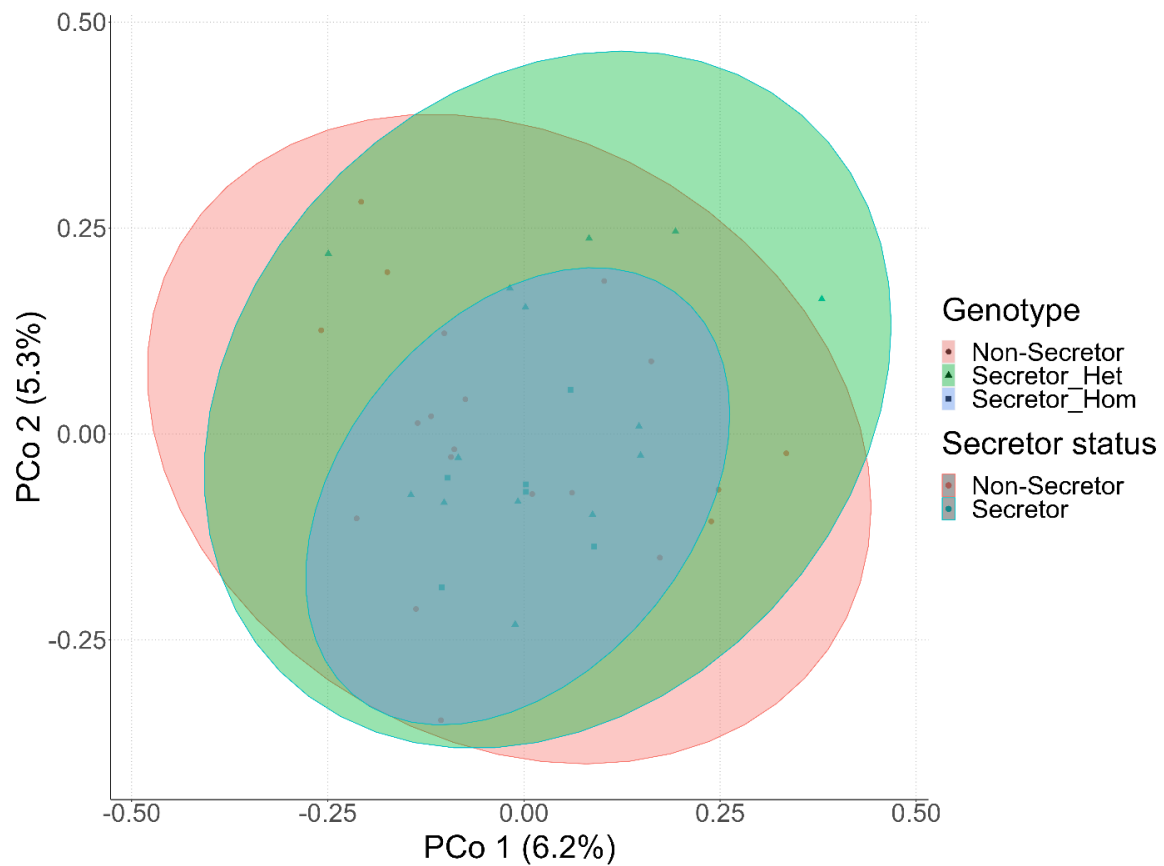


Figure 12 – Beta diversity represented by a Principal Coordinate Analysis (PCoA) based on the Bray-Curtis distance between samples/donors. Sample points are coloured according to secretor status and their shape represents *FUT2* genotype. The fill colour of the ellipse represents the *FUT2* genotype, while its contour is coloured according to the secretor status.

2.4. Discussion

It has been previously reported that *Bacteroidetes* and *Firmicutes*, followed by *Proteobacteria*, *Tenericutes*, *Actinobacteria* and *Verrucomicrobia* are the most dominant phyla of the human gut microbiota, accounting for up to 90% of the total microbial population¹¹⁴. The results shown in Figure 4 are thus very similar in terms of dominant phyla. Furthermore, our results in Figure 5 show that there were no significant

differences between the abundances of these major phyla in the faecal microbiota of secretors and non-secretors, as previously reported by Davenport *et al.* (2016)⁹³.

Regarding the abundance of our taxa of interest, lactobacilli and bifidobacteria are minor components of the typical gut microbiota according to the literature, representing around 0.3% and 2-14% of the healthy adult microbiome, respectively^{49,67,68}. Our overall relative abundance results for lactobacilli are comparable to those in the current literature, as seen in Figure 7, but we did not observe differences depending on the secretor status. Rausch *et al.* (2011) used 16S rRNA gene sequencing data from gut mucosa samples to show that there was an increased abundance of lactobacilli associated with healthy secretors, which they hypothesised could be explained by the fact that some lactobacilli species possess adhesins specifically targeting ABO blood group antigens in the mucosa^{51,115}. The differences between our results are then most likely due to the different sampling sites, namely faeces in this study and mucosa in the study by Rausch *et al.* (2011).

Regarding bifidobacteria, their relative abundance in our cohort, shown in Figure 6, is two orders of magnitude lower than the values reported in the literature. This is most likely due to the use of V1-V2 primers for 16S rRNA gene sequencing at the IKMB. These primers are not well suited for amplification of 16S rRNA genes of bifidobacteria and it has been previously reported that they can even lead to “a complete loss of identification of *Bifidobacterium*”¹¹⁶. Therefore, a statement regarding the abundance of bifidobacteria according to the secretor status cannot be reliably made because these primers probably caused an underrepresentation of these bacteria in our study. At the same time, previous research has shown conflicting results regarding the relative abundance of bifidobacteria as a function of the secretor status. On the one hand, two studies have shown that secretors have higher abundance and diversity of

bifidobacteria in their faecal content compared to non-secretors^{34,92}. On the other hand, Davenport *et al.* (2016) found no significant differences in faecal bifidobacteria abundance associated with the secretor status using 16S rRNA gene sequencing and primers targeting on the V4 hypervariable region⁹³. Given the larger sample size of the latter study, it is likely that we would have obtained similar non-significant results in this study if the primers chosen were able to adequately capture bifidobacteria.

Overall, we can conclude that based on our particular type of analysis and using this particular cohort, we did not observe significant differences in the relative abundance of lactobacilli and bifidobacteria between secretors and non-secretors. However, we cannot exclude the possibility that such differences could exist in reality, given our relatively small sample size, the low abundance of both taxa in the human microbiome, and the non-optimised choice of primers to properly detect bifidobacteria. Therefore, to obtain higher quality evidence, we suggest using different primers for rRNA gene sequencing, using a comprehensive sequencing technique such as shotgun metagenomics and a larger sample size. Quantitative polymerase chain reaction (qPCR) could also be a way to assess abundances in parallel with sequencing in order to obtain more accurate results and absolute quantification as well¹¹⁷.

Despite the lack of differences in the relative abundance of lactobacilli, when we focused on the distribution of the different lactobacilli genera present in our samples, we observed distinct compositions depending on the secretor status, as shown in Figure 8. Secretors presented mainly the genera *Limosilactobacillus* and *Ligilactobacillus*, while non-secretors showed a wide variety of genera, mainly *Limosilactobacillus*, *Levilactobacillus*, *Latilactobacillus*, *Lactiplantibacillus* and *Lacticaseibacillus*. Interestingly, the two most abundant genera in secretors are typically autochthonous (e.g. *L. mucosae*, *L. ruminis*), which is an unusual feature

among the lactobacilli species, which are mainly transient¹¹⁸. This could indicate host-microbe co-adaptation in the secretor gut, retaining more autochthonous lactobacilli species, as compared to the non-secretors, which in this study present mostly transient lactobacilli species from probiotics or food origin. However, it is important to note that the correlation between these two lactobacilli genera and secretors has yet to be further investigated and that the number of lactobacilli counts per donor was low in this analysis, preventing more robust conclusions.

Finally, regarding microbiome diversity analyses, no differences were observed between secretors and non-secretors in terms of alpha diversity (both in terms of species richness and evenness) using the number of observed ASVs, the Chao1 estimator and the Shannon index as metrics (Figure 9, Figure 10 and Figure 11, respectively). The study by Davenport *et al.* (2016) used five different diversity metrics to examine alpha diversity, namely Faith's phylogenetic diversity, the number of observed OTUs, the Chao1 richness estimator, the Gini index and the Shannon diversity index, and found that no alpha diversity metric was statistically significantly between secretors and non-secretors⁹³. Gampa *et al.* (2017) also found in their study that secretors and non-secretors did not have significant differences in alpha diversity at the OTU level¹¹⁹.

Beta diversity was lower in homozygous secretors compared to the other genotype groups, as seen in Figure 12, and thus the microbiota of homozygous secretors seemed to be more homogeneous than that of heterozygous secretors or non-secretors. However, this difference was not significant according to PERMANOVA. Furthermore, it is important to note that the number of homozygous secretors in our study (6 donors) is lower than that of heterozygous secretors and non-secretors (14 and 20 donors, respectively), which could influence the results as commonly used

measures of beta diversity do not take sample size into consideration¹²⁰. The study by Davenport *et al.* (2016) found no significant differences in beta diversity associated with the secretor status for any of the metrics they investigated (UniFrac and Bray-Curtis distance), which is consistent with our results⁹³. However, Rausch *et al.* (2011) found in their study on mucosal samples from Crohn's disease patients that microbial community variation depends not only on disease status but also on disease-*FUT2* genotype interactions⁵¹. Therefore, future research could benefit not only from methodological improvements as discussed above, but also from focusing on the interaction between *FUT2* genotype and associated diseases, especially as it has been hypothesised that fucosylation may be a protective mechanism during acute or chronic stress, and thus the secretor status may be more relevant under such conditions²⁹.

Overall, this study found no significant results for the first research question of this doctoral project, which hypothesised that lactobacilli and bifidobacteria would be more abundant in secretors. Regarding the diversity of lactobacilli and bifidobacteria and whether more fucose-utilising bacteria were present in secretors (second research question), there was a different lactobacilli composition depending on secretor status. However, as this analysis was carried out at the genus level due to an insufficient number of ASVs assigned at the species level, it was not possible to assess the presence of fucose-utilising bacteria in our cohort, as this is a strain-specific property of certain lactobacilli and bifidobacteria species. The next studies focused on a more in-depth analysis using a culture-based approach to address the third research question and look more closely at the fucose catabolic properties of these taxa.

3. Fast identification method for screening bacteria from faecal samples using Oxford Nanopore Technologies MinION sequencing

Ana Sofia G. Borges¹, Meghna Basu^{2,3}, Erik Brinks¹, Corinna Bang⁴, Gyu-Sung Cho^{1*}, John F. Baines^{2,3}, Andre Franke⁴ and Charles M.A.P. Franz¹

Affiliations

¹Department of Microbiology and Biotechnology, Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Hermann-Weigmann-Straße 1, 24103 Kiel, Germany

²Section of Evolutionary Medicine, Institute for Experimental Medicine, Christian-Albrechts-University Kiel, UKSH, Campus Kiel, Michaelisstraße 5, 24105 Kiel, Germany

³Max Planck Institute for Evolutionary Biology, August-Thienemann-Straße 2, 24306 Plön, Germany

⁴Institute of Clinical Molecular Biology, Christian Albrechts University of Kiel, Rosalind-Franklin-Straße 12, 24105 Kiel, Germany

*Corresponding Author. Contact: Gyusung.Cho@mri.bund.de

Published in *Current Microbiology* (2023)

DOI: 10.1007/s00284-023-03201-7

3.1. Contextualization of the publication within the thesis

As shown in the previous chapter, lactobacilli and bifidobacteria can be particularly difficult to study through culture-independent analyses due to their low abundance in human faecal samples. Even when the primers used in 16S rRNA gene sequencing are adjusted or shotgun metagenomics are used to describe the diversity and abundance of these taxa instead, the results are usually not very statistically powerful¹²¹. However, given the important role of these taxa for gut health and their probiotic properties, it is still crucial to study them.

For this particular project, it was necessary to isolate lactobacilli and bifidobacteria from faecal samples to answer our research questions. However, when faecal samples are cultured under non-selective conditions, all different kinds of bacteria can grow. Selective conditions can be used to restrict the types of bacteria that grow in the culture media. Certain bacteria have unique features, making it possible to design growth conditions that will exclusively target those particular species. However, this is not the case for our taxa of interest, especially given their great diversity. Nevertheless, there are certain conditions (e.g. selective media, temperature, anaerobic conditions) that can increase the likelihood of culturing lactobacilli and bifidobacteria from human faecal samples.

After culturing bacteria, they must be correctly and unambiguously identified to ensure that the taxa of interest have been isolated. Many of the methods used to identify bacteria are time-consuming, expensive or highly specific for bacteria with certain characteristics¹²². Since we wanted to isolate a very diverse group of species, we developed a rapid, inexpensive and broad-spectrum method for bacterial identification to overcome these methodological challenges. This method, coupled with a few growth conditions that favour the growth of our taxa of interest, allowed us to obtain a

considerable number of lactobacilli and bifidobacteria isolates for further studies, as will be shown in this chapter.

An important side note is that in this publication long donor codes and media names were used to name the isolates. However, throughout the remaining chapters of the thesis, long donor codes have been abbreviated to the last three digits and media names have been abbreviated to the first letter to simplify descriptions and plots. For example, a strain named 10693850MRS7 in this publication is referred to as 850M7 in the other chapters of the thesis.

3.2. Abstract

Most bacterial identification methods require extensive culturing, strain purification and DNA extraction protocols. This leads to additional expenses and time lags when isolating specific bacteria from complex microbiological ecosystems. This study aimed to develop a fast and robust method for identification of lactobacilli, bifidobacteria and *Bacteroides* in human faecal samples. Bacteria from faecal samples were cultured anaerobically on selective media. Sonication-based DNA extraction was performed, followed by almost complete 16S rRNA gene polymerase chain reaction amplification and MinION sequencing with the Flongle adapter. Sequence analysis was performed using NanoCLUST, while RStudio was used for graphics. For 110 of the 125 colonies investigated, 100% of reads were attributed to a single species, while the remaining 15 colonies consisted of mixtures of up to three different species. The proposed bacterial identification method is advantageous for isolating particular bacteria for which there are no exclusively selective media, as it avoids lengthy colony purification and DNA purification methods, and yields a quick colony identification with high accuracy. This method can thus be used for directly screening for pure cultures of target microorganisms, and is suitable for the identification of bacteria in culturomics studies.

3.3. Introduction

Non-selective microbiological growth media can support the growth of a wide variety of microorganisms by supplying a rich nutrient source. On the other hand, selective media often include growth inhibiting substances that select for the types of bacteria that can grow in the presence of these and inhibit or eliminate the growth of unwanted microbiota¹²³. The use of selective media is often necessary for targeted isolation of specific bacteria that occur in complex microbiological ecosystems such as food, plants, soil, medical samples and the gastrointestinal tract. The drawback of selective media is often the lack of exclusive selectivity, meaning that for a group of microorganisms, the isolates obtained from the media are not always those targeted by the selective conditions.

For example, the selective medium de Man, Rogosa and Sharpe (MRS)¹²⁴ is widely used for the isolation of lactic acid bacteria (LAB), but other Gram-positive, acid-tolerant bacteria such as bifidobacteria and certain *Bacillus* spp. can grow on this medium, as can some yeasts and moulds¹²⁵. Furthermore, many *Bifidobacterium* media available rely on the MRS agar base with various additives (e.g. cysteine hydrochloride, mupirocin, lithium chloride or raffinose) that make it selective for these bacteria. Lithium propionate MRS agar (LP-MRS) was previously shown to support the growth of bifidobacteria, while LAB such as *Lactobacillus delbrueckii* and *Streptococcus thermophilus* were inhibited¹²⁶. Another example of selective media is the *Bacteroides*-specific *Bacteroides fragilis* bile-esculin (BBE) agar for the isolation of *Bacteroides* spp., which contains the selective agents gentamicin (which inhibits facultative anaerobes) and bile (which inhibits most Gram-positive bacteria and anaerobic microorganisms other than the *Bacteroides fragilis* group). However, some

Enterococcus, *Klebsiella* and *Fusobacterium* strains may grow on BBE agar, though they may be distinguished by different colony characteristics¹²⁷.

To obtain pure cultures for storage in culture collections, characterization and comparisons of strains, there is still a need for purification of isolates which is generally done by repeated rounds of isolating colonies, growing these in liquid media and streaking these out for purity onto agar media. This is both time consuming and costly in terms of media and equipment. The isolation from agar media then usually requires further identification of the strains either by 16S rRNA gene sequencing, or other suitable methods such as matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) strain typing¹²².

In recent years, considerable attention has been given to culturomics, which is based on a better understanding of the bacterial environments and adaptation of this knowledge to optimize growth media by e.g. adding new elements into the culture media, isolating microcolonies or increasing the number of culture conditions¹²³. It has been particularly useful for the study of the gut microbiota together with metagenomic studies, which have unravelled a large variety of gut-associated microorganisms that had previously not been described, due to their failure to grow on laboratory media¹²². Such culturomics studies generally rely on MALDI-TOF or 16S rRNA gene polymerase chain reaction (PCR) to identify bacteria^{122,128}.

This study aimed to develop a method which would allow for the fast isolation and identification of bacterial colonies obtained from human faecal samples. This method relied on the use of selective media to isolate LAB, bifidobacteria and *Bacteroides* spp. present in faecal samples. Colonies on plates were picked, grown in broth media and directly identified by using Oxford Nanopore Technologies (ONT) sequencing and bioinformatic analyses, which also allowed an insight into the purity of the isolates.

Depending on the application (e.g. culturomics, genomics), this method can be advantageous in terms of time and costs when compared to alternatives.

3.4. Materials and Methods

3.4.1. Isolation of bacteria from human faeces

Healthy blood donors from Northern Germany were recruited by mail and asked to provide stool samples in diverse buffer solutions. Transport of samples was done by mail within 24 hours after sample collection. Immediately after their arrival at the Microbiome laboratory of the IKMB, the stool samples were aliquoted and stored at -80 °C until further processing. The study was conducted in accordance with the Declaration of Helsinki, and ethical approval was granted by the ethics committee at Kiel University (AZ A103/14, D590/22). Six randomly chosen faecal samples from this study collective were transferred to an A45 anaerobic workstation (Don Whitley Scientific, Meintrup, Herzlake, Germany) with an atmosphere consisting of 10% hydrogen, 10% carbon dioxide and 80% nitrogen at 37°C. Inside the anaerobic chamber, the faecal samples were thawed and 100 µl of faeces were added to 900 µl of quarter-strength Ringers solution (QSRS) (Merck, Darmstadt, Germany) and diluted in a ten-fold dilution series in QSRs. 100 µl aliquots of these dilutions were plated out onto MRS agar (Merck), LP-MRS agar (MRS containing 3 g/L sodium propionate and 2 g/l lithium chloride)¹²⁶ and BBE agar¹²⁷ to isolate LAB, bifidobacteria and *Bacteroides* spp., respectively. All dilutions were plated out by spread plating. The QSRs and all culture media plates were previously incubated for 24 h in the anaerobic chamber to allow to equilibrate to anaerobic conditions. Plates were incubated in the anaerobic chamber at 37°C for 48 h, after which colonies from plates of the highest dilutions which showed well-separated single colonies were randomly picked for isolating bacteria.

Ten colonies were picked per medium per donor. Single colonies were transferred to Hungate tubes containing 10 ml of anaerobic broth media of the same type as the plate from which the colonies were picked for lactobacilli and *Bifidobacteria*, or in supplemented Brain Heart Infusion Supplemented (BHIS) medium for *Bacteroides*¹²⁹. The broth media used for inoculation were previously prepared anaerobically by flushing with N₂ gas at 0.5 bar in the Hungate tubes. After autoclaving, the Hungate tubes with broth media were incubated for 24 h in the anaerobic chamber to equilibrate to anaerobic conditions. Single colonies were incubated anaerobically for 2-5 days at 37°C depending on the turbidity of the medium, so that only well-grown cultures were used for DNA isolation.

3.4.2. DNA isolation and 16S rRNA gene PCR amplification

A 10 µl aliquot of each culture was removed after incubation in the anaerobic chamber and added to 990 µl of sterile bi-distilled water in Eppendorf cups. The bacterial cultures were placed in an Emmi-H40 ultrasonicator (EMAG AG, Mörfelden-Walldorf, Germany) filled with water and ice, and sonicated at a half wave operating frequency of approx. 40 kHz (total power 440 W, ultrasonic power max. 240 W) for 2.5 min, after which samples were put on ice for 1 min¹³⁰. This procedure was repeated six times. Ice was regularly added to the ultrasonicator as it melted during the sonication process. After the last sonication cycle, the sample was kept on ice for 5 min to allow the debris to settle and 10 µl of supernatant was then collected to serve as DNA template for 16S rRNA gene amplification by PCR. The PCR reaction was performed in a 50 µl volume and consisted of 10 µl of template DNA, 2.5 µl of each forward and reverse primers (final concentration 0.5 µM), 6 µl of dNTPs (final concentrations 600 µM each), 10 µl of 5x *Taq* polymerase buffer and 1 unit of Q5 high-fidelity DNA polymerase (New England Biolabs, Massachusetts, United States). The primers used were specific for

amplification of the almost complete 16S rRNA gene and included the modified forward 27F (5'-AGR GTT TGA TCM TGG CTC AG-3') and the 1492R (5'-TAC CTT GTT ACG ACT T-3') primers^{122,131}. A set of barcoded 16S rRNA gene primers were used to directly apply the PCR products for ONT sequencing. Therefore, barcodes and spacer sequences were added to the 27F and 1492R primers as described elsewhere¹³².

DNA was amplified by PCR (initial denaturation, 94°C for 3 min; 32 cycles of 94°C for 30 s, 56°C for 30 s and 72°C for 1 min 10 s, followed by a final extension at 72°C for 5 min). 5 µl of each PCR product was subjected to electrophoresis on a 1.5% TAE [Tris-acetate- ethylenediaminetetraacetic acid (EDTA)] gel in 1 x TAE buffer for 1 h at 100 V and bands were checked for products of the expected size (ca. 1500 bp) under UV light using the GelRed Nucleic Acid Stain (Merck). The PCR products were purified using the Mag-Bind® TotalPure NGS Kit (Omega) with a ratio of 1:1 (v/v). Purified PCR products were then quantified using the Qubit™ dsDNA BR Assay Kit (Invitrogen™) and a Tecan Spark multimode microplate reader (Tecan) followed by equimolar pooling of the samples.

3.4.3. ONT MinION 16S rRNA gene sequencing and data analysis

200 ng of the equimolar pool were used for sequencing on an ONT MinION Flongle (ONT, United Kingdom) following the “Genomic DNA by Ligation (SQK-LSK109)” protocol for Flongle flow cells (Version: GDE_9063_v109_revA1_14Aug2019). These flow cells allow for the sequencing of up to 96 samples per run. The final library was quantified using the Qubit™ dsDNA HS Assay Kit (Invitrogen™) and a Qubit 3.0 Fluorometer (Invitrogen™). A final library concentration of 20 fmol was loaded on the Flongle flow cell and sequenced for 24 hours.

Once the raw data files were obtained, basecalling was done using the GPU Guppy Basecalling Software (version 3.4.4) with a nvidia RTX 2080. Q-score was assessed using pycoQC (v2.5.2) and was between 8.57 and 10.49. The resulting fastq files were demultiplexed using Porechop (v0.2.4) with customized barcodes¹³². The demultiplexed files can be found in the BioProject database with the accession number PRJNA914123. Q-score was checked using fastqc (v0.11.9) and Geneious Prime (2023.0.1) and values varied between Q13,7 and Q19. The demultiplexed fastq files were further analysed using NanoCLUST (v1.0dev) with standard clustering step parameters¹³³ and species identification of resulting consensus sequences was verified with EZTaxon¹¹³. RStudio (R 4.2.0) and the ggplot2 package were used to graphically represent the resulting taxonomic classification^{106,111}. The overall method is illustrated in Figure 13.

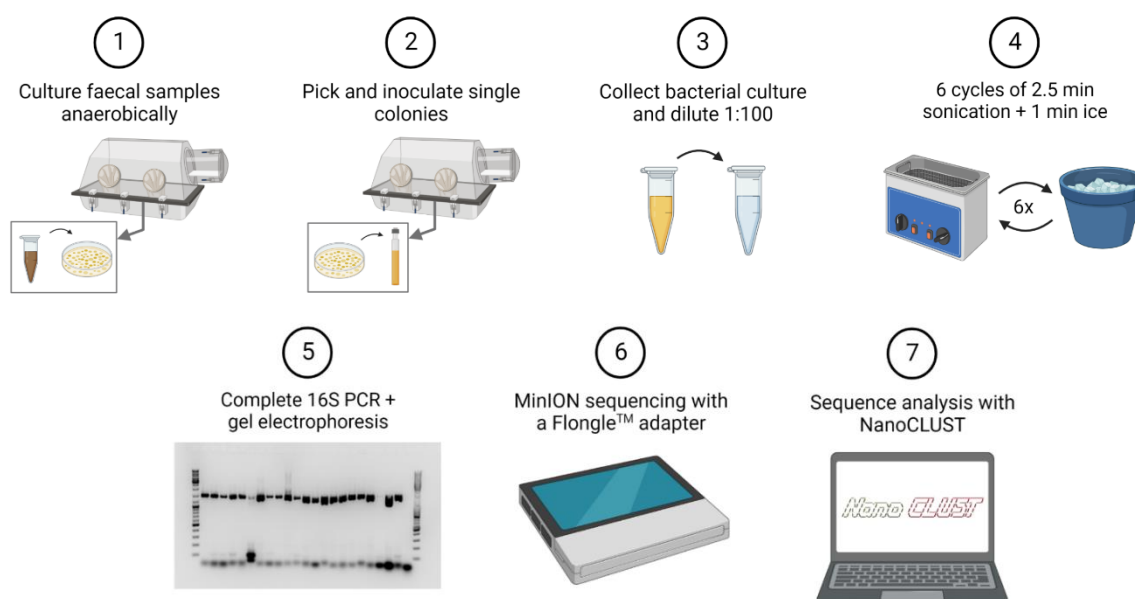


Figure 13 – Schematic representation of the developed method. Figure created with BioRender.com.

3.4.4. Illumina whole genome sequencing (WGS) and identity confirmation

In order to check the species identification obtained with our 16S rRNA gene-based method at the whole genome level, we randomly selected four isolates (10670603MRS5, 10685605MRS4, 10693850MRS2, and 10693850MRS7) and sequenced their genomic DNA using Illumina short read sequencing. Isolates 10670603MRS5 and 10693850MRS7 were sequenced using the Illumina TruSeq Nano DNA LT library Preparation Kit (Illumina, Munich, Germany), while isolates 10685605MRS4 and 10693850MRS2 were sequenced using the Illumina DNA Prep Tagmentation Kit (Illumina) according to the manufacturer's protocol. All isolates were sequenced on an Illumina MiSeq sequencer using the MiSeq Reagent Nano Kit v2 (Illumina) (500-cycles) for paired end sequencing with 2x251 cycles according to the manufacturer's instructions, but on separate runs for the different protocols.

The raw WGS data were trimmed using the Trimmomatic pipeline (v. 0.32; parameters: Phred=33, sliding window=4:15, leading=3 and minlen=45)¹³⁴ and *de novo* assembly was performed using the SPAdes pipeline (v. 3.10.0; parameters: --pe, kmer: 21, 33, 55, 77; --careful)¹³⁵. After *de novo* assembly, contig sequences that were shorter than 500 bp or contaminated with spiked PhiX sequence were removed using the BBDuk pipeline¹³⁶. The average nucleotide identity (ANI) and the digital DNA-DNA hybridization (dDDH) of WGS data were performed using the OrthoANI pipeline (v. 1.2) with USEARCH tool as default parameters¹³⁷ and the Genome-to Genome Distance Calculator using formula 2, respectively^{138–140}.

3.5. Results and Discussion

The results of the sequencing of 180 isolates (60 for each agar medium) are shown in Figure 14 according to the agar medium the colonies were isolated from. Of the 60

colonies isolated from MRS agar medium, the majority could be identified as *Bifidobacterium* spp. (n=24), lactobacilli (different genera based on the reclassification of the genus *Lactobacillus*⁴⁷) (n=19) and *Enterococcus* spp. (n=12), while only few were identified as *Streptococcus* (n=2), *Weissella* (n=1), *Alteribacillus* (n=1) and *Dialister* (n=1). Regarding the LP-MRS agar used to isolate bifidobacteria, the majority of the colonies could be identified as *Bifidobacterium* (n=22), lactobacilli (n=18) and *Enterococcus* (n=16), while one of each could be identified as *Blautia*, *Nitrosomonas*, *Pantoea* and *Ruminococcus*. In case of the BBE agar for isolation of *Bacteroides* spp., the bacteria were identified as *Bacteroides* (n=42), *Parabacteroides* (n=13), *Phocaeicola* (n=3) and *Coproccoccus* (n=2) (Figure 14).

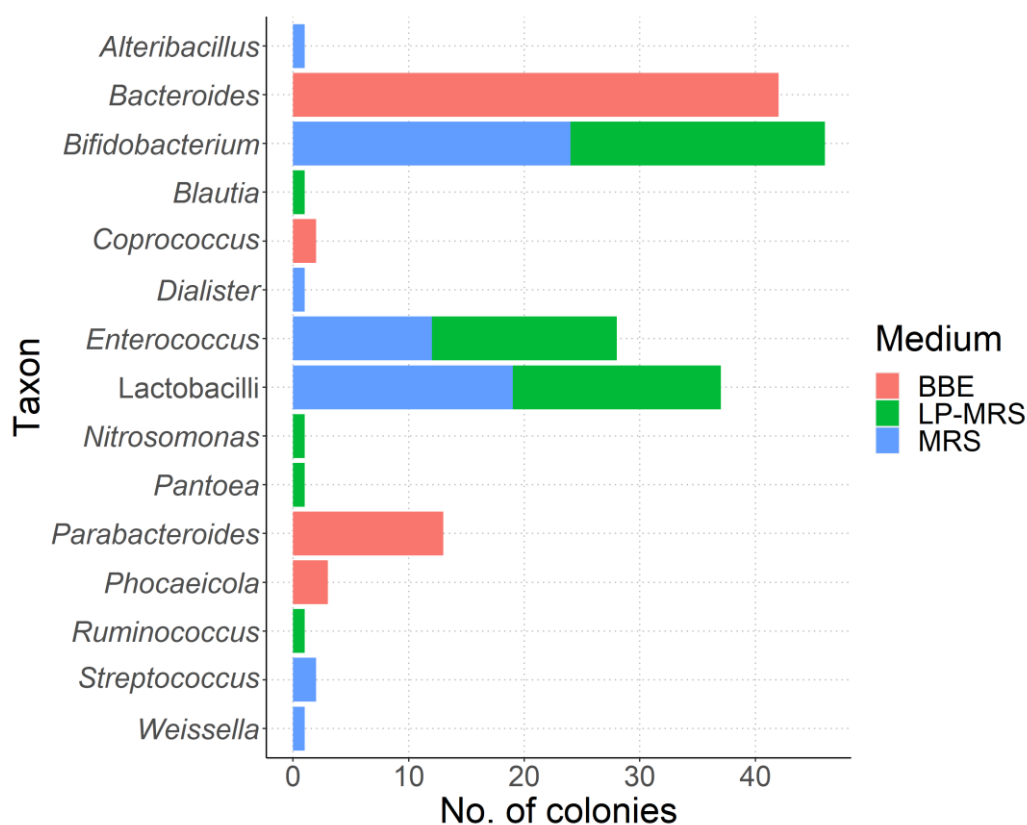


Figure 14 – 16S rRNA gene-based identification of 180 colonies from faecal samples according to culture media. In case of impure colonies only the taxon with highest number of sequence reads was included. Red colour corresponds to colonies cultured

in BBE medium, green for colonies cultured in LP-MRS and blue for colonies cultured in MRS.

For analysis of the predominant species, in the case of impure colonies where more than one genus/species occurred, only the species with the highest numbers of sequence reads were included. The predominant *Bacteroides* (*Ba.*) spp. isolated from BBE agar were *Ba. fragilis* and *Ba. xylanisolvens*, while *Ba. uniformis*, *Ba. ovatus* and *Ba. thetaiotamicron* were also isolated frequently. Only a few isolates of *Ba. caccae* and *Ba. cellulosilyticus* could be obtained (Figure 15). Among the lactobacilli (*L.*), the genera *Lactobacillus*, *Lacticaseibacillus*, *Latilactobacillus* and *Ligilactobacillus* were identified. Interestingly, *L. acidophilus* could only be isolated from MRS medium (Figure 15), while the other lactobacilli species could be isolated from both MRS and LP-MRS. *L. sakei* was most frequently isolated, while *L. rhamnosus*, *L. acidophilus*, *L. curvatus*, and *L. ruminis* could be isolated less frequently. The *Enterococcus* (*E.*) spp. isolated from both LP-MRS and MRS included mainly *E. faecalis* and *E. faecium*, while bacteria from 2 colonies were identified as *E. mundtii*. Bifidobacteria could be isolated from both MRS and LP-MRS in most cases, except for *Bifidobacterium* (*B.*) *dentium*, for which only few isolates were obtained. The most frequently isolated bifidobacteria were *B. longum*, *B. adolescentis* and *B. animalis*, while *B. bifidum* and *B. dentium* were isolated rarely (Figure 15).

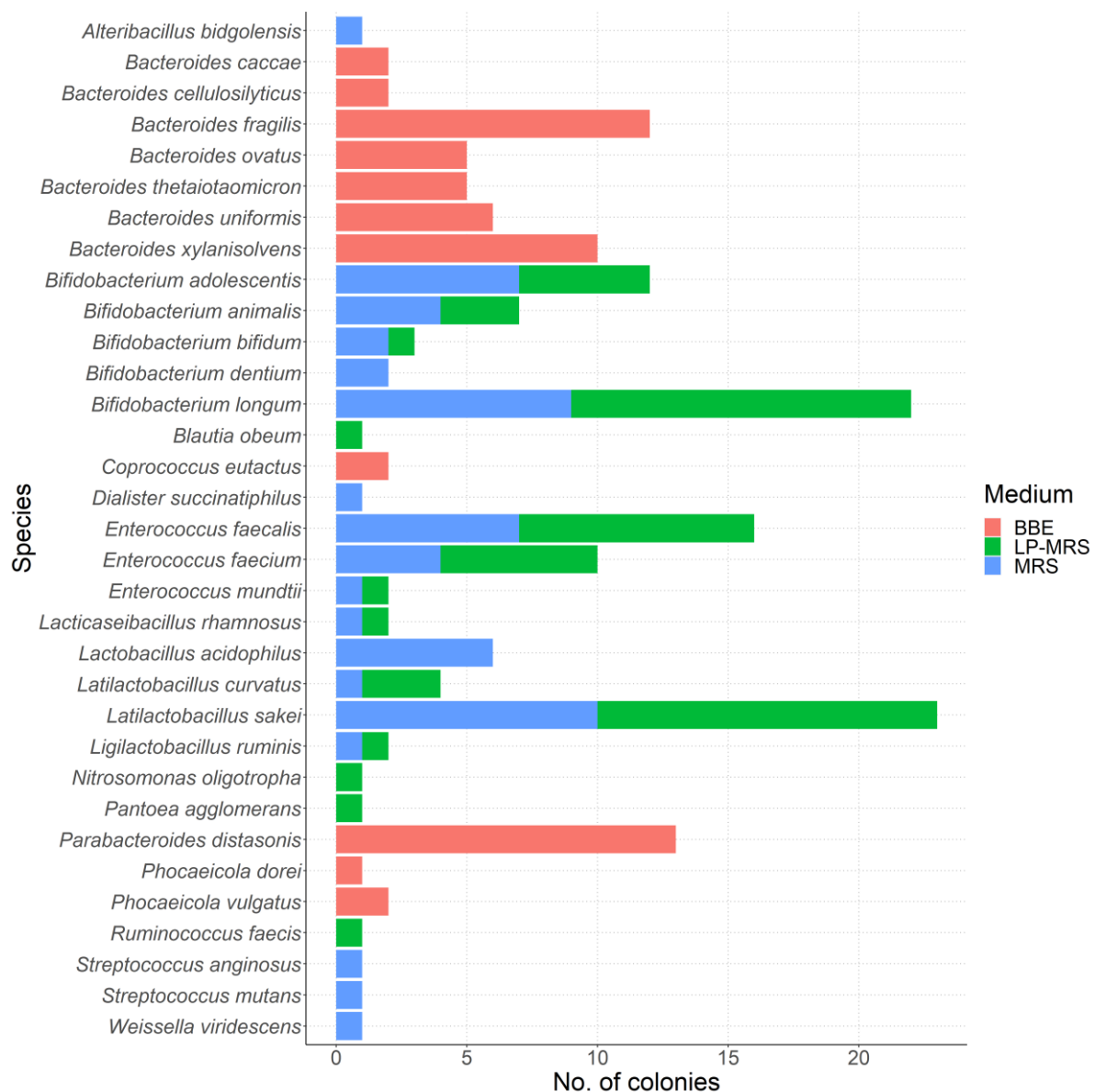


Figure 15 – Species-level 16S rRNA gene-based identification of 180 colonies from faecal samples according to culture media. In case of impure colonies only the species with highest number of sequence reads was included. Red colour corresponds to colonies cultured in BBE medium, green to colonies cultured in LP-MRS and blue to colonies cultured in MRS.

Although the aim of the study was not an in-depth characterization of the LAB, bifidobacteria and *Bacteroides* bacterial communities isolated from healthy human

faeces, the microorganisms isolated were typical of this environment. Of the LAB isolated in this study, the majority of lactobacilli identified (*L. sakei*, *L. rhamnosus*, *L. acidophilus*, *L. curvatus*) may be considered to be of food origin and associated with westernized populations, as described previously¹²¹. For bifidobacteria, the most frequently isolated species were *B. longum*, *B. adolescentis* and *B. animalis*. This is in agreement with the literature, that shows that *B. adolescentis* and *B. longum* are predominant bifidobacteria in adults⁶⁷. The *Bacteroides* spp. identified have also been described to be part of the human faecal community¹⁴¹.

The non-*Bacteroides* bacteria isolated from BBE agar were *Parabacteroides distasonis*, *Phocaeicola vulgatus*, *Phocaeicola dorei* and *Coprococcus eutactus*. The non-lactobacilli isolated from MRS agar included *Alteribacillus bidgolensis*, *Dialister succinatiphilus*, *Streptococcus anginosus*, *Streptococcus mutans* and *Weissella viridescens*. The non-bifidobacteria isolated from LP-MRS included *Blautia obeum*, *Nitrosomonas oligotropha*, *Pantoea agglomerans* and *Ruminococcus faecis* (basonym of *Mediterraneibacter faecis*). As expected, our results show that the selective media used for the isolation of lactobacilli, bifidobacteria and *Bacteroides* are not exclusively selective and supported the growth of other faecal bacteria such as *Dialister*, *Alteribacillus*, *Blautia*, *Nitrosomonas*, *Pantoea*, *Coprococcus*, *Phocaeicola* and *Parabacteroides* spp.

The purity of the colonies was determined by ONT MinION sequencing, and defined as the percentage of reads attributed to the most frequently identified species in the colony. A colony with 100% of reads allocated to a specific species was considered as pure. This purity level was considered sufficient for deciding whether the colony should be further examined for WGS in future investigations. Of the 180 colonies investigated, 125 were identified as lactobacilli, bifidobacteria or *Bacteroides*. Out of these, 110

showed 100% purity (88%; Figure 16, Supplementary Table 1), while 15 of the 125 sequenced colonies were <100% pure, and for these usually a second (or in one case) a third species could be identified (Supplementary Table 1). Out of these 15 mixed colonies, four showed purity between 50-60%, two colonies each showed purity between 60- 70%, 70-80% and 80-90%, and five colonies presented purity between 90-100% (Figure 16). The NanoCLUST plots illustrate samples that were 100% pure indicating a single species, or less than 100% purity, in which two or even three species could be detected (Figure 17). As examples, it is shown that sequencing reads from colony 10685605MRS4 with 100% purity could be assigned to only *B. bifidum* (Figure 17A), while colony 10667674BBE1 could be identified as *Ba. fragilis* (86.69% purity) and *Akkermansia municipihila* (Figure 17B) and colony 10672907LP-MRS8 was determined to be a mix of *B. adolescentis* (96.48% purity), *Rubrobacter bracharensis* and *Alteribacillus bidgolensis* (Figure 17C). Figure 17D shows the proportions of the identified bacteria in these three examples in further detail. It should be noted that often two clusters were obtained for a single species, which correspond to the forward and reverse sequences of the almost complete 16S rRNA gene Flongle sequencing result.

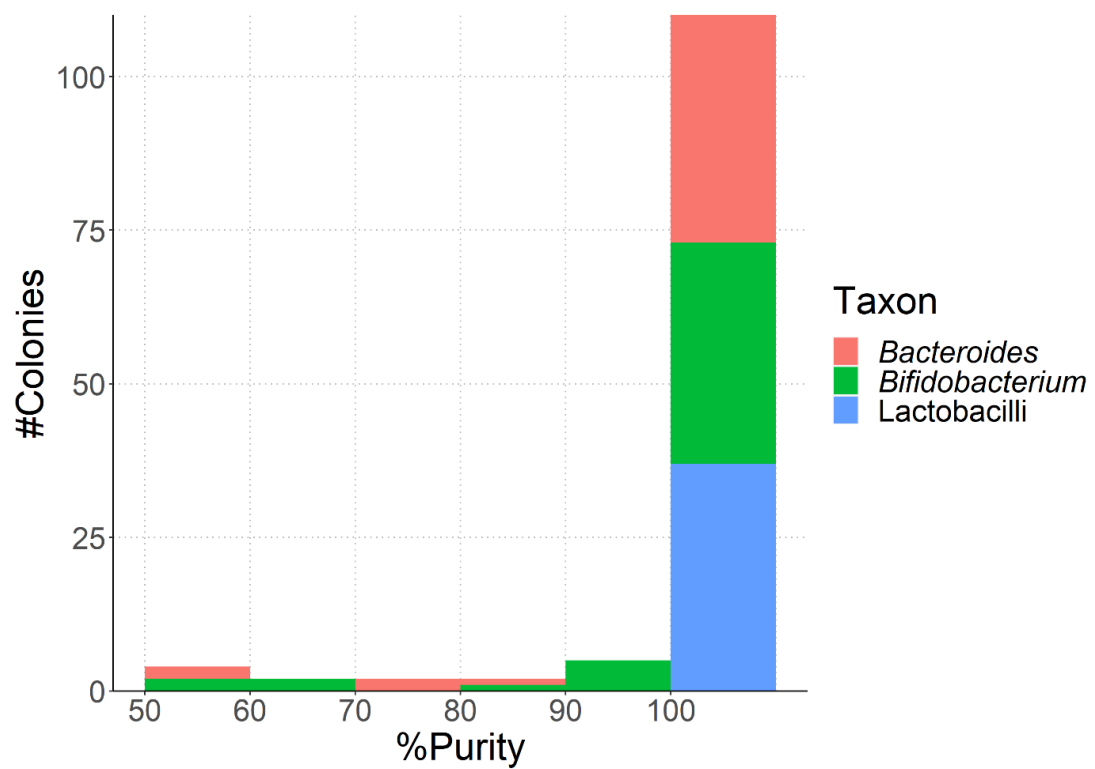


Figure 16 – Purity of the 125 colonies that were identified as lactobacilli, bifidobacteria and *Bacteroides*. Red colour corresponds to colonies identified as *Bacteroides*, green to bifidobacteria and blue to lactobacilli.

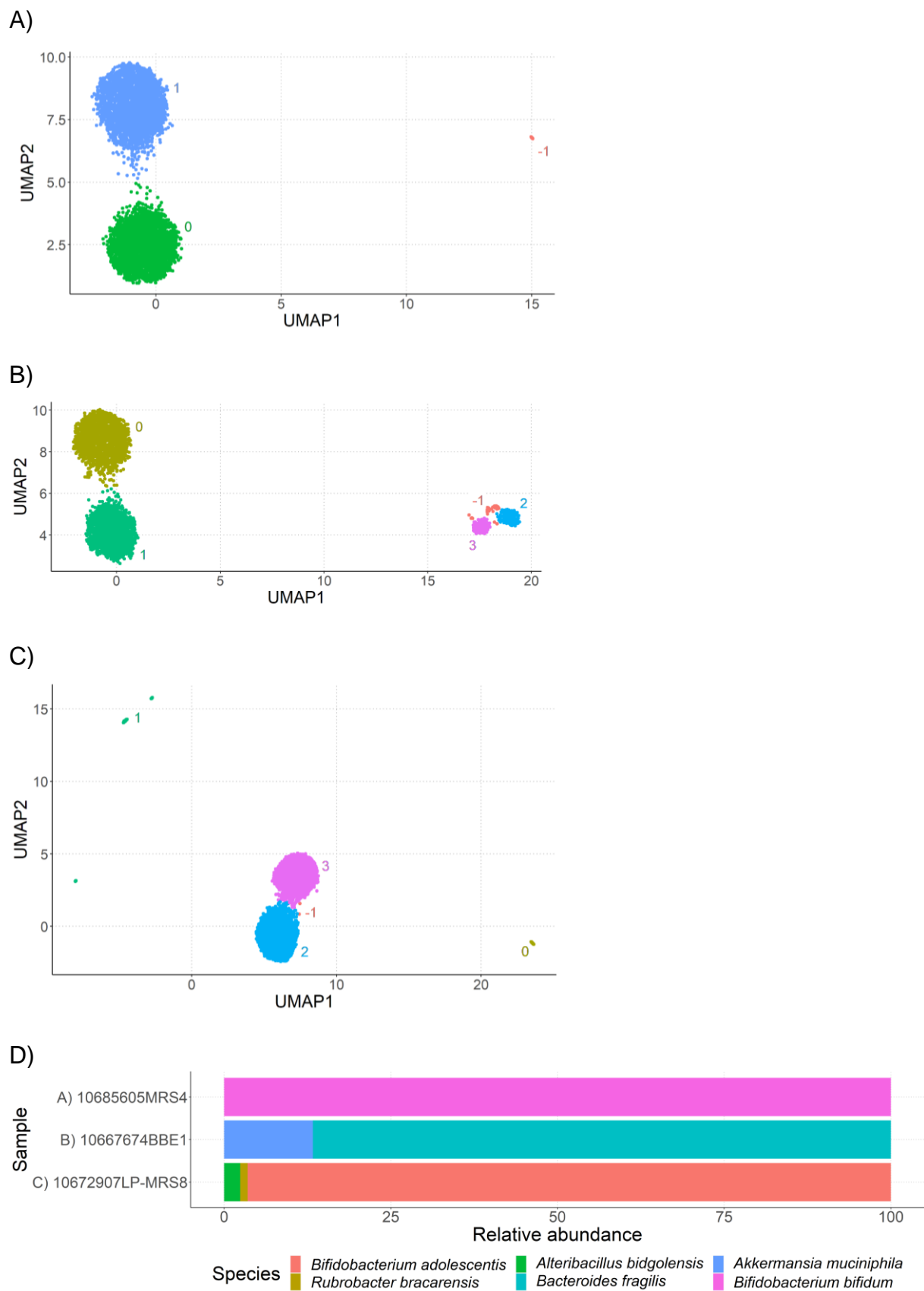


Figure 17 – NanoCLUST plots obtained after sequencing of selected colonies showing clustering of sequenced reads and read percentage. A) Isolate 10685605MRS4,

identified as *B. bifidum* with 100% purity, for which clusters 0 (green) and 1 (blue) were identified as *B. bifidum*. B) Sample 10667674BBE1, identified as *Ba. fragilis* with 86.69% purity, for which clusters 0 (golden) and 1 (green) were identified as *Ba. fragilis* and clusters 2 (blue) and 3 (pink) as *Akkermansia muciniphila*. C) Isolate 10672907LP-MRS8, identified as *B. adolescentis* with 96.48% purity, for which cluster 0 (golden) was identified as *Rubrobacter bracharensis*, cluster 1 (green) as *Alteribacillus bidgolensis* and clusters 2 (blue) and 3 (pink) as *B. adolescentis*. D) Percentage of reads attributed to each species present in each isolate. Clusters named “-1” correspond to unclassified reads that are filtered by NanoCLUST at later stages of the analysis¹³³.

In addition, we randomly selected four isolates (10670603MRS5, 10685605MRS4, 10693850MRS2, and 10693850MRS7) for precise strain identification through WGS by Illumina short read sequencing. Strain identification was performed by calculation of the ANI and the dDDH with closely related type strains, i.e. *L. acidophilus* DSM 20079^T (acc. no. CP020620), *B. bifidum* LMG 11041^T (acc. no. JGYO01000000), *B. dentium* Bd1^T (acc. no. JDUY00000000) and *L. rhamnosus* DSM 20021^T (acc. no. AZCQ01000000). When compared with the type strains, the ANI and dDDH values exceeded the cut-off values for species delineation and indicated the very same identification to NanoCLUST with ONT MinION sequencing (Supplementary Table 2 and 3, and Supplementary Figure 1).

Our aim was to identify bacteria present in complex samples and to select specific species for which there are no exclusively selective media. Both of our goals were achieved successfully and efficiently by combining and adapting several existing technologies. Firstly, this method requires no more than two culturing steps, namely

the culturing of the samples on agar plates and the single-step culturing of the selected resulting colonies in broth medium. Secondly, our DNA extraction method worked for all bacterial taxa and it took less than 25 minutes regardless of the number of samples (limited only by space in the ultrasonicator) and without increasing the handling workload. This is highly advantageous compared to all other methods that we tested, especially for Gram-positive bacteria. Bead beating and heat-based techniques were unsuccessful and kit-based extractions are more expensive and require a higher workload.

Furthermore, 16S rRNA gene sequencing has been shown to provide an accurate bacterial taxonomic classification¹⁴². The 16S rRNA gene is around 1550 bp and it includes both conserved regions, which allow for the design of universal primers for different bacterial taxa, and variable regions that allow for reliable bacterial taxonomic identification¹⁴³. Although ONT MinION sequencing is known to be error-prone, resulting in sequences of approximately 95% accuracy¹⁴⁴, it allows for long-read sequencing. Therefore, 16S rRNA gene sequencing using a long-read sequencing platform can provide an accurate taxonomic classification when compared to short-read platforms such as Illumina because it does not require any assembling steps. Naturally, WGS data can provide better species resolution especially between closely related species, but that would be at the expense of time, resources and costs spent on colonies that are not the focus of our future studies, such as the abundant *Enterococcus* strains that we found. Nonetheless, we performed WGS on four randomly selected isolates to check for the accuracy of the identification and we obtained consistent results between the two techniques. Finally, NanoCLUST has outperformed state-of-the-art software both in terms of identification accuracy and abundance profile estimation at the species-level when classifying mock

communities¹³³. This analysis pipeline is based on an unsupervised read clustering step followed by the generation of a consensus sequence, which is sufficient for accurate species identification through BLAST classification and greatly increases the speed of the pipeline compared to alternatives that perform BLAST on every read. In this way, the bacteria obtained from the colonies could be accurately identified by matching the consensus sequences to the NCBI Refseq database¹³³. Additionally, NanoCLUST provides an insight into the number of reads of each cluster, which allows the user to have an overview of the purity of the isolate.

Alternative methods such as MALDI-TOF require pure colonies, costly equipment such as the MALDI-TOF mass spectrometer, and extensive databases of protein profiles, often developed in-house^{122,145}. In contrast, in this study identification could be achieved with minimal culturing, comparatively low-cost equipment and 16S rRNA gene sequencing by MinION Flongle sequencing, which is superior when compared to conventional Sanger sequencing regarding cost and time. More specifically, our method offers a cost per sample below 5€ from the beginning to the end of the pipeline, and it can take less than 2 weeks to identify 96 samples, whereas with other methods it could take around 4 weeks due to the need for pure colonies, high-workload DNA extraction methods and slower sequencing procedures. The method described here also serves to circumvent the need for sub-culturing by allowing a rapid assessment of whether the cultures are pure for utilisation in further studies. If not, it allows decisions on whether to further purify the bacteria present in the colonies, depending on further studies, which may include taxonomical or genomic investigations, as well as investigations of functional properties. In addition, the method may be used to identify and culture specific target bacteria to be added to culture collections. The method may

also present a valuable addition to future culturomics studies, possibly negating the need for MALDI-TOF MS identification techniques.

3.6. Conclusion

The method here described allows for an accurate bacterial identification as well as an insight into the purity of the isolates. Therefore, it can be useful for having a fast, unexpensive and robust overview of the bacteria extracted from complex microbiological samples before selecting them for further studies.

3.7. Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s00284-023-03201-7>.

3.8. Funding

Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Research and Training Group 2501 on Translational Evolutionary Research (RTG 2501 TransEvo), Collaborative Research Center (CRC) 1182 “Origin and Function of Metaorganisms”, Cluster of Excellence 2167 “Precision Medicine in Chronic Inflammation (PMI)” (Grant number EXC2167 to John F. Baines and Andre Franke) and the DFG Research Unit 5042 “miTarget”.

3.9. Data Availability

The datasets and code generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

3.10. Conflict of interest

The authors declare no conflict of interest.

3.11. Use of the developed method for the overarching project

We aimed to isolate 20 colonies from each of the 40 donors, specifically 10 colonies from MRS medium and 10 from LP-MRS medium, in order to collect sufficient lactobacilli and bifidobacteria isolates for comparative genomics.

In total, we identified 833 colonies with the developed method as shown in Figure 18, 60 of which were obtained from a different culturing strategy specific for *Bacteroides* as previously explained. These 60 colonies were not used for further studies within this doctoral project as the focus was on bifidobacteria and lactobacilli.

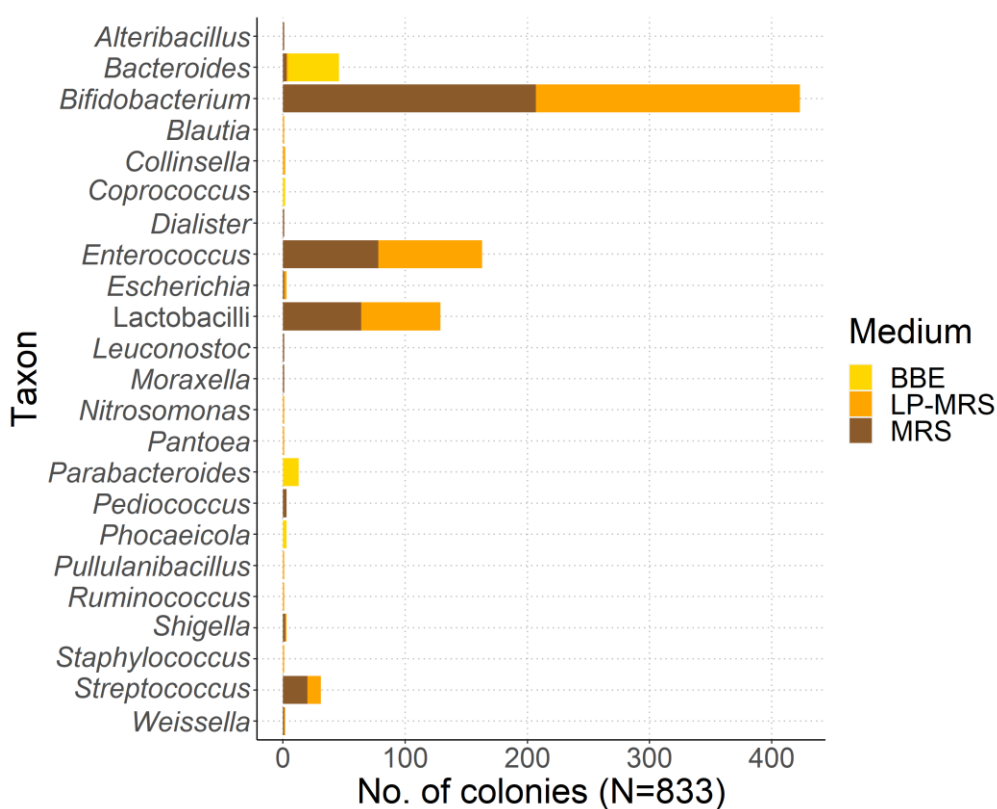


Figure 18 – 16S rRNA gene-based identification of 833 colonies from faecal samples according to culture media.

It can be observed in Figure 18 that the LP-MRS and the MRS media were only partially selective. While both definitely allowed us to obtain a considerable number of desired isolates, LP-MRS did not select for bifidobacteria over lactobacilli as previously demonstrated¹²⁶, and MRS allowed for an appreciable growth of bifidobacteria.

The 833 isolates were also analysed in terms of purity. 674 out of 833 colonies (81%) had 100% of reads attributed to a single species and the remaining consisted of mixtures of up to 19 different species (Figure 19).

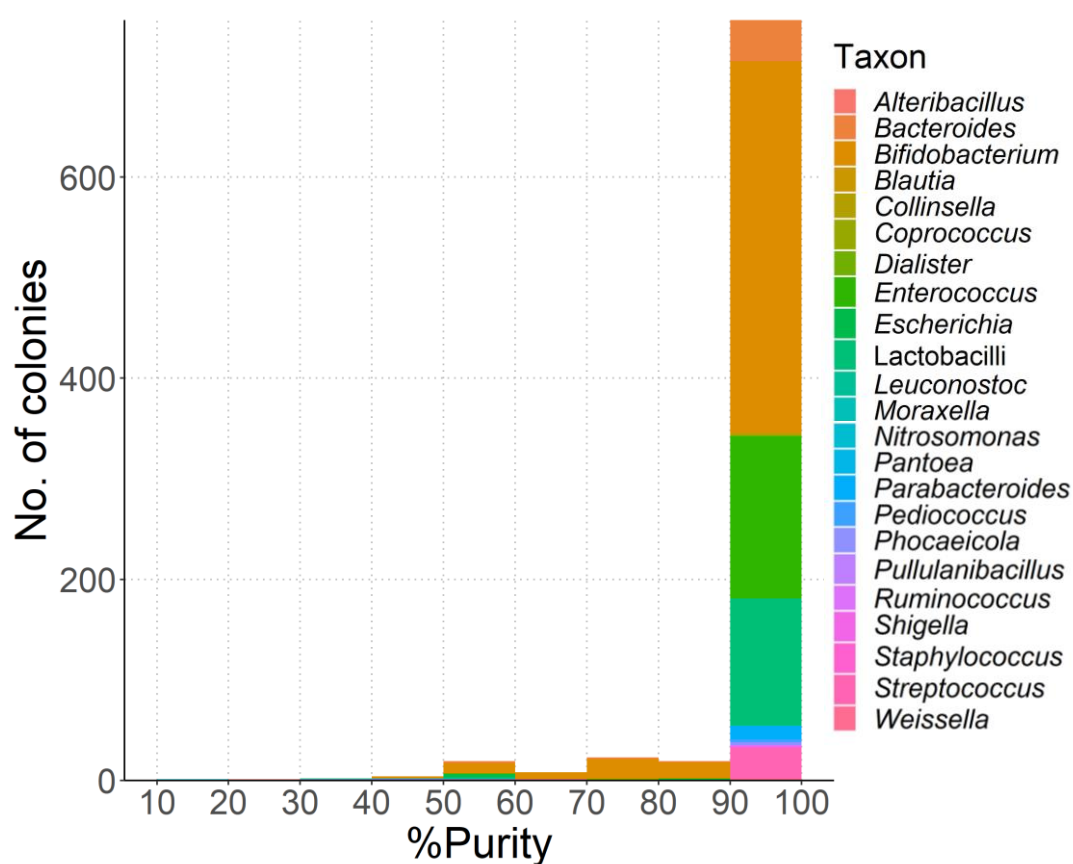


Figure 19 – Purity of the 833 isolates identified with the developed method.

Having been very successful in isolating a considerable number of isolates belonging to the taxa of interest, we used these for our further studies, namely the comparative genomics analyses, which will be described in Chapter 4.

4. Comparative genomic analysis of lactobacilli and bifidobacteria based on secretor status

Ana Sofia G. Borges¹, Erik Brinks¹, Gyu-Sung Cho¹ and Charles M.A.P. Franz^{1*}

Affiliations

¹Department of Microbiology and Biotechnology, Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Hermann-Weigmann-Straße 1, 24103 Kiel, Germany

*Corresponding Author. Contact: charles.franz@mri.bund.de

Manuscript in preparation.

4.1. Introduction

Having identified almost all 800 colonies isolated for this project, the taxa of interest were selected for WGS and used for comparative genomic analyses.

The hypothesis of this project was that by comparing the genomes of bacteria isolated from secretors with those from non-secretors, it would be possible to find evidence of host adaptation dependent on the secretor status. For example, genes and/or functions associated with fucose metabolism could be more abundant in bacteria from secretors than from non-secretors. It could also be that if bacteria of the same species were analysed according to their genomic data, clusters would form based on the secretor status. Either of these possibilities, if true, would indicate that host adaptation according to the secretor status occurs even in the absence of differences in diversity and abundance of gut microbial taxa as seen in the results of the 16S rRNA gene amplicon sequencing data analysis. This comparative genomic analysis as a function of the secretor status has not been previously been performed in lactobacilli and bifidobacteria, and thus represents novel research. Therefore, this study focused on investigating this hypothesis using several different comparative genomics metrics, which will be explained throughout this chapter.

4.2. Materials and methods

4.2.1. Selection of isolates for WGS

Of the 773 colonies included in the project, not all were further characterised by WGS for further studies. First of all, only colonies identified as lactobacilli and bifidobacteria were fully sequenced. Then, the following selection process was devised: if several colonies belonged to the same species, came from the same donor and were cultured in the same medium, only a maximum of two colonies would be fully sequenced. These

two colonies would be randomly selected from those with the highest purity level. If the selected colonies had a purity level below 90%, they were purified in the anaerobic chamber before WGS using the traditional plate streaking method.

4.2.2. Illumina WGS and bacterial identification

WGS was performed according to 3.4.4, with the exception that some isolates were sequenced on an Illumina NextSeq 500 sequencer using the NextSeq 500/550 Mid Output Kit v2.5 (300 Cycles) for paired end sequencing with 2x149 cycles according to the manufacturer's instructions. ANI and dDDH-based bacterial identification was also performed according to 3.4.4, using ANI $\geq 97\%$ and dDDH values $\geq 70\%$ as species-level cut-off values¹⁴⁶.

The remaining 27 colonies of the project were cultured, purified with the streaking method, and then fully sequenced, skipping the 16S rRNA gene sequence-based identification step and therefore the selection process as well. Preliminary bacterial identification of these isolates was conducted using KmerFinder 3.2 (v. 3.0.2.; database version 2022-07-11)^{147–149} and final identification according to ANI and dDDH was performed as described in 3.4.4.

4.2.3. Contamination control

An additional quality control was performed on the samples specifically to test for contamination levels. For this step, the raw data were used to run the Aquamis pipeline¹⁵⁰, where contamination assessment is performed by ConFindr (v 0.7.4)¹⁵¹. In case Aquamis could not produce the final contamination report, its resulting fasta files were used to run the quality tool EvalG from the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) web-based tool^{152,153}, which is based on a re-implementation of the checkM algorithm¹⁵⁴. According to previous literature¹⁵⁵, we

classified our samples as follows: very highly contaminated (>15% contamination), highly contaminated (>10% to ≤15%), moderately contaminated (>5% to ≤10%), slightly contaminated (≤5%), and not contaminated (0%). We decided to use only the not contaminated and slightly contaminated isolates for further studies. Additionally, when the pipelines failed a sample for reasons other than contamination, we also accepted their suggestion and excluded that sample from further analyses. Examples of such reasons include low coverage depth, low/high total length and low completeness.

4.2.4. Genome annotation

The genomes of the isolates in this study were annotated using Prokka with parameters `--Kingdom` and `--gcode 11` (v. 1.14.0)¹⁵⁶ and the BV-BRC web-based tool with default parameters^{152,153}.

4.2.5. Dereplication

During the initial faecal culturing stage, it is relatively common to obtain clonal isolates. To address this issue, the samples were dereplicated in analyses that could be affected by this bias. To do this, we used `dereplicator.py` as described previously¹⁵⁷ to remove assemblies that were too closely related as measured by pairwise distance. We used a minimum pairwise distance of 0.01 for our analysis, as this still retained isolates of the same species if they were sufficiently different, but excluded those that could potentially be clonal isolates.

4.2.6. Comparative genomics and data analysis

Statistical significance tests were performed on abundance tables using the Wilcoxon rank-sum test¹⁰⁹ and p-values were adjusted using the FDR correction¹¹⁰. These calculations were performed using the ggpubr package¹⁰⁸.

Pan-genome calculations were performed using Panaroo¹⁵⁸ and its default parameters. Genes were classified as core genes if present in $\geq 99\%$ of the strains, soft core genes if present in $\geq 95\%$ of the strains but $< 99\%$, shell genes if present in $\geq 15\%$ of the strains but $< 95\%$, and cloud genes if present in $< 15\%$ of the strains. Geneious Prime (v. 2023.2.1)¹⁵⁹ was used to map the *L. rhamnosus* GG *fuc* operon⁵² to our genomes to validate Panaroo's results specifically for this operon.

The web-based tool eggNOG-mapper (v. 2.0.1) was used for additional functional annotation of coding sequences (CDSs)¹⁶⁰, according to the database for COGs¹⁶¹. As in previous studies^{84,162}, relative frequencies of COGs were calculated by dividing the counts of each category for each genome per secretor status by the total number of clustered proteins of each genome. When a gene was assigned to two categories, these were counted separately. Comparisons between groups were performed using the Wilcoxon rank-sum test¹⁰⁹, and p-values were adjusted using the FDR correction¹¹⁰. These calculations were performed using the ggpubr package¹⁰⁸.

The vegan¹⁶³ package was used to calculate dissimilarity matrices for the PCoAs. PERMANOVA¹¹² and permutational analysis of multivariate dispersions (PERMDISP)¹⁶⁴ were used to test the significance of the PCoAs for metadata variables without missing data. Heatmaps were plotted using the ComplexHeatmap package¹⁶⁵. RStudio (R 4.2.0)¹⁰⁶ and the ggplot2 package¹¹¹ were used for overall data analysis and visualization.

4.3. Results and discussion

4.3.1. Abundance of lactobacilli and bifidobacteria isolates according to secretor status

Table 2 and Table 3 show the number of lactobacilli and bifidobacteria, respectively, that were sequenced from the 800 isolated colonies.

Table 2 – ANI and dDDH based identification of the sequenced lactobacilli isolates.

Species identification	Secretor	Non-secretor	Total
<i>Lacticaseibacillus casei</i>	5	10	15
<i>Lacticaseibacillus rhamnosus</i>	1	8	9
<i>Lactiplantibacillus paraplantarum</i>	0	3	3
<i>Lactiplantibacillus plantarum</i> subsp. <i>plantarum</i>	2	11	13
<i>Lactobacillus acidophilus</i>	0	3	3
<i>Lactobacillus gasseri</i>	0	2	2
<i>Lactobacillus helveticus</i>	0	1	1
<i>Lactobacillus paragasseri</i>	1	1	2
<i>Latilactobacillus curvatus</i>	1	3	4
<i>Latilactobacillus sakei</i> subsp. <i>carnosus</i>	15	8	23
<i>Latilactobacillus sakei</i> subsp. <i>sakei</i>	0	3	3
<i>Ligilactobacillus ruminis</i>	4	0	4
<i>Levilactobacillus brevis</i>	1	7	8
<i>Limosilactobacillus mucosae</i>	14	2	16
<i>Limosilactobacillus oris</i>	0	2	2
<i>Limosilactobacillus vaginalis</i>	0	2	2

<i>Loigolactobacillus coryniformis</i> subsp. <i>torquens</i>	0	1	1
Total	44	67	111

Table 3 – ANI and dDDH based identification of the sequenced bifidobacteria isolates.

Species identification	Secretor	Non-secretor	Total
<i>Bifidobacterium adolescentis</i>	17	26	43
<i>Bifidobacterium angulatum</i>	8	0	8
<i>Bifidobacterium animalis</i> subsp. <i>lactis</i>	4	0	4
<i>Bifidobacterium bifidum</i>	15	8	23
<i>Bifidobacterium dentium</i>	4	0	4
<i>Bifidobacterium longum</i> subsp. <i>longum</i>	61	38	99
Total	109	72	181

As explained in 4.2, a quality control step specific for contamination detection was applied after species identification. This reduced the number of isolates available for future studies, as shown in Table 4 and Table 5. While for bifidobacteria a considerable number of isolates remained after excluding potentially contaminated samples, for lactobacilli only 56% of the purified isolates showed no contamination or slight contamination ($\leq 5\%$). This was a substantial reduction in sample size, which could be attributed either to batch contamination, as many of these isolates were sequenced on the same batch, or to misidentification of certain contigs due to closely related species, as most contaminations were due to mixtures of different lactobacilli species. Another contributing factor was our conservative approach in excluding of outliers in terms of coverage depth, total length and completeness. For future studies beyond this project, the exact cause of this sample loss should be investigated more closely, starting by

exploring the possibility of salvaging some of these outliers and presumed lactobacilli mixtures, for example by experimenting with different assemblers and parameters. If this does not improve the sample size, it is also possible to re-run the sequencing batch and purify these colonies again. However, due to time constraints, in this study the comparative genomic analyses were carried out using the remaining samples.

Table 4 – Lactobacilli isolates with very low or no contamination used for further analyses.

Species identification	Secretor	Non-secretor	Total
<i>Lacticaseibacillus casei</i>	3	0	3
<i>Lacticaseibacillus rhamnosus</i>	0	2	2
<i>Lactiplantibacillus plantarum</i> subsp. <i>plantarum</i>	2	10	12
<i>Lactobacillus acidophilus</i>	0	3	3
<i>Lactobacillus helveticus</i>	0	1	1
<i>Latilactobacillus sakei</i> subsp. <i>carnosus</i>	12	8	20
<i>Latilactobacillus sakei</i> subsp. <i>sakei</i>	0	3	3
<i>Ligilactobacillus ruminis</i>	2	0	2
<i>Limosilactobacillus mucosae</i>	14	2	16
Total	33	29	62

Table 5 – Bifidobacteria isolates with very low or no contamination used for further analyses.

Species identification	Secretor	Non-secretor	Total
<i>Bifidobacterium adolescentis</i>	14	14	28
<i>Bifidobacterium angulatum</i>	8	0	8
<i>Bifidobacterium animalis</i> subsp. <i>lactis</i>	3	0	3
<i>Bifidobacterium bifidum</i>	15	8	23
<i>Bifidobacterium dentium</i>	2	0	2
<i>Bifidobacterium longum</i> subsp. <i>longum</i>	59	33	92
Total	101	55	156

The lactobacilli isolated in this study, as shown in Table 4, are mostly those typically found in the human gut microbiome^{48,49}. Although most gut lactobacilli are considered to be allochthonous, we also isolated some autochthonous isolates, e.g. *L. rhamnosus*⁴⁹. Some lactobacilli that are likely to be of food origin were also isolated. *L. helveticus* is commonly used in large numbers as a starter culture for the production of cheese, *L. plantarum* subsp. *plantarum* is a part of the natural microbiota of certain fermented plant materials (e.g. sauerkraut) and the two *L. sakei* subspecies are used as starter cultures for the production of fermented raw meat products (e.g. salami), which are all common German foods⁴⁷.

Table 4 shows no relevant differences in terms of total isolates found in secretors (n=33) and non-secretors (n=29), and a Wilcoxon rank-sum test with FDR correction confirmed this with a non-significant p-value of 0.94. It was unfortunately not possible to test for significant differences between secretors and non-secretors for specific species due to their low numbers and concentrated presence in certain donors.

However, despite these factors and the strong sampling bias (we only isolated 800 colonies when there are $0.9\text{-}10^{11}$ bacteria per gram of wet stool¹⁶⁶), some trends could be observed. On the one hand, both *L. casei* and *L. ruminis* were found in low numbers in secretors, but not in non-secretors. On the other hand, *L. rhamnosus*, *L. acidophilus*, *L. helveticus* and *L. sakei* subsp. *sakei* were found in low numbers in non-secretors, but not isolated from secretors. It is interesting to note this trend in *L. rhamnosus*, as some of its strains have the ability to use fucose⁵², but it may be that these isolates from non-secretors are not those particular strains. *L. mucosae* was isolated more frequently from secretors (n=14) than non-secretors (n=2), which is an interesting result given that this species can ferment fucose, although the mechanisms have not yet been clearly identified⁵⁵. *L. sakei* subsp. *carnosus* was also found more often in secretors (n=12) than in non-secretors (n=8). Finally, *L. plantarum* subsp. *plantarum* was isolated more frequently from non-secretors (n=10) than from secretors (n=2).

As shown in Table 5, all the bifidobacteria isolated in this study are among those typically isolated from human faecal samples⁷⁰. We observed a striking difference between the number of *Bifidobacterium* isolates found in secretors and non-secretors. Although there is a sampling bias, it is interesting to note that we obtained twice as many *Bifidobacterium* isolates from secretors (n=101) compared to non-secretors (n=55), which is a significant difference ($p=0.015$). For some species, namely *B. angulatum*, *B. animalis* subsp. *lactis* and *B. dentium*, there were in fact no isolates from non-secretors, while they were found in low numbers in secretors. Regarding the remaining species obtained, there were no differences in the number of *B. adolescentis* isolates found in secretors and non-secretors, but there were higher numbers of *B. bifidum* and *B. longum* subsp. *longum* in secretors than in non-secretors, which are two species that have fucosidases and/or fucose degradation pathways. Specifically, as previously explained, *B. bifidum* has been shown to be more abundant in secretors

than in non-secretors⁹² and to have one extracellular $\alpha(1,2)$ -fucosidase, while certain *B. longum* subsp. *longum* strains have been shown to have a gene cluster dedicated to the import of fucosylated molecules, fucose metabolism and with two intracellular α -fucosidases⁸². *B. longum* subsp. *longum* was also the only species in this study that was almost ubiquitous in our donors, allowing us to test for significant differences between secretors and non-secretors despite the sampling bias. However, the p-value was 0.05 and therefore not significant.

After contamination detection, comparative genomic analyses were performed as a function of secretor status. Figure 20 shows an overview of the isolate selection process of each step of the project.

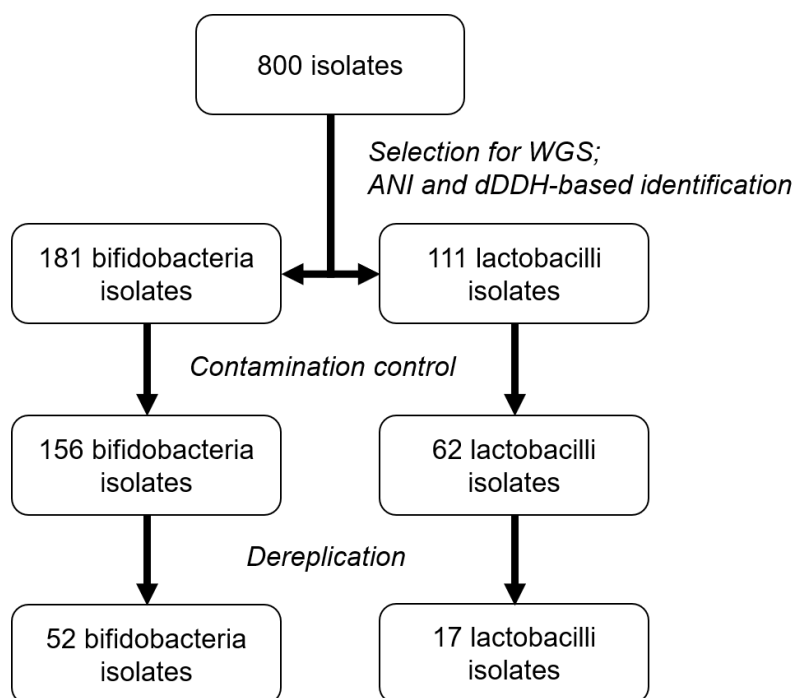


Figure 20 – Overview of the isolates remaining after each methodological step.

In each subsection it will be indicated whether all isolates or dereplicated data were used for each analysis. Table 6 and Table 7 show the number of isolates remaining

after dereplication of lactobacilli and bifidobacteria, respectively, as a function of the secretor status of the donor.

Table 6 – Dereplicated lactobacilli isolates with very low or no contamination.

Species identification	Secretor	Non-secretor	Total
<i>Lacticaseibacillus casei</i>	2	0	2
<i>Lacticaseibacillus rhamnosus</i>	0	1	1
<i>Lactiplantibacillus plantarum</i> subsp. <i>plantarum</i>	0	1	1
<i>Lactobacillus acidophilus</i>	0	1	1
<i>Lactobacillus helveticus</i>	0	1	1
<i>Latilactobacillus sakei</i> subsp. <i>carnosus</i>	2	3	5
<i>Latilactobacillus sakei</i> subsp. <i>sakei</i>	0	2	2
<i>Ligilactobacillus ruminis</i>	2	0	2
<i>Limosilactobacillus mucosae</i>	1	1	2
Total	7	10	17

Table 7 – Dereplicated bifidobacteria isolates with very low or no contamination.

Species identification	Secretor	Non-secretor	Total
<i>Bifidobacterium adolescentis</i>	8	7	15
<i>Bifidobacterium angulatum</i>	2	0	2
<i>Bifidobacterium animalis</i> subsp. <i>lactis</i>	1	0	1
<i>Bifidobacterium bifidum</i>	1	0	1
<i>Bifidobacterium dentium</i>	2	0	2
<i>Bifidobacterium longum</i> subsp. <i>longum</i>	18	13	31
Total	32	20	52

4.3.2. Comparative genomics of lactobacilli

In this subsection the comparative genomic analyses performed on lactobacilli as a group will be discussed. This allows comparisons with the literature, as they were previously considered to be a single genus, but it must be remembered that this is an analysis of a taxon with family-like diversity. Naturally, in the future such analyses can be carried out for each of the new genera, especially if a sufficient number of isolates is available for each of them.

Both before and after dereplication, the collection of lactobacilli of this study did not show any core genes, which highlights the diversity of this group (Figure 21A and B). However, this could also be partially caused by Panaroo's default parameters, which are considered quite conservative¹⁵⁸. Future analyses using different parameters may yield some core genes. It could also be observed that genes previously classified as shell genes were reclassified as cloud genes by dereplication. Dereplication resulted in the loss of similar strains and therefore certain genes were less shared across the remaining isolates.

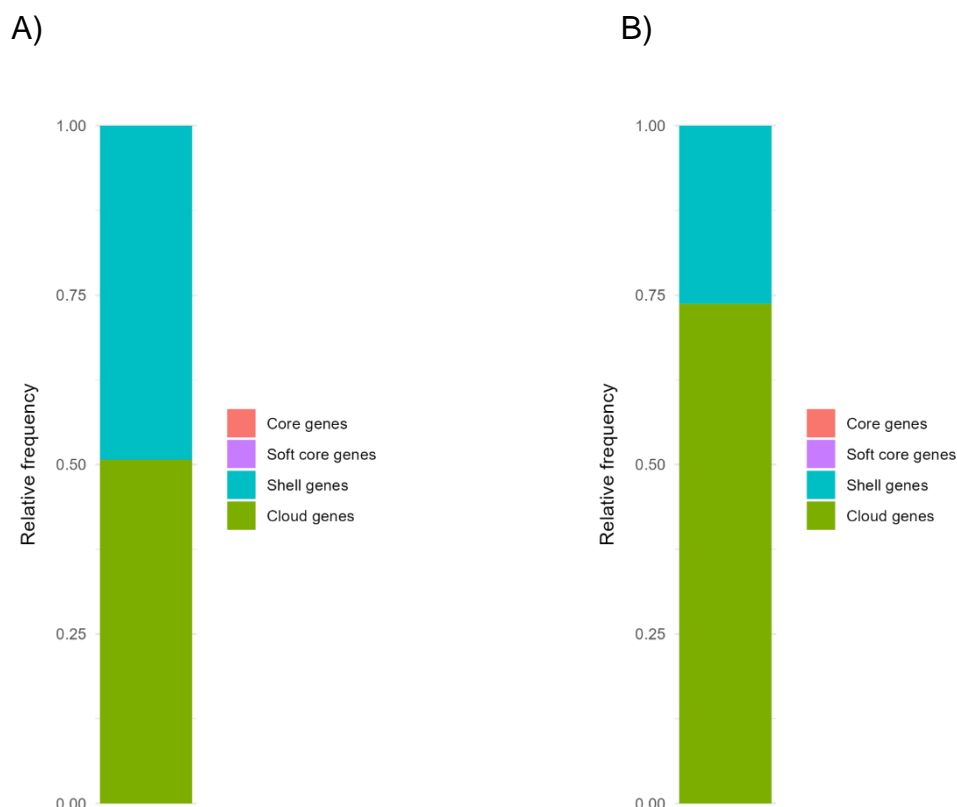


Figure 21 – Summary pan-genome statistics of the Panaroo analysis of lactobacilli isolates. A) Results for all the lactobacilli strains (n=62): 0 core genes, 0 soft core genes, 6,530 shell genes, and 6,718 cloud genes out of 13,248 total genes. B) Results for the dereplicated lactobacilli isolates (n=17): 0 core genes, 0 soft core genes, 2,766 shell genes, and 7,758 cloud genes out of 10,524 total genes.

In the study by Inglin *et al.* (2018), the pan-genome of 98 fully sequenced genomes and 234 genomes of lactobacilli from five different species (*L. reuteri*, *L. delbrueckii*, *L. plantarum*, *L. rhamnosus* and *L. helveticus*) contained 20,800 genes, of which 266 were core genes⁶³. These results are very different from ours, as we did not obtain any core genes for our collection of lactobacilli, either in the analysis with all isolates or in the analysis with the dereplicated isolates. This may be due to the fact that 6 different new genera from 9 different species/subspecies were included in both analyses, resulting in higher taxonomic diversity and fewer shared genes. However, Inglin *et al.*

(2018) found a much larger pan-genome of 20,800 genes⁶³, whereas in this study 13,248 total genes were found for the analysis with all isolates and 10,524 total genes in the dereplicated data. This could be due to the fact that our study used 62 and 17 genomes for the complete and dereplicated analyses, respectively, whereas Inglin *et al.* (2018) used a much larger number of genomes for their analysis, thus finding more genes in their data⁶³.

In the study by Sun *et al.* (2014), 73 core genes were identified out of 44,668 gene families contained in 213 lactobacilli genomes from a variety of different species⁶⁴. This study highlights the diversity of this taxon by the very low percentage of core genes out of the total number of genes, which were not even found in our study. The high number of gene families found compared to our study can be attributed to the much higher number of genomes included. Overall, the study by Sun *et al.* (2014) and our own study highlight the diversity of this group and the interest in performing pan-genome analyses with our data in the future at the new genera level or at the species level⁶⁴.

Using the dereplicated data, we then annotated gene clusters according to the COG categories, whose distribution is shown in Figure 22. The most abundant category is S (function unknown), followed by categories K (transcription), J (translation), G (carbohydrate metabolism and transport), E (amino acid metabolism and transport) and L (replication, recombination and repair). This distribution is similar to that previously described in the literature for this taxon^{167–172}. By applying Wilcoxon rank-sum tests with FDR correction, we verified that there are no significant differences in all functional categories depending on the secretor status.

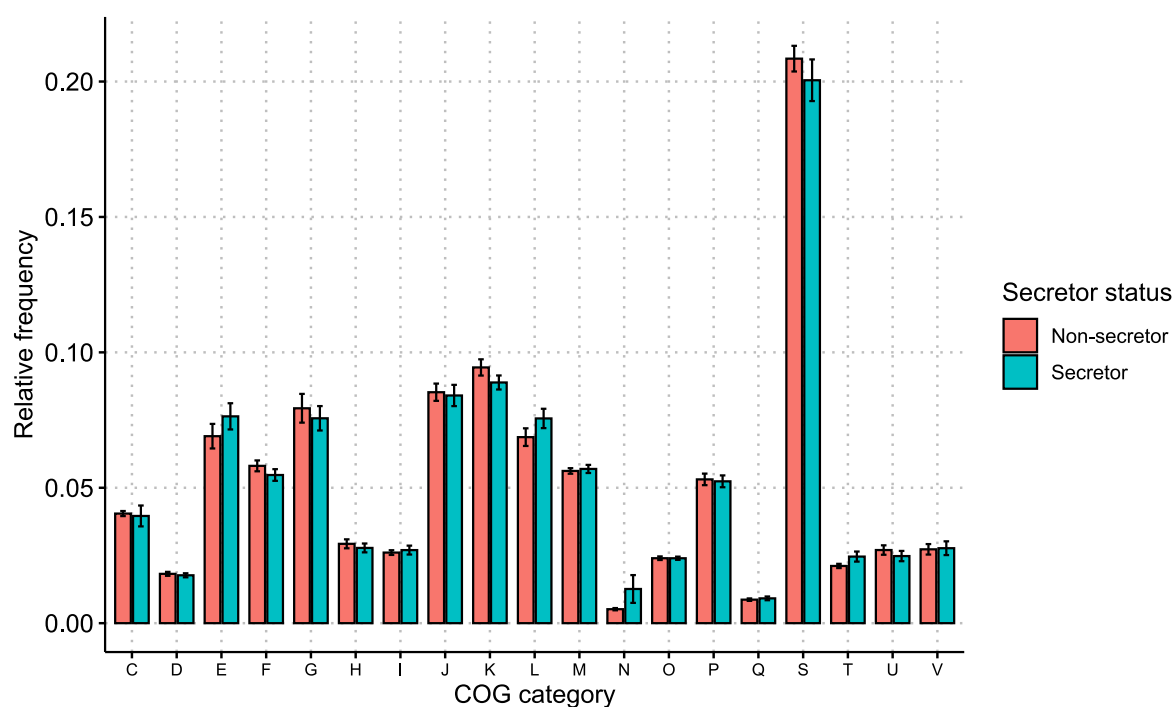


Figure 22 – Distribution of functional categories in the 17 dereplicated lactobacilli genomes from donors with different secretor status. The COG categories are abbreviated with capital letters, namely: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control and mitosis; E, amino acid metabolism and transport; F, nucleotide metabolism and transport; G, carbohydrate metabolism and transport; H, coenzyme metabolism; I, lipid metabolism; J, translation; K, transcription; L, replication, recombination and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, post-translational modification, protein turnover, chaperone functions; P, inorganic ion transport and metabolism; Q, secondary structure; R, general functional prediction only; S, function unknown; T, signal transduction; U, intracellular trafficking and secretion; V, defence mechanisms; W, extracellular structures; Z, cytoskeleton. Only categories present in our data are shown.

We then focused on finding fucose-related genes in all of the lactobacilli strains. When Panaroo's output was filtered accordingly, there were few fucose-related genes present in the lactobacilli isolates. Various strategies were then used to confirm the completeness of Panaroo's results for these particular genes. First, it was checked that all previously described fucose-related genes are larger than the 500 bp cutoff used by the BBDuk pipeline in 3.4.4¹³⁶. Secondly, all synonyms for the gene names were investigated, but added no value. Finally, the well-established *fuc* operon from *L. rhamnosus* GG was mapped to the genomes of lactobacilli isolates of this study, but there were no new matches for any isolate, regardless of using raw data or assembled data. Therefore, the analyses were continued using Panaroo's original output. However, it is important to bear in mind that research on fucose-related pathways in lactobacilli is still in its infancy. Therefore, some metabolism pathways may not have been identified yet, while certain genes have been predicted but not properly identified and characterised.

Figure 23 shows that only the *fucP* gene could be identified in 16 out of 62 lactobacilli genomes. All *fucP*-positive isolates correspond to *L. mucosae*, which has previously been shown to be able to utilise fucose, but not large fucosylated molecules⁵⁵. However, the fucose catabolic mechanisms of this bacterial species have not yet been identified.

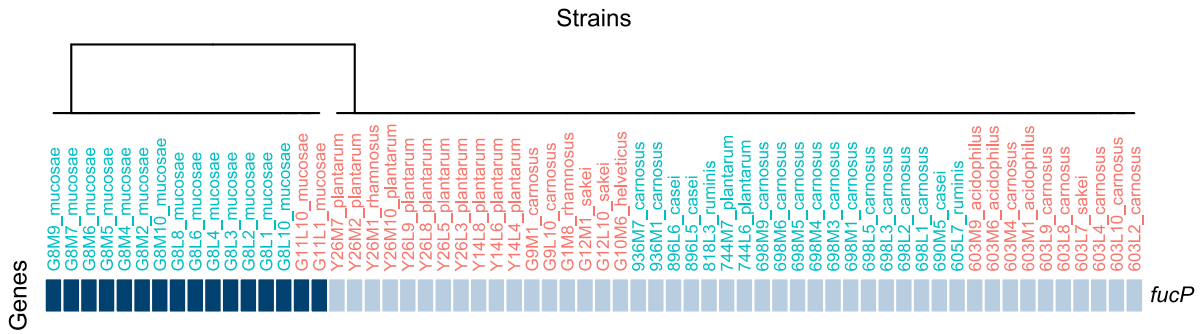


Figure 23 – Presence/absence heatmap of fucose-related genes in 62 lactobacilli genomes. The species name was added to the isolate name to facilitate data interpretation. Clustering was performed according to fucose-related genes only. In the heatmap cells, present genes are in dark blue and absent genes are in light blue. Regarding strain names, those from secretors are in blue and those isolated from non-secretors are in salmon red.

The *fucP* gene is an L-fucose proton symporter (also called L-fucose permease) that can also transport other sugars such as L-galactose and D-arabinose, but at a lower rate^{173,174}. This gene may be present to transport fucose that is later degraded by yet unidentified mechanisms, or it may be present to transport those other sugars. However, regardless of its function in *L. mucosae*, at least it can be concluded that its presence does not appear to be dependent on the secretor status of the host, although the number of *L. mucosae* isolates is not well balanced between secretors (n=14) and non-secretors (n=2). We also found no differences according to secretor status in any of the remaining isolates from different lactobacilli species.

4.3.3. Comparative genomics of bifidobacteria

After running Panaroo on bifidobacteria, we found that there were only 43 core genes in our collection of isolates, out of a total of 12,766 gene clusters (Figure 24A). After

dereplicating from 156 to 52 isolates, these values did not change much, with 44 core genes out of 10,348 total genes (Figure 24B). Most of the genes lost by dereplication were former cloud genes, which could indicate loss of unique genes. Interestingly, in this case the proportion of shell genes increased with dereplication. This may be due to the fact that dereplication reduced the number of *B. bifidum* isolates from 23 to a single isolate, and thus genes that are present in the remaining species but not in *B. bifidum* can be classified as shell genes rather than cloud genes in the analysis of the dereplicated data, as *B. bifidum* is a minor percentage of the total number of dereplicated strains analysed.

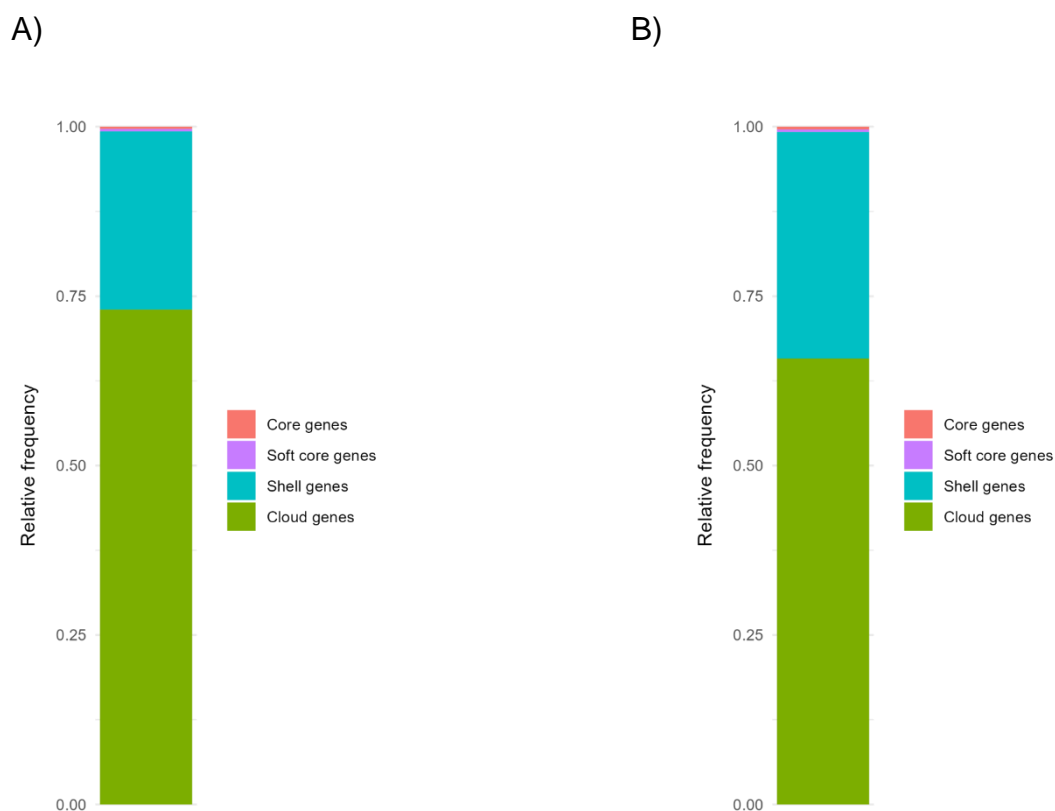


Figure 24 – Summary pan-genome statistics of the Panaroo analysis of Bifidobacterium isolates. A) Results for all the bifidobacteria isolates (n=156): 43 core genes, 46 soft core genes, 3,356 shell genes, and 9,321 cloud genes out of a total of 12,766 genes. B) Results for the dereplicated isolates (n=52): 44 core genes, 38 soft core genes, 3,458 shell genes, and 6,808 cloud genes out of 10,348 total genes.

There have been several studies on the pan-genome of bifidobacteria. A 2012 study investigated 19 bifidobacteria strains from 9 different species of human origin and found 724 core genes out of 6,980 genes¹⁷⁰. Another study based on 45 bifidobacteria type strains from different origins (e.g. human, animal, dairy, sewage) found 402 core genes out of more than 20,000 total genes¹⁷⁵. Sharma *et al.* (2018) reported 613 core genes out of 30,293 total genes when examining 215 bifidobacteria strains from 44 different species and different hosts, including non-human origin¹⁷⁶. In the study by Rodriguez *et al.* (2020), 129 bifidobacteria strains from several different species and also from different hosts had 438 core genes, 115 soft core genes, 1,802 shell genes, and 24,550 cloud genes out of a total of 26,905 genes⁷³.

This genus has an open pan-genome, so the total number of genes does not stabilise as more genomes are added⁹⁸. Given that 156 bifidobacteria genomes were analysed and then dereplicated into 52 representative genomes, the total of 12,766 and 10,348 gene clusters, respectively, is a relatively small pan-genome size compared to the literature. However, most studies investigated bifidobacteria from multiple origins, and it is known that bifidobacteria have co-evolved with their hosts⁷³, thus adding different origins will most likely result in a larger pan-genome. The study by Lukjancenko *et al.* (2012) has reported on strains of human-origin only, but included 9 species and some were from oral microbiome samples, whereas we only examined 6 species of stool origin¹⁷⁰.

Regarding the core genome, bifidobacteria usually have a core genome of 400-500 genes, as reported above. However, the results of this study are an order of magnitude lower than those reported in the literature. This is most likely due to Panaroo's default parameters, which are considered conservative¹⁵⁸. For example, Panaroo's default protein family sequence identity threshold is 70%, whereas all the studies summarised

above used an identity threshold of 50%^{73,98,158,170,175,176}. For more accurate comparisons with the current literature, it may be interesting to repeat the analyses using more sensitive parameters in future research.

Subsequently, gene clusters of the dereplicated data were annotated according to the COG categories. The resulting COG distribution can be observed in Figure 25. The category with the highest relative frequency was S (function unknown), followed by G (carbohydrate metabolism and transport), E (amino acid metabolism and transport), K (transcription), J (translation) and L (replication, recombination and repair). This distribution was very similar to previously published results^{98,170–172,176–178} and there was a strong focus on carbohydrate metabolism as expected⁹⁷. Furthermore, there were no significant differences across all functional categories depending on secretor status, as assessed by Wilcoxon rank-sum tests with FDR correction.

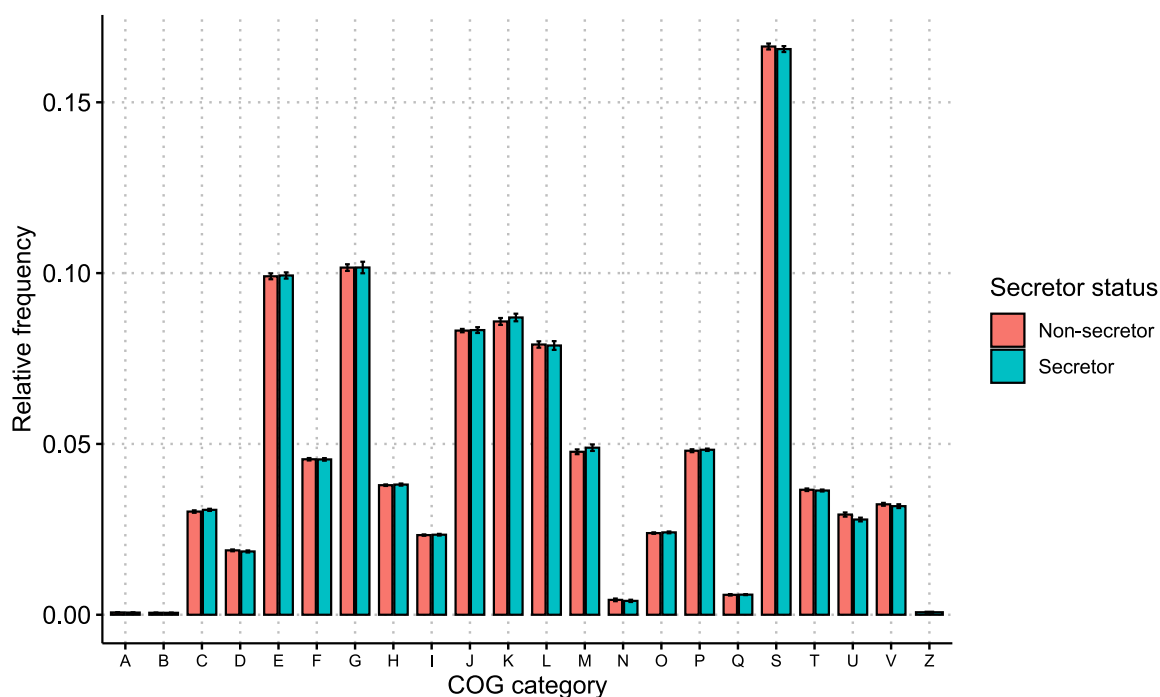


Figure 25 – Distribution of functional categories in 52 dereplicated bifidobacteria genomes from donors with different secretor status. COG categories are abbreviated

with capital letters as explained in Figure 22. Only categories present in our data are shown.

The presence/absence of fucose-related genes in all the bifidobacteria isolates was the next focus of this investigation. The same troubleshooting methods as described in 4.3.2 were applied to find as many fucose-related genes as possible. Ideally, it would have been possible to also map a bifidobacteria *fuc* operon to the bifidobacteria isolates of this study, but as this was not readily accessible, we used the *L. rhamnosus* GG *fuc* operon. As this did not yield any new matches, the analysis was continued using the original Panaroo output, as shown in Figure 26.

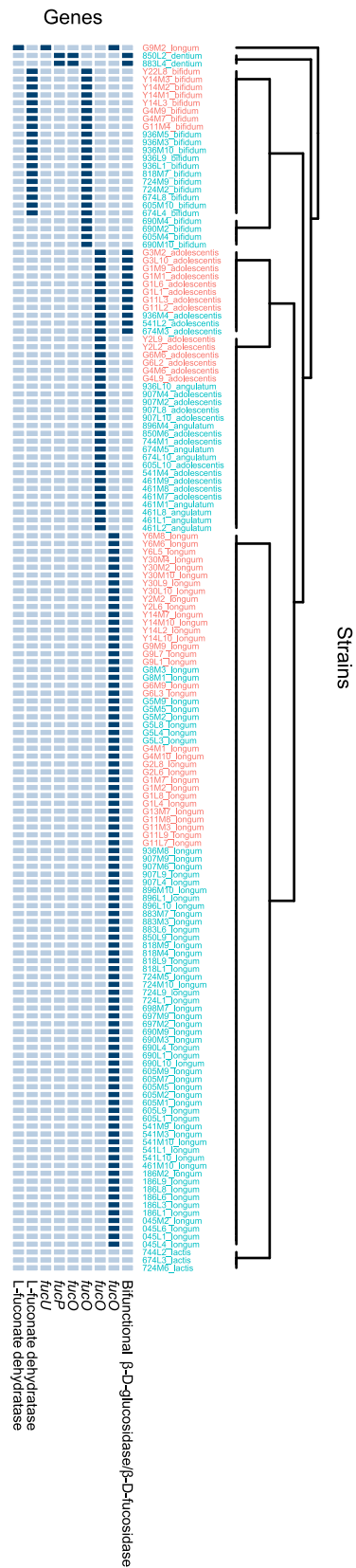


Figure 26 – Presence/absence heatmap of fucose-related genes in 156 bifidobacteria genomes. The species name was added to the isolate name to facilitate data

interpretation. Clustering was performed according to fucose-related genes only. In the heatmap cells, present genes are in dark blue and absent genes in light blue. Regarding strain names, those from secretors are in blue and those isolated from non-secretors are in salmon red.

It could be observed that *fucO* was present in the genome of most of the 156 isolates regardless of secretor status. This gene encodes a lactaldehyde reductase, which is part of both the L-rhamnose and L-fucose catabolic pathways¹⁷⁹. This gene was only not found in the three *B. animalis* subsp. *lactis* isolates, which all stemmed from secretors. In all other isolates, *fucO* could be ubiquitously identified, but as four different genetic variants depending on the species: one for *B. longum* subsp. *longum*, one for both *B. adolescentis* and *B. angulatum*, one for *B. bifidum* and one for *B. dentium*. It is interesting that *B. adolescentis* and *B. angulatum* exhibited the same variant, because in phylogenies based on the core genome, *B. angulatum* is actually closer to the *B. longum* species than to *B. adolescentis*¹⁸⁰.

It could also be observed that there was a cluster of *B. adolescentis* that had both the *fucO* gene and a bifunctional β -D-glucosidase/ β -D-fucosidase. There are more isolates from non-secretors (n=8) than from secretors (n=3) in this *B. adolescentis* cluster, and in the cluster with only the *fucO* gene there are more strains from secretors (n=11) than from non-secretors (n=6). However, as this bifunctional glucosidase/fucosidase is not active on L-fucose residues¹⁸¹, these small differences may not be related to secretor status.

Regarding *B. bifidum*, most of the strains considered possessed the *fucO* gene and an L-fuconate dehydratase (possibly the *fucE* gene⁸⁹), regardless of secretor status. According to the current literature, *B. bifidum* is unable to degrade fucose⁸³. However,

L-fuconate dehydratase was almost ubiquitously present in the strains of this study and has been reported to also be present in other *B. bifidum* strains¹⁸². It could be that it is either used on a different substrate (e.g. this enzyme can also act on D-arabinonate, but at a slower rate¹⁸³), or that it is the result of HGT from other gut bacteria and has remained in the genome of certain strains of this species over time, as has been reported for other *Bifidobacterium* species⁸². In our bacterial collection there are also four *B. bifidum* strains isolated from secretors that have only the *fucO* gene.

The *B. dentium* isolates in this study have the *fucP* and *fucO* genes, and a bifunctional β -D-glucosidase/ β -D-fucosidase. All strains were isolated from secretors, which precludes a comparison between secretors and non-secretors. Nevertheless, this species has been shown to be unable to grow in fucose but to have some ability to degrade salivary mucins, demonstrating its adaptation to the oral cavity^{184,185}. It has also been shown to possess a putative fucosidase. Therefore, the genes found in this study may be the result of the ability to degrade fucose-containing salivary mucins to some extent and then excrete them¹⁸⁶. The *fucP* gene, an L-fucose proton symporter, can also be used to transport sugars other than fucose as mentioned above^{173,174}.

Finally, there was one *B. longum* subsp. *longum* strain (G9M2) that was isolated from a non-secretor that had the *fucO* and *fucU* genes and an L-fuconate dehydratase (possibly the *fucE* gene⁸⁹). These genes may be part of a strain-specific fucose catabolic pathway, as some *B. longum* subsp. *longum* strains can indeed degrade fucose^{82,84}. However, the fucose catabolic pathway was not fully present, so it is possible that either there are other genes responsible for the missing pathway steps that have not yet been identified, or that these genes are actually the result of HGT as previously hypothesized⁸².

Overall, the implication of these results for this study is that the presence and/or absence of fucose-specific genes does not seem to be dependent on secretor status. However, for those clusters in which some of these genes are present regardless of secretor status, it would be interesting to investigate at the transcriptomic level whether their expression is dependent on secretor status. For example, it could be that the *fuc* operon of bifidobacteria is inducible, similar to what has previously been seen for the N-acetylgalactosamine (GalNAc) operon of *E. coli*^{187,188}. Additionally, it would be interesting to obtain isolates of *B. dentium*, *B. angulatum* and *B. animalis* subsp. *lactis* from non-secretors to compare with our isolates from secretors.

4.3.4. Comparative genomics of *Bifidobacterium longum* subsp. *longum*

The focus of this study then shifted to analysing isolates at the species level for more accurate comparative genomics. *B. longum* subsp. *longum* was the only (sub)species for which a high number of isolates was obtained from both secretors (n=59) and non-secretors (n=33) and from several different donors. As it has also been reported to include strains capable of utilising fucose, this subspecies was analysed first. In the future, it would be interesting to look into other species and/or taxonomic groups.

First, the distribution of core and accessory genes were investigated. As the absolute and relative frequencies were almost identical between the analysis with all isolates (n=92) and the one with the dereplicated data (n=31), only the latter will be described. Figure 27 shows that 29.12% of the genes belong to the core genome, 1.06% are soft core genes, 22.31% are shell genes and 47.51% are cloud genes.

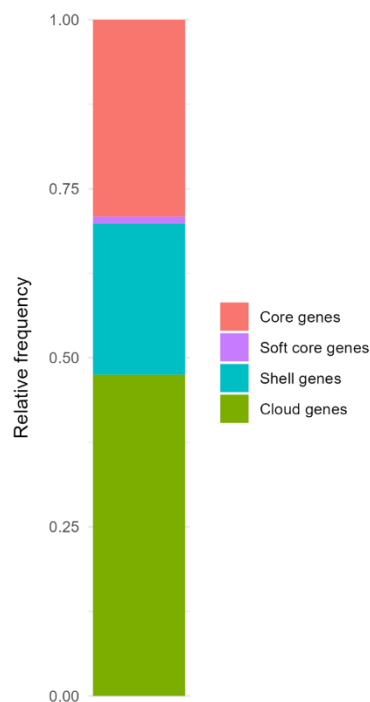


Figure 27 – Summary pan-genome statistics of the Panaroo analysis of the 31 dereplicated *B. longum* subsp. *longum* isolates (n=31): 1,373 core genes, 50 soft core genes, 1,052 shell genes, and 2,240 cloud genes out of a total of 4,715 genes.

In the study by Díaz *et al.* (2021), 115 *B. longum* subsp. *longum* strains from humans of different ages, probiotic and bovine origin were shown to have 999 core genes (11.52%), 187 soft core genes (2.16%), 1,238 shell genes (14.28%) and 6,246 cloud genes (72.04%) out of 8,670 total genes⁸⁴. Compared to this study with 31 dereplicated strains representing 92 isolates, the core genome was smaller, probably due to the inclusion of more strains and from different sources, whereas this study focused on a cohort of German human adults. The total pan-genome is larger, probably also due to the variety of sources and the higher number of strains included⁸⁴.

Albert *et al.* (2019) found that 43 *B. longum* subsp. *longum* strains from a cohort of human adults had 761 core genes (6.04%), 376 soft core genes (2.98%), 1,194 shell genes (9.47%) and 10,278 cloud genes (81.51%) out of a total of 12,609 genes¹⁸⁹. The

strains of that study seem to have much more variability than the isolates from this study and those of Díaz *et al.* (2021)⁸⁴, as the number of isolates is not too different from the number of dereplicated strains of our study. However, in the study by Albert *et al.* (2019), eight strains were isolated from vaginal swab samples and were outliers in the analyses performed¹⁸⁹, despite a previous study stating that bifidobacteria from intestinal or vaginal sources do not differ significantly¹⁷⁷. This may have introduced more variability into their pan-genomic analysis, as samples from the same site but from different individuals tend to be more similar than samples from different sites but from the same individual¹⁹⁰. Nevertheless, the small core genome and large pan-genome of the Albert *et al.* (2019) study are quite different from the results of our study and those of Díaz *et al.* (2021)^{84,189}.

Secondly, the gene presence/absence output from Panaroo was investigated for all *B. longum* subsp. *longum* isolates and plotted in a PCoA. Figure 28 shows that there is a slight clustering according to secretor status for this particular subspecies, and in Figure 29 a slight clustering according to *FUT2* genotype can also be observed. However, PCo1 and PCo2 only explain a small proportion of the dissimilarity between samples (12.0% and 8.0% respectively). This is typical of microbiome data, where there are usually several factors influencing the data¹⁹¹, and even those with a more pronounced effect still only explain a small proportion of the dissimilarity of the data. Although other principal coordinates can be checked, the influence of all factors on the data cannot be fully perceived simultaneously in a PCoA and corresponding plots due to its reduced dimensionality.

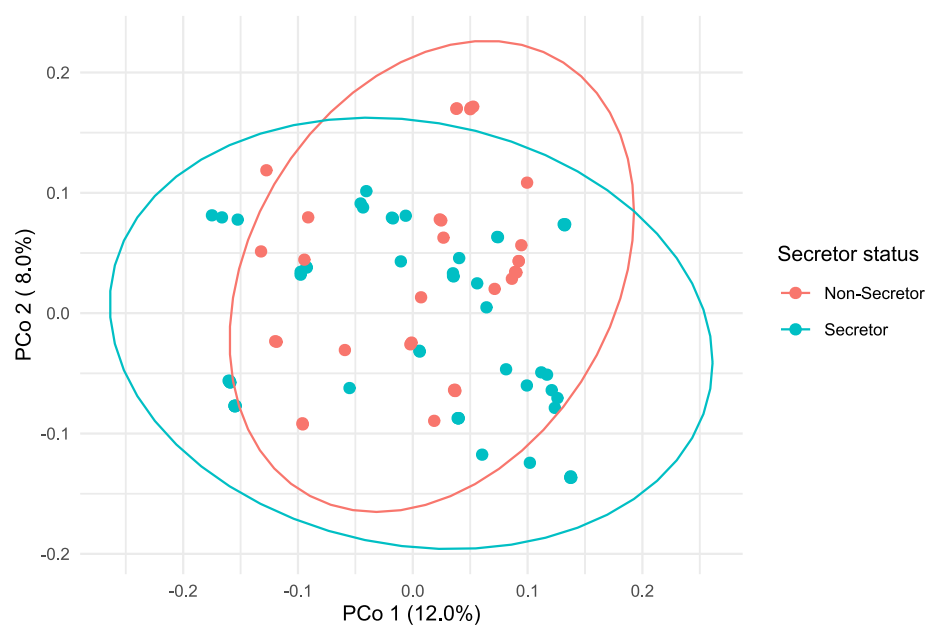


Figure 28 – PCoA based on gene presence/absence data of all *B. longum* subsp. *longum* isolates (n=92), coloured according to secretor status.

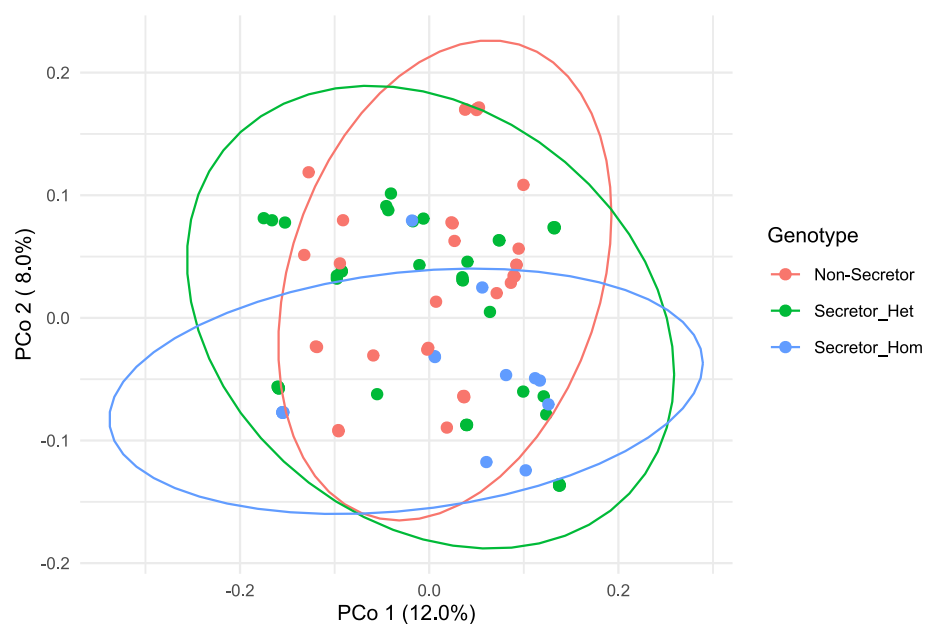


Figure 29 – PCoA based on gene presence/absence data of all *B. longum* subsp. *longum* isolates (n=92), coloured according to *FUT2* genotype.

Therefore, a PERMANOVA was performed to test the overall significance and variance explained by each data variable, as well as a PERMDISP to check whether a significant PERMANOVA p-value was caused by either the dispersion and/or centroid position. A summary of these statistics is given in Table 8. With the exception of culture medium, all other data variables had a highly significant p-value according to the PERMANOVA, despite the lack of notable clustering in the PCoA. Of these significant variables, ABO blood group had a relatively high R² value (9.476% explained variance), but also a significant PERMDISP p-value ($p=0.003$), showing that in this particular case the dispersion of the data points is at least partly responsible for the significant results, although the centroids could still be significantly distinct. The corresponding PCoA plot, coloured according to ABO blood group, is shown in Figure 30. For all other variables (secretor status, *FUT2* genotype, donor and sex), the PERMDISP p-value is not significant, so it appears that the centroid position is significantly different for the groups tested and thus there is significant clustering. However, it is also important to look at the PERMANOVA R² value, which is a measure of the variance explained. Variables such as secretor status, *FUT2* genotype and sex have a relatively small effect on the variance of the data (3.696%, 2.024% and 3.292% explained variance, respectively). This is consistent with the slight clustering in their PCoA plots, shown in Figure 28, Figure 29 and Figure 31, respectively. However, the donor variable has a large effect on the variance of the data (65.160%), which can also be seen in its PCoA (Figure 32). This is consistent with the current literature in that for the same site (stool in this case), interpersonal variability has a strong effect on the data across microbiome studies¹⁹⁰, suggesting host specialisation and possibly host adaptation.

Table 8 – PERMANOVA and PERMDISP statistics for each of the data variables tested. Significant results are marked as follows: (*) for $0.01 < p\text{-value} \leq 0.05$, (**) for $0.001 < p\text{-value} \leq 0.01$, (***) for $p\text{-value} \leq 0.001$.

	PERMANOVA p-value	PERMANOVA R2	PERMDISP p-value
Secretor status	0.001 (***)	0.03696	0.796
<i>FUT2</i> genotype	0.001 (***)	0.02024	0.536
Donor	0.001 (***)	0.65160	0.532
Medium	0.840	0.00172	0.114
ABO blood group	0.001 (***)	0.09476	0.003 (**)
Sex	0.001 (***)	0.03292	0.250

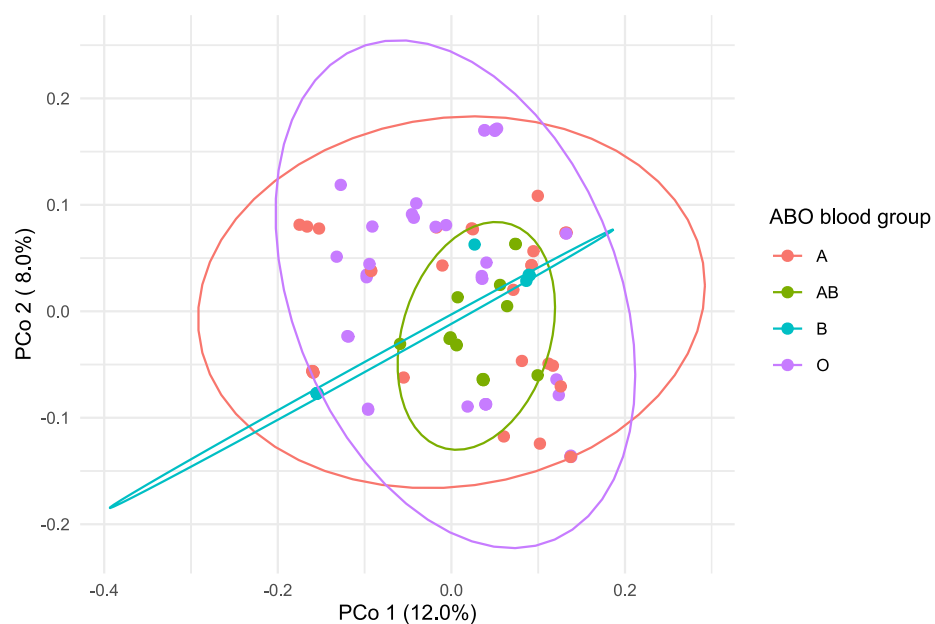


Figure 30 – PCoA based on gene presence/absence data of all *B. longum* subsp. *longum* isolates (n=92), coloured according to ABO blood group.

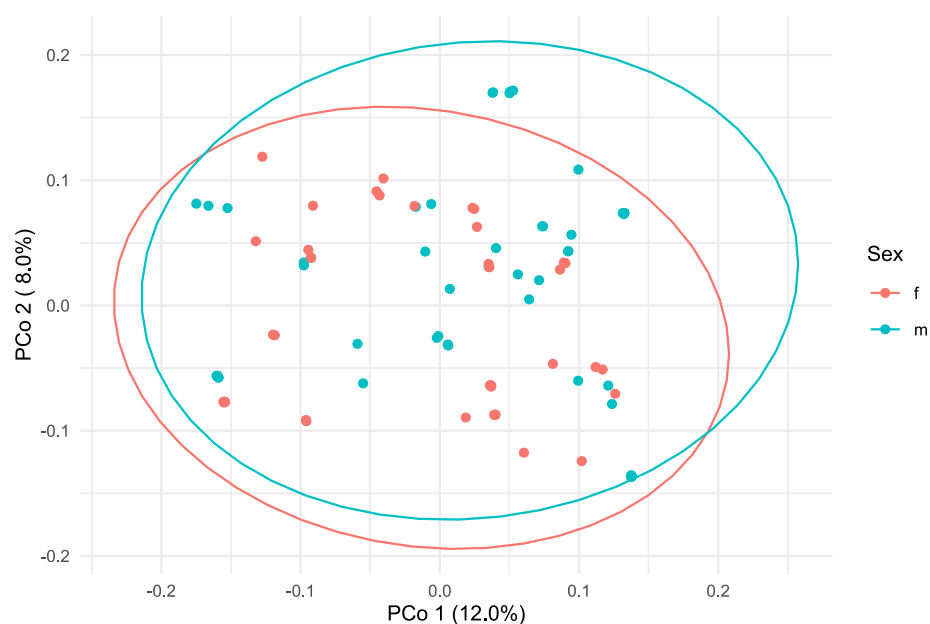


Figure 31 – PCoA based on gene presence/absence data of all *B. longum* subsp. *longum* isolates (n=92), colored by sex.

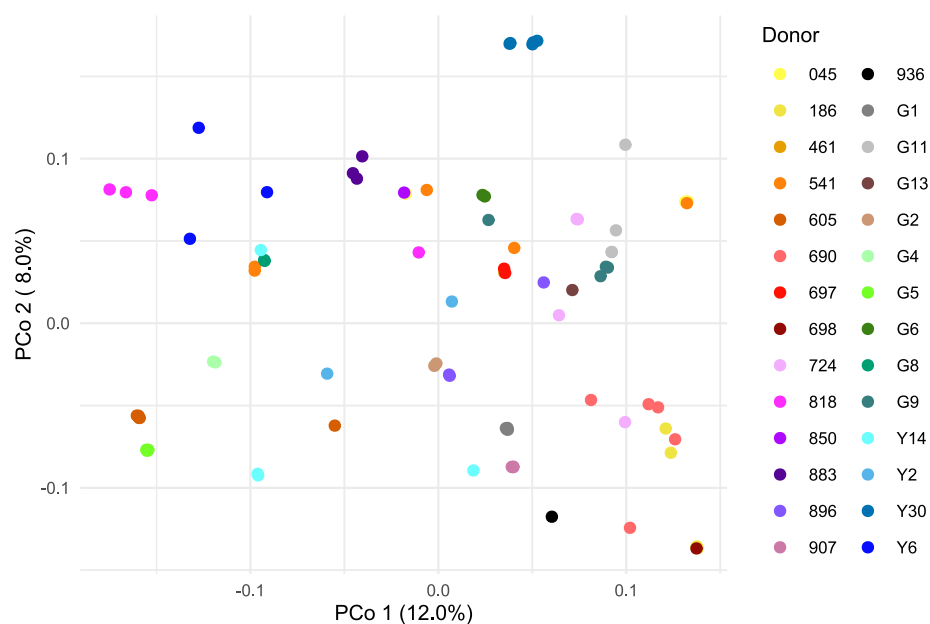


Figure 32 – PCoA based on gene presence/absence data of all *B. longum* subsp. *longum* isolates (n=92), coloured according to donor. No ellipses were drawn due to the low number of data points for certain donors.

Despite all the significant results, it is important to note that the PCoAs shown so far were performed on all *B. longum* subsp. *longum* isolates without dereplication. As there could be some inherent bias caused by closely related strains or even the occasional clonal isolate, we also performed the same statistical tests and PCoA on the 31 dereplicated *B. longum* subsp. *longum* isolates to test for significance and clustering according to the different data variables in unequivocally unique isolates. Figure 33 shows that, once again, the secretor status does not yield highly distinguishable clusters. For the *FUT2* genotype (Figure 34), isolates from homozygous secretors cluster much more closer together than the other two groups. Naturally, it could be due to the lower number of isolates in this particular group, or it could be a tendency of homozygous secretors, as we have already observed a similar trend in the 16S rRNA gene amplicon sequencing data analysis.

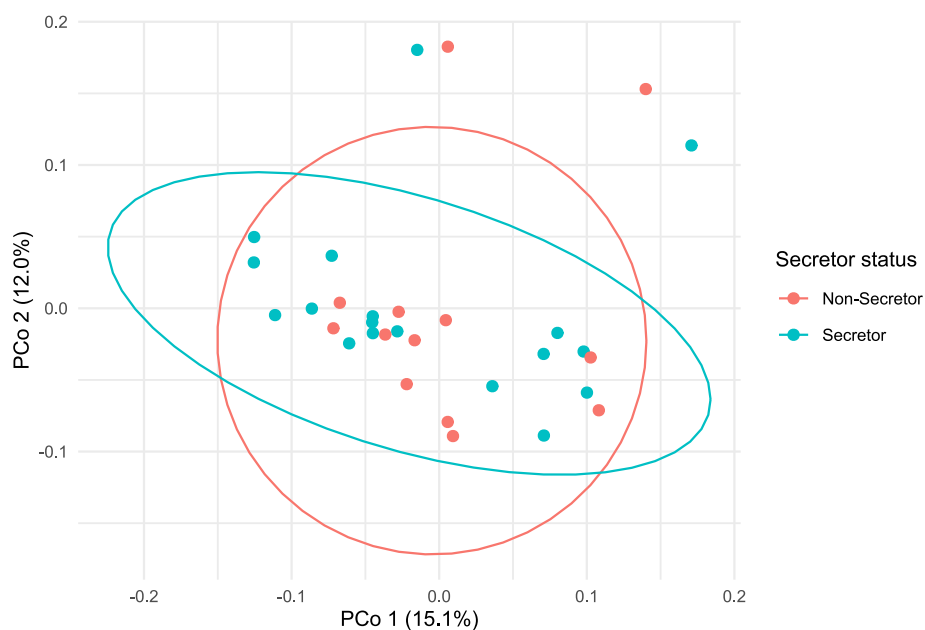


Figure 33 – PCoA based on gene presence/absence data of the dereplicated *B. longum* subsp. *longum* isolates (n=31), coloured according to secretor status.

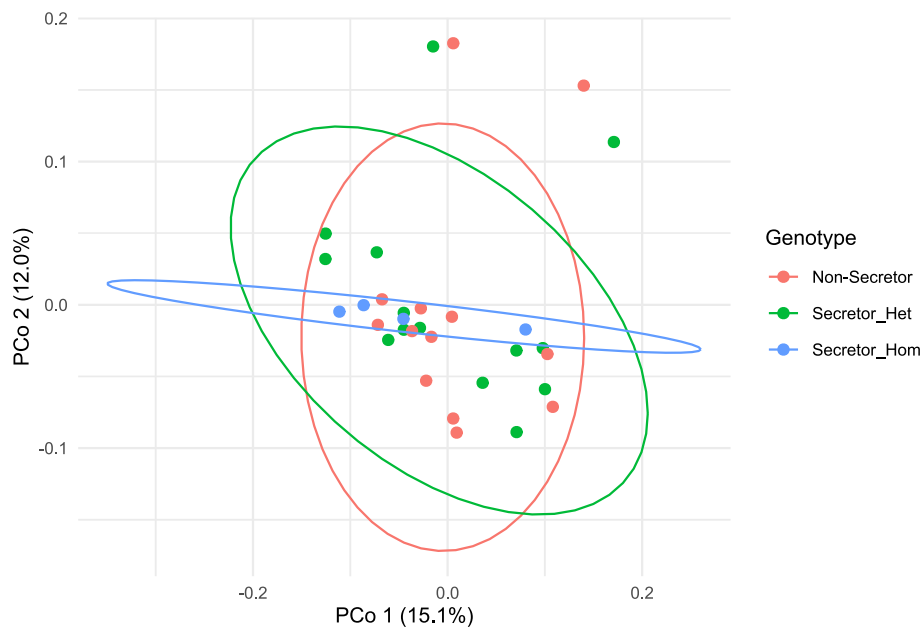


Figure 34 – PCoA based on gene presence/absence data of the dereplicated *B. longum* subsp. *longum* isolates (n=31), coloured according to *FUT2* genotype.

The PERMANOVA and PERMDISP statistical tests were then performed. Once again, these tests are particularly important when the first principal coordinates do not explain a significant percentage of the dissimilarity of the data (PCo1 15.1% and PCo2 12.0%). As can be seen in Table 9, most of the significant results were lost due to dereplication, highlighting the importance of avoiding redundancy. The only variable that remained significant was, as expected, the donor variable, which explains 69.418% of the variance. However, this time its PERMDISP p-value was also significant, meaning that the observed significance can be at least partially explained by the dispersion. In fact, when we look at the corresponding PCoA (Figure 35), we can observe that dereplication causes most donors to have only one isolate, with rare exceptions such as donors 186, 541, 724 and Y14, for which two isolates are represented. Of these isolates, those from donors 186 and 724 cluster very closely together, but those from donors 541 and Y14 do not. Therefore, it is difficult to make a definitive statement about

the influence of the donor variable on *B. longum* subsp. *longum*, as isolates from the same donor tend to be so similar that most are lost by dereplication. This may be due to host specialisation or to the culturing of clonal isolates. In future studies dereplication could be performed with different parameters and it could also be useful to isolate more isolates of this subspecies per donor in order to get a higher sample size per donor.

Table 9 – PERMANOVA and PERMDISP statistics for each of the data variables tested. Significant results are marked as follows: (*) for $0.01 \leq \text{p-value} \leq 0.05$, (**) for $0.001 \leq \text{p-value} \leq 0.01$, (***) $\text{p-value} \leq 0.001$.

	PERMANOVA p-value	PERMANOVA R2	PERMDISP p-value
Secretor status	0.109	0.03495	0.890
<i>FUT2</i> genotype	0.459	0.02488	0.442
Donor	0.008 (**)	0.69418	0.002 (**)
Medium	0.383	0.02610	0.571
ABO blood type	0.165	0.09171	0.100
Sex	0.201	0.03036	0.510

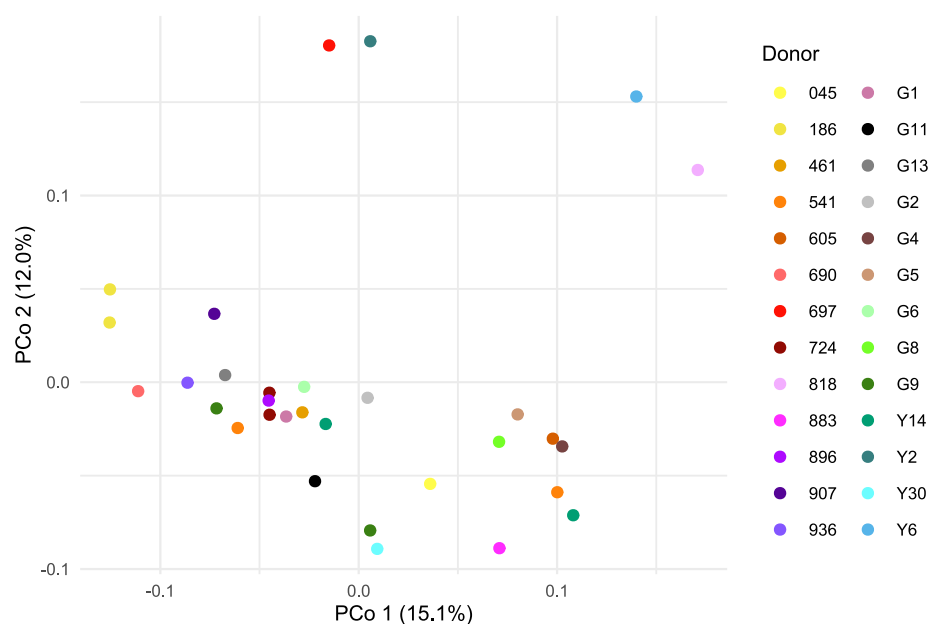


Figure 35 – PCoA based on gene presence/absence data of the dereplicated *B. longum* subsp. *longum* isolates (n=31), coloured according to donor. No ellipses were drawn due to low number of data points for certain donors.

The gene clusters of the dereplicated *B. longum* subsp. *longum* data were then annotated according to the COG categories. Figure 36 shows very similar results to those previously obtained for the whole *Bifidobacterium* genus, as would be expected given that this subspecies represented the majority of the isolates. Similarly, the category with the highest relative frequency was S (function unknown), followed by G (carbohydrate metabolism and transport), E (amino acid metabolism and transport), K (transcription), J (translation) and L (replication, recombination and repair). These results are consistent with those most commonly found in the literature^{84,192}, and again there is a large proportion of carbohydrate metabolism genes compared to other gut bacteria, as is typical for bifidobacteria⁹⁷. However, in the study by Odamaki *et al.* (2018) on *B. longum* subsp. *longum* isolates from an adult Japanese population, COG X (mobilome: prophages, transposons) showed a high relative frequency¹⁹³. In our

study, the set of *B. longum* subsp. *longum* isolates did not have counts for the functional category X, similar to other studies on this subspecies^{84,189}. However, the presence of prophages in bifidobacteria is strain-dependent¹⁹², and thus this difference could be due to the different isolates investigated. Overall, significant differences according to secretor status could not be observed, as assessed by Wilcoxon rank-sum tests with FDR correction.

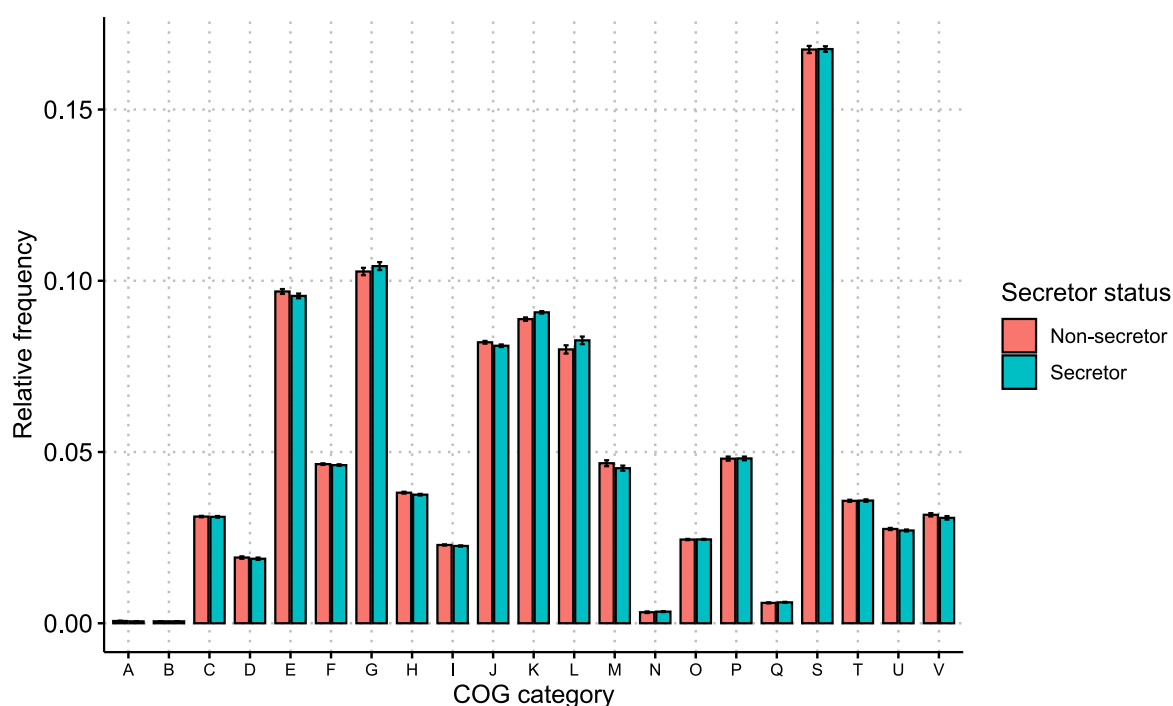


Figure 36 – Distribution of functional categories in 31 dereplicated *B. longum* subsp. *longum* genomes from donors with different secretor status. COG categories are abbreviated with capital letters as explained in Figure 22. Only categories present in our data are shown.

Finally, fucose-specific genes were investigated for all isolates of this subspecies, as shown in Figure 37. As already seen in Figure 26, it could be observed that *fucO* is present in all strains of subspecies *longum*. As explained in 4.3.3, this gene encodes a lactaldehyde reductase that is involved in the L-rhamnose and L-fucose catabolic

pathways¹⁷⁹. Additionally, there is one strain (G9M2) that also possessed the *fucU* gene and an L-fuconate dehydratase. These genes may be part of a strain-specific mechanism for fucose degradation that has yet to be fully characterised, or a result of HGT from other bacterial species⁸². This set of genes has been reported for several strains of subspecies *longum* in the study by Díaz *et al.* (2021)⁸⁴. However, our strains do not contain fucosidases or the *fucC* gene, which have been reported for certain *B. longum* subsp. *longum* strains⁸⁴.

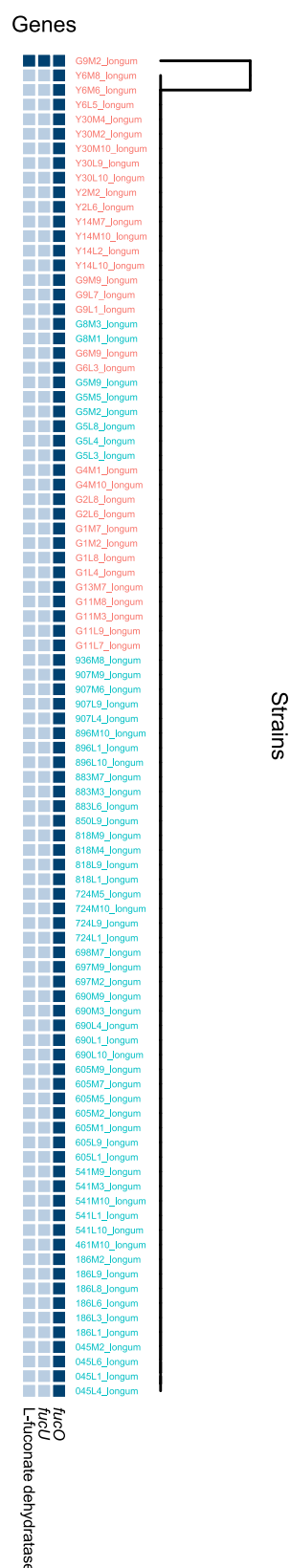


Figure 37 – Presence/absence heatmap of fucose-related genes in 92 *B. longum* subsp. *longum* genomes. The species name was added to the isolate name to facilitate data interpretation. Clustering was performed according to fucose-related genes only.

In the heatmap cells, present genes are in dark blue, absent genes in light blue. Regarding strain names, those from secretors are in blue and those isolated from non-secretors are in salmon red.

4.4. Conclusion

In this study, a similar number of lactobacilli isolates was obtained from secretors and non-secretors, but more *L. mucosae* isolates stemmed from secretors and more *L. plantarum* subsp. *plantarum* isolates stemmed from non-secretors. Bifidobacteria were significantly more commonly isolated from secretors than non-secretors. Secretors also had a more diverse bifidobacterial community, including species such as *B. angulatum*, *B. animalis* subsp. *lactis* and *B. dentium*. However, it is important to note that only a subset of all gut bacteria was sampled, implying a strong sampling bias. For example, when a species is not present in one of the groups studied, it does not necessarily mean that the species does not exist in that group, but that it was simply not cultured for that group in this particular project, which could be due to chance effects or perhaps a minimal presence or absence of that species in that group.

The pan-genome analysis of lactobacilli as a group highlighted the importance of doing such an investigation for each new genus in the future, as the isolates were too diverse to identify a core genome. This analysis would also benefit from less conservative parameters, as is also the case for the bifidobacteria pan-genome analysis. Our PCoA results from *B. longum* subsp. *longum* reiterate the importance of inter-individual variability in microbiome data, as the donor parameter itself proved to explain a high proportion of the variance of the isolates. Finally, we observed no functional differences

according to secretor status for lactobacilli, bifidobacteria and *B. longum* subsp. *longum*, as well as no secretor status-dependent differences in fucose-related genes.

Although there is still much to explore in our data, the preliminary results suggest that the secretor status does not have a significant influence on the genomes of these probiotic-like taxa, and no substantial evidence for host adaptation dependent on secretor status could be obtained. Further studies could be performed to obtain more conclusive results, such as analysing whole genome arrangements¹⁵⁸, analysing SNP variations with a genome-wide association study (GWAS)¹⁹⁴, and performing transcriptomics experiments to account for potentially inducible operons and differential gene expression¹⁸⁷, as suggested above.

5. Challenging the diet culture through an evolutionary perspective

Ana Sofia G. Borges^{1,2,+}, Freya A. Pappert^{1,3,*,+} and Olivia Roth^{1,3}

Affiliations

¹DFG Research Training Group for Translational Evolutionary Research

²Department of Microbiology and Biotechnology, Max Rubner-Institut Kiel, Hermann-Weigmann-Str. 1, D-24103 Kiel, Germany

³Zoological Institute, Christian-Albrechts-Universität Kiel, Am Botanischen Garten 1-9, 24118 Kiel, Germany

⁺These authors contributed equally to this work.

*Corresponding author. Contact: fpappert@zoologie.uni-kiel.de

Manuscript in preparation.

5.1. Contextualization of the manuscript within the thesis

As explained throughout this thesis, microbiome research is still in its infancy and must overcome several challenges, such as small sample sizes and high individual variability, before it can reach the clinical setting³⁷. However, this does not stop probiotics from being marketed in various formats as a panacea and, as a result, interest is growing among the general public. While there are certain medical conditions for which probiotics have been shown to be useful (e.g. *C. difficile*-induced diarrhoea¹⁹⁵), several of the probiotic products currently on the market do not live up to the claimed health benefits¹⁹⁶. Even when it comes to the secretor status, a patent application was filed in 2009 claiming to have found secretor status-specific probiotic strains of bifidobacteria¹⁹⁷, despite the research still being far from clear in 2024, as shown in this doctoral study. Furthermore, there is often a discrepancy between the information on the label and the actual content of probiotics, while at the same time some contain resistance and virulence genes¹⁹⁸. Given the lack of adequate regulation of these products¹⁹⁹ and the high public demand, the probiotics sector has become a dangerously profitable area.

This issue is part of a much bigger problem, namely deception in wellness trends. The wellness industry is now three times more profitable than the pharmaceutical industry, yet it is often not adequately regulated and its claims are often not properly substantiated²⁰⁰. A famous example is the weight loss industry, which has failed to provide sustainable solutions over the years²⁰¹, but has nevertheless been incredibly successful in misleading many into thinking that the solution to their problems is just one diet away²⁰².

The following manuscript will address the challenges posed by a diet-obsessed society and the evolutionary mismatch between modern weight loss goals and the way our

bodies have evolved to store energy. This manuscript was written in the Translational Evolutionary Biology course of the RTG 2501 TransEvo. A subsection on the gut microbiome has been added to link the main topic of this opinion piece and this doctoral thesis. We will also discuss the non-linear relationship between BMI and health, despite its use as a common metadata parameter, as seen in the human stool donor cohort of this doctoral project.

5.2. Abstract

Diet culture drives many people to extreme dieting or unhealthy behaviours to control weight with the promise of a healthier body, whilst also stigmatising those in larger bodies. The explanation for the presumably worse health that a large proportion of our dieting society experiences relies on an evolutionary mismatch between modern restrictive diets and the biological response to food deprivation. Our societal weight-centric mindset thus exposes us to physiological consequences emerging from largely ineffective and unsustainable dietary restriction. Insights from evolutionary adaptation may pave the way towards an improvement of individual health by including weight-inclusive approaches such as Health at Every Size[®] (HAES[®]).

5.3. Introduction

Restrictive diets are commonly promoted to fight high obesity rates through weight loss. This is achieved by caloric restriction (CR), which can include imposed rules on what, how and when to eat. CR entails an evolutionary conserved stress response to low nutrient intake. In countless animal studies CR has been shown to alter cell metabolism, prevent age-related diseases, prevent inflammation and overall prolong the healthy lifespan^{203–205}. Nonetheless, there is still little research on humans about

such outcomes and CR can unintentionally turn into dietary restriction, when macro- and micro-nutrients are removed leading to deficiencies and malnutrition²⁰⁶.

The battle against obesity and the pursuit of health endorsed the colloquially called “diet culture”, i.e. the societal trend of glorifying slimness and equating it to health, success, attractiveness and sometimes even moral superiority²⁰⁷. Between 2010-2015 about 42% of adults worldwide were trying to lose weight, of which 98% reported that their main motive was “to improve health and prevent future disease” and 80% to “improve appearance”²⁰⁸. This thin-ideal is popularised in the news, social media, diet foods, weight loss advertisements and fitness trends. In 2021 the global market for weight loss products and services was 254.9 billion \$ and is predicted to reach 377.3 billion dollars by 2026²⁰⁹. Diet culture is presumably causing an alarming rise in eating disorders²¹⁰, whilst also normalizing disordered eating²¹¹ and increasing weight stigma²¹². Whilst new perspectives are growing nowadays, as the body positivity movement, we are conversely still advised to control our body weight^{213,214}.

It is alarming that we lack evidence that the imposed diets lead to sustained weight loss. Numerous studies show that individuals that lose weight under CR tend to regain it over the next few years (Figure 38)²⁰¹, up to the point that one to two-thirds of dieters actually put on more weight than at the start²¹⁵. The few successful weight loss cases report higher preoccupation around shape and weight, higher binge eating frequency²¹⁶, with changes in health markers (e.g. fasting blood glucose, cholesterol, triglyceride levels) being small and uncorrelated with weight change²¹⁷. Follow-up after the diet is usually only recorded up to two years since its start²¹⁷, which is not indicative of sustained weight loss. Moreover, people are frequently attracted by the initial thrill of losing weight quickly, but when the results are not long-lasting, the next diet is attempted. This corresponds to yo-yo dieting, which leads to a cyclic pattern of losing

and regaining weight resulting in a net increase in body weight and cardiovascular risk²¹⁸.

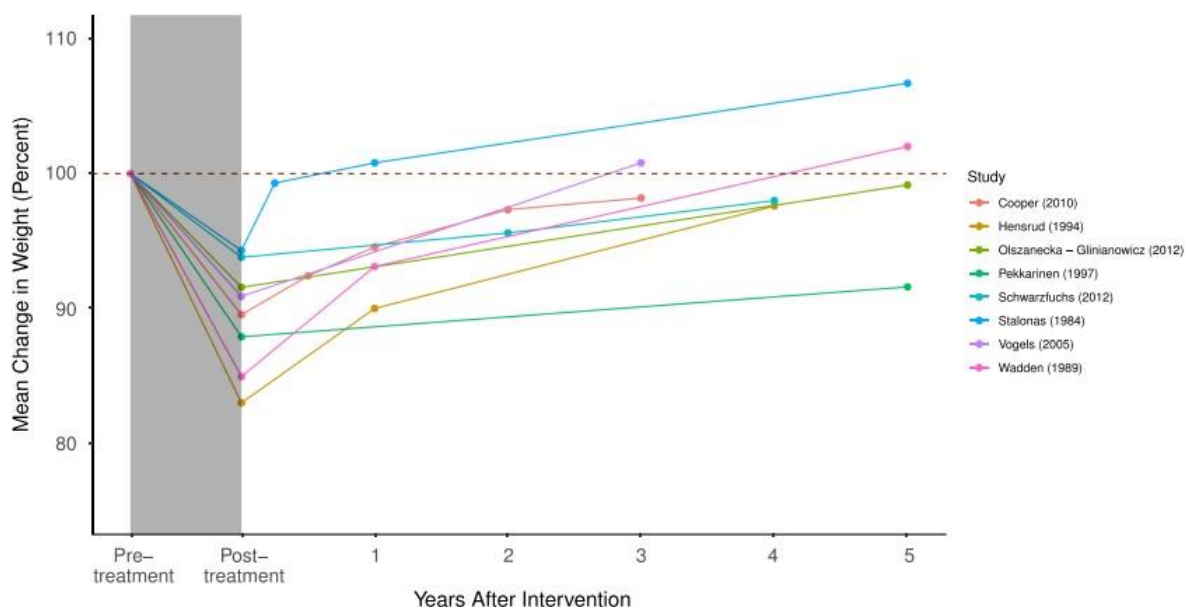


Figure 38 – Illustration of the general unsuccessful outcome of diets across several studies in a 5-year period. Figure from Nordmo *et al.* (2020)²⁰¹.

The reason why the majority of diets fail and may not necessarily mean better health has an evolutionary background. The aim of this perspective article is to delve into some of the mechanisms why humans have not evolved to lose fat, but rather have adapted a plethora of strategies to maintain a set-point weight and avoid starvation^{219,220}. However, there is a mismatch between what society perceives as a “normal and healthy” body and our individual evolved phenotype. In such light, it is proposed that shifting the current weight-centric paradigm to a Health at Every Size® (HAES®)-aligned approach, an alternative focusing on promoting healthy habits rather than weight loss, would foster populational and individual health improvements by working within an evolutionary framework²⁰².

5.4. Discussion

5.4.1. “Nothing in biology makes sense except in the light of evolution” - Theodosius Dobzhansky (1900-1975)

5.4.1.1. Set-point theory

Humans have evolved in constantly changing environments accompanied by energy uncertainty (e.g. seasonality, extreme environmental events, habitat displacement, diverse ecological challenges)²²¹, the real survivors being those who could efficiently store the surplus calories they consumed^{220,222}. That is why our set-point weight, which is hypothesized to be an individual weight range that is biologically predetermined, is much better protected on its lower end²²³, whilst the upper end may be subject to drift, depending on stochastic environmental factors (e.g. predator-prey interactions, infectious diseases)^{220,224}. When we lose weight, evolutionary conserved metabolic adaptations that help protect our set-point weight may prove counterproductive. The total daily energy expenditure is the number of calories used per day, and about 70% of it corresponds to the basal metabolic rate (BMR), which is the number of calories needed to perform vital functions²²⁵. Upon dieting, the BMR decreases in order to conserve energy. Specifically, for every kilogram of weight lost, energy expenditure decreases by ~30 kcal/day (regardless of the degree of adiposity of individuals), while appetite increases by ~100 kcal/day due to hormonal changes^{219,226}. The latter encompass an altered secretion of appetite-related hormone levels such as insulin, leptin, ghrelin and cortisol, increasing the sensation of hunger²²⁷. The decline in BMR and most hormonal shifts persist for at least a year after the period of active restriction^{228,229}.

The Minnesota Starvation Experiment studied the effects of CR to pinpoint the best dietary rehabilitation method for food deprived victims of World War II. In this study, 32 healthy men went through a semi-starvation period of six months on about 1600 kcal/day while keeping a physically active lifestyle²³⁰. Appreciable recovery in the subsequent rehabilitation period could only be achieved by eating about 4000 kcal/day for some months²³¹. Despite its small sample size and limited reproducibility due to ethical reasons, this longitudinal experiment revealed the impact of long-term food deprivation²³⁰. Besides leading to extreme hunger, CR can also cause a decrease in body temperature, diminished libido²³², menstrual irregularities²³³ and a decrease in thyroid hormones²³⁴, among other physiological consequences, together with a vast impact on mental well-being²³¹. Such survival mechanisms are maladaptive when one aims to lose weight with an evolutionary mismatch between dieting and our body interpreting the rapid calorie loss as starvation²³².

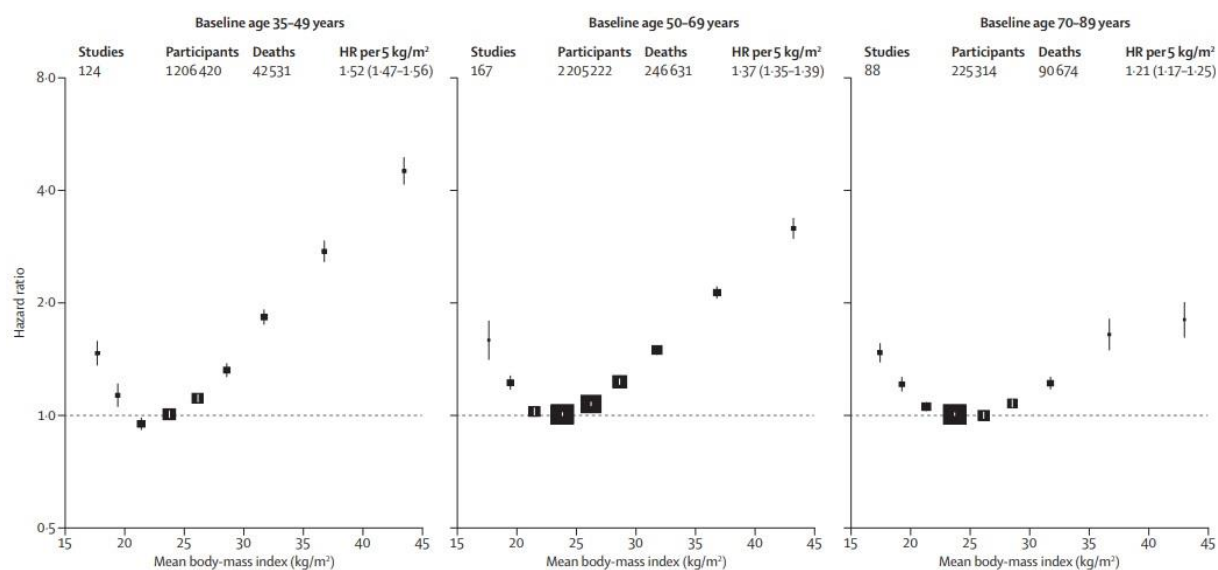
5.4.1.2. Misconceptions about adiposity and health

Many organisms have evolved mechanisms to store their energy reserves efficiently as lipids to survive periods of nutrient scarcity²³⁵. Fat is divided into white and brown adipose tissue, the first being important for the storage of energy and functioning as an endocrine tissue, and the latter essential to regulate thermogenesis^{224,235}. It is an organ with evolutionary conserved metabolic pathways that mainly provides energy for growth, reproduction and immune function²³⁶.

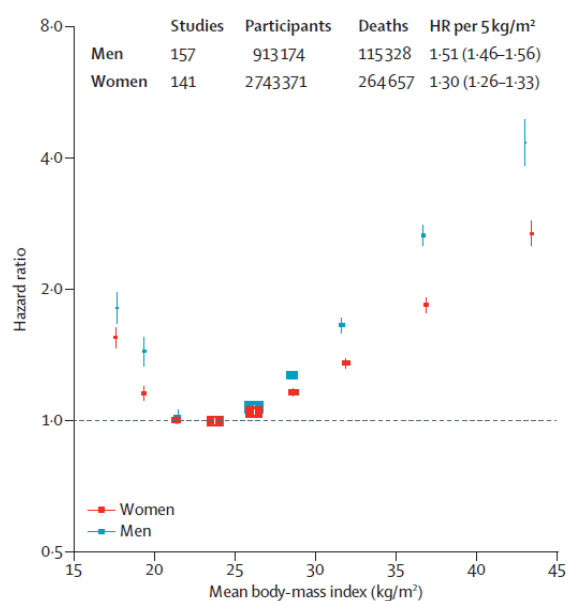
Fat often receives an exclusively negative connotation because it is associated with poor health outcomes in obesity research^{237,238}. However, fat provides a survival advantage in certain circumstances, including old age and particular medical conditions such as chronic obstructive pulmonary disease and renal cell carcinoma^{239,240}. Moreover, the relationship between BMI and health outcome is non-

linear and rather complex^{241,242}. This is even more pronounced when accounting for population, age and sex (Figure 39A and B)²⁴³, e.g. the BMI associated with a lower hazard ratio (HR) for all-cause mortality shifts upwards with increasing age (Figure 39A) and at the higher end of the BMI scale, women have a lower HR compared to men (Figure 39B)²⁴³. Additionally, the HR for BMI 18.5 and 30 is very similar, which is surprising considering that the latter is deemed a medical concern and the first as healthy. This suggests that health can be found in a broader range than usually thought, and that both extremes are associated with a higher HR. Furthermore, regardless of their BMI, individuals that practice several healthy behaviours have a similar HR (Figure 39C)²⁴⁴. Altogether this implies that both health and adiposity are incredibly multifaceted and polygenic in nature²²⁴, being shaped by sex, age, utero environment, ethnicity, economic and social status^{221,245}. Finally, we note that the BMI is a measurement that does not represent individual health accurately, but that is nonetheless still used in medical practice and research²⁴⁶.

A)



B)



C)

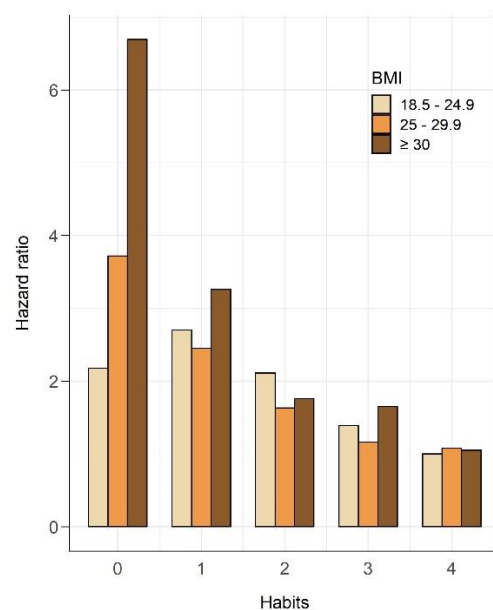


Figure 39 – A) HR for all-cause mortality depending on age and BMI. B) HR for all-cause mortality depending on sex and BMI. Data of A) and B) are derived from three geographical regions (East Asia, Europe, and North America), the meta-analysis was restricted to never-smokers without pre-existing chronic diseases, and areas of squares are proportional to the information content²⁴³. Figures A) and B) are from Angelantonio *et al.* (2016)²⁴³. C) HR for all-cause mortality depending on BMI and on

the number of healthy behaviours adopted (daily consumption of five servings of vegetables and fruits, physical activity minimum 12 times a month, moderate alcohol consumption, and not smoking). Figure generated from data in Matheson *et al.* (2012)²⁴⁴.

Health is determined by a combination of factors, some of which are not modifiable, e.g. our genetic makeup, that has been selected over the course of human evolution. Other components influencing health can only be reshaped through societal changes, including access to medical care and socioeconomic factors²⁴⁷. One striking example of this is that 10% of the UK households would have to spend around 70% of their disposable income to meet dietary guidelines²⁴⁸. After removing all of these elements, only around 30% of our health is actually influenced by individual behaviours, such as those highlighted in Figure 39C²⁴⁷. Weight is also determined by a multitude of factors, such as development, genetics, age, ethnicity and metabolism, together with modifiable factors such as food and physical activity²⁴⁹. For instance, genetics alone accounts for 40% to 70% of the inter-individual variability in BMI²⁵⁰.

5.4.1.3. Population genetics

Despite humans having adapted common physiological survival strategies regarding their adipose tissue, there is still great phenotypic variation between and within populations in how much and where fat is stored²⁵¹. Major population genetic variability seems to have occurred when *Homo sapiens* migrated out of Africa over 1.8 million years ago to other continents, adapting to different environmental niches through natural selection and being subjected to gene flow and random genetic drift²²⁴. Not only specific adaptation to particular ecosystems and diets have created a unique genetic history, but also historical context of populations with particular cultural norms of mating

rituals (e.g. body type preference) and religious fasting behaviours could underlie ethnic-specific variations in weight²²⁴. Pleiotropic and trade-off effects could partly explain why alleles that seem to be of no use in our current modern environment and contribute to weight gain remain in the human gene pool. Moreover, whether adiposity is stored more viscerally or subcutaneously, and whether it accumulates more as lower-body fat (seen more often in women) or in the abdominal area is not a personal choice^{251,252}. As such, those who are preoccupied with their body shape and try to achieve cultural beauty standards may in vain attempt this through poor nutritional choices and sometimes dangerous diets²⁵³.

5.4.1.4. Transgenerational inheritance

A parent's life experience can alter the phenotype of their offspring through transgenerational non-genetic inheritance, aiming to better prepare them to the expected parental habitat and so improve their fitness²⁵⁴. Phenotypic plasticity is a fast acclimatisation to environmental conditions and can happen within and between generations, such as transgenerational plasticity (TGP). These rapid adaptations rely predominantly on epigenetic modifications, in contrast to genetic adaptations, which require mutation accumulation over generations^{255,256}. An environmental cue that can induce TGP is CR^{245,257}. In this context, one possible reason for today's developmental mismatch in high-income countries may be that during the world wars our ancestors were exposed to a habitat with less readily available food, while their offspring were born in an industrialized society with highly processed meals^{221,245,257}. In turn, TGP may have caused acclimatisation to poor food conditions, which no longer apply today, thus leading to the parent-offspring mismatch that contributes to an increased probability of obesity. The Dutch famine in World War II is an illustration of this, as progeny from these mothers had a higher prevalence of obesity as adults²⁵⁸. Interestingly, such

developmental mismatches may be countered faster than expected precisely thanks to TGP, without relying on mutations to occur in nutrient-sensitive genes or on favourable genes to be selected and spread throughout human populations²⁵⁹. With this in mind, we hypothesize that parents who have repeatedly dieted may influence the phenotype of their offspring to have a higher likelihood of obesity.

5.4.1.5. Gut microbiome

The gut microbiome, which has co-evolved with humans for millions of years²⁶⁰, can influence body weight through mechanisms such as modulation of gut permeability, production of metabolites that mediate satiety and insulin resistance, and regulation of energy harvested from food²⁶¹. It is evolutionarily advantageous to have gut bacteria that can extract more energy from the same food because it can be used as fuel by the host^{262–264}. However, today these bacteria contribute to the evolutionary mismatch between current weight loss ideals, our modern plentiful environment and the way our bodies have evolved to store energy^{262–264}.

The main topic of this doctoral project may be an example of this evolutionary mismatch. One study on a particular *FUT2* SNP in a Kazakh population showed that homozygous secretors had a lower risk of being overweight²⁶⁵. However, another study on the same *FUT2* SNP in a Caucasian population reported that non-secretors were associated with a lower BMI compared to secretors²⁶⁶. In addition, a study focusing on a different *FUT2* SNP in post-bariatric surgery patients reported that homozygous secretors have increased levels of propionate (a SCFA whose precursor can be derived from fucose catabolism), which can be a source of energy and thus potentially lead to the increased waist-hip ratio seen compared to the other groups²⁶⁷. Two studies then suggest that secretors may be more likely to have a higher BMI or waist-hip ratio, most likely due to the increased energy sources derived from fucose utilisation^{266,267}.

Interestingly, it is the non-secretor status that is usually associated with an altered gut microbiome composition and several medical conditions that involve an altered microbiome (e.g. IBD)²³. This may be an example of how a lower weight does not equate to a healthier state. It may also be an illustration of how bacteria that harvest more energy from food are often vilified, when in fact they can support the host by providing more energy from otherwise indigestible material, for example in the form of SCFAs, which are important for human health²⁶⁸. Nevertheless, the contradictory nature of the results of these studies prevents further conclusions on this subject until more research is conducted. Body weight is highly complex and polygenic²²⁴ and thus research cannot focus on a single gene, SNP or population, but must instead look at the multiple factors that influence body weight and their interactions, to avoid confounding.

The influence of the gut microbiome on body weight, and the fact that most microbiome research is in its early stages, provide a breeding ground for misleading wellness trends. While gut microbiome research is indeed promising, it is very often stripped of nuance by diet culture and used to sell yet another weight loss product without proper evidence^{269,270}. The gut microbiome has been shown to be very stable throughout life²⁷¹ and we are still far from implementing microbiome-targeted therapies and guidelines in a clinical setting^{272,273}. On the one hand, therapies targeting the microbiome, such as faecal transplantation and probiotics, with rare exceptions, still need stronger evidence and streamlining before they can be implemented in a safe, accessible and efficient way^{272,273}. On the other hand, restrictive dietary changes are incredibly difficult to sustain long-term²⁰¹ and the associated microbiome changes seem to be transient²⁷⁴.

Furthermore, diet trends such as drastic restriction of dietary macronutrients (e.g. carbohydrates) and gluten-free diets are often marketed as “detoxes” or “cleanses” for the gut microbiome, but actually result in a reduction in its diversity^{275,276}. This can be particularly concerning given that there has already been a decrease in gut microbial diversity throughout human evolution²⁷⁷. This is likely to have been caused by dietary changes such as increased meat consumption, introduction of starch-rich diets, and use of fire for food production, which reduced the selective pressure to maintain certain bacteria, such as those responsible for digesting resistant plant materials²⁷⁸. This decrease in microbial diversity has accelerated in recent years, probably due to practices such as the consumption of processed starches and increased hygiene measures^{278,279}. Reduced microbial diversity is associated with autoimmune disorders, metabolic damage and increased risk of infection²⁷⁹. Further reduction of microbiome diversity through dieting may lead to even more serious consequences.

Among all the preliminary research on the gut microbiome, one of the few robust findings is that eating a diverse range of foods benefits the gut microbiome²⁷⁵. Given that restrictive diets lead to reduced microbiome diversity^{275,276}, this supports the idea that health-promoting behaviours, such as eating a variety of foods, including fruit and vegetables, as highlighted in Figure 39C²⁴⁴, are what drive health, rather than thinness and restriction as promoted by diet culture.

5.4.2. “Insanity is doing the same thing over and over again, but expecting different results.” - Rita Mae Brown (1944-)

Considering the side effects of diet culture, a more inclusive and nuanced approach is Health at Every Size® (HAES®). It advocates for societal changes, such as normalization of body diversity and equitable healthcare for all individuals, regardless of body size. This includes doing enjoyable physical activity and eating for well-

being²⁸⁰. One common framework for the latter is Intuitive Eating (IE), which is a non-diet approach to nutrition, focusing on eating based on internal cues, rather than external conditions such as diet plans and food rules. It promotes interoceptive awareness and responsiveness, i.e. recognizing internal body sensations, such as hunger and fullness, and responding to them in order to meet physical and psychological needs²⁸¹. Another key characteristic of IE is the unconditional permission to eat whatever food is desired, by removing moral judgements attached to dietary choices²⁸¹, so as to be more aligned with individual nutritional needs and feeding behaviours. IE also acknowledges that not everyone has access to fresh and healthy food²⁸², and that certain medical conditions may require dietary adjustments (e.g. food allergies, diabetes, IBD)^{283,284} and can be relied on as a supportive and non-judgmental framework.

HAES[®] promotes equitable physical health benefits in comparison to weight loss diets, with the advantages of longer lasting effects and promotion of psychological well-being^{202,285,286}. When HAES[®] and IE were compared to traditional diets, the participants on the HAES[®] program maintained their weight and improved their physical health markers in the first year, while dieters lost weight and presented health improvements. However, in the second year, dieters regained the lost weight and little health improvements remained, whereas these were sustained in the HAES[®] participants²⁸⁷. Two other studies reported that while the diet group lost weight and the weight-neutral group did not, both presented comparable health benefits^{288,289}. This reinforces the idea that health improvements can be achieved by incorporating healthy habits, rather than by losing weight, as previously suggested in Figure 39C²⁴⁴. Regarding body size changes, a recent systematic review showed that IE and similar approaches lead to comparable weight loss to diet programs in short-term studies²⁹⁰. Therefore, not only does diet culture wrongly associate health benefits with weight loss, but it is also

inefficient in its own goal when compared to alternatives that specifically do not target weight loss. Additionally, IE predicts longitudinally improved mental health and decreased disordered eating behaviours such as binge eating²⁹¹, possibly by nullifying the restriction mindset and weight preoccupation promoted by diet culture.

From an evolutionary perspective, the weight maintenance associated with HAES® and IE is more in accordance with the set-point theory and thus we hypothesize that these alternatives could help people settle in their current phenotypically preferred set-point weight range, whilst preventing poor health outcomes caused by weight cycling. We also believe that HAES® and IE could boost libido by diminishing body image concerns^{211,292} and improve female reproductive health by reducing the incidence of amenorrhea, which can be found among those under CR²⁹³. Finally, we hypothesize that these alternatives may benefit future generations by preventing the food deprivation-induced physiological adaptations described above, since humans are phenotypically plastic and able to rapidly adapt to changing environments thanks to epigenetic changes^{256,259}, nongenetically transmitted parental effects²⁵⁴ and even gene-culture co-evolution²⁹⁴. Future research on these topics will help provide sounder evidence to support weight-inclusive alternatives, together with more studies comparing the weight-centric and weight-neutral approaches, preferably with a long-term follow-up (e.g. 5-year-long) and extensive metadata to avoid the influence of confounders (e.g. socioeconomic status and human genetics).

5.5. Conclusion

Diet culture tends to promote polarizing views about health that are often incorrect, as it is a complex subject that requires a nuanced approach. The aim of these considerations is not to berate people who want to lose weight (a common desire when living in a culture that rewards thinness), but rather to show that health and weight are

two different concepts that are not linearly correlated. Moreover, weight loss dieting does not guarantee health benefits, may involve certain health risks and can be inefficient in the long-term.

Fundamental insights are obtained through comprehension of the evolutionary mismatch between modern weight loss trends and our physiology not having adapted to distinguish intentional dietary restriction from food scarcity. Non-diet and weight-inclusive approaches that follow HAES[®] principles, like IE, are evidence-based alternatives that have presented more sustainable health outcomes when compared to traditional diets²⁰² and that move away from calorie-counting and weight-related worries to actual health and nourishment promotion. We consider that the research is pointing towards the need for a paradigm shift to a weight-inclusive approach²⁰² and evolutionary medicine can help reframe our understanding of weight and health.

5.6. Funding

This work was supported by the Research and Training Group 2501 on Translational Evolutionary Research (RTG 2501 TransEvo).

5.7. Acknowledgements

We would like to thank Charles Franz for his support, motivation and corrections in the review process of the manuscript. Permission for the following acknowledgment was obtained from Charles Franz.

5.8. Disclaimer

The views expressed in this article are those of the authors and do not represent the official position or reflect the views of the Max Rubner-Institut, the Federal Research Institute of Nutrition and Food in Germany.

6. General discussion

6.1. Main findings

This thesis investigated whether the host secretor status influences gut microbiome taxa with probiotic potential, namely bifidobacteria and lactobacilli. The main findings are outlined below:

➤ In Chapter 2, the composition of faecal samples from the cohort studied was examined using 16S rRNA gene amplicon sequencing data. No significant differences between secretors and non-secretors were found in terms of phylum abundance, lactobacilli abundance, bifidobacteria abundance and alpha diversity. An interesting finding was that secretors presented mainly lactobacilli of the genera *Limosilactobacillus* and *Ligilactobacillus*, two typically autochthonous genera, whereas non-secretors presented a variety of lactobacilli, including several typically transient genera, suggesting possible host adaptation in secretors that does not seem to occur in non-secretors. Finally, homozygous secretors appeared to be more homogeneous than heterozygous secretors and non-secretors in terms of beta diversity, although the difference was not statistically significant.

➤ In Chapter 3, a method for accurate 16S rRNA gene-based bacterial identification was developed to be able to select only bifidobacteria and lactobacilli for further studies. The method is fast, especially the DNA extraction step (<25 min), inexpensive (<5€/sample) and can be applied to both Gram-negative and Gram-positive bacteria. Since it relies on sequencing of the (almost) complete 16S rRNA gene using MinION Nanopore sequencing technology, bacterial identification is also more accurate than other techniques based on short-read sequencing such as Illumina-based sequencing¹⁴². It is particularly advantageous for isolating bacteria for which media are not exclusively selective, which allowed targeted selection of the bacteria to be further investigated in this study. It also does not require colony

purification and allows the purity of the strains to be assessed at the data analysis stage.

➤ In Chapter 4 the WGS data generated by the developed method were analysed. A similar number of lactobacilli isolates was obtained from secretors and non-secretors, but bifidobacteria were significantly more often isolated in secretors than in non-secretors, and secretors had a more diverse bifidobacterial community. In terms of genomic analyses, we observed no significant differences according to secretor status in terms of fucose-specific genes and functional COG categories in lactobacilli, bifidobacteria and *B. longum* subsp. *longum* in particular. Our PCoA results from *B. longum* subsp. *longum* were largely explained by the donor variable, and there was no significant clustering of gene presence/absence data associated with secretor status or *FUT2* genotype. However, we observed that isolates from homozygous secretors were more homogeneous than bacteria from the other two groups.

➤ In Chapter 5 the non-linear relationship between body weight and health was explored^{241,242}, as well as how body weight is influenced by several different factors with an evolutionary background, including set-point weight theory²²³, population genetics²²⁴, transgenerational inheritance²⁵⁹ and the gut microbiome²⁶¹. These highlighted points result in an evolutionary mismatch between our physiology and modern weight loss goals. Consequently, dieting can be inefficient long-term and pose several health risks^{201,215}. Finally, we suggested focusing on health-promoting behaviours instead of appearance- and/or weight-based concerns through inclusive approaches such as HAES® and IE, which have been shown to lead to sustainable positive health outcomes^{202,285,286}.

6.2. Hypotheses testing

The aim of this project was to investigate the abundance, diversity and genomic differences of lactobacilli and bifidobacteria as function of the host secretor status. The initial hypotheses were studied as follows:

➤ The first hypothesis was that secretors would have a higher relative abundance of lactobacilli and bifidobacteria, presumably due to the ability of certain strains to use fucose as an adhesion site and/or energy source^{34,92}. The 16S rRNA gene amplicon sequencing data analysis confirmed the null hypothesis, meaning that there were no significant differences between secretors and non-secretors. However, the primers were not ideal for bifidobacteria¹¹⁶, so a future study with optimized primers could be warranted. The WGS data analysis confirmed the null hypothesis for lactobacilli, but a significant difference could be observed in the abundance of bifidobacteria, which were more frequently isolated from secretors. However, due to sampling bias, further studies would be needed to confirm this finding.

➤ The second hypothesis was that the diversity of species found in secretors would include more fucose-metabolising bacteria compared to non-secretors^{34,92}. In the 16S rRNA gene amplicon sequencing data analysis, there were not enough ASVs assigned to the species level to perform such an analysis. However, the different lactobacilli genera present in our samples were investigated and it could be observed that secretors had more autochthonous genera than non-secretors, who in turn presented more diverse genera. However, it is difficult to make considerations regarding differences between secretors and non-secretors in fucose-catabolizing bacteria, as this is a strain-dependent property^{52,91} and there was no strain-level and insufficient species-level information in this analysis. In the WGS data analysis, bifidobacteria were much more abundant in secretors, and perhaps as a result more

isolates of fucose-metabolizing bacteria were obtained from secretors, although further strain-level analyses would be needed to confirm this. Bifidobacteria were also more diverse in secretors. Interestingly, a different mix of lactobacilli species was observed for secretors and non-secretors, but both groups were quite diverse, which contradicts the results of the 16S rRNA gene amplicon sequencing data analysis, possibly due to the low number of isolates sequenced and the associated sampling bias. However, more *L. plantarum* subsp. *plantarum* isolates were obtained from non-secretors and more *L. mucosae* isolates (bacteria which can metabolise fucose⁵⁵) were obtained from secretors, which would have been expected from the 16S rRNA gene amplicon sequencing data analysis. Overall, there are some indications that there may be more fucose-metabolising bacteria in secretors, but the evidence is not clear. It would be interesting to assess these results with shotgun metagenomics and strain-level data to better assess this hypothesis.

➤ The third hypothesis was that there would be genomic variations according to the secretor status in these two taxa. Specifically, it was hypothesised that secretors would have more genes dedicated to fucose utilisation and/or have a higher relative frequency of functional categories associated with fucose catabolism. If the secretor status played an important role in species-level genomic variation, it could also be expected that a corresponding clustering of genomic data would be observed. In our WGS data analysis, we did not find significant differences according to secretor status with respect to fucose-metabolising genes and functional category assignment. However, when analysing the gene presence/absence data of *B. longum* subsp. *longum*, homozygous secretors tended to cluster closer together. This group also clustered closer together in the beta diversity analysis based on the 16S rRNA gene sequencing data, although this was not statistically significant. It would be interesting to explore these differences further with a larger sample size. The only significant

clustering variable for *B. longum* subsp. *longum* was the stool donor variable, demonstrating the importance of inter-individual differences in microbiome research¹⁹⁰.

Overall, it appears that the secretor status does not have a significant effect on gut lactobacilli and bifidobacteria, but more research would be needed to confirm this, for example by implementing the suggestions recommended in subsection 6.3.

6.3. Future research

The results of this study could be complemented by further investigations on different levels. Firstly, in terms of the metadata of this cohort, it would be interesting to look at the remaining parameters, namely BMI and YOB. These variables were incomplete for our cohort, but by excluding those donors for whom the information was not available, we could investigate their influence on our data. As we have seen in Chapter 5, BMI is not an ideal measure of individual health²⁴⁶, so the assessment of such analyses would have to be performed carefully. YOB would be particularly promising to test for bifidobacteria, as their composition and abundance change significantly throughout our lifetime⁶⁷. Another possibility would be to look at interactions between variables. For example, it has been shown that gut microbiota variations may depend on ABO blood groups in secretors but not in non-secretors¹¹⁹. It would also be interesting to take into consideration technical parameters, such as how long the faecal samples were frozen before processing, culturing batch and sequencing run. Additionally, it would be good to complete the metadata of this study with other parameters that have been shown to strongly influence the composition of the gut microbiome, i.e. drugs²⁹⁵, transit time²⁹⁶, stool consistency²⁹⁷ and time of stool collection, since the microbiome depends on the circadian rhythm²⁹⁸.

Secondly, there are analyses that can be performed on the data obtained in this study to complement its findings. For example, it would be interesting to also look at the results of the 16S rRNA gene amplicon sequencing data analysis disaggregated by other metadata variables. Regarding the WGS data, more analyses could be performed using Panaroo's output, such as looking at the core/accessory gene distribution per secretor status, the phylogeny of the isolates, or whole genome arrangements¹⁵⁸ as suggested previously. This output could also potentially be improved by using gene prediction pipelines such as Prodigal²⁹⁹, as several fucose-related genes belonging to lactobacilli and bifidobacteria have not yet been properly characterised. Panaroo's output could also be used to perform a GWAS¹⁹⁴ and analyse the potential association between microbial genetic variants and host secretor status. In addition, functional differences dependent on the secretor status could be tested with more descriptive databases than the COG categories, such as the KEGG pathway database³⁰⁰. It would also be interesting to carry out the analyses that we performed for *B. longum* subsp. *longum* for other species found in our project. In cases where the sample size would be too small, we could group isolates according to their phylogenetic relatedness. For example, it would be possible to group lactobacilli according to the new genera. In addition, it could be promising to look more systematically at the *FUT2* genotype, as it was seen that homozygous secretors tended to be more homogeneous as a group than heterozygous individuals, both in the beta diversity results of the 16S rRNA gene amplicon sequencing data analysis and in the *B. longum* subsp. *longum* gene presence/absence analysis. Although these results were not statistically significant, increasing the sample size to an even number of individuals for all groups would benefit such an analysis and allow us to investigate whether there are significant differences between homozygous and heterozygous secretors.

Thirdly, there are a few improvements that could be made to better the quality of the analyses. As mentioned above, the 16S rRNA gene amplicon sequencing data analysis would benefit from optimized primers, especially to properly capture bifidobacteria¹¹⁶. It would also be recommendable to examine the lactobacilli isolates that were not included due to contamination levels over 5% in order to recover as many as possible and increase the size of the isolate collection of this study. In addition, increasing the number of participating donors, especially homozygous secretors, would improve the power of the statistical tests.

Finally, there are several follow-up studies that could be performed. Shotgun metagenomics data are available for this cohort, and their analysis may allow gaining new insights by exploring SNP variations and potential resulting amino acid changes that may be dependent on host secretor status^{301,302}. This is particularly important in this case, as fucose catabolism mechanisms in lactobacilli and bifidobacteria are strain-dependent^{52,91} and shotgun metagenomics can provide strain-level resolution^{301,302}. As mentioned above, transcriptomic studies would also be promising when fucose-related genes are present independent of secretor status, as their expression might be inducible¹⁸⁷. Another way to focus on lactobacilli and bifidobacteria that catabolize fucose would be to develop a fucose-based medium to promote their growth. However, developing media for fastidious bacteria such as lactobacilli is a challenge in itself⁴⁴. Bifidobacteria do not have such extensive growth requirements, but only a limited number of isolates can grow in minimal media³⁰³. Additionally, many strains of these taxa have either only fucosidases or only the *fuc* operon, and may only thrive in the presence of fucose through symbiosis with other microbes^{56,57,89,91}. It may therefore be quite complicated to develop such a medium. Furthermore, as strain-specific mechanisms for fucose catabolism in bifidobacteria are more abundant in strains from infants⁹¹, it would be interesting to repeat this study in a

cohort of infants with varying *FUT2* genotype. Another idea would be to look at mucosal samples, as lactobacilli have been shown to have significantly different abundance in the mucosa depending on secretor status⁵¹, and the composition of bifidobacteria is different depending on whether they are isolated from stool or mucosal samples³⁰⁴. Overall, it is known that the faecal microbiota does not provide a complete picture of the gut microbiome^{305–307}, thus it would be ideal to complement our studies with mucosal samples. One last idea would be to repeat the analyses performed in this project once the fucose catabolic pathways of lactobacilli and bifidobacteria have been further described, as one of the main limitations in approaching this topic was the lack of proper characterisation of fucose-specific genes in lactobacilli and bifidobacteria.

6.4. Overall perspective

This project is one of many research studies that aim to find ways to tailor probiotic administration to specific host factors in the hope of improving their efficacy³⁷. Currently, probiotics have mostly mixed results³⁰⁸, with a few exceptions⁴² (e.g. *C. difficile*-induced diarrhoea¹⁹⁵), and ideally it would be possible to personalise their administration in order to optimise their effects. This is a challenging task given the diversity and uniqueness of each individual's gut microbiome, but it is a worthwhile endeavour given the several promising links between the microbiome and disease states, as well as its therapeutic potential^{309–311}. Furthermore, even if the results of a study are negative, it is still incredibly important to discard hypotheses and prevent non-evidence-based commercial exploration, which can often happen in unregulated fields³⁰⁸. Therefore, despite the current mixed results of probiotics, it is crucial to continue researching, not only to gain helpful insights for better therapies, but also to debunk possible misinformation.

This project focused on a particular human genotype and its possible influence on taxa with probiotic potential. Not all lactobacilli and bifidobacteria are considered probiotics, but several probiotic bacteria belong to these taxa³⁷. Probiotic properties are strain-specific³⁷, but it is not possible to culture only the potentially probiotic bacteria when stool samples are cultured. Fucose catabolic pathways are also strain-specific^{52,91}, but it would be very time-consuming or even impossible to develop a fucose-based culture medium for isolating fucose-utilising lactobacilli because of their fastidious growth requirements⁴⁴. Therefore, we proceeded with the almost “blind” culturing strategy adopted in this study and were able to isolate several lactobacilli and bifidobacteria isolates thanks to the developed method. Culturomics studies are relevant as a complement to non-culture-based studies because they allow to further characterise isolates and capture minor populations that would otherwise be below computational detection thresholds^{312,313}. We aimed to achieve a combination of these methods by analysing both 16S rRNA gene amplicon sequencing data from stool samples and WGS data from our cultured isolates. In the future, it would be interesting to study strain-level variation, as not only probiotic properties but also fucose catabolic pathways are strain-dependent.

Lactobacilli and bifidobacteria have co-evolved with their hosts over time ^{44,71–73}. However, in this study there was no clear evidence that gut lactobacilli and bifidobacteria co-evolved with their hosts specifically according to the human secretor status. There were only two possible indicators of such secretor status-based co-evolution. First, more autochthonous lactobacilli ASVs were found in secretors, possibly indicating some degree of host adaptation and maintenance of a more stable lactobacilli community in secretors⁴⁴. Second, bifidobacteria were isolated more often from secretors than from non-secretors, including some potential fucose utilizers. However, without strain-level analyses, growth studies in fucose and taking into

account sampling bias, it is not possible to attribute such differences to the secretor status with certainty, especially given that so many of the fucose-catabolizing bifidobacteria described in the literature have been isolated mostly from infants. Further research is therefore needed in this area.

Given the lack of robust evidence to support personalisation of probiotics according to host secretor status, no clinical recommendation can be made other than to exercise caution regarding secretor status-based probiotics. The medical principle of "first, do no harm" should guide clinicians, researchers and, hopefully, the general public in the use of emerging therapies that aim to improve gut health. Whether it is personalised probiotics or another fad diet (as discussed in Chapter 5), both can be highly profitable for businesses given the lack of adequate regulation and the simultaneous unmet medical needs of the population. By advocating for evidence-based medical recommendations, we can mitigate the risks associated with unproven interventions and prioritise the well-being of the population, which should be at the heart of research, clinical guidelines and public health initiatives.

References

1. Cani, P. D. Human gut microbiome: Hopes, threats and promises. *Gut* **67**, 1716–1725 (2018).
2. Turnbaugh, P. J. *et al.* The Human Microbiome Project. **449**, (2007).
3. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. 1–14 (2016) doi:10.1371/journal.pbio.1002533.
4. Oren, A. & Garrity, G. M. Valid publication of the names of forty- two phyla of prokaryotes. (2021) doi:10.1099/ijsem.0.005056.
5. Quigley, E. M. M. Gut Bacteria in Health and Disease. **9**, 560–569 (2013).
6. Inman, M. How Bacteria Turn Fiber into Food. *PLoS Biol* **9**, e1001227 (2011).
7. Li, H., He, J. & Jia, W. The influence of gut microbiota on drug metabolism and toxicity. *Expert Opin Drug Metab Toxicol* **12**, 31–40 (2016).
8. Ubeda, C., Djukovic, A. & Isaac, S. Roles of the intestinal microbiota in pathogen protection. *Clin Transl Immunology* **6**, (2017).
9. Ram, E. *et al.* Impact of the Gut Microbiota Balance on the Health – Disease Relationship : The Importance of Consuming Probiotics and Prebiotics. 1–28 (2021).
10. Kriss, M., Hazleton, K. Z., Nusbacher, N. M., Martin, C. G. & Lozupone, C. A. Low diversity gut microbiota dysbiosis: drivers, functional implications and recovery. *Curr Opin Microbiol* **44**, 34–40 (2018).
11. Gomaa, E. Z. Human gut microbiota / microbiome in health and diseases : a review. *Antonie Van Leeuwenhoek* **113**, 2019–2040 (2020).

12. Caparrós, E. *et al.* Dysbiotic microbiota interactions in Crohn's disease. *Gut Microbes* **13**, (2021).
13. Dahal, R. H., Kim, S., Kim, Y. K., Kim, E. S. & Kim, J. Insight into gut dysbiosis of patients with inflammatory bowel disease and ischemic colitis. *Front Microbiol* **14**, (2023).
14. Shen, Z.-H. *et al.* Relationship between intestinal microbiota and ulcerative colitis: Mechanisms and clinical application of probiotics and fecal microbiota transplantation. *World J Gastroenterol* **24**, 5–14 (2018).
15. Durack, J. & Lynch, S. V. The gut microbiome : Relationships with disease and opportunities for therapy. **216**, 20–40 (2018).
16. Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).
17. Bubier, J. A., Chesler, E. J. & Weinstock, G. M. Host genetic control of gut microbiome composition. *Mammalian Genome* **32**, 263–281 (2021).
18. Schneider, M., Al-Shareffi, E. & Haltiwanger, R. S. Biological functions of fucose in mammals. *Glycobiology* **27**, 601–618 (2017).
19. Becker, D. J. & Lowe, J. B. Fucose: Biosynthesis and biological function in mammals. *Glycobiology* **13**, (2003).
20. Garber, J. M., Hennet, T. & Szymanski, C. M. Significance of fucose in intestinal health and disease. *Mol Microbiol* **115**, 1086–1093 (2021).
21. Wiciński, M., Sawicka, E., Gębalski, J., Kubiak, K. & Malinowski, B. Human milk oligosaccharides: Health benefits, potential applications in infant formulas, and pharmacology. *Nutrients* **12**, (2020).

22. Arzamasov, A. A. & Osterman, A. L. Milk glycan metabolism by intestinal bifidobacteria: insights from comparative genomics. *Critical Reviews in Biochemistry and Molecular Biology* vol. 57 562–584 Preprint at <https://doi.org/10.1080/10409238.2023.2182272> (2022).
23. Taylor, S. L., McGuckin, M. A., Wesselingh, S. & Rogers, G. B. Infection's Sweet Tooth: How Glycans Mediate Infection and Disease Susceptibility. *Trends Microbiol* **26**, 92–101 (2018).
24. Rühlemann, M. C. *et al.* Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nat Genet* **53**, 147–155 (2021).
25. Azad, M. B., Wade, K. H. & Timpson, N. J. FUT2 secretor genotype and susceptibility to infections and chronic conditions in the ALSPAC cohort. *Wellcome Open Res* **3**, 65 (2018).
26. Kulkarni, D. G. & Venkatesh, D. Non-secretor status; a predisposing factor for vaginal candidiasis. *Indian J Physiol Pharmacol* **48**, 225–9 (2004).
27. Yang, P., Li, H.-L. & Wang, C.-Y. FUT2 Nonfunctional Variant: A “Missing Link” Between Genes and Environment in Type 1 Diabetes? *Diabetes* **60**, 2685–2687 (2011).
28. Hu, M. *et al.* Fucosyltransferase 2: A Genetic Risk Factor for Intestinal Diseases. *Front Microbiol* **13**, (2022).
29. Pickard, J. M. & Chervonsky, A. V. Intestinal Fucose as a Mediator of Host–Microbe Symbiosis. *The Journal of Immunology* **194**, 5588–5593 (2015).
30. Pickard, J. M. *et al.* Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness. *Nature* **514**, 638–641 (2014).

31. Koda, Y., Tachida, H., Soejima, M., Takenaka, O. & Kimura, H. Ancient origin of the null allele se428 of the human ABO-secretor locus (FUT2). *J Mol Evol* **50**, 243–248 (2000).
32. Johnson, O. L., Tobler, R., Schmidt, J. M. & Huber, C. D. Fluctuating selection and the determinants of genetic variation. *Trends in Genetics* **39**, 491–504 (2023).
33. Galeev, A. *et al.* The role of the blood group-related glycosyltransferases FUT2 and B4GALNT2 in susceptibility to infectious disease. *International Journal of Medical Microbiology* **311**, 151487 (2021).
34. Wacklin, P. *et al.* Faecal microbiota composition in adults is associated with the FUT2 gene determining the secretor status. *PLoS One* **9**, (2014).
35. Hill, C. *et al.* Expert consensus document: The international scientific association for probiotics and prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat Rev Gastroenterol Hepatol* **11**, 506–514 (2014).
36. Probiotics in food. <https://www.fao.org/3/a0512e/a0512e.pdf> (2006).
37. Butel, M. J. Probiotics, gut microbiota and health. *Med Mal Infect* **44**, 1–8 (2014).
38. Franz, C. M. A. P., Huch, M., Abriouel, H., Holzapfel, W. & Gálvez, A. Enterococci as probiotics and their implications in food safety. *Int J Food Microbiol* **151**, 125–140 (2011).
39. Martinović, A., Cocuzzi, R., Arioli, S. & Mora, D. *Streptococcus thermophilus*: To Survive, or Not to Survive the Gastrointestinal Tract, That Is the Question! *Nutrients* **12**, 2175 (2020).

40. Wang, Y. *et al.* Probiotic Potential of *Leuconostoc pseudomesenteroides* and *Lactobacillus* Strains Isolated From Yaks. *Front Microbiol* **9**, (2018).
41. Butel, M. J. Probiotics, gut microbiota and health. *Med Mal Infect* **44**, 1–8 (2014).
42. Rondanelli, M. *et al.* Using probiotics in clinical practice: Where are we now? A review of existing meta-analyses. *Gut Microbes* **8**, 521–543 (2017).
43. Mäkivuokko, H. *et al.* Isolation of bifidobacteria for blood group secretor status targeted personalised nutrition. *Microb Ecol Health Dis* **23**, (2012).
44. Duar, R. M. *et al.* Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiol Rev* **41**, S27–S48 (2017).
45. Walter, J., Britton, R. A. & Roos, S. Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. *Proceedings of the National Academy of Sciences* **108**, 4645–4652 (2011).
46. Frese, S. A. *et al.* The Evolution of Host Specialization in the Vertebrate Gut Symbiont *Lactobacillus reuteri*. *PLoS Genet* **7**, e1001314 (2011).
47. Zheng, J. *et al.* A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol* **70**, 2782–2858 (2020).
48. Rossi, M. *et al.* Mining metagenomic whole genome sequences revealed subdominant but constant *Lactobacillus* population in the human gut microbiota. *Environ Microbiol Rep* **8**, 399–406 (2016).
49. Heeney, D., Gareau, M. & Marco, M. Intestinal *Lactobacillus* in health and disease, a driver or just along for the ride? *Curr Opin Biotechnol* (2018).

50. Azad, M. A. K., Sarker, M., Li, T. & Yin, J. Probiotic Species in the Modulation of Gut Microbiota: An Overview. *Biomed Res Int* **2018**, (2018).
51. Rausch, P. *et al.* Colonic mucosa-associated microbiota is influenced by an interaction of crohn disease and FUT2 (Secretor) genotype. *Proc Natl Acad Sci U S A* **108**, 19030–19035 (2011).
52. Becerra, J. E., Yebra, M. J. & Monedero, V. An L-fucose operon in the probiotic *Lactobacillus rhamnosus* GG is involved in adaptation to gastrointestinal conditions. *Appl Environ Microbiol* **81**, 3880–3888 (2015).
53. Rodríguez-Díaz, J., Rubio-del-Campo, A. & Yebra, M. J. *Lactobacillus casei* ferments the N-Acetylglucosamine moiety of fucosyl- α -1,3-N-acetylglucosamine and excretes L-fucose. *Appl Environ Microbiol* **78**, 4613–4619 (2012).
54. Rodríguez-Díaz, J., Monedero, V. & Yebra, M. J. Utilization of natural fucosylated oligosaccharides by three novel α -L-fucosidases from a probiotic *Lactobacillus casei* strain. *Appl Environ Microbiol* **77**, 703–705 (2011).
55. Jia, Y. *et al.* Comparative genomics analysis of *Lactobacillus mucosae* from different niches. *Genes (Basel)* **11**, (2020).
56. Zúñiga, M., Monedero, V. & Yebra, M. J. Utilization of host-derived glycans by intestinal *Lactobacillus* and *Bifidobacterium* species. *Front Microbiol* **9**, (2018).
57. Nogacka, A. M., Cuesta, I., Gueimonde, M. & de los Reyes-Gavilán, C. G. 2-Fucosyllactose Metabolism by *Bifidobacteria* Promotes *Lactobacilli* Growth in Co-Culture. *Microorganisms* **11**, 2659 (2023).
58. Kononova, S., Litvinova, E., Vakhitov, T., Skalinskaya, M. & Sitkin, S. Acceptive immunity: The role of fucosylated glycans in human host–microbiome interactions. *Int J Mol Sci* **22**, (2021).

59. Nishiyama, K., Sugiyama, M. & Mukai, T. Adhesion properties of lactic acid bacteria on intestinal mucin. *Microorganisms* **4**, (2016).
60. Vernikos, G., Medini, D., Riley, D. R. & Tettelin, H. Ten years of pan-genome analyses. *Curr Opin Microbiol* **23**, 148–154 (2015).
61. Brockhurst, M. A. *et al.* The Ecology and Evolution of Pangenomes. *Current Biology* **29**, R1094–R1103 (2019).
62. Gong, Y., Li, Y., Liu, X., Ma, Y. & Jiang, L. A review of the pangenome: how it affects our understanding of genomic variation, selection and breeding in domestic animals? *J Anim Sci Biotechnol* **14**, 73 (2023).
63. Inglin, R. C., Meile, L. & Stevens, M. J. A. Clustering of Pan- and Core-genome of *Lactobacillus* provides Novel Evolutionary Insights for Differentiation. *BMC Genomics* **19**, 1–15 (2018).
64. Sun, Z. *et al.* Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nat Commun* **6**, (2015).
65. Zhang, Q. *et al.* Comparative genomics of *Lactobacillus crispatus* from the gut and vagina reveals genetic diversity and lifestyle adaptation. *Genes (Basel)* **11**, (2020).
66. O’Callaghan, A. & van Sinderen, D. Bifidobacteria and their role as members of the human gut microbiota. *Front Microbiol* **7**, (2016).
67. Arboleya, S., Watkins, C., Stanton, C. & Ross, R. P. Gut bifidobacteria populations in human health and aging. *Front Microbiol* **7**, 1–9 (2016).

68. Odamaki, T. *et al.* Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. *BMC Microbiol* **16**, 90 (2016).
69. Mattarelli, P. & Biavati, B. *Species in the Genus Bifidobacterium*. (2018). doi:10.1016/B978-0-12-805060-6/00002-8.
70. Mättö, J. *et al.* Genetic heterogeneity and functional properties of intestinal bifidobacteria. *J Appl Microbiol* **97**, 459–470 (2004).
71. Turrone, F., Milani, C., Van Sinderen, D. & Ventura, M. Bifidobacteria: Ecology and Coevolution With the Host. in *The Bifidobacteria and Related Organisms* 213–220 (Elsevier, 2018). doi:10.1016/B978-0-12-805060-6.00012-0.
72. Turrone, F. *et al.* Bifidobacteria and the infant gut: an example of co-evolution and natural selection. *Cellular and Molecular Life Sciences* **75**, 103–118 (2018).
73. Rodriguez, C. I. & Martiny, J. B. H. Evolutionary relationships among bifidobacteria and their hosts and environments. *BMC Genomics* **21**, 1–12 (2020).
74. O'Connell Motherway, M. *et al.* A Bifidobacterial pilus-associated protein promotes colonic epithelial proliferation. *Mol Microbiol* **111**, 287–301 (2019).
75. Garrido, D., Kim, J. H., German, J. B., Raybould, H. E. & Mills, D. A. Oligosaccharide binding proteins from bifidobacterium longum subsp. infantis reveal a preference for host glycans. *PLoS One* **6**, (2011).
76. Liu, J., Zheng, M., Zhang, C. & Xu, D. Amide resonance in the catalysis of 1,2- α -l-fucosidase from bifidobacterium bifidum. *Journal of Physical Chemistry B* **117**, 10080–10092 (2013).

77. Turróni, F. *et al.* Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proc Natl Acad Sci U S A* **107**, 19514–19519 (2010).
78. Nagae, M. *et al.* Structural basis of the catalytic reaction mechanism of novel 1,2- α -L-fucosidase from *Bifidobacterium bifidum*. *Journal of Biological Chemistry* **282**, 18497–18509 (2007).
79. Katayama, T. *et al.* Molecular cloning and characterization of *Bifidobacterium bifidum* 1,2- α -L-fucosidase (AfcA), a novel inverting glycosidase (glycoside hydrolase family 95). *J Bacteriol* **186**, 4885–4893 (2004).
80. Ashida, H. *et al.* Two distinct α -L-fucosidases from *Bifidobacterium bifidum* are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates. *Glycobiology* **19**, 1010–1017 (2009).
81. Centanni, M., Ferguson, S. A., Sims, I. M., Biswas, A. & Tannock, G. W. *Bifidobacterium bifidum* ATCC 15696 and *bifidobacterium breve* 24b metabolic interaction based on 2-O-Fucosyl-lactose studied in steady-state cultures in a fed-batch-style chemostat. *Appl Environ Microbiol* **85**, (2019).
82. Garrido, D. *et al.* A novel gene cluster allows preferential utilization of fucosylated milk oligosaccharides in *Bifidobacterium longum* subsp. *longum* SC596. *Sci Rep* **6**, (2016).
83. Bunesova, V., Lacroix, C. & Schwab, C. Fucosyllactose and L-fucose utilization of infant *Bifidobacterium longum* and *Bifidobacterium kashiwanohense*. *BMC Microbiol* **16**, 1–12 (2016).

84. Díaz, R., Torres-Miranda, A., Orellana, G. & Garrido, D. Comparative genomic analysis of novel bifidobacterium longum subsp. Longum strains reveals functional divergence in the human gut microbiota. *Microorganisms* **9**, (2021).
85. Higgins, M. A. & Ryan, K. S. Generating a fucose permease deletion mutant in Bifidobacterium longum subspecies infantis ATCC 15697. *Anaerobe* **68**, (2021).
86. Sela, D. A. *et al.* Bifidobacterium longum subsp. infantis ATCC 15697 α -fucosidases are active on fucosylated human milk oligosaccharides. *Appl Environ Microbiol* **78**, 795–803 (2012).
87. Dedon, L. R., Özcan, E., Rani, A. & Sela, D. A. Bifidobacterium infantis Metabolizes 2'Fucosyllactose-Derived and Free Fucose Through a Common Catabolic Pathway Resulting in 1,2-Propanediol Secretion. *Front Nutr* **7**, (2020).
88. Zabel, B. *et al.* Novel Genes and Metabolite Trends in Bifidobacterium longum subsp. infantis Bi-26 Metabolism of Human Milk Oligosaccharide 2'-fucosyllactose. *Sci Rep* **9**, (2019).
89. Arzamasov, A. A. & Osterman, A. L. Milk glycan metabolism by intestinal bifidobacteria: insights from comparative genomics. *Crit Rev Biochem Mol Biol* **57**, 562–584 (2022).
90. Hobbs, M. E. *et al.* Discovery of an L-Fucono-1,5-lactonase from cog3618 of the Amidohydrolase Superfamily. *Biochemistry* **52**, 239–253 (2013).
91. Bottacini, F., Van Sinderen, D. & Ventura, M. Omics of bifidobacteria: Research and insights into their health-promoting activities. *Biochemical Journal* **474**, 4137–4152 (2017).
92. Wacklin, P. *et al.* Secretor genotype (FUT2 gene) is strongly associated with the composition of bifidobacteria in the human intestine. *PLoS One* **6**, (2011).

93. Davenport, E. R. *et al.* ABO antigen and secretor statuses are not associated with gut microbiota composition in 1,500 twins. *BMC Genomics* **17**, (2016).
94. Turpin, W. *et al.* FUT2 genotype and secretory status are not associated with fecal microbial composition and inferred function in healthy subjects. *Gut Microbes* **9**, 357–368 (2018).
95. Bottacini, F. *et al.* Comparative genomics of the genus *Bifidobacterium*. *Microbiology (N Y)* **156**, 3243–3254 (2010).
96. Freitas, A. C. & Hill, J. E. *Bifidobacteria* isolated from vaginal and gut microbiomes are indistinguishable by comparative genomics. *PLoS One* **13**, 1–16 (2018).
97. Milani, C. *et al.* Genomics of the genus *Bifidobacterium* reveals species-specific adaptation to the glycan-rich gut environment. *Appl Environ Microbiol* **82**, 980–991 (2016).
98. Milani, C. *et al.* Genomic encyclopedia of type strains of the genus *Bifidobacterium*. *Appl Environ Microbiol* **80**, 6290–6302 (2014).
99. Borges, A. S. G. *et al.* Fast Identification Method for Screening Bacteria from Faecal Samples Using Oxford Nanopore Technologies MinION Sequencing. *Curr Microbiol* **80**, 101 (2023).
100. Blutspendedienst des Bayerischen Roten Kreuzes. Verteilung der Blutgruppen. <https://www.blutspendedienst.com/blutspende/blut-blutgruppen/verteilung-der-blutgruppen> (2024).
101. Trautmann, T. *et al.* The Impact of Oral Sodium Chloride Supplementation on Thrive and the Intestinal Microbiome in Neonates With Small Bowel Ostomies: A Prospective Cohort Study. *Front Immunol* **11**, (2020).

102. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).
103. Malte Rühlemann. DADA2 16S workflow with AR. https://github.com/mruehlemann/ikmb_amplicon_processing/blob/master/dada2_16S_workflow_with_AR.R (2019).
104. Benjamin Callahan. A DADA2 workflow for Big Data. <https://benjjneb.github.io/dada2/bigdata.html> (2022).
105. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, (2013).
106. RStudioTeam. RStudio: Integrated Development Environment for R. <http://www.rstudio.com/> (2022).
107. Rosenqvist, T. *et al.* Succession of bacterial biofilm communities following removal of chloramine from a full-scale drinking water distribution system. *NPJ Clean Water* **6**, (2023).
108. Alboukadel Kassambara. ggpubr: ‘ggplot2’ Based Publication Ready Plots (R package version 0.6.0). <https://rpkgs.datanovia.com/ggpubr/> (2023).
109. Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biometrics Bulletin* **1**, 80 (1945).
110. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).

111. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
112. Anderson, M. J. Permutational Multivariate Analysis of Variance (PERMANOVA). in *Wiley StatsRef: Statistics Reference Online* 1–15 (Wiley, 2017). doi:10.1002/9781118445112.stat07841.
113. Yoon, S. H. *et al.* Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* **67**, 1613–1617 (2017).
114. Jethwani, P. & Grover, K. Gut Microbiota in Health and Diseases – A Review. *Int J Curr Microbiol Appl Sci* **8**, 1586–1599 (2019).
115. Watanabe, M. *et al.* Identification of a new adhesin-like protein from *Lactobacillus mucosae* ME-340 with specific affinity to the human blood group A and B antigens. *J Appl Microbiol* **109**, 927–935 (2010).
116. Abellan-Schneyder, I. *et al.* Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. *mSphere* **6**, (2021).
117. Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A. & Korpela, K. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLoS One* **15**, (2020).
118. Walter J. Discovering the molecular foundations of *Lactobacillus* autochthony in the gastrointestinal tract. *Japanese Journal of Lactic Acid Bacteria* **19**, 9–20 (2008).
119. Gampa, A., Engen, P. A., Shobar, R. & Mutlu, E. A. Relationships between gastrointestinal microbiota and blood group antigens. *Physiol Genomics* **49**, 473–483 (2017).

120. Zou, Y. The calculation of β -diversity for different sample sizes. *Biodiversity Science* **29**, 790–797 (2021).
121. Pasolli, E. *et al.* Large-scale genome-wide analysis links lactic acid bacteria from food with the gut microbiome. *Nat Commun* **11**, (2020).
122. Lagier, J. C. *et al.* Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol* **1**, 16203 (2016).
123. Bonnet, M., Lagier, J. C., Raoult, D. & Khelaifia, S. Bacterial culture through selective and non-selective conditions: the evolution of culture media in clinical microbiology. *New Microbes New Infect* **34**, 100622 (2020).
124. De Man, J. C., Rogosa, M. & Sharpe, M. S. A Medium for the Cultivation of Lactobacilli. *J Appl Microbiol* **23**, 130–135 (1960).
125. Ashraf, R. & Shah, N. P. Selective and differential enumerations of *Lactobacillus delbrueckii* subsp. *bulgaricus*, *Streptococcus thermophilus*, *Lactobacillus acidophilus*, *Lactobacillus casei* and *Bifidobacterium* spp. in yoghurt—a review. *Int J Food Microbiol* **149**, 194–208 (2011).
126. Vinderola, C. G. & Reinheimer, J. A. Culture media for the enumeration of *Bifidobacterium bifidum* and *Lactobacillus acidophilus* in the presence of yoghurt bacteria. *Int Dairy J* **9**, 497–505 (1999).
127. Livingston, S. J., Kominos, S. D. & Yee, R. B. New medium for selection and presumptive identification of the *Bacteroides fragilis* group. *J Clin Microbiol* **7**, 448–453 (1978).
128. Diakite, A. *et al.* Optimization and standardization of the culturomics technique for human microbiome exploration. *Sci Rep* **10**, 9674 (2020).

129. Bacic, M. K. & Smith, C. J. Laboratory Maintenance and Cultivation of Bacteroides Species. *Curr Protoc Microbiol* 1–24 (2008) doi:10.1002/9780471729259.mc13c01s9.Laboratory.
130. Jamal, M. *et al.* Ultra-High Efficient Colony PCR for High Throughput Screening of Bacterial Genes. *Indian J Microbiol* **57**, 365–369 (2017).
131. Frank, J. A. *et al.* Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* **74**, 2461–2470 (2008).
132. Krych, L. *et al.* DNA enrichment and tagmentation method for species-level identification and strain-level differentiation using ON-rep-seq. *Commun Biol* **2**, 369 (2019).
133. Rodríguez-Pérez, H., Ciuffreda, L. & Flores, C. NanoCLUST: a species-level analysis of 16S rRNA nanopore sequencing data. *Bioinformatics* **37**, 1600–1601 (2021).
134. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
135. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455–477 (2012).
136. BBDuk Guide - U.S. Department of Energy Joint Genome Institute. <https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbdduk-guide/>.
137. Lee, I., Ouk Kim, Y., Park, S. C. & Chun, J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* **66**, 1100–1103 (2016).

138. Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L. & Goker, M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res* **50**, D801–D807 (2022).
139. Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P. & Goker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, 60 (2013).
140. Meier-Kolthoff, J. P. & Göker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* **10**, (2019).
141. Zafar, H. & Saier, M. H. Gut Bacteroides species in health and disease. *Gut Microbes* **13**, 1–20 (2021).
142. Matsuo, Y. *et al.* Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION nanopore sequencing confers species-level resolution. *BMC Microbiol* **21**, 35 (2021).
143. Clarridge J. E., 3rd. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* **17**, 840–862 (2004).
144. Santos, A., van Aerle, R., Barrientos, L. & Martinez-Urtaza, J. Computational methods for 16S metabarcoding studies using Nanopore sequencing data. *Comput Struct Biotechnol J* **18**, 296–305 (2020).
145. Faron, M. L., Buchan, B. W. & Ledebøer, N. A. Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry for Use with Positive Blood Cultures: Methodology, Performance, and Optimization. *J Clin Microbiol* **55**, 3328–3338 (2017).

146. Stackebrandt, E. & Goebel, B. M. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *Int J Syst Evol Microbiol* **44**, 846–849 (1994).
147. Hasman, H. *et al.* Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. *J Clin Microbiol* **52**, 139–146 (2014).
148. Larsen, M. V. *et al.* Benchmarking of methods for genomic taxonomy. *J Clin Microbiol* **52**, 1529–1539 (2014).
149. Clausen, P. T. L. C., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* **19**, (2018).
150. Deneke, C. *et al.* Species-specific quality control, assembly and contamination detection in microbial isolate sequences with aquamis. *Genes (Basel)* **12**, (2021).
151. Low, A. J., Koziol, A. G., Manninger, P. A., Blais, B. & Carrillo, C. D. ConFindr: Rapid detection of intraspecies and cross-species contamination in bacterial whole-genome sequence data. *PeerJ* **2019**, (2019).
152. Wattam, A. R. *et al.* Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res* **45**, D535–D542 (2017).
153. Olson, R. D. *et al.* Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res* **51**, D678–D689 (2023).

154. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043–1055 (2015).
155. Cornet, L. *et al.* Consensus assessment of the contamination level of publicly available cyanobacterial genomes. *PLoS One* **13**, (2018).
156. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
157. Wick R R; Holt K E. Assembly Dereplicator. <https://github.com/rrwick/Assembly-Dereplicator> (2023).
158. Tonkin-Hill, G. *et al.* Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* **21**, (2020).
159. Geneious Prime 2023.2.1. <https://www.geneious.com>.
160. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* **38**, 5825–5829 (2021).
161. Tatusov, R. L. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**, 33–36 (2000).
162. Botelho, J. *et al.* Phylogroup-specific variation shapes the clustering of antimicrobial resistance genes and defence systems across regions of genome plasticity in *Pseudomonas aeruginosa*. *EBioMedicine* **90**, 104532 (2023).
163. Oksanen J *et al.* vegan: Community Ecology Package. <https://CRAN.R-project.org/package=vegan> (2022).

164. Anderson, M. J. Distance-Based Tests for Homogeneity of Multivariate Dispersions. *Biometrics* **62**, 245–253 (2006).
165. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
166. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. 1–14 (2016) doi:10.1371/journal.pbio.1002533.
167. Xiao, Y., Zhao, J., Zhang, H., Zhai, Q. & Chen, W. Mining genome traits that determine the different gut colonization potential of *Lactobacillus* and *Bifidobacterium* species. *Microb Genom* **7**, (2021).
168. Rajput, A. *et al.* Pangenome analysis reveals the genetic basis for taxonomic classification of the Lactobacillaceae family. *Food Microbiol* **115**, 104334 (2023).
169. Koduru, L. *et al.* Systematic evaluation of genome-wide metabolic landscapes in lactic acid bacteria reveals diet- and strain-specific probiotic idiosyncrasies. *Cell Rep* **41**, 111735 (2022).
170. Lukjancenko, O., Ussery, D. W. & Wessenaar, T. M. Comparative Genomics of *Bifidobacterium*, *Lactobacillus* and Related Probiotic Genera. *Microb Ecol* **63**, 651–673 (2012).
171. Satti, M., Tanizawa, Y., Endo, A. & Arita, M. Comparative analysis of probiotic bacteria based on a new definition of core genome. *J Bioinform Comput Biol* **16**, 1840012 (2018).
172. Pokusaeva, K., Fitzgerald, G. F. & van Sinderen, D. Carbohydrate metabolism in *Bifidobacteria*. *Genes Nutr* **6**, 285–306 (2011).

173. Sugihara, J., Sun, L., Yan, N. & Kaback, H. R. Dynamics of the L-fucose/H⁺ symporter revealed by fluorescence spectroscopy. *Proceedings of the National Academy of Sciences* **109**, 14847–14851 (2012).
174. Zhu, Y. & Lin, E. C. C. An evolvant of *Escherichia coli* that employs the L-fucose pathway also for growth on L-galactose and D-arabinose. *J Mol Evol* **23**, 259–266 (1986).
175. Sun, Z. *et al.* Comparative Genomic Analysis of 45 Type Strains of the Genus *Bifidobacterium*: A Snapshot of Its Genetic Diversity and Evolution. *PLoS One* **10**, e0117912 (2015).
176. Sharma, V., Mobeen, F. & Prakash, T. Exploration of Survival Traits, Probiotic Determinants, Host Interactions, and Functional Evolution of *Bifidobacterial* Genomes Using Comparative Genomics. *Genes (Basel)* **9**, 477 (2018).
177. Freitas, A. C. & Hill, J. E. *Bifidobacteria* isolated from vaginal and gut microbiomes are indistinguishable by comparative genomics. *PLoS One* **13**, (2018).
178. Milani, C. *et al.* *Bifidobacteria* exhibit social behavior through carbohydrate resource sharing in the gut. *Sci Rep* **5**, 15782 (2015).
179. González, R., Klaassens, E. S., Malinen, E., de Vos, W. M. & Vaughan, E. E. Differential Transcriptional Response of *Bifidobacterium longum* to Human Milk, Formula Milk, and Galactooligosaccharide. *Appl Environ Microbiol* **74**, 4686–4694 (2008).
180. Lugli, G. A. *et al.* Investigation of the Evolutionary Development of the Genus *Bifidobacterium* by Comparative Genomics. *Appl Environ Microbiol* **80**, 6383–6394 (2014).

181. Nunoura, N., Ohdan, K., Yano, T., Yamamoto, K. & Kumagai, H. Purification and Characterization of β -D-Glucosidase (β -D-Fucosidase) from *Bifidobacterium breve* clb Acclimated to Cellobiose. *Biosci Biotechnol Biochem* **60**, 188–193 (1996).
182. Zhurina, D. *et al.* Complete Genome Sequence of *Bifidobacterium bifidum* S17. *J Bacteriol* **193**, 301–302 (2011).
183. Yuen, R. & Schachter, H. L-Fucose Metabolism in Mammals. I. Pork Liver L-Fuconate Hydro-lyase. *Can J Biochem* **50**, 798–806 (1972).
184. Ventura, M. *et al.* The *Bifidobacterium dentium* Bd1 Genome Sequence Reflects Its Genetic Adaptation to the Human Oral Cavity. *PLoS Genet* **5**, e1000785 (2009).
185. Engevik, M. A. *et al.* *Bifidobacterium dentium* Fortifies the Intestinal Mucus Layer via Autophagy and Calcium Signaling Pathways. *mBio* **10**, (2019).
186. Klein, A. *et al.* The broad diversity of neutral and sialylated oligosaccharides derived from human salivary mucins. *Biochemistry* **31**, 6152–6165 (1992).
187. Yang, H. *et al.* ABO genotype alters the gut microbiota by regulating GalNAc levels in pigs. *Nature* **606**, 358–367 (2022).
188. Tokeson, J. P., Garges, S. & Adhya, S. Further inducibility of a constitutive system: ultra-induction of the gal operon. *J Bacteriol* **173**, 2319–2327 (1991).
189. Albert, K., Rani, A. & Sela, D. A. Comparative Pangenomics of the Mammalian Gut Commensal *Bifidobacterium longum*. *Microorganisms* **8**, 7 (2019).
190. Costello, E. K. *et al.* Bacterial Community Variation in Human Body Habitats Across Space and Time. *Science* (1979) **326**, 1694–1697 (2009).

191. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* (1979) **352**, 560–564 (2016).
192. O'Callaghan, A., Bottacini, F., O'Connell Motherway, M. & van Sinderen, D. Pangenome analysis of *Bifidobacterium longum* and site-directed mutagenesis through by-pass of restriction-modification systems. *BMC Genomics* **16**, 832 (2015).
193. Odamaki, T. *et al.* Genomic diversity and distribution of *Bifidobacterium longum* subsp. *longum* across the human lifespan. *Sci Rep* **8**, 85 (2018).
194. Power, R. A., Parkhill, J. & de Oliveira, T. Microbial genome-wide association studies: lessons from human GWAS. *Nat Rev Genet* **18**, 41–50 (2017).
195. Al Sharaby, A., Abugoukh, T. M., Ahmed, W., Ahmed, S. & Elshaikh, A. O. Do Probiotics Prevent *Clostridium difficile*-Associated Diarrhea? *Cureus* (2022) doi:10.7759/cureus.27624.
196. Senok, A. C., Ismaeel, A. Y. & Botta, G. A. Probiotics: facts and myths. *Clinical Microbiology and Infection* **11**, 958–966 (2005).
197. Wacklin, P., Mättö, J. & Mäkituokko, H. Use of blood group status II. (2009).
198. Gundogdu, A. *et al.* Culture-independent assessment of probiotic supplement consistency in commercially available probiotic supplements. *Food Biosci* **53**, 102709 (2023).
199. Rijkers, G. T. *et al.* Guidance for Substantiating the Evidence for Beneficial Effects of Probiotics: Current Status and Recommendations for Future Research1–3. *J Nutr* **140**, 671S-676S (2010).

200. *Global Spa & Wellness Economy Monitor*. https://globalwellnessinstitute.org/wp-content/uploads/2018/06/global-spa-and-wellness-economy-monitor-2014_fix4182016.pdf (2014).
201. Nordmo, M., Danielsen, Y. S. & Nordmo, M. The challenge of keeping it off, a descriptive systematic review of high-quality, follow-up studies of obesity treatments. *Obesity Reviews* **21**, 12949 (2020).
202. Bacon, L. & Aphramor, L. Weight Science: Evaluating the Evidence for a Paradigm Shift. *Nutr J* **10**, 9 (2011).
203. Sinclair, D. A. Toward a unified theory of caloric restriction and longevity regulation. *Mech Ageing Dev* **126**, 987–1002 (2005).
204. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186**, 243–278 (2023).
205. de Cabo, R. & Mattson, M. P. Effects of Intermittent Fasting on Health, Aging, and Disease. *New England Journal of Medicine* **381**, 2541–2551 (2019).
206. Speakman, J. R. & Mitchell, S. E. Caloric restriction. *Mol Aspects Med* **32**, 159–221 (2011).
207. Harrison, C. *Anti-Diet - Reclaim Your Time, Money, Well-Being, and Happiness Through Intuitive Eating*. (Little, Brown Spark, 2019).
208. Santos, I., Sniehotta, F. F., Marques, M. M., Carraça, E. V. & Teixeira, P. J. Prevalence of personal weight control attempts in adults: a systematic review and meta-analysis. *Obesity Reviews* **18**, 32–50 (2017).
209. BCC Research. *Global Weight Loss Products and Services Market 2021-2026*. (2021).

210. Galmiche, M., Déchelotte, P., Lambert, G. & Tavalacci, M. P. Prevalence of eating disorders over the 2000–2018 period: a systematic literature review. *Am J Clin Nutr* **109**, 1402–1413 (2019).
211. Linardon, J. & Mitchell, S. Rigid dietary control, flexible dietary control, and intuitive eating: Evidence for their differential relationship to disordered eating and body image concerns. *Eat Behav* **26**, 16–22 (2017).
212. Puhl, R. M. & Heuer, C. A. Obesity Stigma: Important Considerations for Public Health. *Am J Public Health* **100**, 1019–1028 (2010).
213. Gracia-Arnaiz, M. Fat bodies and thin bodies. Cultural, biomedical and market discourses on obesity. *Appetite* **55**, 219–225 (2010).
214. Cohen, R., Newton-John, T. & Slater, A. The case for body positivity on social media: Perspectives on current advances and future directions. *J Health Psychol* **26**, 2365–2373 (2021).
215. Mann, T. *et al.* Medicare’s search for effective obesity treatments: Diets are not the answer. *American Psychologist* **62**, 220–233 (2007).
216. Feller, S. *et al.* What distinguishes weight loss maintainers of the German Weight Control Registry from the general population? *Obesity* **23**, 1112–1118 (2015).
217. Tomiyama, A. J., Ahlstrom, B. & Mann, T. Long-term Effects of Dieting: Is Weight Loss Related to Health? *Soc Personal Psychol Compass* **7**, 861–877 (2013).
218. Rhee, E.-J. Weight Cycling and Its Cardiometabolic Impact. *J Obes Metab Syndr* **26**, 237–242 (2017).

219. Leibel, R. L., Rosenbaum, M. & Hirsch, J. Changes in Energy Expenditure Resulting from Altered Body Weight. *New England Journal of Medicine* **332**, 621–628 (1995).
220. Speakman, J. R. The evolution of body fatness: trading off disease and predation risk. *Journal of Experimental Biology* **221**, 167254 (2018).
221. Wells, J. C. K. The evolution of human adiposity and obesity: where did it all go wrong? *Dis Model Mech* **5**, 595–607 (2012).
222. Chakravarthy, M. V. & Booth, F. W. Eating, exercise, and “thrifty” genotypes: connecting the dots toward an evolutionary understanding of modern chronic diseases. *J Appl Physiol* **96**, 3–10 (2004).
223. Farias, M. M., Cuevas, A. M. & Rodriguez, F. Set-Point Theory and Obesity. *Metab Syndr Relat Disord* **9**, 85–89 (2011).
224. Qasim, A. *et al.* On the origin of obesity: identifying the biological, environmental and cultural drivers of genetic risk among human populations. *Obesity Reviews* **19**, 121–149 (2018).
225. Trexler, E. T., Smith-Ryan, A. E. & Norton, L. E. Metabolic adaptation to weight loss: implications for the athlete. *J Int Soc Sports Nutr* **11**, (2014).
226. Polidori, D., Sanghvi, A., Seeley, R. J. & Hall, K. D. How Strongly Does Appetite Counter Weight Loss? Quantification of the Feedback Control of Human Energy Intake. *Obesity* **24**, 2289–2295 (2016).
227. Chao, A. M., Jastreboff, A. M., White, M. A., Grilo, C. M. & Sinha, R. Stress, cortisol, and other appetite-related hormones: Prospective prediction of 6-month changes in food cravings and weight. *Obesity* **25**, 713–720 (2017).

228. Rosenbaum, M., Hirsch, J., Gallagher, D. A. & Leibel, R. L. Long-term persistence of adaptive thermogenesis in subjects who have maintained a reduced body weight. *Am J Clin Nutr* **88**, 906–912 (2008).
229. Nymo, S. *et al.* Investigation of the long-term sustainability of changes in appetite after weight loss. *Int J Obes* **42**, 1489–1499 (2018).
230. Dulloo, A. G. Physiology of weight regain: Lessons from the classic Minnesota Starvation Experiment on human body composition regulation. *Obesity Reviews* **22**, 1–19 (2021).
231. Kalm, L. M. & Semba, R. D. They Starved So That Others Be Better Fed: Remembering Ancel Keys and the Minnesota Experiment. *J Nutr* **135**, 1347–1352 (2005).
232. Dirks, A. J. & Leeuwenburgh, C. Caloric restriction in humans: Potential pitfalls and health concerns. *Mech Ageing Dev* **127**, 1–7 (2006).
233. Elias, S. G. *et al.* Menstruation during and after caloric restriction: The 1944–1945 Dutch famine. *Fertil Steril* **88**, 1101–1107 (2007).
234. Weinsier, R. L. *et al.* Do adaptive changes in metabolic rate favor weight regain in weight-reduced individuals? An examination of the set-point theory. *Am J Clin Nutr* **72**, 1088–1094 (2000).
235. Birsoy, K., Festuccia, W. T. & Laplante, M. A comparative perspective on lipid storage in animals. *J Cell Sci* **126**, 1541–1552 (2013).
236. Coelho, M., Oliveira, T. & Fernandes, R. State of the art paper Biochemistry of adipose tissue: an endocrine organ. *Archives of Medical Science* **2**, 191–200 (2013).

237. Bays, H. E. *et al.* Pathogenic potential of adipose tissue and metabolic consequences of adipocyte hypertrophy and increased visceral adiposity. *Expert Rev Cardiovasc Ther* **6**, 343–368 (2008).
238. Aronne, L. J., Nelinson, D. S. & Lillo, J. L. Obesity as a disease state: A new paradigm for diagnosis and treatment. *Clin Cornerstone* **9**, 9–29 (2009).
239. Yamauchi, Y. *et al.* Paradoxical association between body mass index and in-hospital mortality in elderly patients with chronic obstructive pulmonary disease in Japan. *Int J Chron Obstruct Pulmon Dis* **9**, 1337 (2014).
240. Petrelli, F. *et al.* Association of Obesity With Survival Outcomes in Patients With Cancer. *JAMA Netw Open* **4**, e213520 (2021).
241. Sturgiss, E., Jay, M., Campbell-Scherer, D. & van Weel, C. Challenging assumptions in obesity research. *BMJ* **359**, j5303 (2017).
242. Bhaskaran, K., dos-Santos-Silva, I., Leon, D. A., Douglas, I. J. & Smeeth, L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3·6 million adults in the UK. *Lancet Diabetes Endocrinol* **6**, 944–953 (2018).
243. Di Angelantonio, E. *et al.* Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *The Lancet* **388**, 776–786 (2016).
244. Matheson, E. M., King, D. E. & Everett, C. J. Healthy Lifestyle Habits and Mortality in Overweight and Obese Individuals. *The Journal of the American Board of Family Medicine* **25**, 9–15 (2012).

245. Stanner, S. A. *et al.* Does malnutrition in utero determine diabetes and coronary heart disease in adulthood? Results from the Leningrad siege study, a cross sectional study. *BMJ* **315**, 1342–1348 (1997).
246. Nuttall, F. Q. Body Mass Index. *Nutr Today* **50**, 117–128 (2015).
247. Booske, B. C., Athens, J. K., Kindig, D. A. & Remington, P. L. *Different Perspectives for Assigning Weights to Determinants of Health*. (2010).
248. Scott, C., Sutherland, J. & Taylor, A. *Affordability of the UK' s Eatwell Guide*. (2018).
249. Institute of Medicine. *Weight Management*. (National Academies Press, Washington, D.C., 2004). doi:10.17226/10783.
250. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
251. Nazare, J.-A. *et al.* Ethnic influences on the relations between abdominal subcutaneous and visceral adiposity, liver fat, and cardiometabolic risk profile: the International Study of Prediction of Intra-Abdominal Adiposity and Its Relationship With Cardiometabolic Risk/Intra-Abdominal Adiposity. *Am J Clin Nutr* **96**, 714–726 (2012).
252. Carroll, J. F. *et al.* Visceral Fat, Waist Circumference, and BMI: Impact of Race/ethnicity. *Obesity* **16**, 600–607 (2008).
253. Whyte, H. E. & Findlay, S. Dieting in adolescence. *Paediatr Child Health* **9**, 487–491 (2004).

254. Donelson, J. M., Salinas, S., Munday, P. L. & Shama, L. N. S. Transgenerational plasticity and climate change experiments: Where do we go from here? *Glob Chang Biol* **24**, 13–34 (2018).
255. Skinner, M. K., Manikkam, M. & Guerrero-Bosagna, C. Epigenetic transgenerational actions of environmental factors in disease etiology. *Trends in Endocrinology & Metabolism* **21**, 214–222 (2010).
256. Duncan, E. J., Gluckman, P. D. & Dearden, P. K. Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? *J Exp Zool B Mol Dev Evol* **322**, 208–220 (2014).
257. Gluckman, P. D., Hanson, M. A. & Low, F. M. Evolutionary and developmental mismatches are consequences of adaptive developmental plasticity in humans and have implications for later disease risk. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20180109 (2019).
258. Stein, Z. *Famine and Human Development: The Dutch Hunger Winter of 1944-1945*. (Oxford University Press, 1975).
259. Rey, O., Danchin, E., Mirouze, M., Loot, C. & Blanchet, S. Adaptation to Global Change: A Transposable Element–Epigenetics Perspective. *Trends Ecol Evol* **31**, 514–526 (2016).
260. Moeller, A. H. *et al.* Cospeciation of gut microbiota with hominids. *Science (1979)* **353**, 380–382 (2016).
261. Lee, C. J., Sears, C. L. & Maruthur, N. Gut microbiome and its role in obesity and insulin resistance. *Ann N Y Acad Sci* **1461**, 37–52 (2020).
262. Carmody, R. N., Sarkar, A. & Reese, A. T. Gut microbiota through an evolutionary lens. *Science (1979)* **372**, 462–463 (2021).

263. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
264. Boekhorst, J. *et al.* Stool energy density is positively correlated to intestinal transit time and related to microbial enterotypes. *Microbiome* **10**, 223 (2022).
265. Razbekova, M. *et al.* Genetic factors associated with obesity risks in a Kazakhstani population. *BMJ Nutr Prev Health* **4**, 90–101 (2021).
266. de Luis, D., Izaola, O. & Primo, D. Effect of the variant rs602662 of FUT2 gene on anthropometric and metabolic parameters in a Caucasian obese population. *Eur Rev Med Pharmacol Sci* **26**, 7169–7175 (2022).
267. Komorniak, N. *et al.* Can the FUT 2 Gene Variant Have an Effect on the Body Weight of Patients Undergoing Bariatric Surgery?—Preliminary, Exploratory Study. *Nutrients* **12**, 2621 (2020).
268. Bäckhed, F., Ley, R. E., Sonnenburg, J. L. & Gordon, J. I. The human intestinal microbiota and its relationship to energy balance. *Scand J Food Nutr* **50**, 121–123 (2006).
269. Hanage, W. P. Microbiology: Microbiome science needs a healthy dose of scepticism. *Nature* **512**, 247–248 (2014).
270. Slashinski, M. J., McCurdy, S. A., Achenbaum, L. S., Whitney, S. N. & McGuire, A. L. “Snake-oil,” “quack medicine,” and “industrially cultured organisms:” biovalue and the commercialization of human microbiome research. *BMC Med Ethics* **13**, 28 (2012).
271. Faith, J. J. *et al.* The Long-Term Stability of the Human Gut Microbiota. *Science* (1979) **341**, (2013).

272. Yadav, M. & Chauhan, N. S. Microbiome therapeutics: exploring the present scenario and challenges. *Gastroenterol Rep (Oxf)* **10**, (2022).
273. Ilan, Y. Why targeting the microbiome is not so successful: can randomness overcome the adaptation that occurs following gut manipulation? *Clin Exp Gastroenterol* **Volume 12**, 209–217 (2019).
274. Leeming, E. R., Johnson, A. J., Spector, T. D. & Le Roy, C. I. Effect of Diet on the Gut Microbiota: Rethinking Intervention Duration. *Nutrients* **11**, 2862 (2019).
275. Heiman, M. L. & Greenway, F. L. A healthy gastrointestinal microbiome is dependent on dietary diversity. *Mol Metab* **5**, 317–320 (2016).
276. Reddel, S., Putignani, L. & Del Chierico, F. The Impact of Low-FODMAPs, Gluten-Free, and Ketogenic Diets on Gut Microbiota Modulation in Pathological Conditions. *Nutrients* **11**, 373 (2019).
277. Moeller, A. H. *et al.* Rapid changes in the gut microbiome during human evolution. *Proceedings of the National Academy of Sciences* **111**, 16431–16435 (2014).
278. Doms, S., Hermes, B.-M. & Baines, J. F. Evolutionary Perspectives on the Human Gut Microbiome. in *The Gut Microbiome in Health and Disease* 67–78 (Springer International Publishing, Cham, 2018). doi:10.1007/978-3-319-90545-7_5.
279. Moeller, A. H. The shrinking human gut microbiome. *Curr Opin Microbiol* **38**, 30–35 (2017).
280. Penney, T. L. & Kirk, S. F. L. The Health at Every Size Paradigm and Obesity: Missing Empirical Evidence May Help Push the Reframing Obesity Debate Forward. *Am J Public Health* **105**, e38–e42 (2015).

281. Tribole, E. Intuitive Eating: Research Update. *Sports, Cardiovascular, and Wellness Nutrition Pulse* **36**, (2017).
282. Diehl, J. A., Heard, D., Lockhart, S. & Main, D. S. Access in the Food Environment: A Health Equity Approach Reveals Unequal Opportunity. *J Plan Educ Res* **40**, 69–81 (2020).
283. Feinman, R. D. *et al.* Dietary carbohydrate restriction as the first approach in diabetes management: Critical review and evidence base. *Nutrition* **31**, 1–13 (2015).
284. Pigneur, B. & Ruemmele, F. M. Nutritional interventions for the treatment of IBD: current evidence and controversies. *Therap Adv Gastroenterol* **12**, 175628481989053 (2019).
285. Schaefer, J. T. & Magnuson, A. B. A Review of Interventions that Promote Eating by Internal Cues. *J Acad Nutr Diet* **114**, 734–760 (2014).
286. Tylka, T. L. *et al.* The Weight-Inclusive versus Weight-Normative Approach to Health: Evaluating the Evidence for Prioritizing Well-Being over Weight Loss. *J Obes* **2014**, 1–18 (2014).
287. Bacon, L., Stern, J. S., Van Loan, M. D. & Keim, N. L. Size Acceptance and Intuitive Eating Improve Health for Obese, Female Chronic Dieters. *J Am Diet Assoc* **105**, 929–936 (2005).
288. Bacon, L. *et al.* Evaluating a ‘non-diet’ wellness intervention for improvement of metabolic fitness, psychological well-being and eating and activity behaviors. *Int J Obes* **26**, 854–865 (2002).

289. Mensinger, J. L., Calogero, R. M., Stranges, S. & Tylka, T. L. A weight-neutral versus weight-loss approach for health promotion in women with high BMI: A randomized-controlled trial. *Appetite* **105**, 364–374 (2016).
290. Fuentes Artilles, R. *et al.* Mindful eating and common diet programs lower body weight similarly: Systematic review and meta-analysis. *Obesity Reviews* **20**, 1619–1627 (2019).
291. Hazzard, V. M. *et al.* Intuitive eating longitudinally predicts better psychological health and lower use of disordered eating behaviors: findings from EAT 2010–2018. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity* **26**, 287–294 (2021).
292. Pujols, Y., Meston, C. M. & Seal, B. N. The Association Between Sexual Satisfaction and Body Image in Women. *J Sex Med* **7**, 905–916 (2010).
293. Huhmann, K. Menses Requires Energy: A Review of How Disordered Eating, Excessive Exercise, and High Stress Lead to Menstrual Irregularities. *Clin Ther* **42**, 401–407 (2020).
294. Creanza, N., Kolodny, O. & Feldman, M. W. Cultural evolutionary theory: How culture evolves and why it matters. *Proceedings of the National Academy of Sciences* **114**, 7782–7789 (2017).
295. Vich Vila, A. *et al.* Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. *Nat Commun* **11**, 362 (2020).
296. Procházková, N. *et al.* Advancing human gut microbiota research by considering gut transit time. *Gut* **72**, 180–191 (2023).

297. Vandeputte, D. *et al.* Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut* **65**, 57–62 (2016).
298. Frazier, K. & Chang, E. B. Intersection of the Gut Microbiome and Circadian Rhythms in Metabolism. *Trends in Endocrinology & Metabolism* **31**, 25–36 (2020).
299. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
300. Kanehisa, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32**, 277D – 280 (2004).
301. Costea, P. I. *et al.* metaSNV: A tool for metagenomic strain level analysis. *PLoS One* **12**, 1–9 (2017).
302. Nayfach, S., Rodriguez-Mueller, B., Garud, N. & Pollard, K. S. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Res* **26**, 1612–1625 (2016).
303. Poch, M. & Bezkorovainy, A. Growth-Enhancing Supplements for Various Species of the Genus *Bifidobacterium*. *J Dairy Sci* **71**, 3214–3221 (1988).
304. Turrone, F. *et al.* Exploring the Diversity of the *Bifidobacterial* Population in the Human Intestinal Tract. *Appl Environ Microbiol* **75**, 1534–1545 (2009).
305. Ingala, M. R. *et al.* Comparing Microbiome Sampling Methods in a Wild Mammal: Fecal and Intestinal Samples Record Different Signals of Host Ecology, Evolution. *Front Microbiol* **9**, (2018).

306. Ahn, J.-S. *et al.* Fecal Microbiome Does Not Represent Whole Gut Microbiome. *Cell Microbiol* **2023**, 1–14 (2023).
307. Durbán, A. *et al.* Assessing Gut Microbial Diversity from Feces and Rectal Mucosa. *Microb Ecol* **61**, 123–133 (2011).
308. Suez, J., Zmora, N., Segal, E. & Elinav, E. The pros, cons, and many unknowns of probiotics. *Nat Med* **25**, 716–729 (2019).
309. Singh, T. P. & Natraj, B. H. Next-generation probiotics: a promising approach towards designing personalized medicine. *Crit Rev Microbiol* **47**, 479–498 (2021).
310. Kim, B.-S., Jeon, Y.-S. & Chun, J. Current Status and Future Promise of the Human Microbiome. *Pediatr Gastroenterol Hepatol Nutr* **16**, 71 (2013).
311. Thaiss, C. A. & Elinav, E. The remedy within: will the microbiome fulfill its therapeutic promise? *J Mol Med* **95**, 1021–1027 (2017).
312. Abdallah, R. A. *et al.* The impact of culturomics on taxonomy in clinical microbiology. *Antonie Van Leeuwenhoek* **110**, 1327–1337 (2017).
313. Lagier, J.-C. *et al.* Microbial culturomics: paradigm shift in the human gut microbiome study. *Clinical Microbiology and Infection* **18**, 1185–1193 (2012).