Human gastrointestinal microbiota modulation with dietary prebiotics – an intimate interaction throughout life

Klaudyna Aneta Borewicz

Thesis committee

Promotor
Prof. Dr H. Smidt
Personal chair at the Laboratory of Microbiology
Wageningen University & Research

Other members
Prof. Dr V. Fogliano, Wageningen University & Research
Dr Elaine E. Vaughan, Sensus, Roosendaal, The Netherlands
Dr R. Satokari, University of Helsinki, Finland
Prof. Dr K. Venema, Maastricht University, The Netherlands

This research was conducted under the auspices of the Graduate School VLAG
(Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences)

Human gastrointestinal microbiota modulation with dietary prebiotics – an intimate interaction throughout life

Klaudyna Aneta Borewicz

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 13 June 2018
at 11 a.m. in the Aula.

Klaudyna Aneta Borewicz
Human gastrointestinal microbiota modulation with dietary prebiotics – an intimate interaction throughout life,
240 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)
With references, with summary in English

ISBN: 978-94-6343-277-1
DOI: https://doi.org/10.18174/448525

Table of contents

Chapter 1

General Introduction and Outline of the Thesis

"everything humans need to survive is intimately coupled with the activities of microbes"

Caroline Harwood

The world which is the most familiar to all of us is the one we can see with our eyes. However, beside this highly familiar visible realm, there exists another, very old yet obscure world, which we are now beginning to discover and understand with the help of modern technology. This is the invisible and fascinating world of microorganisms.

Unlocking the secrets of the microbial world

Before I can take you on a journey into the microbial world, it is necessary to mention briefly the current methods without which the research presented in this thesis would not be possible. These methods provided us with a window through which we could "see" and study microbial ecosystems [1]. They allowed us not only to identify and discover new microbial species, but also to study microbial activities and functions within the complex networks of interactions [2].

In the past, the identification and quantification of microbes was done using culture based approaches, which provided a very limited view, as the majority of microbes could not, and still cannot be grown in isolation or outside of their natural habitats [2]. Later, a broad range of molecular methods were designed to help address this limitation. These methods included, for example, the reverse transcription PCR (RT-PCR), quantitative PCR (qPCR), cloning and sequencing, Temperature Gradient Gel Electrophoresis (TGGE) and Denaturing Gradient Gel Electrophoresis (DGGE) and targeted either the 16S ribosomal RNA (rRNA) or its encoding gene. Even though these techniques allowed us to study microbial communities without the need for culturing, they were rather slow, labour intensive and often limited in their scope of the information which they could provide [1].

A second revolution in the molecular (micro)biology started in 2005 when the so-called Next Generation Sequencing (NGS) technologies began to emerge, allowing to produce large amounts of DNA sequence data from multiple samples in parallel at a very high-throughput and depth of analysis [3]. Over the years multiple sequencing chemistries and technologies have been developed for NGS, and nowadays the Illumina system with its HiSeq and MiSeq platforms have become the most popular in the studies on microbial ecosystems [3, 4]. The NGS sequencing process allows various types of DNA templates to be sequenced and the template choice depends on the scientific questions that are being addressed. Commonly used templates include fragmented total (meta)genomic DNA extracted from environmental samples or pure or mixed cultures, PCR amplicons of a specific gene (e.g. a variable region within the 16S rRNA gene), or cDNA reverse transcribed from RNA. The sequencing process starts with the ligation of linker and/or adapter sequences to the DNA template [3]. The resulting library is then amplified, and during amplification the incorporation of each nucleotide is monitored and noted. As a result, many millions of short nucleotide reads are recorded in parallel

producing large amounts of data (hundreds of gigabases (GB) per run) within a period of time that is as little as a few hours to less than three days [3, 5].

When PCR amplicons are used as a template, PCR primers can be barcoded to allow mixtures of different samples to be sequenced during a single run, and then sorted out based on each barcode during data processing [6]. This approach reduces the sequencing cost while maximizing the amount of information produced during a single run, and is often used in microbiota research allowing multiple samples to be analysed simultaneously.

The NGS methods are irreplaceable in modern microbiology; they enable sequencing genomes of individual organisms or entire microbial communities (metagenomics) and they can be used to collect gene expression data when mRNA is used as an initial template (metatranscriptomics - RNA-seq) [4]. Furthermore, the high-throughput approaches also provide a model that can be applied in other fields helping the development of other "omics" methods. These methods can be used to study microbial communities function and activity, for example through the large-scale studies of proteins (proteomics) and metabolic responses (metabolomics). All these methods hinge on the NGS sequencing, while also incorporating more traditional enzymatic and biochemical assays and chromatography techniques. Together, these "omics" approaches provide us with a comprehensive window to study and better understand the invisible microbial world around us.

We live in a microbial world

The very first living organisms on Earth were microscopic, single-celled, nucleus-lacking microbes, called prokaryotes [7]. The prokaryotes, which we can divide into two main groups, the bacteria and archaea, have evolved about four billion years ago, preceding all other microbial forms, such as fungi, viruses and protozoans, and all of the multicellular organisms, which appeared only about one million years ago [7]. Thus, the microbial world dominated our planet for much of its life, and even now, the diversity and numbers of microorganisms found on Earth largely exceed that of all other organisms found in the tree of life [7].

Microbial life occupies nearly all niches on Earth, including the most extreme environments, such as polar ice, hot springs or acidic lakes and rivers. Even though most microbes are too small to be seen with a naked eye, they are omnipresent and intimately tied with our own visible world. Microorganisms can be found everywhere around us - they are in the air we breathe, the water we drink, and in the food we eat. In fact, microbes are an inseparable part of our existence from the moment of birth, until long after we die. Microbial communities also thrive on the surface and inside the bodies of all living organisms, and all animals and plants harbour their own, highly diverse microbial ecosystems. Human bodies, too are covered with highly specialized assemblages of microbes, *our microbiota*, that live associated with all external surfaces (e.g., the skin, the oral cavity, the respiratory tract and the gastrointestinal tract), and also inside of our bodies (e.g., mammary glands). In fact, we could even say that we are more microbial than we are human, as it is estimated that the numbers of microbial cells in our bodies might be equal to, or possibly outnumber our own cells [8, 9].

Even though historically microbes have had a bad reputation for causing infections and diseases, the majority of them are non-pathogenic and their presence is essential to our normal development, health and well-being [7]. As our knowledge about microbial ecology is rapidly increasing, we are now beginning to understand the complexity of these microbial ecosystems, the range of interactions that exist between the members of the microbiota and between the microbiota and its host, and how crucial the microbial world is for our own health and survival.

This thesis will take you on a journey into the invisible microbial world that inhabits the human gastrointestinal (GI) tract. The work presented here comprises a review chapter and five research chapters, and can be divided into three main sections. The first section offers a literature review on the microbiota of the GI tract in adults and the changes in the microbial communities, which have been associated with selected diseases in humans. The second section presents our research findings on the early life development of the GI tract microbiota with particular focus on the role of infant feeding in the first weeks of life. The third section presents our work concerning the effects of novel prebiotic fibers on microbial community structure and function *in vitro*.

Section 1: Our microbiota and its role in health

From the moment of birth our bodies become an open and diverse habitat available for the microbial colonisers to settle in. Each body part has a unique set of characteristics that influence which microbial species can establish themselves and flourish. In the human body the largest microbial populations exists in the GI tract. In fact it is estimated that 95% of all our bacteria reside in the GI tract, with over a thousand different species [10] and a total combined bacterial cell weight of 0.2-2 kg [2, 9, 11]. However, besides bacteria, the GI tract is also home to members of other microbial groups: fungi, archaea, protozoa and viruses, which collectively form the *GI tract microbiota.* Once established, the healthy adult GI tract microbiota forms a stable ecosystem, with the population levels and species compositions in each individual remaining relatively constant over time [12]. This fact is due to *colonization resistance* (also known as competitive exclusion principle), which is one of the ways that microbiota confers health benefits to its host by protecting it from infections by pathogenic invaders [12, 13]. Aside from this protective function, the GI microbial ecosystem is necessary for digestion of food and production of beneficial metabolites, such as short chain fatty acids (SCFA), vitamins (e.g. B12, B5, K), but also breakdown of toxic compounds, proper development of the intestinal wall structure and maintaining intestinal wall integrity [1, 10]. In addition, the GI tract microbiota plays an important role in training the immune system and maintaining its balanced function [14].

In recent years the human GI tract microbiota, in particular its bacterial fraction, has been a subject of extensive research. The main objective has been to develop a better understanding of the link between one's microbiota and his or her health. Up to date, numerous adverse health conditions, both physical and mental, have been linked to microbiota disturbances, and the list is still rapidly expanding (Figure 1).

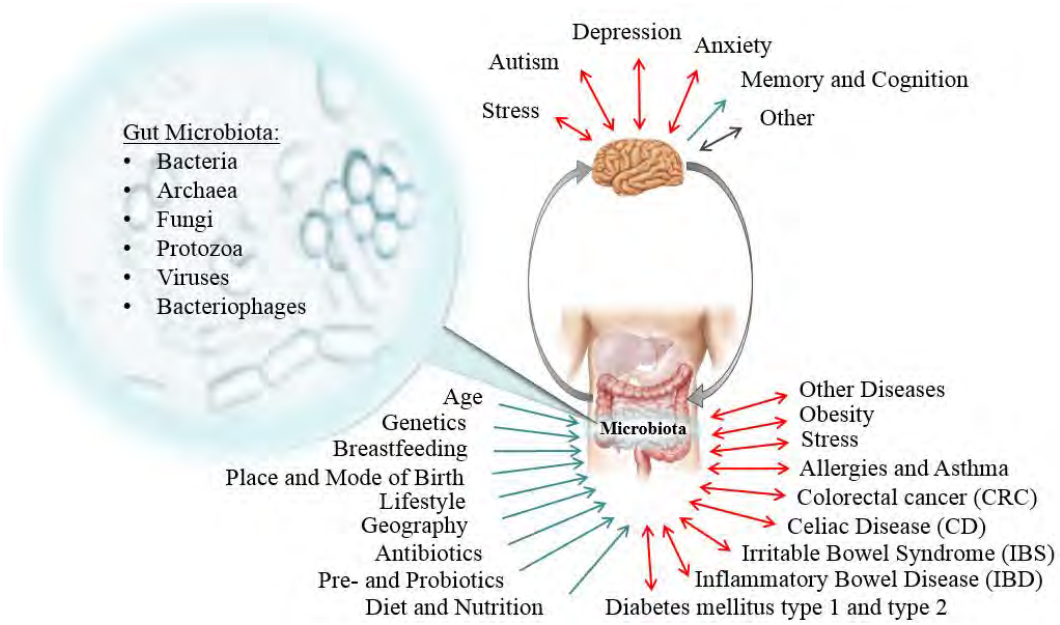


Figure 1. Summary of selected factors and diseases associated with changes in GI tract microbiota structure and function.

In **Chapter Two** of this thesis I provide a literature review presenting our current state of knowledge about microbial communities living in different regions of the adult human GI tract. This review was published in early 2016 in "The Human Microbiome Handbook" edited by Jason Tetro and Emma Allen-Vercoe.

As the application of "omics" methods in microbiological research resulted in a bloom of microbiota related projects worldwide, it also led to nearly an exponential increase in the number of microbiome related scientific publications, as well as increased interest in this topic among industry, health officials and the general public. The main vision behind "The Human Microbiome Handbook" was to present current scientific findings on human GI tract microbiota and associated diseases in an easy and approachable way. The chapter presented in this thesis discusses the microbiota composition and function in healthy adults, and the current status of understanding how disturbances in the GI tract microbiota, also referred to as *dysbioses*, might be linked to various diseases, including GI diseases, obesity, oral diseases, and other adverse health conditions.

Section 2. The microbiota in early life

The early life is a very important period of time not only for our own development, but also the development of the microbial ecosystems that inhabit our body [15]. Many factors, such as environmental exposures, host genetics, or diseases contribute to the shaping of microbial community structure and function, which in turn affects our well-being and health, both during infancy and beyond [1, 16]. During birth we acquire our first microbial residents through contact with our mothers, the medical staff, the environment of a hospital, and/or our home. In the beginning the bacterial communities that occupy various body sites are very

similar, but within days, the microbiota adapts and specializes to thrive at specific body sites [17]. Every time an infant visits a new place or meets a different family member, it receives an invisible welcome gift – a sample of that place's or person's microbiota [15]. Very soon our bodies become home to a unique set of commensal microorganisms, of which most will remain with us for a lifetime.

One of the body sites that undergoes a rapid microbial colonisation in early life is the GI tract [16]. The establishment of the GI tract microbiota is a stepwise succession process, with a general pattern of the initial colonisation with facultative anaerobes (*Escherichia coli*, enterococci and streptococci), followed by obligate anaerobes (*Bacteroides, Clostridium*, and *Bifidobacterium*) [18] eventually leading to a microbial community that is dominated mainly by members of the genus *Bifidobacterium* [7, 15, 19]. The route through which infants acquire this characteristic, bifidobacteria dominated microbial profile is believed to be through breastfeeding [20].

Breastfeeding is a unique trait of all mammals. It has provided a way for the mammalian hosts to co-evolve together with the members of the microbial world and it plays a crucial role in directing a proper microbial colonisation of the GI tract during infancy. In breastfed infants, breastmilk is the sole source of nourishment during the first few months of life. Besides providing the most optimal nutrition for the growing infant, such as lactose, fatty acids, proteins and micronutrients (e.g. nucleotides, vitamins and minerals) [20], it also is a source of microbes [21] and bioactive components [15, 16, 22, 23] including milk glycans, such as free human milk oligosaccharides (HMOs). The health benefits of breastmilk have been summarized in Table 1 [15].

Table 1. Summary of breastmilk benefits in initial microbial colonisation of an infant GI tract

Probiotic	Beneficial bacteria sampled from maternal GI tract and breastmilk
Prebiotic	Human Milk Oligosaccharides (HMOs)
Innate immunity	Antimicrobial peptides, lactoferrin, lysozyme
	Soluble CD14 that recognizes microbial lipopolysaccharides (LPS)
	Soluble Toll Like Receptors (TLR)
Adaptive immunity	Secretory IgA specific to pathogens encountered in maternal GI tract
	Glycans that decoy cell adhesion molecules

HMOs are believed to function as natural *prebiotics,* as they are not accessible to digestion by infant enzymes, but rather are degraded by specific groups of intestinal bacteria [24]. A prebiotic is "a substrate that is selectively utilized by specific microorganisms, conferring a health benefit to the host" [25]. As a result, HMOs facilitate the development and function of a highly specialized microbial ecosystem in the colon of an infant, the so-called milk-oriented microbiota (MOM) [23]. With over 200 different HMO types found in human breastmilk [20, 22], and high variability in the composition of HMOs between mothers and across different lactation stages [26-28], it is likely that these differences could influence the microbiota composition and the colonisation dynamics in the infant GI tract. Testing this hypothesis was one of the objectives of the work presented in the second part of this thesis.

As already mentioned, breastmilk provides the necessary and early link that evolved to guide the development of the early GI tract microbial ecosystem in human infants initiating the cross-talk and leading to the establishment of the peaceful, life-long relationship between a person and its microbes. However, in most modern societies where the fast paced life style and easy access to infant formulas are a norm, mothers often choose to reduce or completely abandon breastfeeding in favour of more convenient use of formulas. In addition, in many cases where breastfeeding is not possible due to a range of different physiological and/or psychosocial reasons, formula feeding is essential [29]. As a result, the breastfeeding-associated human-microbial link is no longer present. Only in recent decades we began to understand how the feeding mode relates to differences in the infant GI tract microbiota, and to recognise possible health consequences of formula feeding on human health, both in infancy and beyond [15]. As the proper development of the immune system is believed to be highly dependent on microbial stimulation, numerous autoimmune disorders are now believed to have their origin in an aberrant development of the GI tract microbiota during early life [1]. For example, previous studies suggested that formula feeding not only could increase the risk of nutritional deficiencies, developing metabolic syndrome and obesity, but also could be an underlying cause of asthma, atopy, coeliac disease, diabetes and other diseases [30-32] (Figure 1).

A multitude of earlier studies, both culture and molecular technique based, confirmed that faeces of breastfed infants are inhabited by higher levels of bifidobacteria, as compared to faeces of formula fed infants [1, 20, 33-38]. With the growing evidence of health benefits of breastfeeding, the main goal of formula producers has been to develop products which could better mimic breastmilk with regard to its nutritional properties and the modulatory prebiotic effect on the infant GI tract microbiota [13, 39, 40]. Thus, today's formulas which lack natural breastmilk HMOs are often fortified with other prebiotics, such as short chain galacto-oligosaccharides (scGOS) and long chain fructo-oligosaccharides (lcFOS) to mimic the bifidogenic effect of the HMOs [24, 41-44]. The basic structures of HMOs, GOS and FOS are shown in Figure 2.

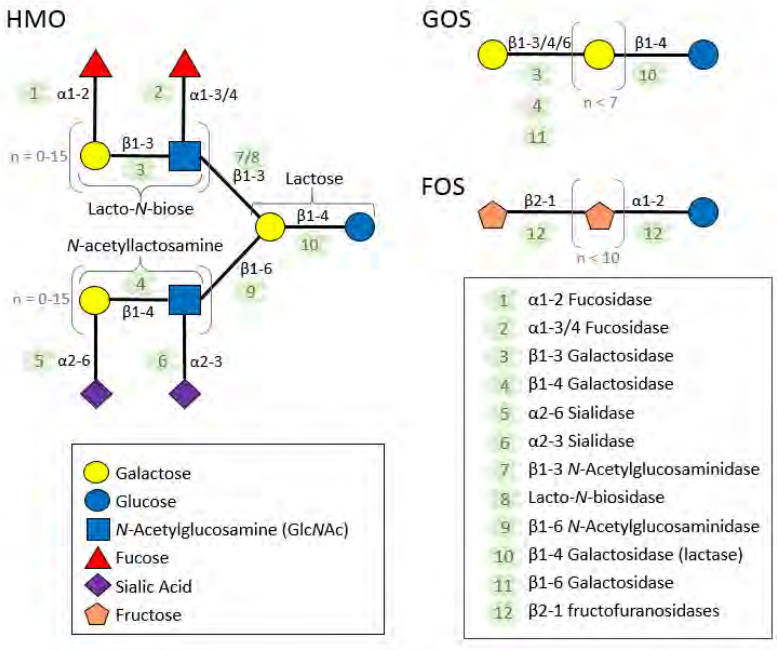


Figure 2. Representative structures of HMO, GOS, and FOS showing their monosaccharide composition, linkage types and potential glycolytic enzymes acting on each linkage type. The schematic structures can be elongated by repeating units in brackets by values indicated with n (adapted from [41]).

The functional alternatives, such as GOS and FOS, are non-branched and do not possess fucose or sialic acid moieties, and thus they lack the structural diversity and the complexity found in HMOs [41]. As the health mediating and functional properties of oligosaccharides largely depend on this structural diversity, the effects of prebiotic fortified formulas need to be taken into consideration. This calls for studies on microbiota composition, patterns of bacterial colonisation, and the immune development in relation to use of prebiotic fortified infant formulas. In particular, there is a need for well-designed long term nutritional and clinical intervention studies that would offer clear answers on the potential long term health risks and benefits of modulating GI tract microbiota during infancy.

The second section of my thesis includes three separate research chapters summarising our findings on the development of the GI microbiota in healthy human infants. In **Chapter Three** we compared the microbiota composition and colonisation patterns of breast- and formula fed infants receiving either the traditional, or modern commercial formulas fortified with GOS and/or FOS. In **Chapter Four** and **Chapter Five** I summarized our work investigating the link between specific breastmilk HMOs and the faecal microbiota composition in breastfed infants, as well as identified microbial networks and key bacterial taxa involved in the intestinal degradation of the most predominant HMOs found in breastmilk.

Section 3: The search for novel prebiotics

The first two sections of this thesis highlight the importance of the GI tract microbiota in human health and investigate how prebiotics could modulate the development and the composition of the microbiota in the infant GI tract. As our understanding of the relationship between human health and GI tract microbiota is rapidly increasing, there is a growing demand for developing novel types of prebiotics with defined functional properties, that would allow us to manipulate the GI tract microbiota to achieve specific health outcomes.

One of the key functions of prebiotics is their ability to induce specific changes in the composition and/or activity of the GI tract microbiota, mainly the *probiotic bacteria*, to benefit host's health [25, 45]. Notably, the concept of probiotic bacteria playing an important role in maintaining health was first defined in the context of healthy breastfed infants whose faeces were highly populated with bifidobacteria and/or lactic acid bacteria [46, 47]. Thus, traditionally, the prebiotic potential has been evaluated mainly based on the substrate's ability to selectively stimulate the growth of these two microbial groups [1, 25, 45].

Isomalto/malto-polysaccharides (IMMPs) are a novel type of indigestible α-glycans derived from starch [48]. The prebiotic potential of IMMPs lays in the presence of α-(1→6) glycosidic linkages, which can comprise more than 90% of all linkages present in the IMMP molecules. The α-(1→6)- rich segments are also found in isomalto-oligosaccharides (IMOs)

and in dextran, both of which are known for their prebiotic properties and their ability to stimulate growth of lactobacilli or bifidobacteria *in vitro* and *in vivo* [49-51]. Therefore, based on the structural similarity between IMMPs, IMOs and dextran, we hypothesized that IMMPs would offer a similar prebiotic effect. To test this hypothesis we used an *in vitro* batch fermentation system with faeces of healthy adults as microbial inoculum (Figure 3).

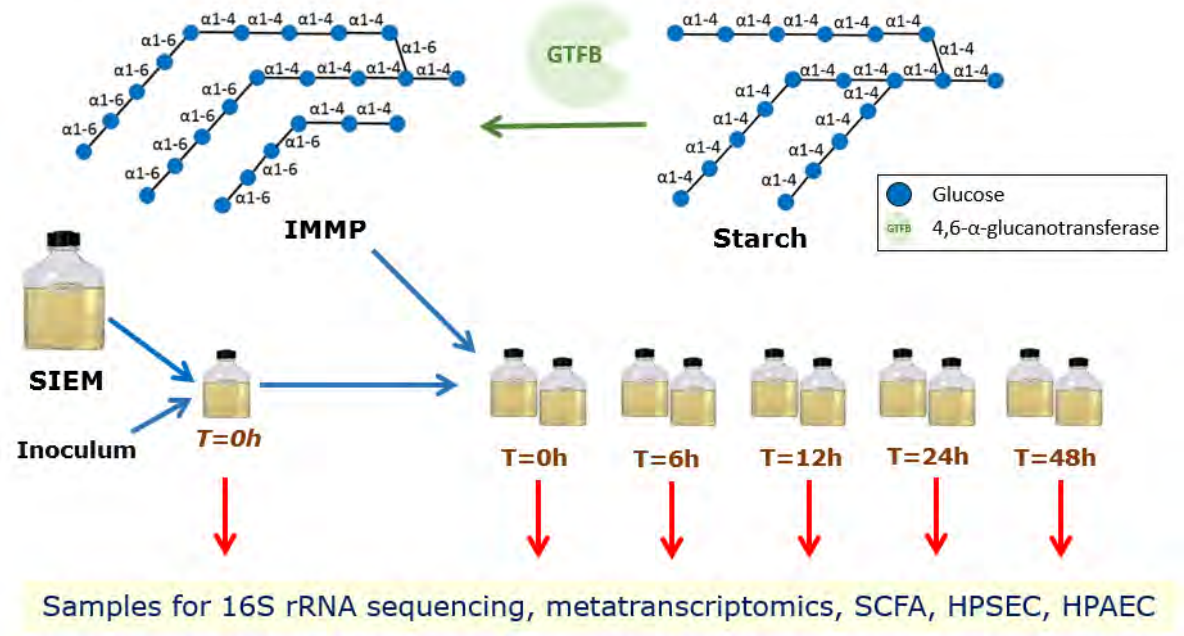


Figure 3. Synthesis of IMMPs and *in vitro* fermentation setup. IMMPs are produced by enzymatic treatment of starch with 4,6-α-glucanotransferase (GTFB) from *Lactobacillus reuteri* 121 [48]. The ratio of α-(1→6) and α-(1→4) glycosidic linkages can vary. Degradation of IMMPs during fermentation was monitored chromatographically. SIEM – Small Intestinal Efflux Medium; Inoculum – microbiota source from human faeces; HPSEC – High Performance Size Exclusion Chromatography; HPAEC - High Performance Anion Exchange Chromatography

We analysed the process of degradation of each of the IMMPs tested, the changes in the relative abundance of different microbes within the community, the metabolic activity of bacteria and the production of health benefiting metabolites including SCFA [48, 52].

It is known that microbes do not act in isolation, but instead they relate with one another through a range of physical and metabolic interactions forming very complex microbial networks [10]. Thus, when a prebiotic substrate stimulates growth and activity of specific bacteria (e.g. butyrogenic bacteria), it may indirectly lead to alterations at the community level, including changes in growth and activity of non-target species. Thus, in the studies on the functional properties of novel prebiotics, both, the interplay between commensal microorganisms and their interactions with the host need to be investigated. One of the ways to investigate the mode of action of prebiotics at the community level is through the analysis of gene expression data using metatranscriptomics [4].

The third section of this thesis includes two chapters and presents results of an *in vitro* batch fermentation study testing the molecular fermentative behaviour of three different IMMP substrates with 27%, 94% and 96% of α-(1→6) glycosidic linkages using a standardized adult

human faecal inoculum as the microbial source. In **Chapter Six** we followed changes occurring in IMMPs during the progression of fermentation and the accompanying changes in microbiota composition. In **Chapter Seven** we looked closer at the gene expression data to evaluate the metabolic activity of the microbial community during the IMMP fermentation *in vitro*.

This thesis ends with a **General Discussion**, which provides some additional insights on the findings presented in **Chapters Two** through **Seven**. In addition, I also discuss some of the limitations, recommendations and perspectives for future studies in the area of GI tract microbiota modulations.

References:

1. Fouhy, F., et al., *Composition of the early intestinal microbiota: knowledge, knowledge gaps and the use of high-throughput sequencing to address these gaps.* Gut Microbes, 2012. **3**(3): p. 203-20.
2. Rajilic-Stojanovic, M., H. Smidt, and W.M. de Vos, *Diversity of the human gastrointestinal tract microbiota revisited.* Environ Microbiol, 2007. **9**(9): p. 2125-36.
3. Kulski, J.K., *Next Generation Sequencing - Advances, Applications and Challenges.* INTECH Open Acess. CC BY 3.0 license, 2016.
4. Bashiardes, S., G. Zilberman-Schapira, and E. Elinav, *Use of Metatranscriptomics in Microbiome Research.* Bioinform Biol Insights, 2016. **10**: p. 19-25.
5. Illumina, I., *Illumina sequencing platforms.* www.illumina.com, 2018.
6. Ramiro-Garcia, J., et al., *NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes.* F1000Research, 2016. **5**: p. 1791.
7. McFall-Ngai, M., et al., *Animals in a bacterial world, a new imperative for the life sciences.* Proc Natl Acad Sci U S A, 2013. **110**(9): p. 3229-36.
8. Savage, D.C., *Microbial ecology of gastrointestinal tract.* Annu Rev Microbiol, 1977. **31**: p. 107-133.
9. Sender, R., S. Fuchs, and R. Milo, *Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans.* Cell, 2016. **164**(3): p. 337-40.
10. Blaut, M., *Ecology and Physiology of the Intestinal Tract*, in *Between Pathogenicity and Commensalism*, U. Dobrindt, J.H. Hacker, and C. Svanborg, Editors. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 247-272.
11. Alison M.Stephen, J.H.C., *The Microbial Contribution to Human Fecal Mass.* J. Med. Microbiol., 1980. **13**: p. 45-56.
12. Walter, J., *Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research.* Appl Environ Microbiol, 2008. **74**(16): p. 4985-96.
13. Parracho, H., A.L. McCartney, and G.R. Gibson, *Probiotics and prebiotics in infant nutrition.* Proc Nutr Soc, 2007. **66**(3): p. 405-11.
14. Belkaid, Y. and T.W. Hand, *Role of the microbiota in immunity and inflammation.* Cell, 2014. **157**(1): p. 121-41.
15. Houghteling, P.D. and W.A. Walker, *Why is initial bacterial colonization of the intestine important to infants' and children's health?* J Pediatr Gastroenterol Nutr, 2015. **60**(3): p. 294-307.
16. Rautava, S., et al., *Microbial contact during pregnancy, intestinal colonization and human disease.* Nat Rev Gastroenterol Hepatol, 2012. **9**(10): p. 565-76.
17. Chu, D.M., et al., *Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery.* Nat Med, 2017. **23**(3): p. 314-326.
18. Jost, T., et al., *New Insights in Gut Microbiota Establishment in Healthy Breast Fed Neonates.* PlosOne, 2012. **7**(8).
19. Matamoros, S., et al., *Development of intestinal microbiota in infants and its impact on health.* Trends Microbiol, 2013. **21**(4): p. 167-73.
20. Choffnes, E.R., et al., *Microbial Ecology in States of Health and Disease: Workshop Summary*, ed. E.R. Choffnes, L. Olsen, and A. Mack. 2014, Washington, DC: The National Academies Press. 548.
21. Pannaraj, P.S., et al., *Association Between Breast Milk Bacterial Communities and Establishment and Development of the Infant Gut Microbiome.* JAMA Pediatr, 2017. **171**(7): p. 647-654.
22. Smilowitz, J.T., et al., *Breast milk oligosaccharides: structure-function relationships in the neonate.* Annu Rev Nutr, 2014. **34**: p. 143-69.
23. Zivkovic, A.M., et al., *Establishment of a Milk- Oriented Microbiota (MOM) in Early Life: How Babies Meet Their MOMs.* Functional Food Reviews, 2013. **5**(1): p. 3–12.
24. Moro, G.B.a.G., *Structural and Functional Aspects of Prebiotics Used in Infant Nutrition.* The Journal of Nutrition, 2008. **138**(9): p. 1818S-1828S.

25. Gibson, G.R., et al., *Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics.* 2017. **14**: p. 491.
26. GV, C., et al., *Changes in carbohydrate composition in human milk over 4 months of lactation.* Pediatrics, 1993. **91**(3): p. 637-41.
27. Austin, S., et al., *Temporal Change of the Content of 10 Oligosaccharides in the Milk of Chinese Urban Mothers.* Nutrients, 2016. **8**(6).
28. Sprenger, N., et al., *Longitudinal change of selected human milk oligosaccharides and association to infants' growth, an observatory, single center, longitudinal cohort study.* PLoS One, 2017. **12**(2): p. e0171814.
29. Giugliani, E.R.J., *Problemas comuns na lactação e seu manejo.* Jornal de Pediatria, 2004. **80**: p. s147-s154.
30. Michaelsen, L.S.-N.a.K.F., *Breast feeding and future health.* Curr Opin Clin Nutr Metab Care 2006 **9**: p. 289–296.
31. Meghan B. Azad PhD, et al., *Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months.* CMAJ, 2013. **185**(5): p. 385-394.
32. Madan, J.C., et al., *Normal neonatal microbiome variation in relation to environmental factors, infection and allergy.* Curr Opin Pediatr, 2012. **24**(6): p. 753-9.
33. Bullen, C.L., P.V. Tearle, and M.G. Stewart, *The Effect of "Humanised" Milks and Supplemented Breast Feeding on The Faecal Flora of Infants* Journal of Medical Microbiology, 1977. **10**(4): p. 403-413.
34. PL, S. and L. A., *The microbial ecology of the large bowel of breast-fed and formula-fed infants during the first year of life.* J Med Microbiol, 1982. **15**(2): p. 189-203.
35. SE, B. and W. BA, *Diet and faecal flora in the newborn: breast milk and infant formula.* Archives of Disease in Childhood, 1989. **64**(12): p. 1672-1677.
36. Kleessen, B., et al., *Influence of two infant formulas and human milk on the development of the faecal flora in newborn infants.* Acta Pediatrica, 1995(84): p. 1347-1356.
37. Kunz, C. and S. Rudloff, *Biological functions of oligosaccharides in human milk.* Acta Paediatr, 1993. **82**: p. 903-12.
38. György, P., R.F. Norris, and C.S. Rose, *Bifidus factor. I. A variant of Lactobacillus bifidus requiring a special growth factor.* Archives of Biochemistry and Biophysics, 1954. **48**(1): p. 193-201.
39. Fomon, S.J., *Infant Feeding in the 20th Century: Formula and Beikost.* Journal of Nutrition, 2001. **131**(2): p. 409S-420S.
40. Stevens, E.E., T.E. Patrick, and R. Pickler, *A history of infant feeding.* J Perinat Educ, 2009. **18**(2): p. 32-9.
41. Garrido, D., D.C. Dallas, and D.A. Mills, *Consumption of human milk glycoconjugates by infant-associated bifidobacteria: mechanisms and implications.* Microbiology, 2013. **159**(Pt 4): p. 649-64.
42. Hascoet, J.M., et al., *Effect of formula composition on the development of infant gut microbiota.* J Pediatr Gastroenterol Nutr, 2011. **52**(6): p. 756-62.
43. Vandenplas, Y., I. Zakharova, and Y. Dmitrieva, *Oligosaccharides in infant formula: more evidence to validate the role of prebiotics.* Br J Nutr, 2015. **113**(9): p. 1339-44.
44. Barile, D. and R.A. Rastall, *Human milk and related oligosaccharides as prebiotics.* Curr Opin Biotechnol, 2013. **24**(2): p. 214-9.
45. Gibson, G.R., *Fibre and effects on probiotics (the prebiotic concept).* Clinical Nutrition Supplements, 2004. **1**: p. 25–31.
46. Tissier, H., *Recherches sur la flora intestinal des nourrissons.* Theses Paris, 1900: p. 1-253.
47. Kechagia, M., et al., *Health benefits of probiotics: a review.* ISRN Nutr, 2013. **2013**: p. 481651.
48. Leemhuis, H., et al., *Isomalto/malto-polysaccharide, a novel soluble dietary fiber made via enzymatic conversion of starch.* J Agric Food Chem, 2014. **62**(49): p. 12034-44.
49. Ketabi, A., L.A. Dieleman, and M.G. Ganzle, *Influence of isomalto-oligosaccharides on intestinal microbiota in rats.* J Appl Microbiol, 2011. **110**(5): p. 1297-306.

50. C.E. Rycroft, M.R.J., G.R. Gibson and R.A. Rastall, *A comparative in vitro evaluation of the fermentation properties of prebiotic oligosaccharides.* Journal of Applied Microbiology, 2001. **91**: p. 878-887.
51. Olano-Martin, E., et al., *In vitro fermentability of dextran, oligodextran and maltodextrin by human gut bacteria.* British Journal of Nutrition, 2007. **83**(03): p. 247-255.
52. Verspreet, J., et al., *A Critical Look at Prebiotics Within the Dietary Fiber Concept.* Annu Rev Food Sci Technol, 2016. **7**: p. 167-90.

Chapter 2

Ecology of the human microbiome

Klaudyna Borewicz, Hauke Smidt

The Human Microbiome Handbook
Edited by: J. Tetro and E. Allen-Vercoe, 2016: p.9-34

Overview

Recent technological and conceptual developments in culture independent approaches targeting bacterial 16S ribosomal RNA (rRNA) genes have offered a new way of looking at microbial ecosystems. This in turn has contributed to the current expansion in the number of research projects aiming at characterizing microbiota composition and function in health and disease. Healthy human microbiota composes of many complex and diverse microbial ecosystems, with estimated 10^{14} microbial cells inhabiting the human body, which is up to 10 times the number of human cells [1]. These microbial ecosystems are also unique between different body sites and between individuals, and this variation in microbial composition can be attributed to many factors including host genetics, environment, diet and early life microbial exposure [2]. Despite taxonomic differences in microbial community structure, the core metabolic and functional pathways carried out by these ecosystems seem to be relatively stable, suggesting that the role of microbiota in health and disease may be largely due to disturbances in microbial function, rather than changes in microbiota composition alone [2].

Microbiota of the gastrointestinal tract

The human gastrointestinal (GI) tract is by far the most densely colonized and best studied microbial ecosystem found in the human body. It is estimated that 1000 to 1500 species of bacteria can inhabit an average adult GI tract, but this number could be even higher [3]. Each person carries approximately 160 bacterial species and about 10 million microbial genes, which give each individual a unique microbial make-up [4]. Host genetics may contribute to these individual variations in microbiota, and it has been shown to be an important factor affecting bacterial community composition and function [5].

Microbial colonization of the GI tract in healthy humans starts at birth and is influenced mainly by the mode of delivery (vaginal vs. Caesarean section) and the method of feeding (breastmilk vs. formula) during infancy [5]. An adult-like microbiota becomes established with introduction of solid foods and begins to resemble microbiota of adults during the first two to three years of life, after which it remains relatively stable throughout adulthood. Diet, infections, antibiotic use and other environmental conditions can temporarily disturb the normal gut microbial ecosystem, however, these disturbances tend to be temporary and in most cases, the microbiota is able to recover back to its former state. Microbial composition changes in elderly, as the diversity and stability of gut microbiota decrease with age [5].

Despite the individual variation in microbial composition, the majority of bacterial species found in the human gut belong to two phyla: Bacteroidetes and Firmicutes [6]. Most species in the phylum Bacteroidetes belong to the class Bacteroidetes, and more specifically to the genera *Bacteroides* and *Prevotella*. Most species in the phylum Firmicutes belong to *Clostridium* clusters IV and XIVa, which include genera *Clostridium*, *Eubacterium* and *Ruminococcus*. Other detected phyla include Proteobacteria, Actinobacteria, Fusobacteria, Spirochaetes, Verrucomicrobia and Lentisphaerae [7]. In addition to bacterial groups, Archaea (methanogens) and eukaryotic microorganisms (fungi) are also part of healthy human gut microbiota.

Metagenomic sequencing data suggests that even with individual differences in microbiota composition, the metabolic pathways remain stable in the GI tract of healthy subjects [2]. This collection of microbes forms a dynamic ecosystem which is known to exert important metabolic, physiological and immunological functions on its host, as well as provide protection from pathogens through so called colonisation resistance [8]. The host, on the other hand offers the microbes a stable environment and nutrients necessary for their survival. The general understanding of the microbial ecosystem function has increased tremendously in the recent years, however, the details are still largely unknown. It is becoming clear that the network of interactions, whether these are positive or negative, is very complex and we are now only at the beginning of understanding the roles of different bacterial groups, and how their functions influence the host.

In order to understand how microbial ecosystems contribute in health and disease, we should first know which microbes comprise the healthy human microbiota. More importantly, we need to ascertain what specific roles they perform and how their presence can impact the host. In the following sections we will first give an overview of the key microbial groups and

their functions in different regions of a GI tract of healthy adults. Later, we will discuss how changes in microbiota correlate with selected types of diseases.

Microbial composition in the GI tract of healthy adults

The human GI tract can be divided in anatomical regions, each characterized by a different set of physicochemical conditions which create a unique environment for microbial growth. The most important factors influencing intestinal microbiota include pH, redox potential, nutrient content, motility and presence of host secretions such as digestive enzymes, bile and mucus. The environment at each anatomical region can be further divided into the luminal content and the mucosal layer. The mucosal layer forms a lining along the GI tract and consists of a single sheet of epithelial cells and an irregular coating of mucus that protects the cells from direct action of host secretions, food and pathogens found in the lumen. The mucosal layer also provides a site of attachment for commensal microbiota. In the following sections, we will describe microbial ecosystems with respect to different regions of the GI tract.

The oral cavity comprises many different niches which provide unique conditions for microbial growth. Most microbes are associated with the mucosal surfaces on the cheeks or tongue, and hard surfaces of teeth, braces or dentures, and there is no resident microbiota in the lumen, because the passage time of food in the mouth is very short. The oral microbial ecosystem is very diverse, with about 10^{12} bacterial cells of about 1000 different species belonging to phyla Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria, Spirochaetes, Synergistetes and Tenericutes, candidate phylum TM7, and the uncultured divisions GN02 and SR1 [8-11]. The relative distribution of each microbial phylum differs between individuals and between location in the mouth [12]. The most predominant genera include *Actinomyces, Streptococcus, Neisseria, Veillonella, Porphyromonas* and *Selenomonas*. In addition, viruses, protozoa, fungi and a small number of methanogenic Archaea are also members of the normal microbiota. The microbial composition at the species level is highly variable between individuals and can be influenced by factors such as age, diet, oral health and hygiene [8].

The upper gastrointestinal tract includes esophagus, stomach and duodenum. In humans, microbial ecosystem composition and function in the upper GI tract are still largely unknown, due to poor accessibility of these areas and the need for invasive procedures in obtaining samples. In the surveys on microbiota of distal esophagus, members of six phyla, namely Firmicutes, Bacteroides, Actinobacteria, Proteobacteria, Fusobacteria, and TM7 were found in the mucosal layer, and most common genera included *Streptococcus, Prevotella* and *Veillonella* [13, 14]. Research shows that the distal esophagus is inhabited by a complex but conserved microbial community, with composition resembling the oral microbiota of the host [13]. Similar to the oral cavity, food does not stay in the esophagus long enough to allow for establishment of resident microbiota. The stomach is the first part of the GI tract that holds food for longer periods of time. Thus, the microbial distribution in the stomach, and in the descending regions of the GI tract, is spatially specific, with different microbes associated with the gastric content and with the mucosal layer [15]. Because of its low pH, which can only be tolerated by certain acid resistant bacteria, the bacterial counts in the stomach content are generally low, with about

10^3 -10^4 bacterial cells per mL [9]. The microbiota of gastric content can vary depending on diet or influx of bacteria from the mouth, esophagus and duodenum, however, these factors affect to a lesser degree the mucosa - associated microbiota, which is protected in the mucus and much more stable [15]. Culture independent studies on stomach microbiota showed that in the mucosal layer Firmicutes, Proteobacteria, Bacteroidetes, and Fusobacteria were the most abundant phyla, and *Streptococcus, Prevotella, Porphyromonas, Neisseria, Haemophilus* and *Veillonella* were common genera, but the distribution of taxa at genus level was highly variable between individuals [16-18]. One of the important, and certainly most well-studied species found in about fifty percent of the human population is *Helicobacter pylori*, which has been associated with gastric diseases, such as gastritis and cancer [15]. The duodenum is the last part of the upper GI tract and the first part of the small intestine, and it is discussed in the next section.

The small intestine is the site where most of the host enzymatic digestion and absorption of nutrients, in particular lipids and simple carbohydrates, takes place. Studies on microbial composition are again very limited, with the majority of findings being based on biopsy specimens in association with various GI disorders. The duodenal lumen forms a unique environment characterized by a low pH, fast passage time, and the presence of antimicrobial bile and digestive enzymes, making it an unfavourable place for microbial growth. No culture independent studies up to date focused on resident microbiota in human duodenal content. On the other hand, biopsy samples provided insight in microbiota in the duodenal mucosa. In a recent study using 16S rRNA gene-targeted HITChip analysis of duodenal biopsies from children, thirteen phylum-like level bacterial groups were detected, and Proteobacteria, Bacilli and Bacteroidetes were the most abundant taxa, with each individual subject showing a different and unique microbial profile [19]. The predominant genus-like groups included *Sutterella wadsworthensis* et rel., *Streptococcus mitis* et rel., *Aquabacterium, Streptococcus intermedius* et rel., *and Prevotella melaninogenica* et rel. [19]. In a study using sequencing of 16S rRNA gene clone libraries, the most abundant phyla detected in biopsies from children and adult subjects were Firmicutes, Proteobacteria, Bacteroidetes, and also Actinobacteria, Fusobacteria and *Deinococcus-Thermus* [20]. Most sequences were classified as *Streptococcus* and *Prevotella* spp. in both age groups, and 5% of sequences that were found only in healthy children could not be assigned to any known genus. Bacterial community richness was higher in the adult group as compared to the juvenile group, with members of *Veillonella*, *Neisseria, Haemophilus, Methylobacterium* and *Mycobacterium* present in adult mucosa. It is interesting to note that overall duodenal microbiota composition seems to resemble the microbiota found in the oral cavity and esophagus, and less so the microbiota found in the lower GI tract [21]. The number of bacterial cells and diversity increase along the intestine, and it is estimated that the jejunum harbours 10^5 -10^6 bacteria per mL of content [9]. An earlier study examining mucosa biopsies of human jejunum showed that *Streptococcus* and *Proteobacteria* were the most abundant taxa, and contributed respectively to 68 and 13 % of all microbiota detected [22]. A more recent study showed that ileostomy effluent samples can provide a good representation of microbial composition in the human jejunum/ proximal-ileum without the need for invasive sampling [23]. The most predominant (common core) taxa in ileostoma effluent, and in jejunum, included Bacilli (*Streptococcus* spp.), *Clostridium* cluster IX

(*Veillonella* spp.), *Clostridium* cluster XIVa and Gammaproteobacteria [23]. Similar findings came from an earlier study on ileostoma effluent, where the most abundant species were members of the Lactobacillales and Clostridiales, mainly *Streptococcus bovis*-related species and the *Veillonella* group, as well as species belonging to *Clostridium* cluster I and *Enterococcus*[24]. However, the ileum associated Bacteroidetes and *Clostridium* clusters III, IV and XIVa were reduced in ileostoma effluent samples. Bacterial numbers increase to about 10^8 -10^9 cells per mL of ileal digesta. Biopsies and catheter-collected lumen samples revealed that the bacterial community in the human ileum is dominated by species belonging to Bacteroidetes and *Clostridium* clusters IV and XIVa and resembles the microbiota found in the colon [9, 22]. Similar to the ileostomy effluent samples, ileum microbiota is also characterized by short and long term fluctuations in microbial profiles within individuals and large inter-individual variability between patients [24].

The large intestine is separated from the small intestine by the ileocecal valve, and it can be divided into cecum, ascending, transversing and descending colon, rectum and the anal canal. The cecum is the first region of the large intestine that receives food from the small intestine. It is also connected with the appendix - a small and rudimentary projection, which in humans has no function in food digestion, but it may play an important role as a reservoir of microbiota and in stabilizing and restoring the colon microbial ecosystem, especially after disturbance, for example due to antibiotic use [25, 26]. Unlike the small intestine, microbial composition and function of the human large intestine has been studied to great extent, mostly because of the ease of collecting faecal samples, and because of the high density of microbial cells, estimated to be around 10^{11} -10^{12} per mL [9]. The most predominant microbial groups found in the human large intestine include *Bacteroides*, members of the various *Clostridium* clusters, *Bifidobacterium,* Enterobacteriaceae and *Eubacterium*. Even though the large intestine can be divided into five anatomical regions, the microbial composition is very uniform, and faecal material seems to represent well the microbiota in the entire region [7]. However, just like in other parts of the GI tract, also in the large intestine there is a large difference between microbial ecosystems found in the lumen and mucosal layer. Faecal samples represent the luminal fraction only, and the mucosal layer is much less explored due to the need for more invasive methods in collecting biopsy samples. Large intestinal microbiota is very diverse, highly unique to each individual and relatively stable over time [27]. Factors such as age, the use of antibiotics, or certain diseases may permanently alter the microbial composition [27]. Recent studies utilizing large cohorts of subjects suggested that the faecal microbiota composition in healthy adults can be categorized into three major enterotypes dominated by different bacterial populations, in particular *Bacteroides, Prevotella* and *Ruminococcus* [28, 29]. These enterotypes are independent of age, ethnicity, gender and body mass. However, this division is still controversial, and some studies failed to detect presence of enterotypes, in both the elderly [30] and in adult research populations [31]. Another large study suggested an alternative to the enterotype theory [27]. The authors noted that in faecal samples of western adults certain bacterial groups, namely *Dialister* spp., *Bacteroides fragilis, Prevotella melaninogenica, P. oralis* and two groups of uncultured Clostridiales cluster I and II, were bimodally distributed in the healthy human population, representing so called “tipping elements” [27]. These bistable bacterial groups were either very abundant or almost absent, and

unstable at their intermediate abundance levels [27]. In addition, the condition of the bistable groups, especially the *Bacteroides* and *Prevotella,* seemed to correlate with the shifts in other bacteria, and as a result they were believed to be driving the overall composition of the colonic ecosystem towards specific enterotypes [27].

Microbial ecosystem function in the GI tract of healthy adults

Metagenomic studies provide insight on the functional potential of microbiota, by analysing microbial genes, collectively known as the microbiome. A recent study reported that each person carries about 10 million bacterial genes in their GI tract, majority of which are involved in bacterial metabolism [4, 32]. Additional information about microbial activity can be obtained from metatranscriptomic, metabolomic and metaproteomic analyses. These approaches provide insight about microbial gene regulation and expression, as well as production of metabolites, proteins, vitamins and regulatory elements. Similar to compositional diversity, there is a large functional variation in different microbial ecosystems, but the core metabolic and functional pathways carried out by the same types of ecosystems seem to be relatively conserved and stable [2]. It is also common for the same metabolic functions to be carried out by different bacterial groups, meaning that correlating the compositional and functional changes in the ecosystem maybe less straightforward, because changes in composition and function of a given microbial ecosystem can be independent from each other [33].

The oral cavity is the first point of contact between microbiota, diet and host. Despite regular influx of food ingested by the host, the majority of nutrients for the oral commensal microbes are derived from glycoproteins present in saliva and gingival crevicular fluid [34]. Complete breakdown of these glycoproteins requires cooperation between different species of bacteria. For example, oral streptococci (e.g. *S. oralis, S. sangiunis*) remove oligosaccharide side chains and break down the protein core by their proteolytic, endopeptidase and glycosidic activity, while other Gram-negative anaerobes (e.g. *Porphyromonas gingivalis, Prevotella intermedia, Prevotella nigrescens, and Peptostreptoccus micros*) further break down proteins into peptides and amino acids [34-36]. Amino acids can be then fermented to short chain fatty acids (SCFA), including branched chain fatty acids, which are further degraded by other bacteria and by methanogenic Archaea [8]. Certain food components, such as gluten or nitrate can also be degraded / transformed by microbial enzymes, and the processes and products are crucial for the health and well-being of the host, while breakdown of these functions can be linked with host diseases [37-39]. As already mentioned, the mouth is an open environment and commensal bacteria create a barrier against colonisation with transient microbes and any opportunistic pathogens that can enter with food or water. An *in vitro* study on oral microbiota from mice provided a good illustration of how the cooperation of different commensal species can leverage a community response to pathogen invasion. The study proposed that cooperation of three different species of oral streptococci were involved, with *S. saprophyticus* sensing the presence of an invader, and initiating the defence pathway, *S. infantis* acting as a mediator and *S. sanguinis* producing hydrogen peroxide and acting as a killer [40]. Besides colonization

resistance, oral microbiota plays an important role in maintaining host – microbe homeostasis, by interacting with host mucosal cells and training the host's immune system to recognize and destroy pathogens, while down-regulating the pro-inflammatory immune response towards the commensal bacteria normally present in the mouth [41].

Upper gastrointestinal tract microbiota function is still not well understood, and most studies to date focused on specific pathogens and their role in the aetiology of different diseases and to lesser extent on the microbial interactions in a healthy ecosystem. Little is known about the ecology of microbiota inhabiting the esophagus and stomach, but its role in colonization resistance and protection from pathogens is likely to be an important one. Normal microbiota generates a microenvironment that can inhibit growth of pathogens, by competing for substrates and binding sites, stimulating host immune responses against invaders and production of antimicrobial substances. For example, *in vitro* and *in vivo* studies using animal models showed that stomach colonization with *H. pylori* is inhibited by the normal commensal microbiota and by probiotic strains of *Lactobacillus*, *Bifidobacterium* and *Saccharomyces,* suggesting the importance of microbial interaction in pathogen resistance [15]. Other studies using human biopsy samples also reported changes in intestinal microbiota associated with gastric cancer, however, the exact function and causality of this association is still being investigated [9]. It is likely that microbial metabolites, bacterial lipopolysaccharides (LPS), lipoproteins, lipoteichoeic acids (LTA), flaggellins and bacterial nucleic acids can interfere with the normal function of gastric mucosa, causing chronic inflammation, changes in mucin production, metaplasia, and eventually can lead to diseases [9, 19]. The functions of the microbiota in the duodenum are still not well understood, but changes in microbial composition between Celiac disease patients and healthy controls suggest that the microbiota plays a role in immune response, inflammation and maintaining gut homeostasis [19, 21]. The homeostasis of gut epithelia relies to a large extent on adequate activation of Toll-like receptors (TLRs), which recognize microbe-associated motifs, regulate the immune response to pathogens, and affect the epithelial barrier by regulating the expression of tight junction proteins, mucin and antimicrobial peptides by the host's intestinal cells [19].

The small intestine is the site where most of the host enzymatic digestion and absorption of energy from the diet takes place. Thus, diet is an important factor modulating microbial function, by selecting bacterial groups that are better equipped to break down different dietary substrates [5]. For example, certain *Lactobacillus* spp. found in duodenum and jejunum had been associated with weight gain and leanness, and differed in their metabolic capacities to break down dietary carbohydrates and fats supplied by the host [5]. The transit time in the small intestine is very short, and *Streptococcus* and *Veillonella* spp., which dominate the microbial ecosystem in the jejunum and ileum, are well adapted to quickly metabolize a variety of available carbohydrates, first to lactate (*Streptococcus*) and then to acetate and propionate (*Veillonella)* [24]. Recent metatranscriptome analysis of ileostoma effluent confirmed high abundance of genes involved in transport and metabolism of diet-derived simple carbohydrates, and linked the task mainly to *Streptococcus* groups [42]. In addition to its function in carbohydrate metabolism, the authors concluded that small intestine microbiota could also play a key role in immune system development and homeostasis. For example, the ileum is

connected with a large mass of gut associated lymphoid tissue (GALT) and Peyer's patches, and commensal bacteria, such as different strains of streptococci, were shown to induce specific immune responses in the host [42]. The close contact between the microbiota and the host cells in the small intestine underlines the current hypothesis that microbially derived metabolites or toxins also modulate gene expression via the gut-brain neural circuit and may influence endocrine function (e.g. secretion of glucagon and incretins) and even show an effect on mood or behaviour of the host [5, 42].

Large intestine microbial ecosystem function has been well studied, mainly due to the ease of collecting faecal samples, but also because it has been known for a long time that colonic microbial processes play an important role in human health. The most direct role is in the digestion and metabolism, as the large intestinal microbiota breaks down indigestible food components and provides the host with an otherwise inaccessible source of energy. It also produces SCFA, which are the main source of energy for colonocytes [43]. In addition, the colonic microbiota is a main source of vitamins K and B12, it prevents colonization by pathogens, and plays an important role in regulating the host's immune responses [5, 43]. A study on the faecal microbiome of healthy Japanese subjects was among the first to explore microbial ecosystem function in the human colon using culture-independent methods. The study revealed that a high proportion of genes present were related to carbohydrate metabolism and transport. The authors also noted an enrichment of peptidases and enzymes for anaerobic pyruvate metabolism and reduction in genes involved in fatty-acid metabolism. There were also high levels of enzymes involved in energy storage, antimicrobial peptide transport and multidrug efflux pump peptides [44]. The authors concluded that these enzymes may help certain commensal microbes to compete with each other and thus, may be essential for maintenance of ecosystem balance. Enzymes for DNA repair were also enriched. On the other hand, there was a low abundance of genes involved in biosynthesis of flagella and chemotaxis and in oxygen take-up [44]. Interestingly, these patterns in gene distribution were not observed in unweaned infants, suggesting that infant microbiota is less complex and thus, microbial ecosystem function is less stable, more dynamic and highly adaptable. In adult microbiota a higher diversity of bacterial species exists with large inter-individual variability in microbial composition, yet there is a shared functional core, which is believed to be stable and much more uniform between individuals [32, 44]. Recently, more in depth analyses showed that there could be functional differences correlating with different enterotypes found in the colon [28]. For example, the *Bacteroides* rich type has more bacterial species that are capable of producing vitamins C, B2, B5 and H. This group is dominated by species that utilize carbohydrate fermentation as the main energy source. On the other hand, the *Prevotella* type showed higher numbers of species producing vitamin B1 and folic acid, and included species that use mucin glycoproteins as a source of energy, similarly to the *Ruminococcus* type [28].

One of the important functions of colonic microbiota that received a lot of attention in recent years is the production of SCFA, and in particular butyrate, by bacteria from *Clostridium* clusters IV and XIVa. The main butyrate producing species are believed to be *Eubacterium rectale* and *Faecalibacterium prausnitzii,* in addition to others in the genera *Coprococcus and Roseburia* [45]. The process provides a great example of synergic interaction between diet,

microbes and host, and the presence of butyrate producers in the colon has been shown to be negatively correlated with functional dysbiosis, reduction of the risk of infections with opportunistic pathogens and the decrease in oxidative stress [5]. Butyrate producers can respond to different environmental conditions, such as diet or pH, and engage different fermentation pathways in which the final products are lactate, formate, hydrogen and carbon dioxide. It has been shown that cross-feeding between bifidobacteria and butyrate producers is also possible: bifidobacteria break down polysaccharides and produce lactate and acetate, which are further utilized by butyrate-producers to form butyrate [45]. Butyrate is known to play an important role in maintaining homeostasis of the intestine. It is the main source of energy for colonocytes and it inhibits expression of pro-inflammatory cytokines in the mucosal layer of intestine [46]. In addition, butyrate has a positive effect on integrity of the mucosal layer by stimulating expression of tight junction proteins, and by inducing production of mucin and antimicrobial peptides [47].

Selected diseases associated with dysbiosis of the intestinal microbiota

Inflammatory Bowel Disease (IBD) is a chronic inflammatory disorder affecting the mucosal layer of the intestines. The two main types of IBD are Crohn's disease and ulcerative colitis. Factors such as genetics, diet, gut permeability, stress and microbiota changes seem to be contributing factors in development of IBD. Both, Crohn's disease and ulcerative colitis are linked with a decrease in microbial diversity [48], reduced levels of Firmicutes, especially *Faecalibacterium prausnitzii* [49, 50], Ruminococcaceae and *Roseburia* [51], and increased levels of Enterobacteriaceae [48], *Bacteroides, Prevotella* [49], adherent-invasive *Escherichia coli*, *Campylobacter concisus*, and enterohepatic *Helicobacter* [52]. One of the characteristics of IBD is a decrease in microbial SCFA production. This can be due to reduction in the abundance of the two main butyrate producing bacteria: *Faecalibacterium* - a commensal bacterium with anti-inflammatory properties, and *Roseburia,* which also produces butyrate from Ruminococcaceae derived acetate [51]. The decrease in butyrate producing bacteria is often accompanied by an increase in sulphate reducing bacteria, which produce toxic hydrogen sulphide. Hydrogen sulphide blocks butyrate utilization by colonocytes and interferes with innate immune defence [47]. In both, Crohn's disease and ulcerative colitis, the leaky epithelial barrier and reduced innate immune defence lead to increased translocation of bacteria through the lamina propria and increase in inflammatory reaction and formation of ulcers. Despite growing evidence that IBD is linked with dysbiosis of intestinal microbiota and with changes in metagenomic pathways, it is still unclear if the changes in microbiota are the cause or the consequence of the intestinal inflammation [47, 51]. Recent studies suggest that mutations in certain host genes coding for receptors involved in bacterial recognition and killing of pathogens, such as TLR4 D299G, TLR1L80P, NOD 2/*CARD15*, Arg702Trp, Gly908Arg and Leu1007, might be an important risk factor in the aetiology of IBD [53-55].

Obesity has many risk factors which together lead to perturbations in energy balance and weight gain and may result in development of other metabolic and cardiovascular diseases [56]. High calorie diet is the main risk factor for obesity, but recent studies show that gut

microbiota may also play a role, for example by breaking down indigestible dietary substrates and providing additional energy to the host [57]. Surplus energy is stored in a form of fat, and excessive fat accumulation leads to weight gain and obesity. For example, it has been shown that germ-free (GF) mice fed diets rich in fat and sugar (the "western diet") did not develop obesity [58]. In turn, transferring gut microbiota of obese mice into GF mice led to rapid increase in body fat, despite of restricted calorie intake [57]. Diet not only provides energy to the host, but it also may affect gut microbiota composition and function, selecting for species that are best adapted to utilize different dietary compounds [59, 60]. These diet-induced changes in the microbial composition and function have been linked with obesity, but the specific effect on different bacterial groups and the causality are less clear [59, 60]. Earlier studies on microbiota of obese humans and mice reported an increased ratio of Firmicutes to Bacteroidetes [61], however, recent studies contradict these findings and suggest that obesity might be linked with a decreased ratio of Firmicutes to Bacteroidetes [62, 63], the increase in other bacterial phyla, for example Actinobacteria [56], or Proteobacteria [64], or overall decrease in microbial diversity [32]. The inconsistencies in these findings might be partially due to differences in research methods used, but could also relate to the differences in host genotypes, which had been shown to influence microbial composition and could predispose certain individuals to developing metabolic conditions [65]. It is also likely that the obesity-associated changes in the gut microbial ecosystem structure and function are more refined than phylum level. Reports on obesity and microbial changes at a genus or species level are still limited, but a recent study found that obese people had higher faecal levels of *Lactobacillus reuteri*, *Bifidobacterium animalis* and *Methanobrevibacter smithii* and lower levels of *B. animalis, L. paracasei,* and *L. plantarum* than the lean controls [66]. Studies on microbiota function in obese and lean subjects show that shifts in relative abundance of microbial phyla are accompanied by changes in metabolic pathways involved in carbohydrate metabolism and SCFA production [56, 57, 62]. The increase in bacterial enzymes involved in degradation of indigestible polysaccharides increases the levels of monosaccharides which become readily available to the host. In addition, there is an increase in production of SCFA, which become converted into triglycerides in the liver [62]. SCFA can activate G-protein coupled receptors (GPR41 and GPR43) in the gut and induce secretion of PYY peptide which decrease intestinal transit, allowing longer time for nutrient uptake [67]. At the current state of research, it is still unclear whether changes in microbiota are a contributing factor causing obesity, or whether microbial dysbiosis is the result of obesity. A recent prospective study suggested that changes in microbiota, such as an increase in *Staphylococcus aureus* and a decrease in *Bifidobacterium* spp. preceded development of obesity in a group of children [68]. Transplantation studies on GF mice provide another line of evidence that the composition and function of the intestinal microbiota is an important factor in aetiology of obesity.

Diabetes mellitus type 1 and type 2 are metabolic diseases characterized by insufficient production of insulin and insulin resistance, respectively, which lead to high blood glucose levels. Recent study on rats and children with type 1 diabetes (TDM1) showed that changes in gut microbiota may play a role in the aetiology of these diseases [69, 70]. Diabetic children had higher numbers of *Clostridium*, *Bacteroides* and *Veillonella* and lower numbers of *Lactobacillus, Bifidobacterium, C. coccoides-E. rectale* group and *Prevotella.* Type 2 diabetes

(TDM2) has been associated with changes in gut microbiota composition and function. In TDM2, the ratio of Firmicutes to Bacteroidetes was reduced and the *Bacteroides-Prevotella* to *C. coccoides-E. rectale* group ratio was increased in patients with elevated plasma glucose levels [71]. In addition, Clostridia levels were lower and Betaproteobacteria levels were higher in the diabetic group, as compared to non-diabetic controls [71]. The mechanism by which altered gut microbiota and diabetes interact is complex and is likely to proceed through a cascade of events. Members of the *C. coccoides/E. rectale* group are the main butyrate producers in the human colon, and butyrate is important for maintaining integrity of intestinal barrier and protecting the host against invasion by opportunistic pathogens and transfer of endotoxins into plasma. Thus, the decrease in *C. coccoides/E. rectale*, and the corresponding increase in the level of Bacteroidetes could be linked with metabolic endotoxemia [72]. Bacteroidetes are Gram-negative bacteria containing lipopolysaccharides (LPS) in their outer membrane. LPS are endotoxins, and increased LPS levels in the colon, as well as changes in gut permeability, result in higher LPS levels in blood serum. LPS had been shown induce production of pro-inflammatory molecules by macrophages and disrupt the function of pancreatic β-cells and insulin secretion [73]. The same study concluded that specific probiotic strains of *Lactobacillus* and *Bifidobacterium* could decrease LPS levels in the colon and inflammation [73]. Similarly, in another study higher levels of *Bifidobacterium* spp. were associated with increase in production of YY and glucagon-like peptide (GLP), reduced gut permeability and decreased LPS in plasma [74].

Colorectal cancer (CRC) is one of the most common forms of cancer [75]. High-calorie diets, rich in animal fat, red and processed meat or alcohol, and low in dietary fiber, whole grains and vegetables are considered a main risk factor [75, 76]. Gut microbiota structure and function is largely dependent on diet, and certain bacterial metabolites are known to be pro-inflammatory and tumor inducing. Normal colonic microbiota composes of members of *Clostridium* clusters IV and XIVa, Lactobacillales, Bifidobacteriales, and Actinomycetales which are believed to have a protective effect in development of CRC because of their role in synthesis of butyrate and other SCFA, as well as conjugated linoleic acids [45, 77]. It has been noted that in CRC patients, there is a significant decrease in Firmicutes, in particular *Roseburia* spp. and *Eubacterium* spp., and as a consequence, a reduction in butyrate production [78]. Butyrate has an important role in reducing inflammation, suppressing pre-cancerous cells and inducing apoptosis of tumor cells [78]. Changes in microbial composition allow opportunistic pathogens, such as *Enterococcus*, *Streptococcus*, and *Escherichia*/*Shigella* to proliferate and cause damage to the gut epithelial cells [79]. There is a wide range of bacterial metabolites which cause damage to the DNA in host's gut epithelial cells and may lead to chromosomal instability and development of CRC. For example, fecapentaenes produced by *Bacteroides* spp., heterocyclic amines produced by *Salmonella typhimurium* and *Enterococcus* spp., and hydrogen sulphide produced by sulphate-reducing bacteria such as *Desulfovibrio*, *Desulfomonas*, *Desulfotomaculum*, *Desulfobulbus, Desulfobacter*, *Desulfococcus*, *Desulfosarcina*, and *Desulfonema* have all been implicated to have a role in development of CRC [76]. Other mechanisms include high superoxide production by *Enterococcus faecalis* [80], the metabolism of 7α-dehydroxylating bacteria of the genera *Eubacterium* and *Clostridium* [81], and formation of colonic lesions by *Streptococcus bovis/gallolyticus* [82]. For

each of the associations listed above, the exact mechanism of action in formation of CRC is still being investigated.

Celiac Disease (CD) is an autoimmune, inflammatory disorder of the small intestine triggered by diet containing gluten proteins found in wheat. It has a genetic component and seems to affect mainly individuals who carry the leukocyte antigen alleles (HLA)-DQ2 or HLA-DQ8 [83]. Since not all genetically predisposed individuals develop the disease, it has been proposed that changes in gut microbiota may also play a role in aetiology of CD [19, 84]. Earlier studies using duodenal biopsy samples reported increased microbial diversity in CD patients, decreased ratio of *Lactobacillus–Bifidobacterium/Bacteroides–E. coli* [84] and higher level of *Bacteroides vulgatus* and *Escherichia coli* in CD patients [85]. Similarly, a more recent study found higher diversity in duodenal mucosa and a lower ratio of Firmicutes / Proteobacteria in children with active CD, as compared to those with non-active disease and healthy controls [52]. The same study reported that CD patients had higher abundance of the families Enterobacteriaceae and Staphylococcaceae, particularly the species *Klebsiella oxytoca*, *Staphylococcus epidermidis*, and *Staphylococcus pasteuri,* and controls had more *Streptococcus anginosus* and *Streptococcus mutans* [52]. Studies on faecal microbiota in CD-predisposed and healthy infants noted major differences between the two groups, however, the conclusions regarding changes in specific bacterial groups were contradicting, with one study finding very low levels of Bacteroidetes in CD infants [86], and the other study reporting reduction in *Bifidobacterium* spp. and *B. longum*, but increase in *B. fragilis* group and *Staphylococcus* spp. [87]. Finally, few studies reported no changes in microbial composition in relation to the disease status, but noted changes in the TLR signalling pathways, which are involved in inflammatory responses and expression of tight junction proteins important in maintaining the integrity of the intestinal mucosa [19, 88]. The causality of the association is still unknown but just like in other metabolic diseases, decrease in commensal populations and the increase in levels of Gram negative or pathogenic bacteria could contribute to the pathogenesis of CD by altering intestinal permeability and inducing inflammation [84]. In addition, recent studies suggest the dysbiosis in oral microbiota could also play a role in the aetiology of this disease [38, 39].

Dental cavities and periodontal disease are two common diseases of the oral cavity. The main risk factors in formation of dental cavities are frequent sugar intake and low saliva production, both of which promote growth of aciduric and acidogenic strains of lactate producing streptoccoci and lactobacilli. On the other hand certain strains had been linked with carries-free status. These include *Streptococcus oligofermentas* which inhibits growth of *S. mutans* [89], and *Porphyromonas catoniae* [90]. In periodontal disease *Porphyromonas gingivalis, Aggregatibacter actinomycetemcomitans, Prevotella intermedia, Tannerella forsythia, Parvimonas micra, Fusobacterium nucleatum* and *Treponema denticola* have all been associated with the disease. However, lactic acid producing bacteria, such as *Streptococcus cristatus, S. salivarius, S. mitis* and *S. sanguinis*, as well as probiotic *Lactobacillus brevis* and *L. reuteri* were shown to attenuate inflammatory markers associated with periodontitis, produce antimicrobial agents and reduce inflammation in periodontal cases [91-93].

Other health conditions have been associated with changes in gut microbiota. Diseases such as Irritable Bowel Syndrome (IBS) [94, 95], antibiotic associated diarrhoea [96], pouchitis [97], necrotizing enterocolitis [98], gastric ulcers, esophagitis [99], Barrett's oesophagus [15] and malnutrition [100] are just some examples of diseases linked with changes in gut microbiota composition and function. However, the impact of gut microbiota on host wellbeing goes beyond its direct effect on the function of the digestive tract. Many other health conditions are now being associated with changes in structure and function of microbial ecosystems in the gut, but also at other body sites. An interesting example comes from the studies on hypertension and the role of oral microbiota. Recent studies on nitrate supplementation and hypertension suggested that facultative anaerobic oral bacteria (in particular *Streptococcus salivarius, S. mitis, S. bovis, Veillonella* spp., *Staphylococcus aureus, S. epidermidis, Nocardia* spp., and *Corynebacterium* spp.) may play an important role in nitrate metabolism, by reducing nitrate to nitrite [37]. Nitrite can be absorbed and converted to nitric oxide, which is essential for maintaining vascular health by reducing hypertension and lowering blood pressure. In a study on healthy subjects, the use of antimicrobial mouth rinse eliminated the beneficial effect of nitrate supplements, suggesting that oral microbiota may contribute to maintaining cardiovascular health [101]. As discussed earlier, the development of the immune system also seems to largely depend on microbiota. The new "hygiene hypothesis" claims that limiting early-life infection impedes natural immune system development and causes predisposition to allergic disease [102], atopic eczema, allergic rhinoconjunctivitis and asthma [103]. Reduced microbial diversity during infancy has been associated with an array of allergic diseases later in life [104]. Also, other autoimmune diseases, such as multiple sclerosis, lupus and rheumatoid arthritis have been shown to be correlated with changes in gut microbiota [9, 105]. Finally, recent studies suggest that certain mental conditions, such as depression, anxiety and autism may all have a microbiota dysbiosis component in their aetiology [94, 106-108].

Acknowledgments

The authors are grateful for financial support through the TKI Agri&Food program as coordinated by the Carbohydrate Competence Center (CCC3; www.cccresearch.nl).

References

1. Savage, D.C., *Microbial ecology of gastrointestinal tract.* Annu Rev Microbiol, 1977. **31**: p. 107-133.
2. Human Microbiome Project, C., *Structure, function and diversity of the healthy human microbiome.* Nature, 2012. **486**(7402): p. 207-14.
3. DiBaise, J.K., et al., *Gut microbiota and its possible relationship with obesity.* Mayo Clin Proc, 2008. **83**(4): p. 460-9.
4. Li, J., et al., *An integrated catalog of reference genes in the human gut microbiome.* Nat Biotechnol, 2014. **32**(8): p. 834-41.
5. Moreno-Indias, I., et al., *Impact of the gut microbiota on the development of obesity and type 2 diabetes mellitus.* Front Microbiol, 2014. **5**: p. 190.
6. Mariat, D., et al., *The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age.* BMC Microbiol, 2009. **9**: p. 123.
7. Gerritsen, J., et al., *Intestinal microbiota in human health and disease: the impact of probiotics.* Genes Nutr, 2011. **6**(3): p. 209-40.
8. Wade, W.G., *The oral microbiome in health and disease.* Pharmacol Res, 2013. **69**(1): p. 137-43.
9. Tlaskalova-Hogenova, H., et al., *The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases.* Cell Mol Immunol, 2011. **8**(2): p. 110-20.
10. Soro, V., et al., *Axenic culture of a candidate division TM7 bacterium from the human oral cavity and biofilm interactions with other oral bacteria.* Appl Environ Microbiol, 2014. **80**(20): p. 6480-9.
11. He, X., et al., *Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle.* Proc Natl Acad Sci USA, 2015. **112**: p. 244-249.
12. Zaura, E., et al., *Defining the healthy "core microbiome" of oral microbial communities.* BMC Microbiol, 2009. **9**: p. 259.
13. Pei, Z., et al., *Bacterial biota in the human distal esophagus.* Proc Natl Acad Sci USA, 2004. **101**(12): p. 4250-5.
14. Fillon, S.A., et al., *Novel device to sample the esophageal microbiome--the esophageal string test.* PLoS One, 2012. **7**(9): p. e42938.
15. Wang, Z.K. and Y.S. Yang, *Upper gastrointestinal microbiota and digestive diseases.* World J Gastroenterol, 2013. **19**(10): p. 1541-50.
16. Stearns, J.C., et al., *Bacterial biogeography of the human digestive tract.* Sci Rep, 2011. **1**: p. 170.
17. Bik, E.M., et al., *Molecular analysis of the bacterial microbiota in the human stomach.* Proc Natl Acad Sci USA, 2006. **103**(3): p. 732-7.
18. Li, X.X., et al., *Bacterial microbiota profiling in gastritis without Helicobacter pylori infection or non-steroidal anti-inflammatory drug use.* PLoS One, 2009. **4**(11): p. e7985.
19. Jing Cheng, et al., *Duodenal microbiota composition and mucosal homeostasis in pediatric celiac disease.* BMC Gastroenterology, 2013. **13**: p. 1-13.
20. Nistal, E., et al., *Differences of small intestinal bacteria populations in adults and children with/without celiac disease: effect of age, gluten diet, and disease.* Inflamm Bowel Dis, 2012. **18**(4): p. 649-56.
21. Wacklin P, et al., *The duodenal microbiota composition of adult celiac disease patients is associated with the clinical manifestation of the disease.* Inflamm Bowel Dis., 2013. **19**(5): p. 934-941.
22. Wang, M., et al., *Comparison of bacterial diversity along the human intestinal tract by direct cloning and sequencing of 16S rRNA genes.* FEMS Microbiol Ecol, 2005. **54**(2): p. 219-31.
23. Zoetendal, E.G., et al., *The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates.* ISME J, 2012. **6**(7): p. 1415-26.

24. Booijink, C.C., et al., *High temporal and inter-individual variation detected in the human ileal microbiota.* Environ Microbiol, 2010. **12**(12): p. 3213-27.
25. Laurin, M., M.L. Everett, and W. Parker, *The cecal appendix: one more immune component with a function disturbed by post-industrial culture.* Anat Rec, 2011. **294**: p. 567–579.
26. Bollinger, R.R., et al., *Biofilms in the large bowel suggest an apparent function of the human vermiform appendix.* J Theor Biol, 2007. **249**(4): p. 826-31.
27. Lahti, L., et al., *Tipping elements in the human intestinal ecosystem.* Nat Commun, 2014. **5**: p. 4344.
28. Arumugam, M., et al., *Enterotypes of the human gut microbiome.* Nature, 2011. **473**(7346): p. 174-80.
29. Benson, A.K., et al., *Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors.* Proc Natl Acad Sci USA, 2010. **107**(44): p. 18933-8.
30. Claesson, M.J., et al., *Gut microbiota composition correlates with diet and health in the elderly.* Nature, 2012. **488**(7410): p. 178-84.
31. Huse, S.M., et al., *A core human microbiome as viewed through 16S rRNA sequence clusters.* PLoS One, 2012. **7**(6): p. e34242.
32. Turnbaugh, P.J., et al., *A core gut microbiome in obese and lean twins.* Nature, 2009. **457**(7228): p. 480-4.
33. Zoetendal, E.G., M. Rajilic-Stojanovic, and W.M. de Vos, *High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota.* Gut, 2008. **57**(11): p. 1605-15.
34. Homer, K.A., R.A. Whiley, and D. Beighton, *Proteolytic activity of oral streptococci.* FEMS Microbiology Letters, 1999. **67**(3): p. 257–260.
35. Wickstrom, C., et al., *Proteolytic degradation of human salivary MUC5B by dental biofilms.* Microbiology, 2009. **155**(Pt 9): p. 2866-72.
36. Bao, G.J., et al., *Proteolytic activities of oral bacteria on ProMMP-9 and the effect of synthetic proteinase inhibitors.* Open Dent J, 2008. **2**: p. 96-102.
37. Hezel, M. and E. Weitzberg, *The oral microbiome and nitric oxide homoeostasis.* Oral Dis, 2013.
38. Helmerhorst, E.J., et al., *Discovery of a novel and rich source of gluten-degrading microbial enzymes in the oral cavity.* PLoS One, 2010. **5**(10): p. e13264.
39. Zamakhchari, M., et al., *Identification of Rothia bacteria as gluten-degrading natural colonizers of the upper gastro-intestinal tract.* PLoS One, 2011. **6**(9): p. e24455.
40. He, X., et al., *The social structure of microbial community involved in colonization resistance.* ISME J, 2014. **8**(3): p. 564-74.
41. Srinivasan, N., *Telling apart friend from foe: discriminating between commensals and pathogens at mucosal sites.* Innate Immun, 2010. **16**(6): p. 391-404.
42. Aidy, S.E., B.v.d. Bogert, and M. Kleerebezem, *The small intestine microbiota, nutritional modulation and relevance for health.* Current Opinion in Biotechnology, 2015. **32**: p. 14-20.
43. Leser, T.D. and L. Molbak, *Better living through microbial action: the benefits of the mammalian gastrointestinal microbiota on the host.* Environ Microbiol, 2009. **11**(9): p. 2194-206.
44. Kurokawa, K., et al., *Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes.* DNA Res, 2007. **14**(4): p. 169-81.
45. Louis, P. and H.J. Flint, *Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine.* FEMS Microbiol Lett, 2009. **294**(1): p. 1-8.
46. Segain, J.-P., et al., *Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease.* 2000. **47**: p. 397-403.
47. Fava, F. and S. Danese, *Intestinal microbiota in inflammatory bowel disease: friend of foe?* World J Gastroenterol, 2011. **17**(5): p. 557-66.
48. Hansen, J., A. Gulati, and R.B. Sartor, *The role of mucosal immunity and host genetics in defining intestinal commensal bacteria.* Curr Opin Gastroenterol, 2010. **26**(6): p. 564-71.
49. Benjamin, J.L., et al., *Smokers with active Crohn's disease have a clinically relevant dysbiosis of the gastrointestinal microbiota.* Inflamm Bowel Dis, 2012. **18**(6): p. 1092-100.

50. Varela, E., et al., *Colonisation by Faecalibacterium prausnitzii and maintenance of clinical remission in patients with ulcerative colitis.* Aliment Pharmacol Ther, 2013. **38**(2): p. 151-61.
51. Morgan, X.C., et al., *Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment.* Genome Biol, 2012. **13**(9): p. R79.
52. Sanchez, E., et al., *Duodenal-mucosal bacteria associated with celiac disease in children.* Appl Environ Microbiol, 2013. **79**(18): p. 5472-9.
53. Cario, E., *Bacterial interactions with cells of the intestinal mucosa: Toll-like receptors and NOD2.* Gut, 2005. **54**(8): p. 1182-93.
54. Pierik, M., et al., *Toll-Like receptor-1, -2, and -6 polymorphisms influence disease extension in inflammatory bowel diseases.* Inflamm Bowel Dis, 2006. **12**: p. 1-8.
55. Ogura, Y., et al., *A frameshift mutation inNOD2 associated with susceptibility to Crohn's disease* Nature, 2001. **411**: p. 603-606.
56. Turnbaugh, P.J., et al., *The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice.* Sci Transl Med, 2009. **1**(6): p. 6ra14.
57. Turnbaugh, P.J., et al., *An obesity-associated gut microbiome with increased capacity for energy harvest.* Nature, 2006. **444**(7122): p. 1027-31.
58. Backhed, F., et al., *Mechanisms underlying the resistance to diet-induced obesity in germ-free mice.* Proc Natl Acad Sci USA, 2007. **104**(3): p. 979-84.
59. Scott, K.P., S.H. Duncan, and H.J. Flint, *Dietary fibre and the gut microbiota.* Nutrition Bulletin, 2008. **33**: p. 201-211.
60. Zhang, C., et al., *Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice.* ISME J, 2010. **4**(2): p. 232-41.
61. Armougom, F., et al., *Monitoring bacterial community of human gut microbiota reveals an increase in Lactobacillus in obese patients and Methanogens in anorexic patients.* PLoS One, 2009. **4**(9): p. e7125.
62. Schwiertz, A., et al., *Microbiota and SCFA in lean and overweight healthy subjects.* Obesity (Silver Spring), 2010. **18**(1): p. 190-5.
63. Duncan, S.H., et al., *Human colonic microbiota associated with diet, obesity and weight loss.* Int J Obes (Lond), 2008. **32**(11): p. 1720-4.
64. Zhang, H., et al., *Human gut microbiota in obesity and after gastric bypass.* Proc Natl Acad Sci U S A, 2009. **106**(7): p. 2365-70.
65. Zoetendal, E.G., et al., *The host genotype affects the bacterial community in the human gastrointestinal tract.* 2001.
66. Million, M., et al., *Obesity-associated gut microbiota is enriched in Lactobacillus reuteri and depleted in Bifidobacterium animalis and Methanobrevibacter smithii.* Int J Obes (Lond), 2012. **36**(6): p. 817-25.
67. Erejuwa, O.O., S.A. Sulaiman, and M.S. Ab Wahab, *Modulation of gut microbiota in the management of metabolic disorders: the prospects and challenges.* Int J Mol Sci, 2014. **15**(3): p. 4158-88.
68. Kalliomaki, M., et al., *Early differences in fecal microbiota composition in children may predict overweight.* Am J Clin Nutr, 2008. **87**: p. 534 –538.
69. Roesch, L.F., et al., *Culture-independent identification of gut bacteria correlated with the onset of diabetes in a rat model.* ISME J, 2009. **3**(5): p. 536-48.
70. Murri, M., et al., *Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study.* BMC Med, 2013. **11**: p. 46.
71. Larsen, N., et al., *Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults.* PLoS ONE, 2010. **5**(2): p. e9085.
72. Qin, J., et al., *A metagenome-wide association study of gut microbiota in type 2 diabetes.* Nature, 2012. **490**(7418): p. 55-60.
73. Rodes, L., *Effect of probiotics Lactobacillus and Bifidobacterium on gut-derived lipopolysaccharides and inflammatory cytokines: an in vitro study using a human colonic microbiota model.* Journal of Microbiology and Biotechnology, 2013. **23**(4): p. 518-526.
74. Cani, P.D., et al., *Changes in gut microbiota control inflammation in obese mice through a mechanism involving GLP-2-driven improvement of gut permeability.* Gut, 2009. **58**(8): p. 1091-103.

75. Chan, A.T. and E.L. Giovannucci, *Primary prevention of colorectal cancer.* Gastroenterology, 2010. **138**(6): p. 2029-2043 e10.
76. Huycke. M. and H.R. Gaskins, *Commensal bacteria, redox stress, and colorectal cancer: mechanisms and models.* Exp Biol Med 2004. **229**: p. 586 - 597.
77. Devillard, E., et al., *Metabolism of linoleic acid by human gut bacteria: different routes for biosynthesis of conjugated linoleic acid.* J Bacteriol, 2007. **189**(6): p. 2566-70.
78. Wang, T., et al., *Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers.* ISME J, 2012. **6**(2): p. 320-9.
79. Martin, H.M., et al., *Enhanced Escherichia coli adherence and invasion in Crohn's disease and colon cancer* Gastroenterology, 2004. **127**(1): p. 80-93.
80. Huycke, M.M., Abrams V, and M. D.R., *Enterococcus faecalis produces extracellular superoxide and hydrogen peroxide that damages colonic epithelial cell DNA.* Carcinogenesis, 2002. **23**: p. 529–536.
81. Wells, J.E.a.P.B.H., *Identification and characterization of a bile acid 7α-dehydroxylation operon in Clostridium sp. strain TO-931, a highly active 7α-dehydroxylating strain isolated from human feces.* Appl Environ Microbiol, 2000. **66**(3): p. 1107-1113.
82. Abdulamir, A.S., R.R. Hafidh, and F. Abu Bakar, *The association of Streptococcus bovis/gallolyticus with colorectal tumors: the nature and the underlying mechanisms of its etiological role.* J Exp Clin Cancer Res, 2011. **30**: p. 11.
83. Trynka, G., C. Wijmenga, and D.A. van Heel, *A genetic perspective on coeliac disease.* Trends Mol Med, 2010. **16**(11): p. 537-50.
84. Nadal, I., et al., *Imbalance in the composition of the duodenal microbiota of children with coeliac disease.* J Med Microbiol, 2007. **56**(Pt 12): p. 1669-74.
85. Schippa, S., et al., *A distinctive 'microbial signature' in celiac pediatric patients.* BMC Microbiol, 2010. **10**: p. 175.
86. Sellitto, M., et al., *Proof of concept of microbiome-metabolome analysis and delayed gluten exposure on celiac disease autoimmunity in genetically at-risk infants.* PLoS One, 2012. **7**(3): p. e33387.
87. De Palma, G., et al., *Influence of milk-feeding type and genetic risk of developing coeliac disease on intestinal microbiota of infants: The PROFICEL study.* PLoS ONE, 2012. **7**(2): p. e30791.
88. Kalliomaki, M., et al., *Expression of microbiota, Toll-like receptors, and their regulators in the small intestinal mucosa in celiac disease.* J Pediatr Gastroenterol Nutr, 2012. **54**(6): p. 727-32.
89. Liu, Y., et al., *Interspecies competition and inhibition within the oral microbial flora: environmental factors influence the inhibition of Streptococcus mutans by Streptococcus oligofermentans.* Eur J Oral Sci, 2012. **120**(3): p. 179-84.
90. Tong, H., *Streptococcus oligofermentans sp. nov., a novel oral isolate from caries-free humans.* International Journal of Systematic and Evolutionary Microbiology, 2003. **53**(4): p. 1101-1104.
91. Riccia, D.N., et al., *Anti-inflammatory effects of Lactobacillus brevis (CD2) on periodontal disease.* Oral Dis, 2007. **13**(4): p. 376-85.
92. Sliepen, I., et al., *Microbial interactions influence inflammatory host cell responses.* J Dent Res, 2009. **88**(11): p. 1026-30.
93. Teughels, W., et al., *Clinical and microbiological effects of Lactobacillus reuteri probiotics in the treatment of chronic periodontitis: a randomized placebo-controlled study.* J Clin Periodontol, 2013. **40**(11): p. 1025-35.
94. Collins, S.M., *A role for the gut microbiota in IBS.* Nat Rev Gastroenterol Hepatol, 2014. **11**(8): p. 497-505.
95. Ohman, L. and M. Simren, *Intestinal microbiota and its role in irritable bowel syndrome (IBS).* Curr Gastroenterol Rep, 2013. **15**(5): p. 323.
96. Young, V.B. and T.M. Schmidt, *Antibiotic-associated diarrhea accompanied by large-scale alterations in the composition of the fecal microbiota.* Journal of Clinical Microbiology, 2004. **42**(3): p. 1203-1206.
97. Angriman, I., M. Scarpa, and I. Castagliuolo, *Relationship between pouch microbiota and pouchitis following restorative proctocolectomy for ulcerative colitis.* World J Gastroenterol, 2014. **20**(29): p. 9665-74.

98. Stewart, C.J., et al., *The preterm gut microbiota: changes associated with necrotizing enterocolitis and infection.* Acta Paediatr, 2012. **101**(11): p. 1121-7.
99. Jensen, E.T., et al., *Early life exposures as risk factors for pediatric eosinophilic esophagitis.* J Pediatr Gastroenterol Nutr, 2013. **57**(1): p. 67-71.
100. Tilg, H. and A.R. Moschen, *Gut microbiota: Malnutrition and microbiota - a new relationship?* 2013. **10**(5): p. 261-262.
101. Govoni, M., et al., *The increase in plasma nitrite after a dietary nitrate load is markedly attenuated by an antibacterial mouthwash.* Nitric Oxide, 2008. **19**(4): p. 333-7.
102. Wold, A.E., *The hygiene hypothesis revised: is the rising frequency of allergy due to changes in the intestinal flora?* Allergy, 1998. **53**(46): p. 20-25.
103. Trompette, A., et al., *Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis.* Nat Med, 2014. **20**(2): p. 159-66.
104. Bisgaard, H., et al., *Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age.* J Allergy Clin Immunol, 2011. **128**(3): p. 646-52 e1-5.
105. Kamada, N., et al., *Role of the gut microbiota in immunity and inflammatory disease.* Nat Rev Immunol, 2013. **13**(5): p. 321-35.
106. Malkki, H., *Neurodevelopmental disorders: Human gut microbiota alleviate behavioural symptoms in a mouse model of autism spectrum disorder.* 2014. **10**(2): p. 60-60.
107. Hsiao, E.Y., et al., *Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders.* Cell, 2013. **155**(7): p. 1451-63.
108. Foster, J.A. and K.A. McVey Neufeld, *Gut-brain axis: how the microbiome influences anxiety and depression.* Trends Neurosci, 2013. **36**(5): p. 305-12.

Chapter 3

The effect of prebiotic fortified infant formulas on microbiota composition and dynamics in early life

Klaudyna Borewicz, Maria Suarez-Diez, Christine Hechler, Roseriet Beijers, Carolina de Weerth, Ilja Arts, John Penders, Carel Thijs, Arjen Nauta, Cordula Lindner, Ellen van Leusen, Hauke Smidt

Manuscript in preparation

Abstract

Gastrointestinal (GI) microbiota composition differs between breastfed and formula fed infants. Today's infant formulas are often fortified with prebiotics to better mimic the functional properties of human milk with respect to its effect on GI microbiota composition and function. We used Illumina HiSeq sequencing of PCR-amplified 16S ribosomal RNA gene fragments to investigate the composition of faecal microbiota in 2-12 week old Dutch infants receiving either breastmilk, infant formulas fortified with prebiotics, or mixed feeding. We compared these results with results obtained from infants fed traditional formulas which did not contain added prebiotics and were commonly used in the Netherlands in 2002-2003. Despite a high natural variability in the faecal microbiota composition of all infants, including those born vaginally and exclusively breastfed, we showed that today's formulas had a strong bifidogenic effect as compared to traditional formulas, and they also resulted in altered patterns of microbial colonisation within the developing infant GI tract. We identified three microbial states in the first 12 weeks of life, with a gradual transition pattern towards a bifidobacteria dominated state in breastfed infants. In infants receiving fortified formulas as their only food this transition was accelerated towards a bifidobacteria dominated state, whereas in infants receiving mixed feeding the transition was delayed.

Introduction

Microbial colonisation of an infant's gastrointestinal (GI) tract starts before or at birth and progresses in a step-wise fashion during the postnatal period [1-3]. Many environmental factors may affect GI microbiota composition and its development during early life [4-8]. These early life exposures and associated GI microbiota perturbations have been linked with changes in immune development leading to potentially serious and lifelong health effects. For example, previous studies suggested that infants fed formula were at higher risk of nutritional deficiencies, asthma, atopy, obesity, developing metabolic syndrome, coeliac disease, diabetes and other diseases, as compared to breastfed infants [5, 7, 9]. Some of the early life exposures cannot be avoided, however, the use of formula feeding is, at least in some cases, a choice made by the parents [10]. Over the last century, when formula feeding became more popular worldwide [11], intensive research led to developing infant formulas that are increasingly similar to human milk with regard to nutrient composition and function. However, they are certainly not identical, and breastmilk with its complex composition still remains a golden standard for infant nutrition [12, 13]. Human milk contains a wide range of compounds with modulatory effect on the infant intestinal microbiota [10, 14]. One of the prevalent and important groups of components are the human milk oligosaccharides (HMOs) that have unique nutritional and functional properties [15]. Their importance as growth factors for a "bifidus flora" was identified more than a hundred years ago [16, 17]. Various studies since then confirmed that the GI microbiota of breastfed infants is dominated by bifidobacteria, as compared to formula fed infants, a fact which has been attributed to the presence of HMOs in the human milk, and their lack in infant formulas [18-23]. Taking into account both, the wide use of infant formulas, and the growing evidence of the importance of GI microbiota for health throughout life, it became clear that the functional prebiotic properties of infant formulas needed to be addressed. Thus, formulas nowadays are often fortified with prebiotics. In European countries these include mostly short chain galacto-oligosaccharides (scGOS) alone, or in a mixture with a chicory root derived inulin containing long chain fructo-oligosaccharides (lcFOS) [24, 25]. Prebiotics mimic the bifidogenic effect of HMOs in human milk and have been associated with improved immunity, bowel function and other health benefiting effects in infants [15, 26, 27]. However, the exact effect of these functional alternatives on the GI tract microbial ecosystem, most importantly with respect to the dynamics of bacterial colonisation in early life are not yet well understood and should be investigated.

Here we present the results of a longitudinal study in which we compared the colonisation patterns of breastfed and formula fed infants. We assessed the microbiota composition in faecal samples from two, six and 12 weeks old infants born between years 2015-2016 and receiving commercial formulas fortified with GOS and/or FOS. We compared those results with the feacal microbiota composition of one month old infants born in 2002-2003 and fed commercial infant formulas purchased during those years. In both studies, infants received formulas that were available on the Dutch market at the time the samples were collected. We show that the new type of formulas have a bifidogenic effect on infant GI microbiota, however, they also result in altered dynamics of bacterial colonisation during the first 12 weeks of life as compared to breastfed infants.

Materials and Methods

Study Description

The analyses described here are part of the BINGO and KOALA birth cohort studies. All infants included in this analysis were healthy, born at term and did not receive oral antibiotic treatment during the study period. The BINGO (Dutch acronym for Biological Influences on Baby's Health and Development) cohort is an ongoing longitudinal study investigating prenatal predictors of infant health and development. This study was approved by the ethical committee of the Faculty of Social Sciences of the Radboud University [ECSW2014-1003-189]. The BINGO study design, infant recruitment criteria and sampling procedures can be found at http://www.bingo-onderzoek.nl/deelname/. Both, the infant faecal samples and breastmilk samples were collected by the mothers within a period of 48 h, when infants were two, six and 12 week old. Breastmilk samples were collected into sterile 30 ml containers. Infant stool samples were collected from a diaper using a sterile stool vial (80×16.5mm; cat#:80.623.022; Sarstedt; Nümbrech, Germany) with a spoon attached to the lid. The mothers were asked to immediately store the milk and the faecal samples in their home freezers (i.e., fresh frozen collection) until collected by the researcher. After collection, samples were stored at -80 ^{o}C until further processing and analysis.

The design, selection criteria and faeces collection procedure of the KOALA Birth Cohort Study (Dutch acronym for: Child, Parents and Health: Lifestyle and Genetic Constitution) have been described elsewhere and the study was approved by the Ethics Committee of the University Hospital of Maastricht [28, 29]. In brief, the KOALA study included two recruitment groups of healthy pregnant women in the South of the Netherlands. The first group (n=2343) was characterised by a conventional lifestyle, whereas families included in the second group (n=491) were considered to have an alternative lifestyle that could involve dietary habits (vegetarian, organic), child-rearing practices and/or low use of antibiotics, and were recruited through alternative channels, such as posters in organic food shops, anthroposophic doctors and midwives. Exclusion criteria were prematurity (birth before 37 weeks of gestation), twins, congenital abnormalities related to growth, and administration of antimicrobial agents before faeces collection. Infant faecal samples were collected by the parents at approximately one month postpartum by removing a sample from a diaper into a sterile tube. Breastmilk samples were collected on the same day by the mothers or research nurses at the participants' homes. Briefly, mothers received a sterile 50 mL tube (Cellstar PP-test tubes, Greiner bio-one, Kremsmünster, Austria) and were instructed to collect the milk sample in the morning, before breastfeeding their child, from the contra-lateral breast (since the last feeding) and to keep the tube in the refrigerator (±4 °C) until it was collected by one of the researchers. If the mother was not able to collect the milk sample by herself (with or without a pumping regimen), an electric breast pump (Medela, Baar, Switzerland) was used with the help of one of the researchers (within the same day). During transport, the milk samples were stored in a cooler (Coleman Company Inc., Breda, the Netherlands) on packed ice (±4 °C) until processing on the same day. The sample was centrifuged (400 × g, 12 min, no brake, 4 °C) to separate the lipid and aqueous fraction. The lipid layer was trimmed off with a pipette and released in plastic storage vials (Sarstedt, Nümbrecht, Germany). The aqueous fraction was

poured in other vials with another pipette. The remaining debris was not used to avoid contamination with cell fragments. All fractions were stored at –80 °C in the European Biobank, Maastricht.

In both studies, parents were asked to complete a questionnaire including information regarding infant's diet (breastmilk only, formula only, or mixed feeding), and the type of infant formula used. In the KOALA study the information about feeding mode referred to the time from birth until sample collection, whereas in the BINGO cohort it referred to feeding mode during each time interval, i.e. from birth to the first time point, and between subsequent time points. All formula fed (FF) infants received commercially available formulas purchased in the Netherlands between 2015-2016 (BINGO cohort), and between 2002-2003 (KOALA cohort). Modern infant formulas used in the BINGO cohort were all fortified with GOS or GOS and FOS, whereas only 17 of 103 FF infants in the KOALA cohort were confirmed to receive formulas containing GOS or GOS and FOS, while remaining infants were fed traditional formulas with no added prebiotics. The prebiotic fortified infant formulas typically contain scGOS (0.24 - 0.50 g/100ml), or mixtures -scGOS:lcFOS (9:1); 0.6 g/100ml [25, 30]. A total of 449 faecal samples were analysed: 210 samples from 77 infants from the BINGO cohort, and 239 samples from the KOALA cohort infants (Table 1). In the BINGO study three children (total 6 samples) were excluded from the analysis due to missing information regarding the feeding mode. The remaining 204 faecal samples were analysed for microbial composition. Eight children from the BINGO study were born via C-section and 66 children were born vaginally (185 faecal samples). The proportion of samples from C-section infants in each feeding mode group was 0.08 for breastfed (BF) and for FF infants, and 0.17 in the mixed fed (MF) group. In total 60 infants could be followed at all three time points. In the KOALA study 239 samples were analysed for microbiota composition from 239 infants at approximately one month of age (mean age = 31 days, *SD*=5). Of all infants, 121 were BF, 103 were FF and 15 were MF. The proportion of samples from C-section infants in each feeding mode group was 0.08 in BF, 0.13 in FF, and 0.12 in MF.

Table1. Sample (BINGO) and infant (KOALA) characteristics

BINGO	**Total N=204 samples (77 infants)**		
Feeding	Breastfed (n=156)	Formula fed (n=25)	Mixed fed (n=23)
2 weeks	52	4	7
6 weeks	54	8	6
12 weeks	50	13	10
Delivery mode			
Vaginal:	139	23	19
C-section:	14	2	4
No record:	3	0	0
Gender			
Male:	84	8	14
Female:	69	17	9
No record:	3	0	0
KOALA	**Total N=239 samples (239 infants)**		
Feeding	Breastfed (n=121)	Formula fed (n=103)	Mixed fed (n=15)
Delivery mode			
Normal vaginal:	100	80	12

Assisted vaginal:	11	4	1
C-section:	10	16	2
No record:	0	3	0
Delivery Place			
Home:	72	36	2
Hospital:	49	63	13
No record:	0	4	0
Gender			
Male:	62	61	10
Female:	59	42	5
Gestation (weeks)			
Mean ±SEM ±SD	40.24 ±1.2 ±1.2	39.85 ±0.1 ±1.3	39.91 ±0.47 ±1.8
Birth Weight (g)			
Mean ±SEM ±SD	3651 ±43.48 ±476	3505 ±44.05 ±445	3543 ±71.27 ±257
Age at collection (d)			
Mean ±SEM ±SD	32.56 ±0.5 ±5.4	31.39 ±0.49 ±4.9	35.67 ±1.3 ±4.9
Health at collection			
Sick:	3	3	0
Not Sick:	118	100	15

DNA extraction

In the KOALA cohort, the total DNA was extracted from the stool samples as previously described [8], using the double bead-beating procedure followed by QIAamp DNA stool mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. In the BINGO cohort, total DNA extraction was done using the Maxwell® 16 Total RNA system (Promega) with Stool Transport and Recovery Buffer (STAR; Roche Diagnostics Corporation, Indianapolis, IN). Briefly, 0.1-0.15g of faecal sample was homogenized in Precellys® 24 homogenizer (Precellyis® 24, Bertin Technologies, France) with 0.25 g of sterilized 0.1 mm zirconia beads and three glass beads (2.5 mm) in 350 µL STAR buffer for 3x1 min at 5.5 ms, with 10 s cooling breaks in RT in-between. Samples were incubated with shaking at 100 rpm for 15 min at 95 ºC and pelleted by 5 min centrifugation at 4 ºC and 14000 × g. Supernatant was removed and the pellets were processed again using 200 µL of fresh STAR buffer. Supernatant was removed, pooled with the first supernatant, and 250 µL was used for purification with the Maxwell® 16 Tissue LEV Total RNA Purification Kit (AS1220) customized for DNA extraction in combination with the STAR buffer following manufacturer's instructions. DNA was eluted with 50 µL of DNAse and RNAse free water (Qiagen, Hilden, Germany). DNA concentrations were measured with a NanoDrop ND-1000 spectrophotometer (NanoDrop® Technologies, Wilmington, DE, USA) and adjusted to 20 ng/µL with DNAse and RNAse free water.

Library preparation and sequencing

The V4 region of 16S ribosomal RNA (rRNA) gene was amplified in duplicate PCR reactions, each in a total volume of 50 µL and containing 5-20 ng (KOALA) or 20 ng (BINGO) of template DNA. Each sample was amplified with 200 nM of uniquely barcoded primers 515F-

n (5'-GTGCCAGCMGCCGCGGTAA-) and 806R-n (5'-RGGATTAGATACCC), 10 µl of 5x HF buffer (Finnzymes, Vantaa, Finland), 200 µM dNTP Mix (Roche Diagnostics GmbH, Mannheim, Germany), 1 U Phusion® Hot Start II High Fidelity DNA Polymerase (Finnzymes, Vantaa, Finland) and 36.5 µL of DNAse and RNAse free water [31]. The amplification program included 30 s initial denaturation step at 98 ºC, following by 25 cycles of denaturation at 98 ºC for 10 s, annealing at 56 ºC for 10 s and elongation at 72 ºC for 10 s, and a final extension at 72 ºC for 7 min. The PCR product presence and size (~290 bp) was confirmed with gel electrophoresis using the Lonza FlashGel® System (Lonza, Cologne, Germany). Seventy unique barcode tags were used in each library and artificial control (Mock) communities were included for quality control. PCR products were purified with the HighPrep® PCR kit (MagBio Genomics, Alphen aan den Rijn, Netherlands), and DNA concentrations were measured with the Qubit® dsDNA BR Assay Kit (Life Technologies, Leusden, Netherlands). Hundred nanograms of each barcoded sample was pooled together, the amplicon pool was concentrated with the HighPrep® PCR kit to a 20 µL volume, the concentration was measured with the Qubit® dsDNA BR Assay Kit and adjusted to 100 ng/µL final concentration. The libraries were sent for adapter ligation and Illumina HiSeq2000 sequencing at GATC-Biotech, Konstanz, Germany.

Data analysis

The 16S rRNA sequencing data was processed and analysed using the NG-Tax analysis pipeline [31]. In brief, libraries were filtered to contain only read pairs with perfectly matching barcodes that were subsequently used to separate reads by sample. Operational taxonomic units (OTUs) were assigned using an open reference approach and SILVA_111_SSU 16S rRNA gene reference database [32]. Diversity analyses were carried out in QIIME on rarefied data with OTU cut-off of 2500 [33, 34], and multivariate analysis was done in Canoco5 [35]. Principal Component Analysis (PCA) and Redundancy Analysis (RDA) were performed using the log transformed genus level relative abundances data obtained from NG-Tax. Alpha diversity indices (Shannon, Chao1, and PD Whole Tree) for each sample were calculated in QIIME using genus level (L6) data obtained with NG-Tax. Alpha diversity index group comparisons were done in GraphPad (GraphPad Prism version 5.00 for Windows, GraphPad Software, San Diego, CA, USA, www.graphpad.com). When values within a group used in the analysis did not pass the Shapiro-Wilk normality test, the Kruskal–Wallis analysis was used to compare diversity indexes. If the values were normally distributed in all groups, Two-Way ANOVA analysis was used for group comparison. Statistical differences in relative abundance of genus level taxa between three different age or feeding modes were assessed with Kruskal–Wallis test using QIIME. For pairwise comparisons the Wilcoxon test was used when possible, otherwise the Kruskal–Wallis test was applied (QIIME). Clusters were identified based on genus level microbial abundance data using Dirichlet Multinomial Mixture (DMM) modelling as described previously [36]. The number of Dirichlet components was selected by inspection of the fit of the model to the count data for varying numbers of components (1 to 7). Goodness of fit was assessed using the Laplace and the Akaike information criteria. Finally, each sample was assigned to the component for which it had the largest fitted value. These analyses were performed in R (version 3.3.1) using the DirichletMultinomial R package [37].

Nucleotide sequences

KOALA data sets cannot be made publicly available due to data confidentiality and the potential to identify individual study participants from the data. Data are available to the research community through the Dataverse repository (URL hdl:10411/CEGPGR) upon request to Prof. C. Thijs of the KOALA Study Management Committee at: Maastricht University, Department of Epidemiology, PO Box 616, 6200 MD Maastricht, The Netherlands, e-mail: c.thijs@maastrichtuniversity.nl, tel: +31(0)43 3882389. Similarly, BINGO data sets cannot be made publicly available due to the data being part of an ongoing longitudinal study. Parts of the data are available to the research community for scientific collaborations upon request to Prof. dr. C. de Weerth at: Radboud University, Department of Developmental Psychology, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands, e-mail: c.deweerth@psych.ru.nl.

Results

A total of 443 samples were included in the analyses: 204 samples (from 74 infants) and 239 samples (from 239 infants) from the BINGO and KOALA cohort, respectively. A total of 28,955,759 sequencing reads were obtained from the BINGO cohort samples, with the per sample counts ranging from 5,215 to 721,990 (Mean = 141,940; *SD*= 126,570), with 95% of samples having at least 20,000 reads. Sequencing of the KOALA cohort samples resulted in a total of 30,132,625 sequencing reads ranging from 1,380 to 448,285 per sample (Mean =126,078, *SD*=84,356), with 95% of samples having at least 25,000 reads. Taxonomic classification of OTUs was done using the NG-Tax pipeline against a customised SILVA database [31], and it resulted in detection of five different phyla, namely Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria, and Verrucomicrobia. At the genus level, the most abundant taxa were *Bifidobacterium, Bacteroides, Streptococcus, Escherichia-Shigella*, and an unassigned genus within the family Enterobacteriaceae (Table S1).

Being overall the most abundant genus-level taxonomic group in infant faecal microbiota, we first assessed potential differences in relative abundance of bifidobacteria with age and different feeding modes. In the BINGO study we observed an age related increase in the average relative abundance of bifidobacteria in the BF infants (Kruskal-Wallis; p=0.01). In contrast, in FF and MF infants the relative abundance of bifidobacteria fluctuated, and the differences were not significant between different age groups (Kruskal-Wallis: p>0.05). In the BINGO study, MF resulted in reduction in relative abundance of bifidobacteria, as compared to FF infants (Kruskal-Wallis: p=0.0078). In KOALA, MF resulted in significant reduction in the relative abundance of bifidobacteria as compared to BF (Kruskal-Wallis: FDR=0.05, p=0.00078), but not in comparison to the FF group (Kruskal-Wallis: p>0.05). In the BINGO study, FF infants at six weeks of age showed nearly a 20% points higher average relative abundance of bifidobacteria as compared to BF infants, a trend not observed in the KOALA study, where the abundance of bifidobacteria in the FF infants was 15% points lower than that in the corresponding BF group.

To identify genus level taxa that were significantly different between BF and FF infants in each study cohort (Figure 1) we used the Wilcoxon test. In the BINGO study 12 genus level groups differed significantly between BF and FF infants when all age groups were analysed together (FDR<0.05, Figure 1), but at six weeks of age only *Blautia* was identified as significantly different (FDR=0.0001) when BF and FF groups were compared. In infants included in the KOALA study, the relative abundances of 19 genus level taxa were statistically different between both feeding types (FDR <0.05, Figure 1).

In both cohorts formula feeding significantly (FDR<0.05) increased relative abundance of *Akkermansia, Enterococcus,* Peptostreptococcaceae *Incertae Sedis,* and Erysipelotrichaceae *Incertae Sedis,* and significantly decreased relative abundance of *Staphylococcus* and *Haemophilus*, as compared to the corresponding BF groups. In addition, in the BINGO cohort only, formula feeding significantly increased relative abundance of *Blautia, Dorea, Granulicatella, Eubacterium, Catenibacterium,* and decreased relative abundance of *Parabacteroides* as compared to the BF group. In the KOALA cohort, formula feeding significantly increased relative abundance of *Barnesiella, Alistipes, Escherichia-Shigella, Veillonella, Flavonifractor, Clostridium,* Lachnospiraceae *Incertae Sedis,* and unidentified genera within the families Ruminococcaceae and Enterobacteriaceae, whereas it decreased *Bifidobacterium, Lactobacillus, Halomonas, Aeribacillus* as compared to the BF group.

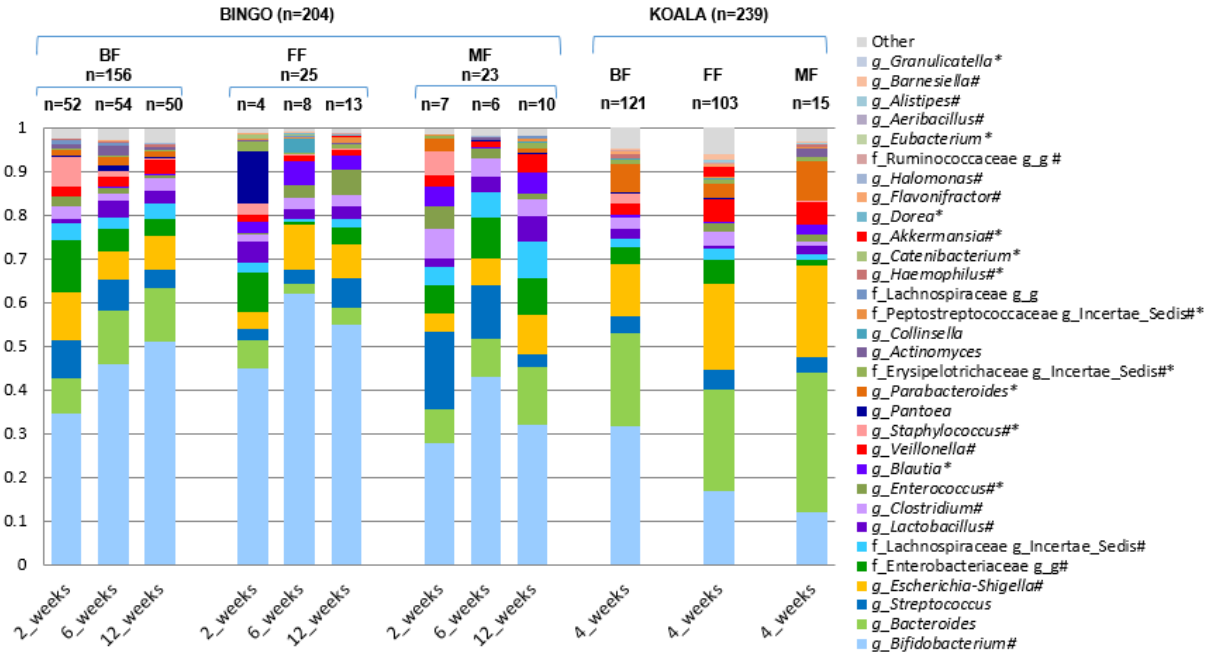
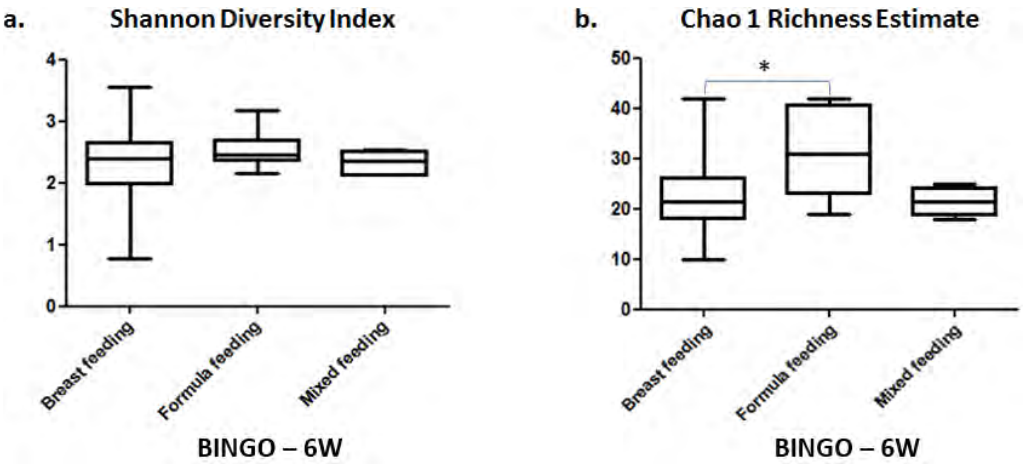


Figure 1. Average relative abundance of genus level taxa in faeces of infants included in the BINGO and KOALA cohorts that were either breastfed (BF), formula fed (FF) or fed both breastmilk and formula (mixed fed, MF). When the taxonomic assignment could not be made at genus level, the lowest classifiable taxonomy assignment is used instead and unidentified genus is indicated with "g_g". Taxa that significantly differ in their relative abundance between BF and FF infants are indicated with * (BINGO) and # (KOALA).

In the BINGO study cohort, there were no statistically significant differences in alpha diversity as estimated with PD Whole Tree between age groups, and between BF infants, FF infants and MF infants (data not shown). When Shannon Diversity Index values were used, the difference was significant between BF and MF infants ($p=0.016$) only at week 12 (Figure 2e), but not at week six (Figure 2a), or week two. There was also a significant difference in Chao1 species richness between BF and FF infants at six weeks (Figure 2b), and 12 weeks of age (Figure 2f), but not in infants at two weeks of age. When we analysed the differences between feeding modes for all age groups combined, there was a significant difference in microbial diversity between BF and MF infants as estimated with Shannon diversity index ($p=0.008$) (Figure 2g), but not with PD Whole Tree ($p=0.227$). Similarly, there was a significant difference in Chao1 species richness between BF and FF infants ($p=0.0002$) for all ages combined (Figure 2h). The fact that differences in bacterial richness and diversity were only observed with Chao1 and Shannon indices but not when the PD Whole Tree index was used suggests that those differences mostly concern closely related taxa.

In the KOALA study cohort there was a statistically significant difference in diversity estimated with the Shannon Diversity Index between BF and FF infants ($p<0.0001$) and between FF and MF infants ($p=0.035$), but not between BF and MF infants (Figure 2c). Comparison of alpha diversity based on the PD Whole Tree estimates showed no significant differences between infants receiving different diets (data not shown). There was a statistically significant difference in Chao1 bacterial richness between BF and FF infants ($p<0.0001$) (Figure 2d). In general, both, the faecal bacterial diversity and richness were higher in FF infants as compared to BF infants, whereas MF infants showed an intermediate phenotype.



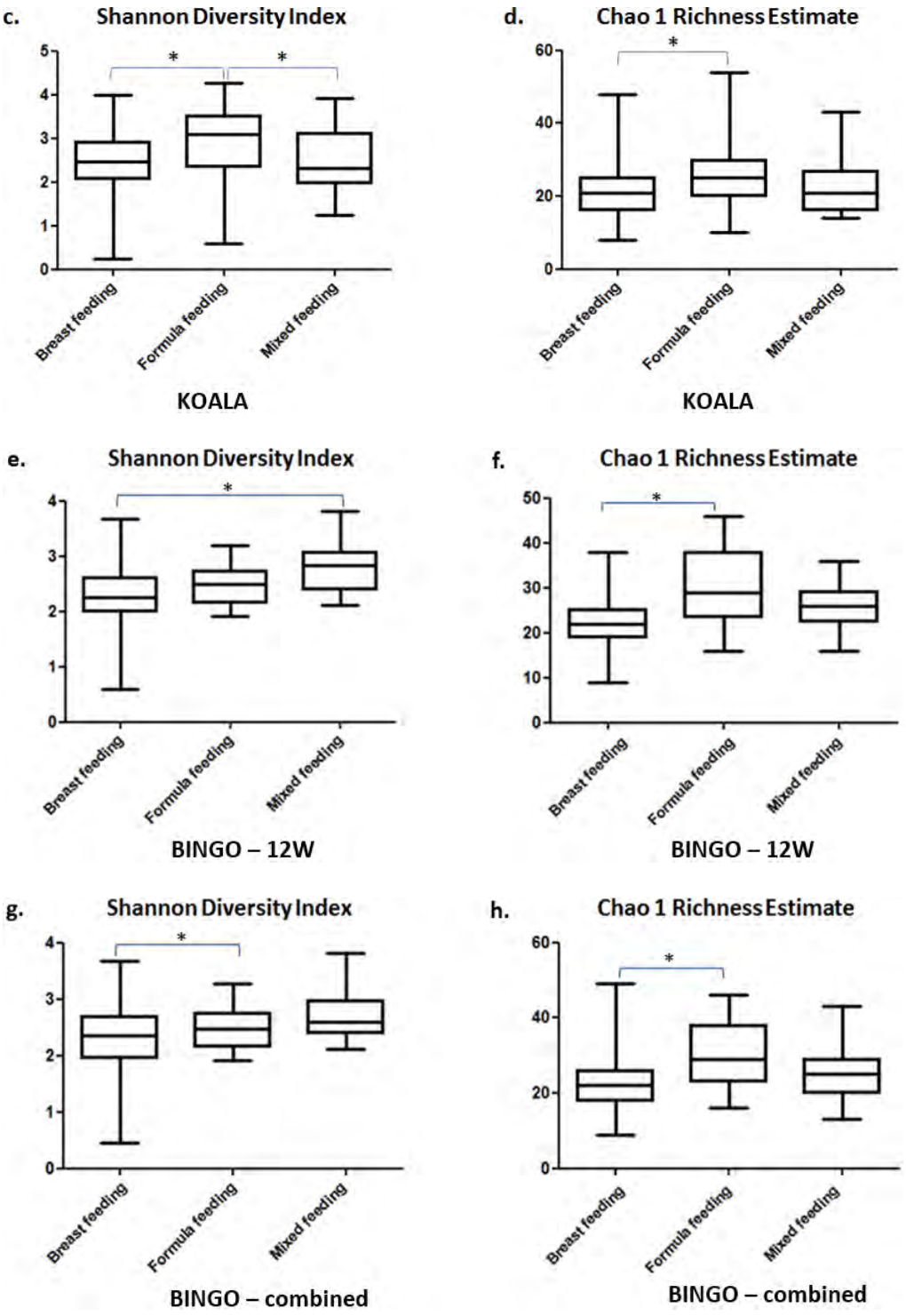


Figure 2. Alpha diversity of faecal microbiota in BF, FF or MF infant groups; a. bacterial diversity in the BINGO study infants at six weeks of age; b. bacterial richness in the BINGO study infants at six weeks of age; c. bacterial diversity in the KOALA study infants at one month of age; d. bacterial richness in the KOALA study infants at one month of age; e. bacterial diversity in the BINGO study infants at 12 weeks of age; f. bacterial richness in the BINGO study infants at 12 weeks of age; g. bacterial diversity in the BINGO study infants at two, six, and 12 weeks of age combined; f. bacterial richness in the BINGO study infants at two, six, and 12 weeks of age combined. *indicates statistically significant difference between groups ($p<0.05$)

In both, BINGO and KOALA sample sets, PCA analysis revealed grouping of samples into FF and BF groups with the MF samples scattered in between (Figure 3a, b, Table S1). This separation was more pronounced among the KOALA cohort infants, and the bacterial taxa driving the separation differed between the studies. Nevertheless, despite a much smaller FF group size in the BINGO cohort, the overall sample distribution pattern on the PCA plots was conserved between the two studies.

a.

b.

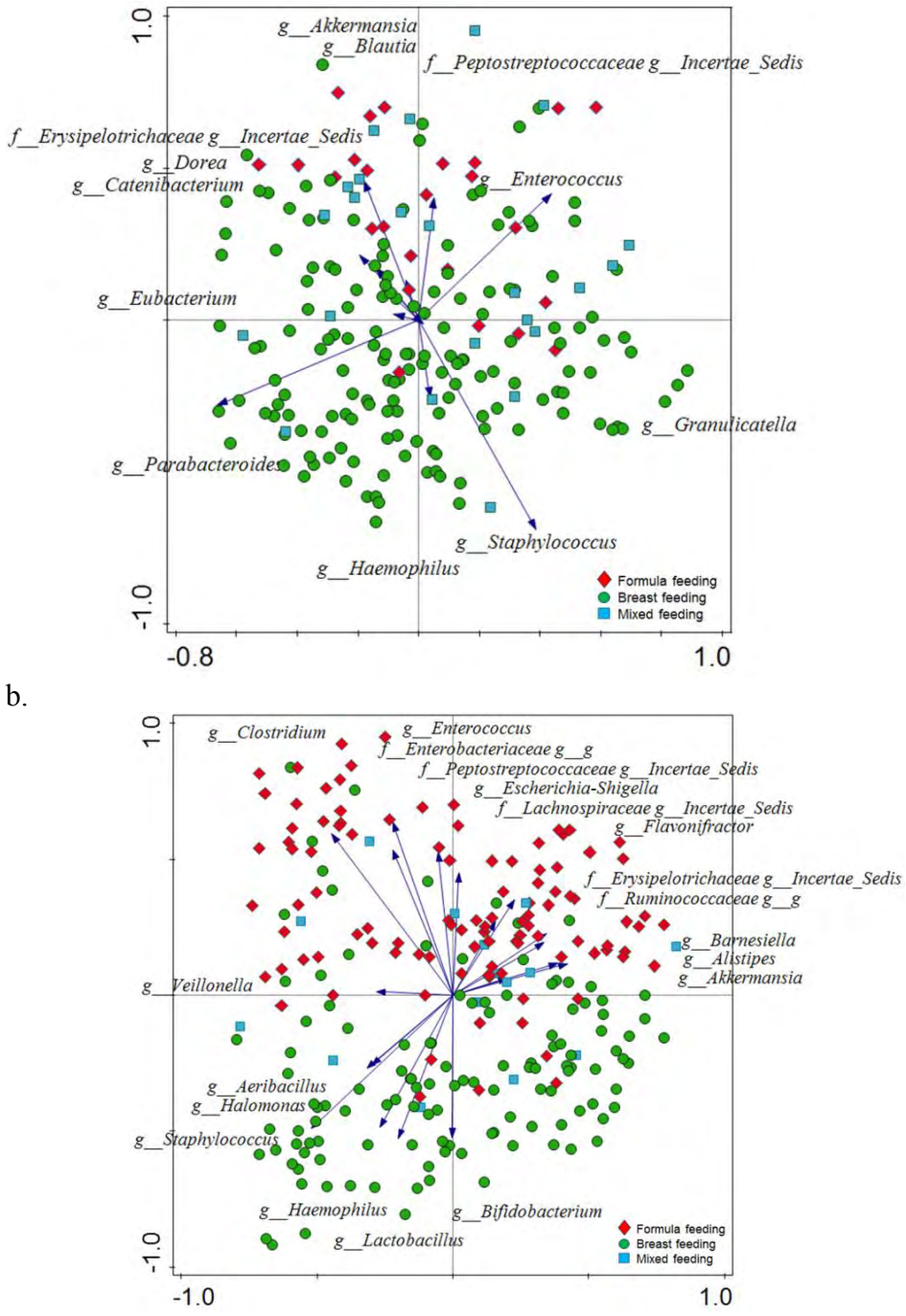
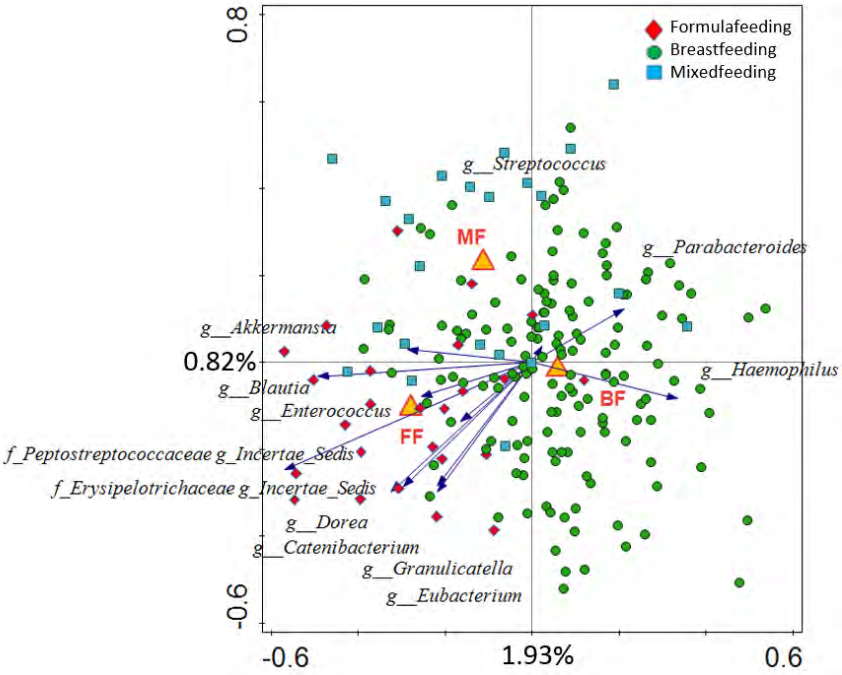


Figure 3. PCA analysis of log transformed genus level relative abundance data. Samples coloured by infant feeding type showing separation between different feeding modes. Microbial groups that significantly differ in relative abundance (Wilcoxon test, FDR $p<0.05$) between BF and FF groups are displayed; a. BINGO study, b. KOALA study

Redundancy analysis (RDA) was used to assess the amount of variation in the microbial composition data which could be explained by feeding mode and additional demographic factors recorded for the two different cohorts. These factors included place and mode of delivery, gender, and medication use (Figure S1a,b). The interactive forward selection method identifies a best subset of variables that explain the variation in the data. In the BINGO cohort, when all samples were analysed together, age, feeding (BF, FF, MF), mode of delivery (C-Section, vaginal), place of delivery (home, hospital, clinic), and medication had a significant effect on microbiota composition (FDR<0.05) and together explained 13.7% of variation. In the KOALA cohort, feeding and delivery mode were selected (FDR<0.05) during interactive forward selection, and they explained 9.5% variation. To investigate the residual effect of feeding mode separately, we repeated the RDA analysis with feeding mode as the main factor and all other factors selected during interactive selection process as covariates. In both, the BINGO cohort and in the KOALA cohort, the effect of feeding was significant (FDR<0.05), and feeding explained 2.9%, and 6.2% of the residual variation, respectively (Figure 4a,b).

a.



b.

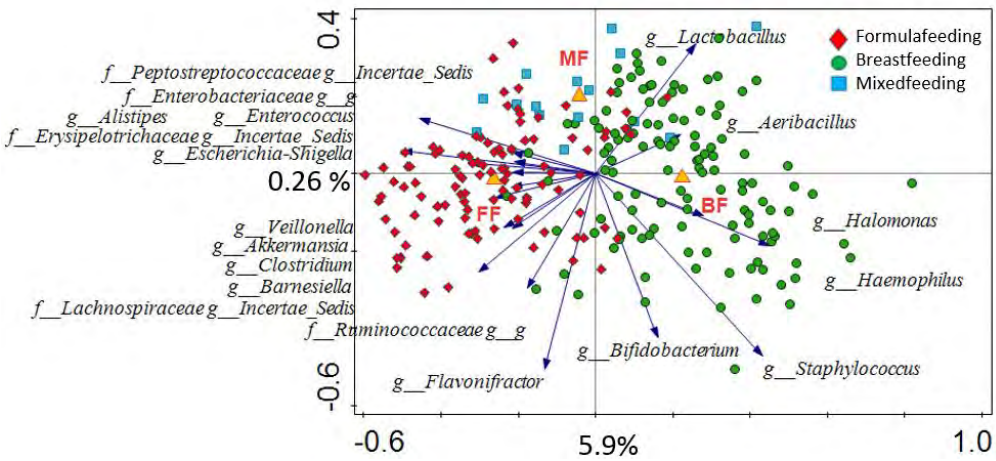


Figure 4. Partial RDA analysis with covariates, using the log transformed genus level relative abundances. Samples coloured by infant feeding type (BF, breastfeeding; FF, formula feeding; MF, mixed feeding) showing separation between different feeding modes. Microbial groups that significantly differ in relative abundance (Wilcoxon test, FDR p<0.05) between BF and FF groups are displayed; a. BINGO study, b. KOALA study

We then applied the same approach at each time point separately to investigate in more detail the effect of different factors (feeding, place and mode of delivery, gender, and medication use) in the BINGO cohort data. The interactive selection showed that at week two, feeding was not a significant factor, but both, the place and the mode of delivery had a significant effect on microbiota (FDR<0.05). At six weeks, the effect of mode of delivery and feeding was significant (FDR<0.05), and RDA analysis using delivery mode as covariate showed that feeding could explain 4.5 % of the residual variation in the microbiota composition. At 12 weeks, the analysis showed a significant effect of feeding (FDR<0.05) and a weak effect of mode of delivery (FDR = 0.056). RDA analysis using delivery mode as covariate showed that feeding could explain 5.9 % of the residual variation in the microbiota composition (Figure 5).

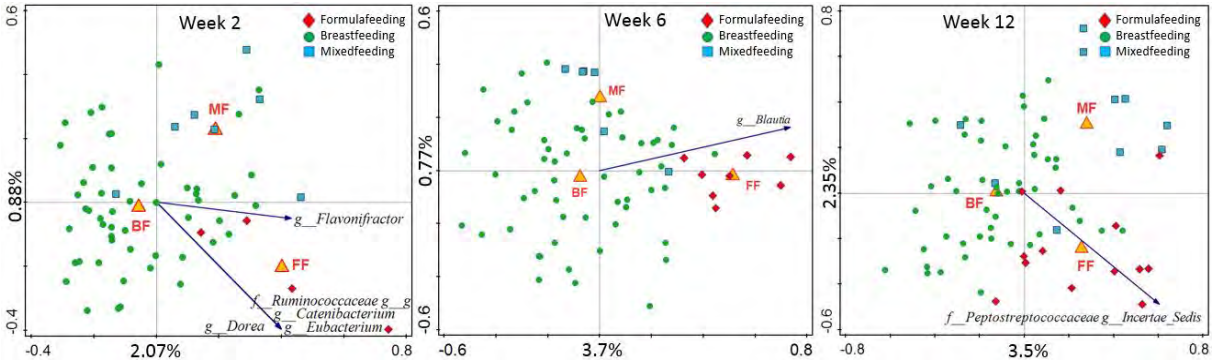


Figure 5. RDA analysis with covariates, using the log transformed genus level relative abundance data at each time point in the BINGO cohort. Samples coloured by infant feeding type (BF, breastfeeding; FF, formula feeding; MF, mixed feeding) showing separation between different feeding modes. Microbial groups that significantly differ in relative abundance (Wilcoxon test, FDR p<0.05) between BF and FF groups at each time point are displayed.

In both studies, RDA analyses showed that breastfeeding and formula feeding resulted in significant differences in microbiota composition (FDR<0.05). In contrast, the effect of mixed feeding was not significant in the KOALA study and in the six week old infants in the BINGO study. However, when samples from all time points in the BINGO study were analysed together, we saw a significant difference between infants receiving MF and the other two feeding groups.

As described above, we observed a large variation in microbiota composition between individual infants in both studies, independent of feeding or delivery mode. Using Dirichlet Multinomial Mixtures (DMM) modelling [36], we assigned samples into three clusters based on the relative abundance of the microbial groups at genus level of classification (Figure 6a).

The clustering was performed independently for both studies. There were some minor differences in the average relative contribution of individual taxa between the three clusters A, B and C obtained from the BINGO and KOALA sample sets, but the overall pattern within each cluster type was preserved (Figure S2). Cluster A contained samples with mixed microbial composition, low relative abundance of *Bifidobacterium* and relatively high proportion of *Streptococcus* and other microbial groups. Cluster B showed high relative abundance of both *Bifidobacterium* and *Bacteroides*, whereas in cluster C *Bifidobacterium* was the dominating genus (Figure 6a). The same cluster pattern was visible when samples were divided into subgroups based on the infants' age or delivery mode (data not shown).

Infant age and feeding mode were associated with cluster assignment of samples in the BINGO study. At two weeks of age, 50% of all samples from BF infants, 75% from FF and 57% of MF infants clustered in group A (Figure 6b), but as the infants aged their faecal microbiota composition was gradually becoming dominated with *Bifidobacterium* (cluster B and C). This gradual transition pattern was clear in BF infants, however, it was distorted in infants who received formula, either as a sole source of food or as supplementary feeding (Figure 6b). In the BINGO cohort at six weeks of age, 87% of FF samples were assigned in *Bifidobacterium* dominated cluster C during the DMM analysis, as compared to only 35% in BF group at that age. Infants who received mixed feeding were more likely to stay in the mixed microbiota cluster A, or the *Bacteroides*/*Bifidobacterium* cluster B, as compared to BF and FF infants. In the subset of the individual infants (n=60) from the BINGO cohort, where samples from all three time points were available, infants were more likely to stay within the same cluster between two and 12 weeks of age, and if they switched to a different cluster group the change was towards the *Bacteroides*/*Bifidobacterium* or *Bifidobacterium* rich clusters B and C (Figure 6c). Of the 13 infants who received formula after the second week of age, seven infants switched into *Bifidobacterium* enriched clusters B or C, while the other six infants remained in the same cluster B or C, or remained in the mixed cluster A. Interestingly, all infants, which remained in cluster A also received mixed feeding. Also, of all infants who were changed from breastfeeding to formula, the change also was associated with a switch from clusters A or B into cluster C. In the KOALA cohort, cluster assignment of BF infants showed a similar result to that of the BINGO cohort at six weeks. However, an opposite trend was found for FF and MF infants, where over 90% of infants could be categorized within clusters A or B, showing low to moderate relative abundance of faecal *Bifidobacterium*.

To discard associations between cluster assignment and other external factors such as delivery mode or place, the clustering analysis was applied to sub-selections of the samples (data not shown). Clustering of the 186 samples from the BINGO study corresponding to the 66 vaginally delivered infants led to the same cluster structure. Same results were obtained when clustering was restricted to the 211 vaginally delivered infants from the KOALA cohort. Restriction of the analysis to the samples from C-section born infants led to no reliable clusters due to the low number of samples. Cluster assignments of samples grouped for other characteristics (delivery place, infant sex or weight at birth) were inspected and compared with overall study distribution, but no significant associations were found ($p>0.05$).

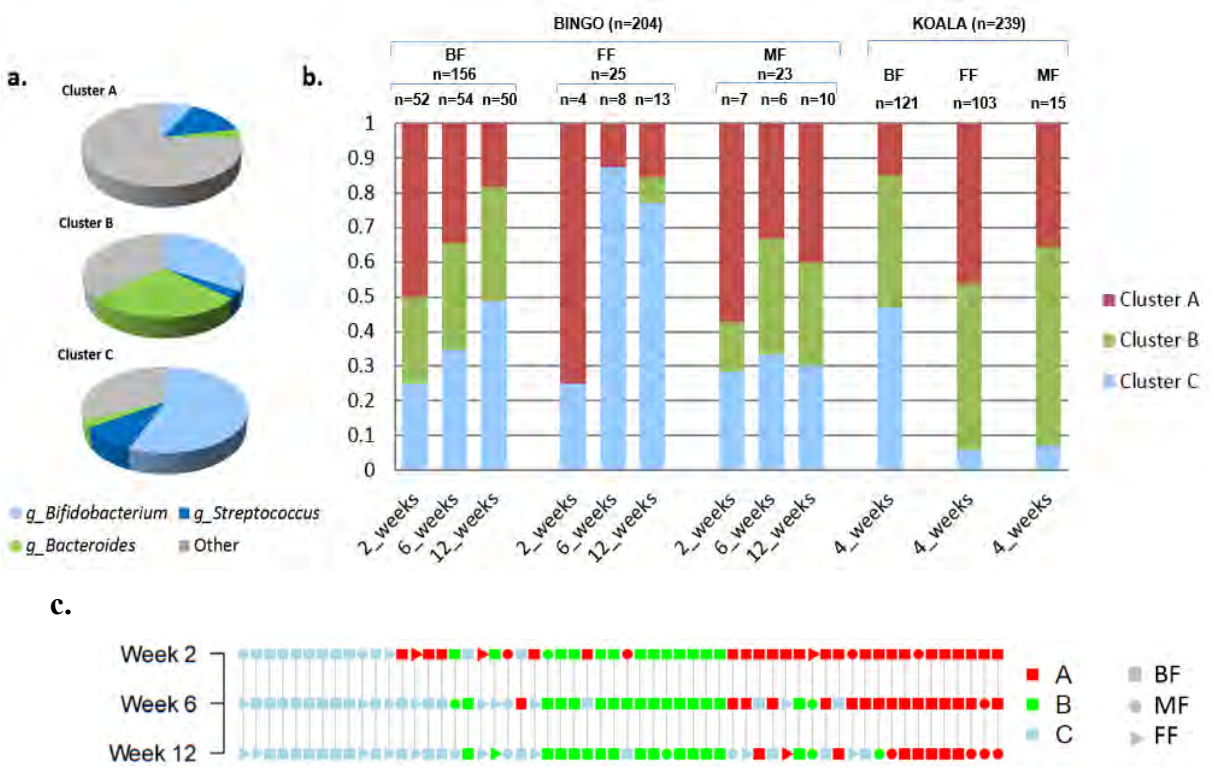


Figure 6. DMM clustering of samples on the basis of faecal microbiota composition at genus level; a. Average relative abundance of microbial groups characteristic to individual cluster A, B and C; b. Fraction of samples from infants receiving different types of feeding (BF, breastfeeding; FF, formula feeding; MF, mixed feeding) within each cluster category; c. Temporal evolution of cluster assignment for infants in the BINGO study cohort, indicating cluster type (red - cluster A, green - cluster B, blue - cluster C), and type of feeding at each time point (square - BF, circle - MF, triangle - FF).

Discussion

The current study examined faecal microbiota composition in healthy infants sampled at two, six and 12 weeks of age from the BINGO study cohort (year 2015-2016) and infants sampled at approximately four weeks of age from the KOALA study cohort (year 2001-2003). All infants received either commercially available formulas, breastmilk or mixed feeding. At the time samples from the KOALA cohort were collected, infant formulas enriched with prebiotics were available on the Dutch market only to a limited extent. Comparing the data obtained from both cohorts, we were able to show that faecal microbiota composition of infants fed more modern types of fortified formulas was more similar to faecal microbiota of BF infants, as compared to FF infants receiving unsupplemented formulas available in years 2001-2003.

The majority of infant formulas available on the Dutch market nowadays are supplemented with GOS and/or FOS. Previous studies showed that the mixture of GOS and/or FOS stimulated growth of *Bifidobacterium* in preterm [38-40] and term infants [41-43], and that GOS/FOS supplementation also resulted in a metabolic activity of the colonic microbiota,

as measured by SCFA and faecal pH, that was similar to the metabolic activity in BF controls [41, 43]. However, many of the microbiological data in these earlier studies were obtained using traditional culture- or molecular probe-based methods, which were often limited in their accuracy or the scope of microbial detection [44]. Here, we have compared samples from two study cohorts obtained more than ten years apart. The use of next generation sequencing in our study allowed us to gain a more comprehensive insight into microbial composition and dynamic patterns at a microbial community level.

In both studies unconstrained multivariate analysis revealed a clear separation of samples relating to the feeding mode (Figure 3). This is in agreement with a number of earlier studies which showed distinctively different faecal microbial profiles when comparing BF, FF and MF infants [44-46]. We found that in both study cohorts feeding mode was associated with faecal microbiota composition (Figure 4), and that bacterial diversity and richness in BF infants was significantly lower than in FF infants, except for the six week old infants from the BINGO cohort for which no difference in diversity was observed (Figure 2). Diversity evaluates both the number of species and the evenness of their distribution. Earlier studies showed that formula feeding was associated with higher faecal microbial diversity, and more adult like microbiota composition [46, 47] and microbiota activity [48]. In the BINGO study, FF infants showed an overall higher number of genus level taxa in their faeces in all age groups as estimated by the Chao1 index. This difference was most prominent at six weeks (31.5, *SD*=8.8 in FF and 22.5 *SD*=6.5 in BF; Figure 2b). Yet, the evenness, or relative contribution of each bacterial group was similar between BF and FF in the BINGO study (Figure 1, 2a), possibly due to more comparable relative abundance of the main groups in response to prebiotics [49]. Despite of the similarities, the relative abundance of 12 genus level groups differed significantly between BF and FF infants in the BINGO study cohort. However, the differences were mainly in the low abundance taxa which accounted for a very small fraction of the total bacteria in the community (Figure 1). In contrast, in the KOALA cohort 19 genus level taxa varied significantly, and the differences in both the species diversity and species richness were significant between FF and BF groups (Figure 2c,d). Together these differentially abundant taxa in the KOALA cohort accounted for more than 60% of the total bacteria in the faeces of either BF or FF infants.

Despite the differences in the type of formula used, in both the BINGO and the KOALA cohorts FF was associated with a significant increase in relative abundance of *Akkermansia, Enterococcus,* Peptostreptococcaceae *Incertae Sedis* and Erysipelotrichaceae *Incertae Sedis*. Higher levels of these microbial groups in FF infants have been reported previously [7, 46, 50]. Interestingly, a similar effect was observed in piglets that were fed dairy milk-based formula and showed significantly higher levels of both, *Akkermansia* and *Enterococcus,* indicating a possible effect of dairy milk compounds that are present in infant formulas [51].

Our results showed few important differences between the BINGO and the KOALA cohorts. *Clostridium* and *Escherichia-Shigella*, both of which include important pathogens, showed higher relative abundance in the FF infants from the KOALA cohort, but not in the BINGO cohort, when compared to the levels found in the corresponding BF groups. Another important difference pertains central bacterial groups, such as *Bifidobacterium* and *Lactobacillus*, whose relative abundance was reduced in the FF group from the KOALA cohort,

but not in the BINGO cohort, as compared to BF infants. These results are in line with earlier findings on the microbiota composition changes due to prebiotic fortification used in formulas [49]. Thus, we can conclude that the formulas used by the participants in the BINGO cohort appear to be much better in mimicking the beneficial properties of human milk with respect to the stimulation of GI tract microbiota that was similar to that found in healthy breastfed infants (Figure 1).

One of the main characteristics of the early life GI tract microbial ecosystems are low diversity and low stability [47, 52, 53]. These characteristics make the microbial ecosystems more vulnerable to environmental disturbance, and may explain the inconsistent patterns of distribution of taxa and high levels of inter-individual variability observed here and in other studies [47, 54]. In our study we observed large differences between infants that could only partially be explained by the available metadata. To disclose any possible more generic patterns in the faecal microbial composition, we applied DMM clustering analysis to all samples from the BINGO and KOALA sets separately. These analyses revealed presence of three distinct clusters with very similar characteristics for the different cohorts (Figure 6a). Furthermore, these clusters were still present when samples from only vaginally delivered breastfed infants were used in the modelling. In addition, these clusters were not associated with other characteristics such as sex, birth weight or delivery place. This implies that other environmental or genetic factors might be driving development of specific microbial assemblages in the infant GI tract [53]. One such factor might be the composition of breastmilk from the individual mother-infant pair. Earlier studies showed that the composition of breastmilk varies between individual mothers and across lactation stage with regard to human milk oligosaccharide (HMO) content and composition [55, 56], milk microbiota [57, 58] and other breastmilk factors which could influence GI colonisation [59].

Presence of three enterotypes has been described earlier for adult faecal microbiota, however, the concept still remains controversial [60, 61]. The large inter-individual variation in faecal microbial composition of infants has been noted before in a number of studies, but none of them indicated existence of equivalent enterotypes in infants [2, 5, 62]. Only recently, it has been suggested that three compositionally distinct human neonatal gut microbiota (NGM) profiles might be present and linked to development of atopy in young infants [62]. However, whether the specific microbial patterns observed in early life are universal, whether they can prompt an individual to develop a given adult enterotype, or if they predispose an infant to any health conditions, remains to be investigated in future longitudinal studies.

The adult like microbiota is established at 2-3 years of age [47, 63], which was beyond the scope of this study. However, the longitudinal design of the BINGO study allowed us to examine the temporal pattern in the development of microbiota in infants from two to 12 weeks of age. Earlier studies indicated that the pattern of microbiota development is non-random [3, 64]. We were interested in the relationship between each of the aforementioned clusters, infant age and feeding mode (Figure 6b). In the BF group there was a clear directional trend, where faecal microbiota gradually transitioned from the “mixed” cluster A to the *Bifidobacterium* dominated clusters B and C as infants got older. This pattern was different in infants receiving prebiotic formula, where exclusive formula feeding, or a change from breastfeeding to formula

were associated with an accelerated shift in microbial community composition to cluster C, already at six weeks of age. This suggests that the establishment of a *Bifidobacterium* dominated ecosystem might be linked with the age of an infant or with breastmilk properties, whereas in the FF infants this maturation of the GI tract microbiota was shaped and accelerated by prebiotic(s). In contrast, the mixed feeding was associated with delayed microbiota development, with more infants at 12 weeks of age falling into the mixed microbiota cluster A, as compared to infants from the BF group (Figure 6b). This mixed state was also characteristic to FF and MF infants receiving formulas with no prebiotics in the KOALA study. Whether there are clinical consequences of the timing of colonization with e.g. *Bifidobacterium* is unknown, but the complex and dynamic structure of the human milk might be an important driving force for the changes in GI microbiota in BF infants, and possibly the underlying reason for the different cluster groups that we observed in our data.

Human milk contains a wide range of compounds of which many can have a modulatory effect on the infant intestinal microbiota, such as carbohydrates including HMOs, immunoglobulins, fatty acids, nucleotides, cytokines, immune cells, lysozymes, lactoferrin and others [10, 59]. In addition, milk also contains its own microbiota [14, 57, 58]. Both, chemical and microbial composition of milk varies between individual mothers and it also changes throughout lactation, being able to adapt to the individual needs of a developing infant and its current health status, all of which influences the bioactive properties of milk [14, 57, 58]. This variability in milk characteristics cannot be mimicked by infant formulas, even when they are fortified with prebiotics. Nevertheless, today's formulas are closer at resembling breastmilk and stimulate bacterial groups which are a hallmark of a healthy infant GI tract. Earlier culture-dependent studies based on colony counts and molecular based methods of identification reported that GOS/FOS stimulated growth of bifidobacteria, but did not affect the total count of *Lactobacillus*, *Bacteroides, Clostridium* species, *Escherichia coli, Enterobacter, Citrobacter, Proteus, Klebsiella,* and *Candida* [38]. These findings were contradicted by later studies showing that prebiotic supplementation resulted in significant increase in lactobacilli and reduction of *Escherichia coli and Clostridium* species and other clinically relevant pathogens [39, 43, 65-67]. Through applying 16S rRNA gene sequencing we were able to provide a more comprehensive view at which bacterial groups varied between different feeding modes. However, this approach did not allow us to further identify bacteria to specific species or strains. A higher resolution of the data may be necessary, as earlier studies on bifidobacteria and lactobacilli reported that prebiotic formulas and breastmilk may stimulate the same bacterial groups, but as it was shown for *Lactobacillus*, they may stimulate different species within the same group [65, 68].

Finally, our findings suggest that the use of new formulas in combination with breastmilk feeding was associated with more mixed microbiota composition and showed lower bifidogenic effect, thus implying a possible interference between the components of the breastmilk and formula. A similar shift towards FF microbiota pattern in infants receiving MF has been noted in the past [69]. Unfortunately, in both cohorts the MF groups were small (23 BINGO samples, 15 KOALA samples), yet this mode of feeding represents a highly realistic scenario in today's infant nutrition and thus, should be further investigated using a larger study

cohort. Such study would require a detailed knowledge of breastmilk composition, as well as information on the types of formulas that the infants received to enable detailed assessment of the effect of each component.

Conclusion

We have compared faecal microbiota composition from two different cohorts of infants born more than 10 years apart and identified the similarities in the underlying structure of the microbial community in both sample sets. Our results showed that the use of today's infant formulas containing prebiotics improved faecal microbiota composition in formula fed infants, compared to infants receiving formula feeding in 2002-2003. However, we also noted that the dynamics of bacterial colonisation during the first 12 weeks of life was altered in infants receiving modern formulas. The long term effects of prebiotic (and probiotic) fortified infant formulas, as well as the effects of mixed feeding should be further investigated.

Funding

This research was performed in the public-private partnership CarboHealth coordinated by the Carbohydrate Competence Center (CCC, www.cccresearch.nl) and financed by participating partners and allowances of the TKI Agri&Food program, Ministry of Economic Affairs of the Netherlands.

Acknowledgements

The authors thank Edoardo Saccenti (Wageningen University, Laboratory of Systems and Synthetic Biology) for helpful discussions.

References

1. Collado, M.C., et al., *Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid.* Sci Rep, 2016. **6**: p. 23129.
2. Valles, Y., et al., *Microbial succession in the gut: directional trends of taxonomic and functional change in a birth cohort of Spanish infants.* PLoS Genet, 2014. **10**(6): p. e1004406.
3. Koenig, J.E., et al., *Succession of microbial consortia in the developing infant gut microbiome.* PNAS, 2010. **108**(1).
4. Fouhy, F., et al., *Composition of the early intestinal microbiota: knowledge, knowledge gaps and the use of high-throughput sequencing to address these gaps.* Gut Microbes, 2012. **3**(3): p. 203-20.
5. Madan, J.C., et al., *Normal neonatal microbiome variation in relation to environmental factors, infection and allergy.* Curr Opin Pediatr, 2012. **24**(6): p. 753-9.
6. Madan, J.C., et al., *Association of Cesarean Delivery and Formula Supplementation With the Intestinal Microbiome of 6-Week-Old Infants.* JAMA Pediatr, 2016. **170**(3): p. 212-9.
7. Meghan B. Azad PhD, et al., *Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months.* CMAJ, 2013. **185**(5): p. 385-394.
8. Penders, J., et al., *Factors Influencing the Composition of the Intestinal Microbiota in Early Infancy.* Pediatrics, 2006. **118**(2): p. 511-521.
9. Michaelsen, L.S.-N.a.K.F., *Breast feeding and future health.* Curr Opin Clin Nutr Metab Care 2006 **9**: p. 289–296.
10. Le Huerou-Luron, I., S. Blat, and G. Boudry, *Breast- v. formula-feeding: impacts on the digestive tract and immediate and long-term health effects.* Nutr Res Rev, 2010. **23**(1): p. 23-36.
11. Stevens, E.E., T.E. Patrick, and R. Pickler, *A History of Infant Feeding.* The Journal of Perinatal Education, 2009. **18**(2): p. 32-39.
12. Fomon, S.J., *Infant Feeding in the 20th Century: Formula and Beikost.* Journal of Nutrition, 2001. **131**(2): p. 409S-420S.
13. Stevens, E.E., T.E. Patrick, and R. Pickler, *A history of infant feeding.* J Perinat Educ, 2009. **18**(2): p. 32-9.
14. Cabrera-Rubio, R., et al., *The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery.* Am J Clin Nutr, 2012. **96**(3): p. 544-51.
15. Barile, D. and R.A. Rastall, *Human milk and related oligosaccharides as prebiotics.* Curr Opin Biotechnol, 2013. **24**(2): p. 214-9.
16. Tissier, H., *Recherches sur la flore intestinale des nourrissons (état normal et pathologique).* Thesis. Paris: G. Carre and C. Naud, 1900.
17. Moro, E., *Morphologische und bakteriologische untersuchungen über die darmbakterien des säuglings: Die bakteriumflora des normalen frauenmilchstuhls.* Jahrbuch Kinderheilkunde, 1900. **61**: p. 686–734.
18. Bullen, C.L., P.V. Tearle, and M.G. Stewart, *The Effect of "Humanised" Milks and Supplemented Breast Feeding on The Faecal Flora of Infants* Journal of Medical Microbiology, 1977. **10**(4): p. 403-413.
19. PL, S. and L. A., *The microbial ecology of the large bowel of breast-fed and formula-fed infants during the first year of life.* J Med Microbiol, 1982. **15**(2): p. 189-203.
20. SE, B. and W. BA, *Diet and faecal flora in the newborn: breast milk and infant formula.* Archives of Disease in Childhood, 1989. **64**(12): p. 1672-1677.
21. Kleessen, B., et al., *Influence of two infant formulas and human milk on the development of the faecal flora in newborn infants.* Acta Pediatrica, 1995(84): p. 1347-1356.
22. P, G., N. RF, and R. CS, *Bifidus Factor. I. A Variant of Lactobucillus bifidus requiring a special growth factor.* Arch Biochem Biophys., 1954. Jan;**48**(1): p. 193-201.
23. Kunz, C. and S. Rudloff, *Biological functions of oligosaccharides in human milk.* Acta Paediatr, 1993. **82**: p. 903-12.
24. Efsa Panel on Dietetic Products Nutrition Allergies, *Scientific Opinion on the essential composition of infant and follow-on formulae.* EFSA Journal, 2014. **12**(7).

25. Allergies, E.P.o.D.P.N., *Commission Delegated Regulation 2016-127 suppl 609-2013 (infant & young child formula).* EFSA Journal, 2015.
26. Vandenplas, Y., I. Zakharova, and Y. Dmitrieva, *Oligosaccharides in infant formula: more evidence to validate the role of prebiotics.* Br J Nutr, 2015. **113**(9): p. 1339-44.
27. Firmansyah, A., et al., *Fructans in the first 1000 days of life and beyond, and for pregnancy.* Asia Pac J Clin Nutr, 2016. **25**(4): p. 652-675.
28. Kummeling, I., et al., *Etiology of atopy in infancy: the KOALA Birth Cohort Study.* Pediatr Allergy Immunol, 2005. **16**(8): p. 679-84.
29. Scheepers, L.E., et al., *The intestinal microbiota composition and weight development in children: the KOALA Birth Cohort Study.* Int J Obes (Lond), 2015. **39**(1): p. 16-25.
30. Vandenplas, Y., E.D. Greef, and G. Veereman, *Prebiotics in infant formula.* Gut Microbes, 2014. **5**(6): p. 681-687.
31. Ramiro-Garcia, J., et al., *NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes.* F1000Research, 2016. **5**: p. 1791.
32. Quast, C., et al., *The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.* Nucleic Acids Res, 2013. **41**(Database issue): p. D590-6.
33. Kuczynski, J., et al., *Using QIIME to analyze 16S rRNA gene sequences from microbial communities.* Curr Protoc Bioinformatics, 2011. Chapter 10: p. Unit 10 7.
34. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data.* Nat Methods, 2010. **7**(5): p. 335-6.
35. Lepš, P.Š.a.J., *Multivariate analysis of ecological data using Canoco 5. In: Canoco support.* Cambridge University Press, UK., 2014: p. 207-229.
36. Ian Holmes, K.H., Christopher Quince, *Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics.* PLoS ONE, 2012. **7**(2): p. 1-15.
37. Morgan, M., *DirichletMultinomial: Dirichlet-Multinomial Mixture Model Machine Learning for Microbiome Data.* R package version 1.18.0., 2017.
38. Boehm, G., et al., *Supplementation of a bovine milk formula with an oligosaccharide mixture increases counts of faecal bifidobacteria in preterm infants.* Arch Dis Child Fetal Neonatal Ed, 2002. **86**: p. 178–181.
39. Kapiki, A., et al., *The effect of a fructo-oligosaccharide supplemented formula on gut flora of preterm infants.* Early Hum Dev, 2007. **83**(5): p. 335-9.
40. Srinivasjois, R., S. Rao, and S. Patole, *Prebiotic supplementation in preterm neonates: updated systematic review and meta-analysis of randomised controlled trials.* Clin Nutr, 2013. **32**(6): p. 958-65.
41. Knol, J., et al., *Colon Microflora in Infants Fed Formula with Galacto- and Fructo-Oligosaccharides: More Like Breast-Fed Infants.* Journal of Pediatric Gastroenterology and Nutrition, 2005. **40**: p. 36–42.
42. Costalos, C., et al., *The effect of a prebiotic supplemented formula on growth and stool microbiology of term infants.* Early Hum Dev, 2008. **84**(1): p. 45-9.
43. Holscher, H.D., et al., *Effects of Prebiotic-Containing Infant Formula on Gastrointestinal Tolerance and Fecal Microbiota in a Randomized Controlled Trial.* Journal of Parenteral and Enteral Nutrition, 2012. **36**(1): p. 95S-105S.
44. Harmsen, H.J.M., et al., *Analysis of Intestinal Flora Development in Breast-Fed and Formula-Fed Infants by Using Molecular Identification and Detection Methods.* Journal of Pediatric Gastroenterology and Nutrition, 2000. **30**(1): p. 61-67.
45. Guaraldi, F. and G. Salvatori, *Effect of Breast and Formula Feeding on Gut Microbiota Shaping in Newborns.* Frontiers in Cellular and Infection Microbiology, 2012. **2**: p. 94.
46. Wang, M., et al., *Fecal Microbiota Composition of Breast-fed Infants is Correlated with Human Milk Oligosaccharides Consumed.* Journal of pediatric gastroenterology and nutrition, 2015. **60**(6): p. 825-833.
47. Yatsunenko, T., et al., *Human gut microbiome viewed across age and geography.* Nature, 2012. **486**(7402): p. 222-227.
48. Bäckhed, F., et al., *Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life.* Cell Host & Microbe, 2015. **17**(5): p. 690-703.

49. Hojsak, I. and A. MocicPavic, *Supplementation of prebiotics in infant formula.* Nutrition and Dietary Supplements, 2014: p. 69.
50. Bergström, A., et al., *Establishment of Intestinal Microbiota during Early Life: a Longitudinal, Explorative Study of a Large Cohort of Danish Infants.* Appl. Environ. Microbiol, 2014. **80**(9): p. 2889-2900.
51. Saraf, M.K., et al., *Formula diet driven microbiota shifts tryptophan metabolism from serotonin to tryptamine in neonatal porcine colon.* Microbiome, 2017. **5**: p. 77.
52. Scholtens, P.A.M.J., et al., *The Early Settlers: Intestinal Microbiology in Early Life.* Annual Review of Food Science and Technology, 2012. **3**(1): p. 425-447.
53. Wopereis, H., et al., *The first thousand days – intestinal microbiology of early life: establishing a symbiosis.* Pediatric Allergy and Immunology, 2014. **25**(5): p. 428-438.
54. Timmerman, H.M., et al., *Intestinal colonisation patterns in breastfed and formula-fed infants during the first 12 weeks of life reveal sequential microbiota signatures.* Scientific Reports, 2017. **7**(1): p. 8327.
55. Coppa, G.V., et al., *Changes in Carbohydrate Composition in Human Milk Over 4 Months of Lactation.* Pediatrics, 1993. **91**(3): p. 637-641.
56. Austin, S., et al., *Temporal Change of the Content of 10 Oligosaccharides in the Milk of Chinese Urban Mothers.* Nutrients, 2016. **8**(6): p. 346.
57. Khodayar-Pardo, P., et al., *Impact of lactation stage, gestational age and mode of delivery on breast milk microbiota.* Journal Of Perinatology, 2014. **34**: p. 599.
58. Fernández, L., et al., *The human milk microbiota: Origin and potential roles in health and disease.* Pharmacological Research, 2013. **69**(1): p. 1-10.
59. Walker, W.A. and R.S. Iyengar, *Breast milk, microbiota, and intestinal immune homeostasis.* Pediatric Research, 2014. **77**: p. 220.
60. Arumugam, M., et al., *Enterotypes of the human gut microbiome.* Nature, 2011. **473**(7346): p. 174-80.
61. Shetty, S.A., et al., *Intestinal microbiome landscaping: insight in community assemblage and implications for microbial modulation strategies.* FEMS Microbiology Reviews, 2017. **41**(2): p. 182-199.
62. Fujimura, K.E., et al., *Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation.* Nat Med, 2016. **22**(10): p. 1187-1191.
63. Castanys-Muñoz, E., M.J. Martin, and E. Vazquez, *Building a Beneficial Microbiome from Birth.* Advances in Nutrition, 2016. **7**(2): p. 323-330.
64. Favier, C.F., et al., *Molecular Monitoring of Succession of Bacterial Communities in Human Neonates.* Applied and Environmental Microbiology, 2002. **68**(1): p. 219-226.
65. Haarman, M. and J. Knol, *Quantitative real-time PCR analysis of fecal Lactobacillus species in infants receiving a prebiotic infant formula.* Appl Environ Microbiol, 2006. **72**(4): p. 2359-65.
66. Moro, G., et al., *Dosage-Related Bifidogenic Effects of Galacto- and Fructooligosaccharides in Formula-Fed Term Infants.* Journal of Pediatric Gastroenterology & Nutrition, 2002. **34**(3): p. 291-295.
67. Knol, J., et al., *Increase of faecal bifidobacteria due to dietary oligosaccharides induces a reduction of clinically relevant pathogen germs in the faeces of formula-fed preterm infants.* Acta Paediatr Suppl, 2005. **94**(449): p. 31-3.
68. Haarman, M. and J. Knol, *Quantitative real-time PCR assays to identify and quantify fecal Bifidobacterium species in infants receiving a prebiotic infant formula.* Appl Environ Microbiol, 2005. **71**(5): p. 2318-24.
69. O'Sullivan, A., M. Farver, and J.T. Smilowitz, *The Influence of Early Infant-Feeding Practices on the Intestinal Microbiome and Body Composition in Infants.* Nutrition and Metabolic Insights, 2015. **8**(Suppl 1): p. 1-9.

Supplemental Figures and Tables

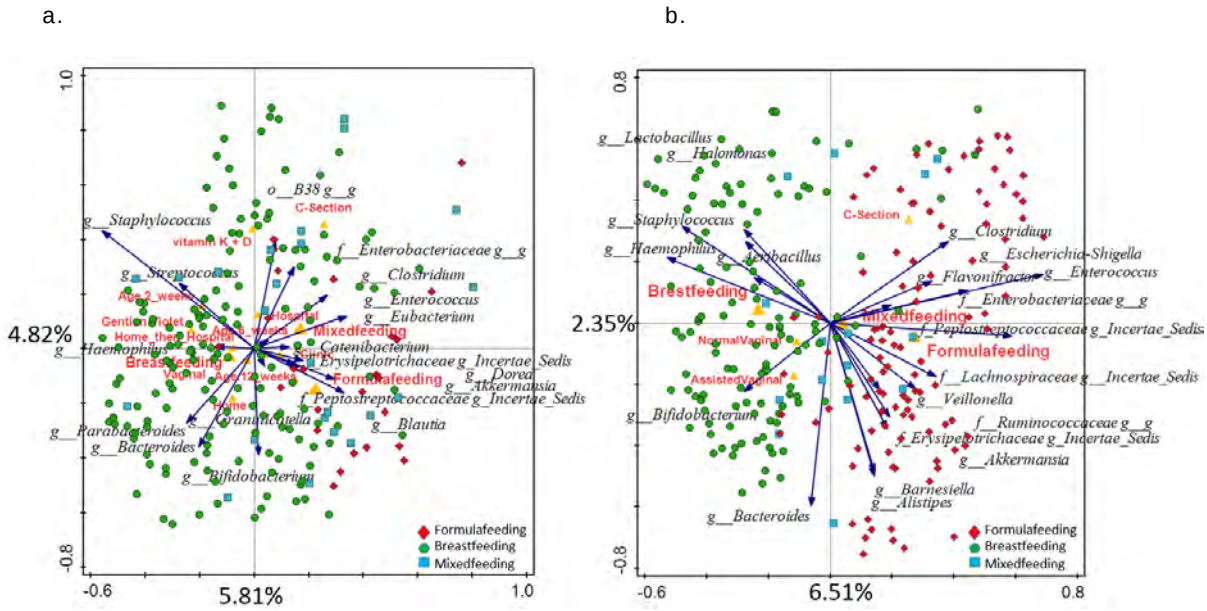


Figure S1. RDA showing factors with a significant effect on faecal microbiota of infants. Sample are colour coded by feeding mode. Displayed taxa include ten best fitting species, in addition to the microbial groups that differed significantly (Wilcoxon test, FDR p<0.05) between BF and FF groups. a. BINGO cohort; b. KOALA cohort

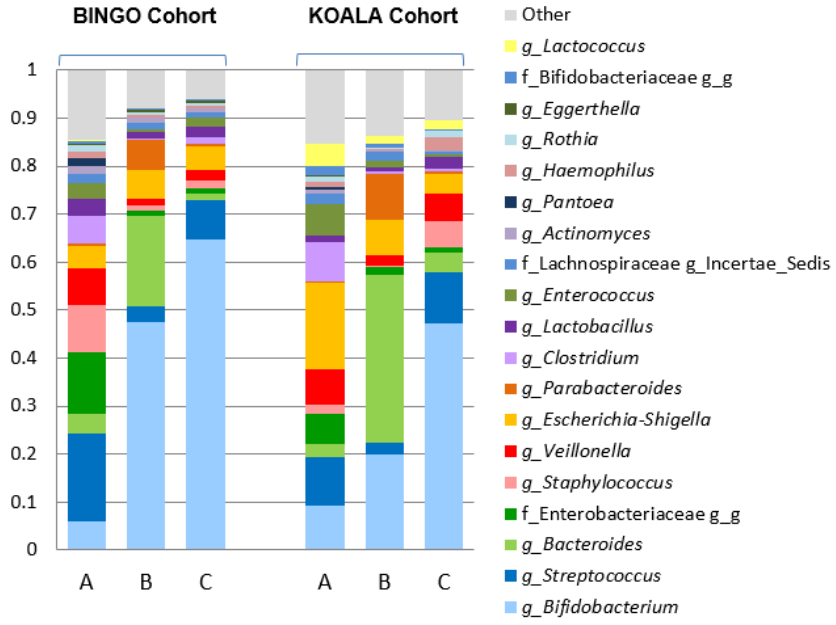


Figure S2. The average relative contribution of the main bacterial genus-level taxa defining each of three clusters obtained by DMM modelling of faecal microbial composition of infants included in the BINGO and KOALA cohorts. When the taxonomic assignment could not be made at genus level, the lowest classifiable taxonomy assignment is used instead.

Table S1. Average relative abundance of major bacterial taxa detected in BINGO and KOALA cohorts. Microbial groups that significantly differ in relative abundance (Wilcoxon test, FDR $p<0.05$) between BF and FF groups are indicated with * (BINGO) and # (KOALA).

Study	**BINGO Cohort**									**KOALA Cohort**		
Feeding Mode	**BF**			**FF**			**MF**			**BF**	**FF**	**MF**
Age (weeks)	**2**	**6**	**12**	**2**	**6**	**12**	**2**	**6**	**12**	**4**	**4**	**4**
g_Bifidobacterium #	0.3467	0.4614	0.5111	0.4519	0.6218	0.5528	0.2806	0.4321	0.3217	0.3192	0.1718	0.1216
g_Bacteroides	0.0831	0.1212	0.1230	0.0652	0.0225	0.0357	0.0759	0.0877	0.1314	0.2134	0.2316	0.3195
g_Streptococcus	0.0844	0.0711	0.0437	0.0238	0.0314	0.0680	0.1788	0.1223	0.0305	0.0375	0.0454	0.0348
g_Escherichia-Shigella #	0.1108	0.0647	0.0769	0.0380	0.1038	0.0790	0.0431	0.0617	0.0912	0.1185	0.1947	0.2095
f_Enterobacteriaceae *g_g* #	0.1199	0.0524	0.0390	0.0910	0.0061	0.0401	0.0634	0.0919	0.0816	0.0396	0.0555	0.0126
f_Lachnospiraceae *Incertae_Sedis* #	0.0390	0.0268	0.0360	0.0229	0.0084	0.0191	0.0416	0.0593	0.0856	0.0188	0.0276	0.0146
g_Lactobacillus #	0.0105	0.0383	0.0271	0.0480	0.0210	0.0278	0.0201	0.0337	0.0564	0.0228	0.0050	0.0204
g_Clostridium #	0.0277	0.0162	0.0290	0.0167	0.0275	0.0249	0.0677	0.0422	0.0392	0.0256	0.0318	0.0076
g_Enterococcus #*	0.0223	0.0133	0.0090	0.0042	0.0295	0.0583	0.0507	0.0237	0.0146	0.0020	0.0194	0.0181
g_Blautia *	0.0002	0.0035	0.0032	0.0245	0.0523	0.0316	0.0468	0.0032	0.0486	0.0053	0.0037	0.0232
g_Veillonella #	0.0237	0.0205	0.0321	0.0167	0.0128	0.0149	0.0260	0.0118	0.0408	0.0274	0.0504	0.0508
g_Staphylococcus #*	0.0683	0.0136	0.0028	0.0257	0.0035	0.0019	0.0535	0.0018	0.0003	0.0218	0.0006	0.0011
g_Pantoea	0.0029	0.0128	0.0009	0.1201	0.0000	0.0020	0.0006	0.0012	0.0031	0.0015	0.0040	0.0001
g_Parabacteroides *	0.0131	0.0182	0.0136	0.0006	0.0002	0.0003	0.0269	0.0011	0.0111	0.0668	0.0316	0.0904
f_Erysipelotrichaceae *Incertae_Sedis* #*	0.0032	0.0036	0.0050	0.0230	0.0027	0.0089	0.0072	0.0000	0.0101	0.0074	0.0109	0.0098
g_Actinomyces	0.0085	0.0225	0.0043	0.0025	0.0009	0.0015	0.0020	0.0070	0.0026	0.0023	0.0005	0.0212
g_Collinsella	0.0014	0.0004	0.0006	0.0000	0.0314	0.0013	0.0000	0.0000	0.0032	0.0011	0.0008	0.0009
f_Peptostreptococcaceae *Incertae_Sedis* #*	0.0001	0.0003	0.0006	0.0019	0.0057	0.0115	0.0013	0.0000	0.0037	0.0001	0.0040	0.0018
f_Lachnospiraceae *g_g*	0.0069	0.0052	0.0004	0.0000	0.0006	0.0005	0.0002	0.0008	0.0071	0.0015	0.0010	0.0000
g_Haemophilus #*	0.0041	0.0057	0.0055	0.0000	0.0000	0.0000	0.0008	0.0000	0.0005	0.0098	0.0004	0.0062
g_Catenibacterium *	0.0000	0.0000	0.0000	0.0090	0.0050	0.0001	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000
g_Akkermansia #*	0.0000	0.0000	0.0000	0.0000	0.0011	0.0050	0.0000	0.0008	0.0003	0.0000	0.0225	0.0008
g_Dorea *	0.0000	0.0001	0.0000	0.0016	0.0026	0.0023	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000
g_Flavonifractor #	0.0000	0.0020	0.0020	0.0014	0.0000	0.0002	0.0002	0.0000	0.0000	0.0034	0.0047	0.0000
g_Halomonas #	0.0003	0.0005	0.0002	0.0000	0.0000	0.0011	0.0002	0.0012	0.0002	0.0007	0.0000	0.0001
f_Ruminococcaceae *g_g* #	0.0000	0.0003	0.0001	0.0017	0.0002	0.0005	0.0000	0.0000	0.0000	0.0033	0.0044	0.0003
g_Eubacterium *	0.0000	0.0000	0.0000	0.0014	0.0004	0.0000	0.0000	0.0000	0.0000	0.0004	0.0002	0.0000
g_Aeribacillus #	0.0001	0.0002	0.0001	0.0000	0.0000	0.0005	0.0000	0.0005	0.0000	0.0004	0.0000	0.0001
g_Alistipes #	0.0002	0.0001	0.0001	0.0000	0.0000	0.0001	0.0000	0.0003	0.0003	0.0013	0.0058	0.0029
g_Barnesiella #	0.0000	0.0000	0.0002	0.0000	0.0003	0.0000	0.0002	0.0000	0.0000	0.0018	0.0145	0.0031
*g_Granulicatella**	0.0000	0.0000	0.0001	0.0000	0.0000	0.0004	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
TOTAL	**0.9775**	**0.9750**	**0.9666**	**0.9918**	**0.9918**	**0.9900**	**0.9877**	**0.9840**	**0.9842**	**0.9545**	**0.9432**	**0.9715**
Other taxa	0.0225	0.0250	0.0334	0.0082	0.0082	0.0104	0.0123	0.0160	0.0158	0.0488	0.0615	0.0285

Chapter 4

The association between infant faecal microbiota composition and the degradation of human milk oligosaccharides in one month old, healthy breastfed infants

Klaudyna Borewicz[*], Fangjie Gu[2*], Edoardo Saccenti, Ilja Arts, John Penders, Carel Thijs, Sander S. van Leeuwen, Arjen Nauta, Cordula Lindner, Ellen van Leusen, Henk A. Schols, Hauke Smidt

[*] These authors contributed equally to this work

Manuscript in preparation

Abstract

In this study we investigated the association between selected human milk oligosaccharides (HMOs) in breastmilk, their intestinal fate, and faecal microbiota composition in healthy, breastfed, one month old infants. The two main objectives were (i) to examine a link between the maternal breastmilk HMOs and the composition of infant faecal microbiota, and (ii) to identify microbial communities involved in the degradation of selected HMOs within the gastrointestinal GI tract of infants. We showed that *Bifidobacterium, Bacteroides, Escherichia-Shigella* and *Parabacteroides* were the predominating genera detected in infant faeces. Infant faecal microbiota composition was associated with gender, mode of delivery, and breastmilk HMOs: Lacto-*N*-fucopentaose I and 2′-Fucosyllactose (2′FL). Dirichlet Multinomial Mixture (DMM) modelling revealed presence of three different patterns in infant faecal microbiota characterised by mixed community structure (Cluster A), high relative abundance of both, *Bifidobacterium* and *Bacteroides* (Cluster B), or high relative abundance of *Bifidobacterium* (Cluster C). There was a significant association between the cluster assignment and an infant's ability to degrade breastmilk HMOs, with complete degradation of HMOs associated with cluster C and non-specific degradation associated with cluster A. Constrained multivariate redundancy analysis indicated a significant association between faecal microbiota composition and gastrointestinal degradation of 2′FL, Lacto-*N*-tetraose and Lacto-*N*-neotetraose, difucosyllactose, 6'Sialyllactose, Lacto-*N*-hexaose, Lacto-*N*-fucopentaose II and Lacto-*N*-fucopentaose III (FDR<0.05). Furthermore, our study showed that degradation of specific HMOs in the infant gastrointestinal tract could be correlated with the statistically significant increase in relative abundance of various phylotypes (OTUs) within the bifidobacteria, and to lesser extent within the *Bacteroides* and lactobacilli.

Introduction

During and after birth, microorganisms from the mother and other environmental sources colonize an infant, and various environmental factors and life events may further shape the microbial communities, making them specific to each body site and to each individual. These microbial ecosystems acquired and developed in early life play an important role in our well-being and health, both during infancy and beyond [1].

One of the body sites that undergoes a rapid microbial colonisation in early life is the gastrointestinal (GI) tract [1]. The anaerobic conditions in the GI tract favour the establishment of bacteria, such as *Bifidobacterium*, *Bacteroides* and *Clostridium* [1]. Besides the absence of oxygen, diet is another key factor that has a strong influence on shaping the GI microbial ecosystem. In mammals, milk evolved not only to provide the most optimal nutrition for the growing infant, but it also contains a broad range of bioactive components that are necessary for the development and maturation of infant's gastrointestinal and immune systems [1-3].

In breastfed infants, breastmilk is the sole source of nourishment during the first few months of life. Breastmilk is a complex biofluid that contains high concentrations of lactose, lipids and milk glycans, including free human milk oligosaccharides (HMOs) [2]. HMOs play an important role in intestinal cell proliferation and maturation, maintaining epithelial barrier function, and protecting the GI tract against bacterial and viral pathogens and toxins [2-5]. Despite being the third most abundant component of human milk, HMOs are not accessible to digestion by infant enzymes [2]. As a result, milk HMOs reach the infant colon, where they are degraded by bacteria. Since not all bacteria have the necessary enzymes to utilize HMOs, these milk glycans facilitate the development of a highly specialized microbial ecosystem dominated by bifidobacteria and *Bacteroides*, while indirectly limiting growth of other bacteria. This prebiotic effect has been demonstrated for selected bacterial species, both *in vitro* [4, 6] and *in vivo* [7], and it has been recognised as one of the key drivers for bacterial species succession in the infant GI tract leading to the development of a relatively simple and stable microbial GI tract ecosystem known as the milk-oriented microbiota (MOM) [3].

Maternal genotype (including e.g. mother's secretor status) determines the HMO composition of breastmilk, and the concentrations of different HMOs vary between individuals and across lactation stages [8-10]. This variability might target the distinct and changing needs of a growing infant and orchestrate the stepwise development of infant GI tract microbiota. Recent developments in glycomics revealed existence of over 200 different HMOs in human breastmilk [2]. The core structures of all HMOs include galactose, glucose and *N*-acetyl-glucosamine, which are further decorated with fucose and/or sialic acid. Based on the presence or absence of sialic acid, HMOs can be classified into two categories: the neutral and the acidic HMOs. In the study reported here we measured 17 highly abundant HMOs, including 12 neutral and five acidic HMOs.

In the light of growing evidence supporting the role of the GI tract microbial ecosystem in health, understanding the biological function of the different HMOs is of a great interest. Previous studies focused mainly on *in vitro* fermentation of HMOs by faecal bacterial inoculum, or by faecal isolates [4], however, the HMO degradation within an infant GI tract is still not

fully understood. Here, we analysed 121 mother-infant pairs to investigate the association between selected breastmilk HMOs and the infant faecal microbiota composition. Our two main research questions were: i) whether there was an association between the composition of HMOs in breastmilk and the composition of faecal microbiota in healthy, breastfed, one month old infants, and; ii) if the degradation of these breastmilk HMOs could be linked to infant faecal microbial communities, and further to specific bacterial taxa found in the infant's GI tract.

Methods

Milk and faecal sample collection

The milk and faecal samples used in this study originated from the KOALA Birth Cohort Study (Dutch acronym for: Child, Parents and Health: Lifestyle and Genetic Constitution). The design, selection criteria and faeces collection procedure have been described elsewhere and the study was approved by the Ethics Committee of the University Hospital of Maastricht [11-13]. Briefly, the KOALA study included two recruitment groups of healthy pregnant women, most of them living in the south of the Netherlands in years 2002-2003. The first group (n=2343) had a conventional lifestyle, whereas the second group (n=491) was considered to have an alternative lifestyle and was recruited through alternative channels, such as posters in organic food shops, anthroposophic doctors and midwives. The alternative lifestyle could involve dietary habits (vegetarian, organic), child-rearing practices and/or low use of antibiotics. Exclusion criteria were prematurity (birth before 37 weeks of gestation), twins, congenital abnormalities related to growth, and administration of antimicrobial agents before faeces collection. All infants included in this study were born healthy, full term, at home or hospital via either vaginal delivery or C-Section. Two infants were reported by the parents as sick during the sample collection day, but none of the infants received antibiotics during the first month of life (Table S1, S2). Infant faecal samples were collected by the parents at approximately one month postpartum from infants' diapers, refrigerated, and shipped via post within one day after collection. Breastmilk samples were collected on the same day by the mothers or research nurses at the participants' homes [13]. Briefly, mothers received a sterile 50 mL tube (Cellstar PP-test tubes, Greiner bio-one, Kremsmünster, Austria) and were instructed to collect the milk sample in the morning, before breastfeeding their child, from the contra-lateral breast (since the last feeding) and to keep the tube in the refrigerator (±4 °C) until it was collected by one of the researchers. If the mother was not able to collect the milk sample by herself (with or without a pumping regimen), an electric breast pump (Medela, Baar, Switzerland) was used with the help of one of the researchers (within the same day). During transport, the milk samples were stored in a cooler (Coleman Company Inc., Breda, the Netherlands) on packed ice (±4 °C) until processing on the same day. The sample was centrifuged (400 × g, 12 min, no brake, 4 °C) to separate the lipid and aqueous fraction. The lipid layer was trimmed off with a pipette and released in plastic storage vials (Sarstedt, Nümbrecht, Germany). The aqueous fraction was poured in other vials with another pipette. The remaining debris was not used to avoid contamination with cell fragments. All fractions were stored at –80 °C in the European Biobank, Maastricht. Only infants who were exclusively

breastfed, and for whom both the faecal and the corresponding maternal breastmilk samples were available were included in the analyses (n=121).

DNA extraction and the analysis of next generation sequencing data

Total DNA was extracted from the stool samples as previously described [14], using the double bead-beating procedure followed by QIAamp DNA stool mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The resulting DNA was used for subsequent PCR amplification of the V4 region of 16S ribosomal RNA (rRNA) genes. Each reaction contained 5-20 ng of template in a total volume of 50 µL. Each sample was amplified with 200 nM of uniquely barcoded primers 515F-n (5'-GTGCCAGCMGCCGCGGTAA-) and 806R-n (5'-RGGATTAGATACCC), 10 µl of 5x HF buffer (Finnzymes, Vantaa, Finland), 200 µM dNTP Mix (Roche Diagnostics GmbH, Mannheim, Germany), 1 U Phusion® Hot Start II High Fidelity DNA Polymerase (Finnzymes) and 36.5 µL of DNAse and RNAse free water [15]. The amplification program included 30 s initial denaturation step at 98 ºC, following by 25 cycles of denaturation at 98 ºC for 10 s, annealing at 56 ºC for 10 s and elongation at 72 ºC for 10 s, and a final extension at 72 ºC for 7 min. The PCR product presence and size (~290 bp) was confirmed with gel electrophoresis using the Lonza FlashGel® System (Lonza, Cologne, Germany). Seventy unique barcode tags were used in each library and artificial control (Mock) communities were included to monitor the quality of PCR amplifications and sequencing (data not shown) [15]. PCR products were purified with HighPrep® PCR kit (MagBio Genomics, Alphen aan den Rijn, Netherlands), and DNA concentrations were measured with Qubit® dsDNA BR Assay Kit (Life Technologies, Leusden, Netherlands). Hundred nanograms of each barcoded sample was added to an amplicon pool and was then concentrated with HighPrep® PCR kit to 20 µL volume. Concentrations of pools were measured with Qubit® dsDNA BR Assay Kit and adjusted to 100 ng/µL final concentration. The libraries were sent for adapter ligation and Illumina HiSeq2000 sequencing at GATC-Biotech, Konstanz, Germany.

HMO Analysis

Seventeen different HMO types were isolated from milk and infant faeces as described earlier [16]. In total, 12 neutral HMOs (2´FL, LNT and LNnT, 3FL, DFL, LNDFHI, LNFPI, LNFPII, LNFPIII, LNFPV, LNH, LNnH) and five acidic HMOs (3´SL, 6´SL, LSTa, LSTb, LSTc) were measured. Their names, classification and chemical structures are summarised in Table S1 [17, 18]. The HMOs were extracted, purified, and quantified by using porous graphitized carbon-ultra high-performance liquid chromatograph - mass spectrometry (HPLC-MS) [17, 18]. 3FL was eluted in a separate fraction and quantified by high performance anion exchange chromatography (HPAEC) [19]. The purification and chromatographic conditions were optimized as described by Gu and co-workers (manuscript submitted).

Data analysis

The 16S rRNA sequencing data was analysed using the NG-Tax analysis pipeline [15]. In brief, libraries were filtered to contain only read pairs with perfectly matching barcodes that were subsequently used to separate reads by sample. Operational taxonomic units (OTUs) were assigned using an open reference approach and SILVA_111_SSU 16S rRNA gene reference

database (https://www.arb-silva.de/) [20]. Microbial composition data was expressed as a relative abundance of each OTU obtained with NG-Tax.

Infants were classified into three distinct microbial cluster types based on genus level microbial abundance data using Dirichlet Multinomial Mixture (DMM) modelling [21]. Briefly, the number of Dirichlet components was selected by inspection of the fit of the model to the count data for varying number of components (1 to 7). Goodness of fit was assessed using the Laplace and the Akaike information criteria. Finally, each sample was assigned to the component for which it had the largest fitted value using the DirichletMultinomial R package [22] in R (version 3.3.1). Microbial composition of each DMM cluster is shown in supplementary Figure S1.

Spearman correlations were calculated using R to evaluate associations between members of the microbial community using the relative abundance data of the 82 different OTUs which were found in more than six of the 121 infants in this study (5%). Network visualization and ClusterONE (clustering with overlapping neighbourhood expansion) analysis were performed using Cytoscape [23].

Redundancy analysis (RDA) was done in Canoco5 [24] using the log transformed OTU level relative abundance data with significance assessed using a permutation test. Explanatory variables used in this multivariate analysis included breastmilk concentrations of HMOs: 2´FL, LNT and LNnT, LNFPIII, LNFPII, LNFPI, LNFPV, LNH, LNnH, LNDFHI, DFL, 6´SL, 3´SL, LSTc, LSTb, LSTa, 3FL (Table S1), delivery mode (normal vaginal, assisted vaginal and C-section), delivery place (home, hospital), gender, gestational age, mother antibiotic use, infant signs of sickness (more specifically, the signs of gastroenteritis including vomiting, fever and diarrhoea) at the time of sample collection, infant age in days, and birth weight. The association between faecal microbiota composition, the assignment of each infant to a specific microbial cluster and the HMO concentrations in corresponding breastmilk samples of the infant's mother were investigated with the Partial Least Squares (PLS) model using MatlabR2107a. The Chi-square test was used to assess the significance of the association between infant gender and infant DMM microbial cluster type, and between mother's secretor status (positive, negative) and infant DMM microbial cluster type.

HMO degradation (consumption) in the infant GI tract was estimated based on profiles in breastmilk and corresponding infant faeces. Based on the degree and types of the HMOs consumed, infants were assigned to categories: "Complete", "Non-specific" and "Specific" (acidic, neutral or other). The Chi-square test was used to assess the significance of the association between consumption category assignment for each HMO and the DMM microbial cluster type of each infant.

Based on the extent to which each individual HMO was consumed (calculated as a ratio of the HMO concentration in infants' faeces and the concentration of the same HMO measured in mothers' milk) infants were divided into tertiles for each individual HMO and divided into "low", "medium", or "high" consumption categories for each HMO. If a given HMO was not detected in milk, the consumption score was not included in the analysis, and if the amount in faeces exceeded the amount detected in milk, the infant was assigned to the "low" category for

that HMO. The association between faecal microbiota composition and the assignment of each infant to a "low", "medium", or "high" consumption category for each HMO were investigated with RDA analysis in Canoco5, with significance assessed using a permutation test [24]. Kruskal-Wallis analysis was performed in QIIME [25, 26] to identify bacterial OTUs that differed significantly between infants who were classified as "high" and "low" consumers for each individual HMO.

Nucleotide sequences

KOALA data sets cannot be made publicly available due to data confidentiality and the potential to identify individual study participants from the data. Data are available to the research community through the Dataverse repository (URL hdl:10411/CEGPGR) upon request to Prof. C. Thijs of the KOALA Study Management Committee at: Maastricht University, Department of Epidemiology, PO Box 616, 6200 MD Maastricht, The Netherlands, e-mail: c.thijs@maastrichtuniversity.nl, tel: +31(0)43 3882389.

Results

HMO Analyses

HMOs in maternal breastmilk and infant faeces were quantified and the minimum, maximum, median, average and standard deviation of the concentrations of each HMO, the HMO type (neutral, fucosylated, and sialylated), and the total amounts were summarised in Table 1. Total concentrations of the measured HMOs in milk ranged from 2.0 to 6.5 mg/mL, and were slightly lower than reported in literature [27], however, our measurements did not include all HMOs types normally present in breastmilk. We also observed large individual variation in the HMO concentrations in both the breastmilk samples and in infant faeces.

Table 1. Average, minimum, maximum, and median concentrations of individual HMOs, classes and total measured HMOs and corresponding standard deviations (SD), in breastmilk and in infants' faeces. For abbreviations of HMOs, please refer to Table S1. NA, not determined.

	Concentrations of HMO, or HMO category									
	maternal breastmilk (µg/mL)					infant faeces (µg/mL)				
HMO	**Min**	**Max**	**Median**	**Average**	**SD**	**Min**	**Max**	**Median**	**Average**	**SD**
3FL	5.0	1098.0	182.0	248.0	222.0	NA	NA	NA	NA	NA
2'FL	0.0	852.8	460.0	372.7	242.3	0.0	240.3	0.5	29.6	61.4
LNT and LNnT	214.2	1806.7	948.0	976.2	319.0	0.0	372.7	15.8	48.6	75.0
LNFPIII	50.7	758.0	243.9	270.1	140.7	0.0	726.7	0.0	41.0	98.0
LNFPII	0.0	1341.5	236.3	339.0	294.3	0.0	549.0	3.1	81.9	125.5
LNFPI	0.0	1493.7	517.2	467.3	367.5	0.0	505.7	0.0	41.8	91.3
LNFPV	0.0	191.4	27.7	41.8	50.1	0.0	75.0	0.0	2.5	8.7
LNH	0.0	313.0	89.5	105.0	64.1	0.0	161.2	0.0	4.4	17.3
LNnH	0.0	299.1	56.1	72.2	56.3	0.0	563.4	0.4	12.1	57.0
LNDFHI	0.0	1856.2	548.3	475.5	388.3	0.0	889.8	28.0	204.3	258.2
DFL	0.0	125.9	42.1	40.4	32.1	0.0	68.4	0.6	9.7	17.0

6´SL	16.6	385.7	97.3	110.8	63.5	0.0	298.5	0.1	18.4	46.7
3´SL	16.8	194.8	91.5	90.7	38.8	0.0	100.0	0.0	4.2	14.9
LSTc	14.7	334.1	98.8	116.2	68.8	0.0	248.5	0.8	28.9	56.8
LSTb	53.2	804.4	244.4	256.2	118.7	0.0	499.9	1.8	52.2	104.5
LSTa	6.3	83.4	24.6	28.2	15.4	0.0	31.7	0.0	2.1	6.1
Sum Neutral	1542.0	5717.3	3064.8	3160.2	824.2	0.0	1671.6	237.9	475.9	523.1
Sum Fucosylated	266.5	4489.4	2042.5	2006.8	737.6	0.0	1591.7	186.1	410.8	463.4
Sum Sialylated	174.1	1273.2	564.7	602.1	210.6	0.0	956.3	3.3	105.8	197.1
Sum Total	1917.4	6545.2	3635.9	3762.3	939.1	0.0	2169.9	267.3	581.7	648.0

Faecal microbiota composition and OTU correlation networks

Illumina HiSeq sequencing of the V4 region of bacterial 16S rRNA genes yielded 14,474,685 high quality reads that passed the quality check and could be assigned to 531 OTUs from 113 genera. In case an OTU could not be classified to a given genus level, it was assigned to the next available taxonomic rank. *Bifidobacterium, Bacteroides, Escherichia-Shigella* and *Parabacteroides* were the predominating genera, with an average relative abundance of 32% (0-91.5%), 21.3% (0-76.7%), 11.8% (0-57.8%) and 6.7% (0-64%), respectively. Genera with the largest number of OTUs included *Bacteroides* (64 OTUs), *Bifidobacterium* (38 OTUs), *Parabacteroides* (31 OTUs), *Lactobacillus* (24 OTUs), and *Streptococcus* (20 OTUs). In total 82 different OTUs were found in more than six infants in the cohort (5% of the population) while the remaining 449 OTUs were only shared by five or fewer infants, and are summarised as “Other” (Table S3).

Faecal microbiota composition of infants in this study cohort was highly variable, yet we could distinguish presence of three universal patterns based on DMM cluster analysis (Figure S1 and Chapter Three of this thesis). These clusters were characterised by microbial communities with a mixed structure (Cluster A), or by communities with either a high relative abundance of *Bifidobacterium* (Cluster C), or a high relative abundance of both *Bifidobacterium* and *Bacteroides* (Cluster B).

In order to investigate statistically significant positive and negative associations between different members within the microbial community we applied Spearman correlation analysis (Figure 1, Table S3). Using the number of node connections and absolute values of Spearman correlation scores as weights, we were able to further identify presence of three microbial network centres using ClusterONE analysis ($p<0.05$). The first network centre (1) consisted of *Bifidobacterium* OTUs and was built around *Bifidobacterium* 418, which shared the highest number of connections of all *Bifidobacterium* OTUs, and was also positively correlating with a number of other OTUs, including those in centre 3. The second centre (2) included *Escherichia-Shigella* 316 positively associated with a few low abundance *Bifidobacterium* OTUs. Finally, the third network centre (3) included a mix of different OTUs, some of which were also linked via positive associations with *Bifidobacterium* 418 and thus centre 1, indicating that microbial groups in centres (1) and (3) might share a symbiotic relationship, possibly related with the presence and activity of *Bifidobacterium* 418. Interestingly, the three centres which we detected with ClusterONE analysis were preserved

across different DMM cluster types (Figure S2), except for centre (2), that was not identified in infants which were classified in the *Bifidobacterium* rich DMM cluster C. In turn, cluster (3) was best developed in infants who were classified in DMM cluster C.

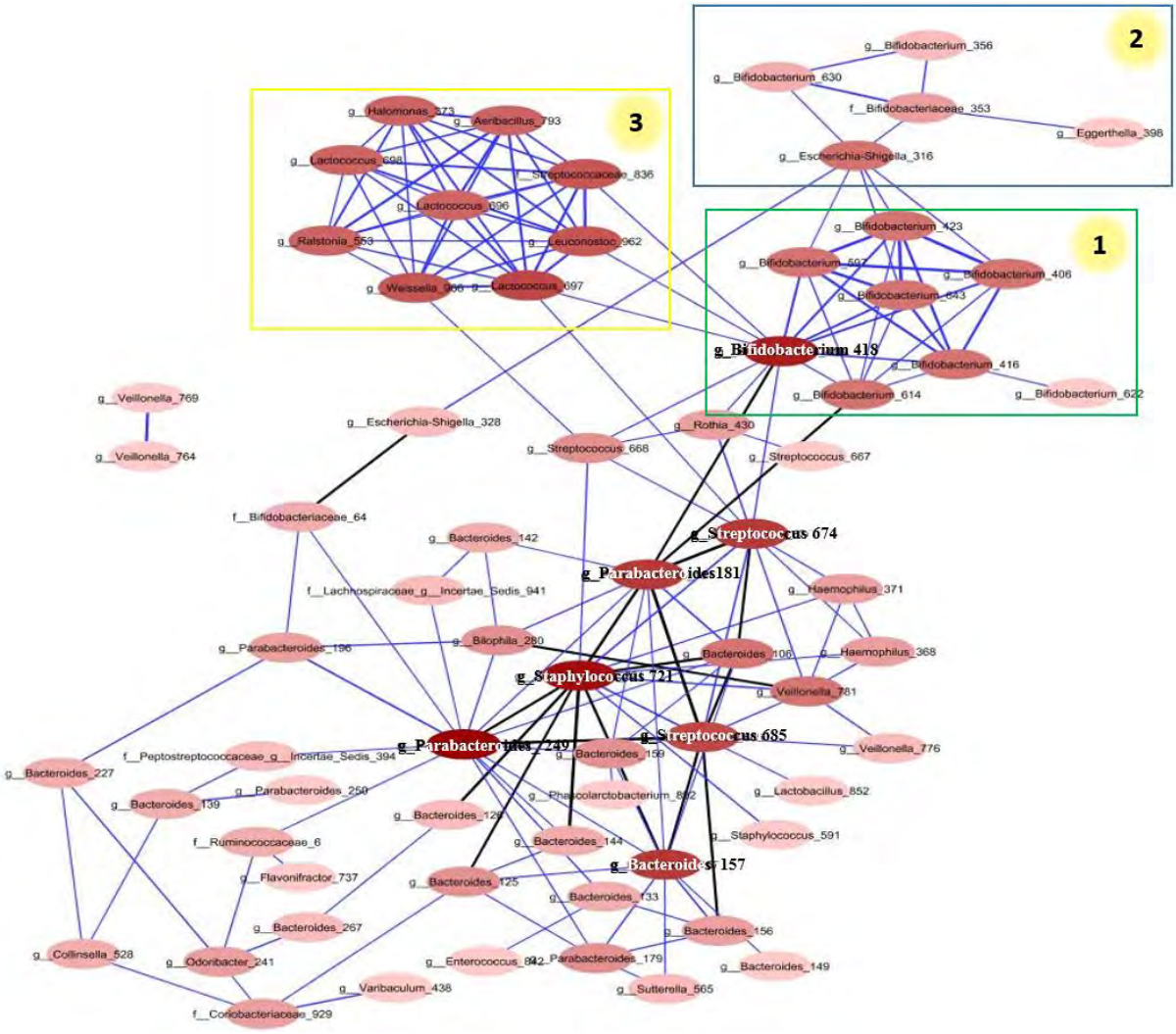


Figure 1. OTU network of statistically significant correlations ($p<0.05$, correlation threshold of ±0.03) identified in the faecal microbial communities based on OTU data from 121 infants included in this study. Node colour intensity is proportional to the number of connections with other nodes. Blue connecting lines show positive associations, black connecting lines show negative associations. Darker shade and thickness of connecting lines indicates higher Spearman correlation score. Darker shade of nodes indicates larger number of connections. The three network centres which were identified as significant ($p<0.05$) in ClusterONE analysis are indicated with yellow, green and blue boxes and are numbered for convenience.

The effect of breastmilk and other factors on faecal microbiota composition

We used RDA to identify factors affecting faecal microbiota composition of infants in our study. Explanatory variables used in this multivariate analysis included breastmilk concentrations of HMOs, mode of delivery, delivery place, gender, gestational age, mothers' antibiotic use, infants' signs of sickness at the time of sample collection, infant age in days, and

birth weight. Together these factors explained 21.7% of the variation in the OTU data. However, only mode of delivery and gender had a significant effect on microbiota composition ($p<0.05$), and 2´FL was borderline significant ($p<0.06$, Figure 2). PLS analysis also showed a significant association between 2´FL (and LNFPI) concentrations in milk with infant microbiota (FDR<0.05, Table S4). When samples were color-coded by infant's DMM cluster type, we also noted that, based on the RDA vector distribution, high levels of breastmilk 2´FL and LNFPI, as well as C-section, were all associated with microbial cluster A, which is characterised by a mixed microbial profile.

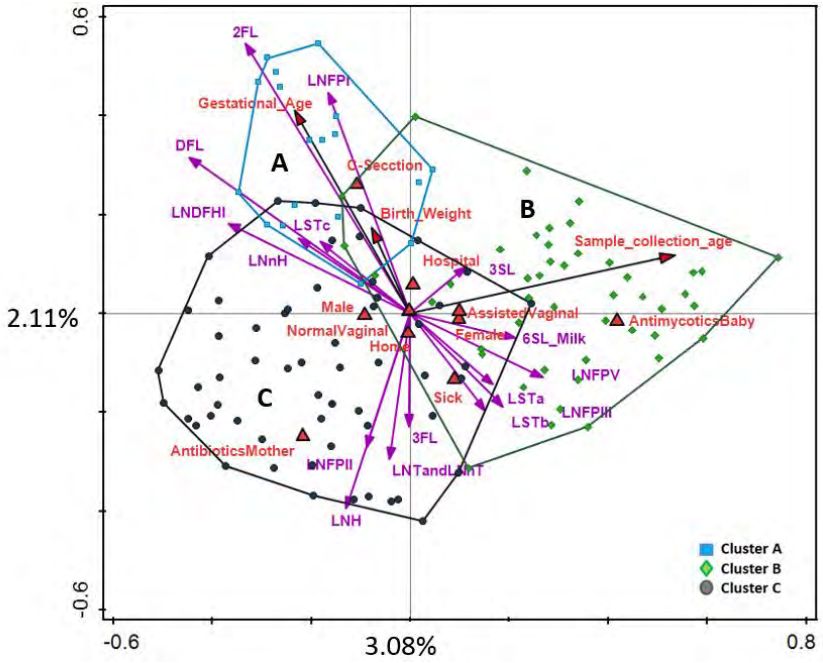


Figure 2. Constrained Analysis (RDA) of different factors and milk HMO levels and their association with the faecal OTU profile of infants. Samples are labelled and enveloped based on the infant assignment to microbial cluster type A, B or C.

The association between breastmilk HMO composition and the assignment of each infant to a specific microbial cluster type were further investigated using RDA (data not shown). The results indicated an association between breastmilk HMO concentrations and the DMM cluster type ($p<0.05$, 3.1% explained). We used Chi-square analysis to test if the mother's secretor status (yes/no), which is known to affect the ability to synthesize (α1-2)-linked fucose, and thus, the amount of neutral HMOs in breastmilk (e.g. 2´FL and LNFPI), had an effect on infant microbiota profiles, as characterised by different DMM cluster types, but the association was not significant ($p=0.08$). Also, infant gender was not significantly associated with any of the DMM microbial cluster types, even though we noted that there were more boys in cluster A (61%) and C (58%), than in cluster B (39%), despite a nearly equal number of boys and girls in the study population (62 and 59, respectively).

<u>The association between infant faecal microbiota composition and HMO degradation</u>

By comparing the HMO profiles in breastmilk and corresponding infant faeces we detected presence of five patterns in the HMO consumption (Figure 3). The first pattern

("Complete") was characterised by low or undetectable amounts of any of the HMOs in infant faeces, suggesting a complete consumption of all HMOs received from the breastmilk (Figure 3a). The second pattern ("Non-specific") showed a faecal HMOs profile that was comparable to that of breastmilk and was of high concentrations, thus implying a non-selective (or broad) and incomplete (or slow) consumption of HMOs by the infant GI tract microbiota (Figure 3b). The third pattern ("Specific") indicated selective consumption of specific HMOs, and was further divided into: "Specific neutral" that showed a high level of neutral HMOs in faeces, meaning that the sialylated (acidic) HMOs (3´SL, 6´SL, LSTa, LSTb, LSTc) were predominantly utilized (Figure 3c), "Specific acidic", which was characterised by the acidic HMO profile of the faeces, meaning that neutral HMOs were predominantly utilized by the infant GI tract microbiota (Figure 3d), and "Specific other", which could not be categorized as neither acidic nor neutral HMOs (data not shown).

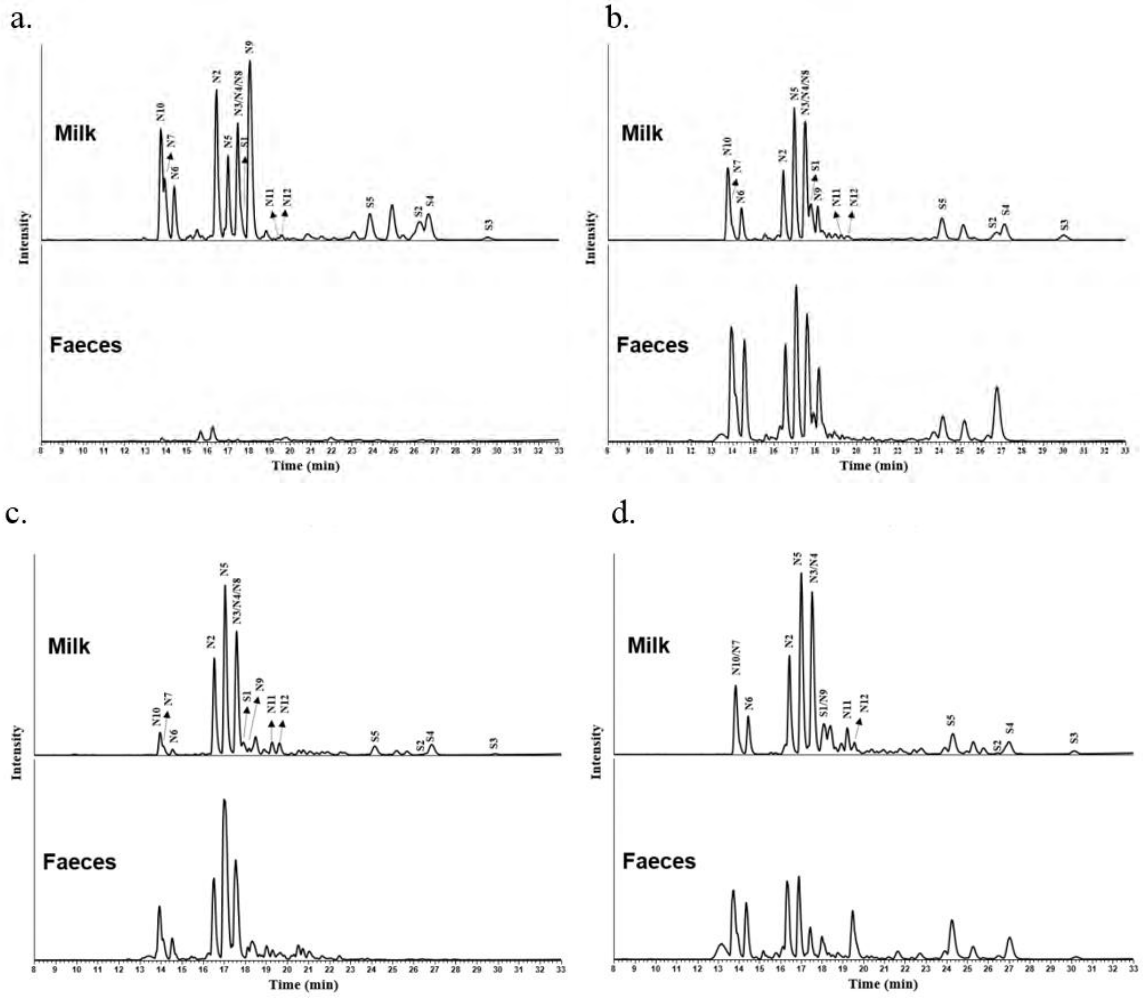


Figure 3. Different utilization patterns of infants defined based on the comparison of HMO profiles in breastmilk and infant faeces. a: complete consumption; b: non-specific; c: specific consumption of acidic HMOs; d: specific consumption of neutral HMOs. Peak assignments are as follows: N2-2´FL, N3-LNT, N4-LNnT, N5-LNFP I, N6-LNFP II, N7-LNFP III, N8-LNFP V, N9-DFL, N10-LNDFH I, N11-LNH, N12-LNnH; S1-6´SL, S2-3´SL, S3-LSTa, S4-LSTb, S5-LSTc.

We used RDA to investigate the association between microbiota composition and different HMO consumption patterns. We noted that "Complete", "Non-specific" and "Specific neutral" consumptions were significantly associated with infant microbiota composition (FDR<0.05), while the association of "Specific acidic" and "Specific other" was not significant. In addition, "Complete" consumption correlated with high relative abundance of bifidobacteria, including the two highly abundant bifidobacterial OTUs 614 and 418 as indicated by vectors on the plot (Figure 4a). Furthermore, the Chi-Square analysis showed a strong and significant association between the frequency of different consumption patterns and each DMM microbial cluster, with 40% of infants who exhibited the mixed microbial profile A also showing a non-specific HMO consumption pattern (Figure 4b).

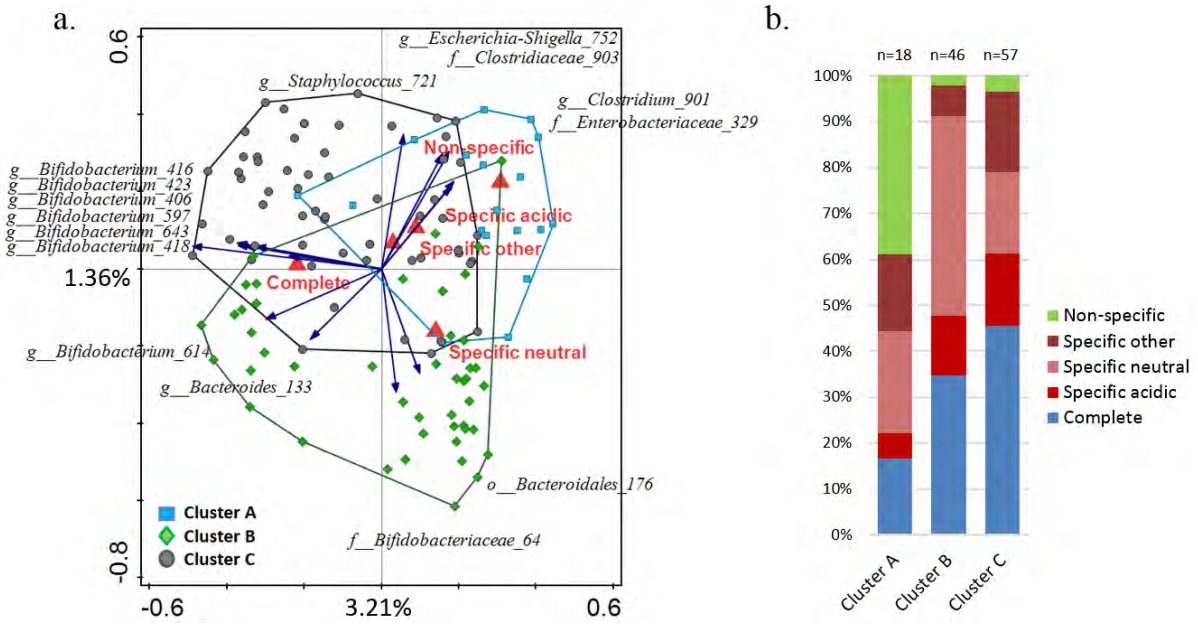


Figure 4. Association of general HMO consumption patterns with microbial cluster types A, B, C; a. RDA showing the association between HMO consumption patterns and microbial OTUs. 15 best fitting OTUs are displayed and samples are color-coded based on their cluster type assignment; b. Segregation of infants based on their HMO consumption pattern in relation to their microbial cluster type classification.

In order to investigate the association between microbiota composition and consumption of specific breastmilk HMOs in more depth, we classified infants as "low", "medium" or "high" consumers for each of the measured HMOs. We then used this classification in the multivariate RDA analysis and showed that the HMO consumption explained 61.5% of variation in microbiota. The microbiota composition was significantly associated with the degradation of 2´FL, LNT and LNnT, DFL, 6´SL, LNH, LNFPII and LNFPIII (FDR<0.05), and there was a trend with the degradation of LSTc, LSTb and 3´SL (FDR=0.07; Figure 5).

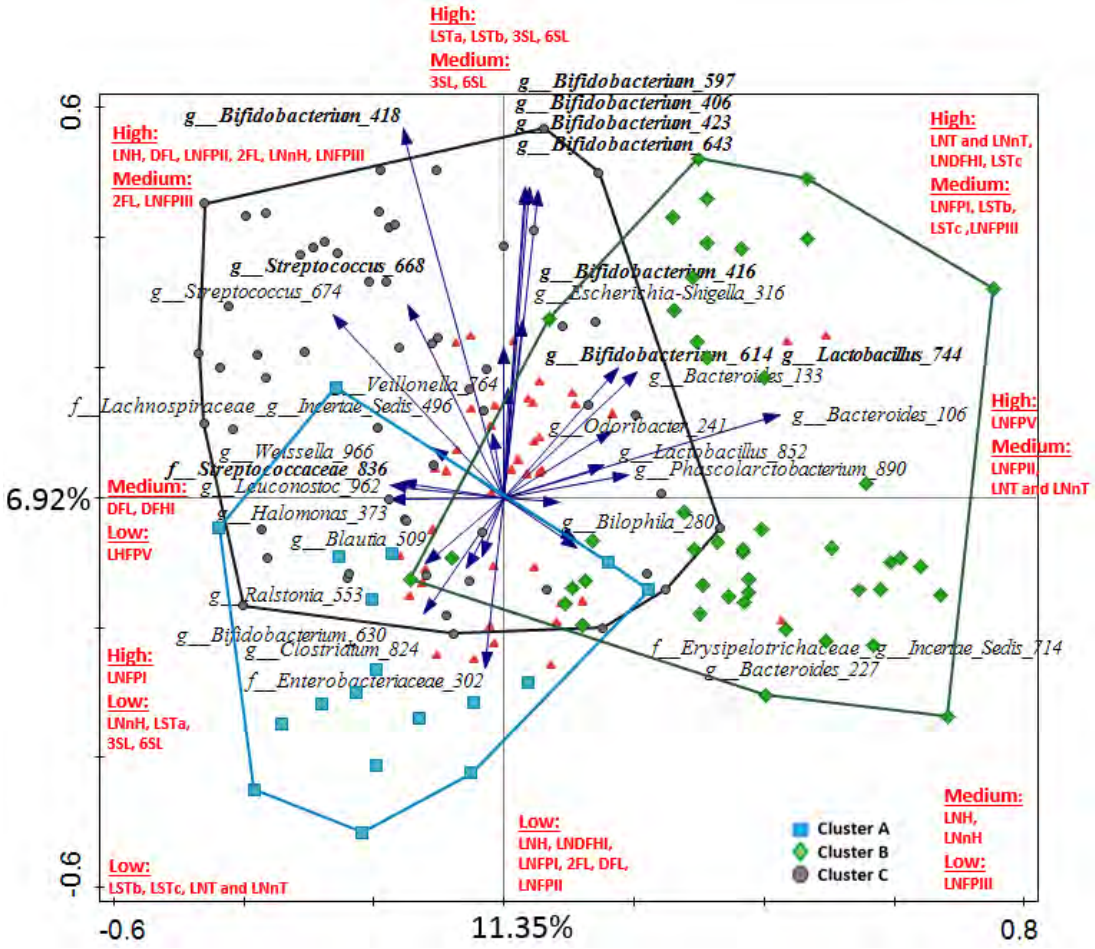


Figure 5. RDA showing the association between the degree of degradation of individual HMOs and microbial OTUs. OTUs which were significantly ($p<0.05$) increased in high-degrading infants for at least one of the HMOs are displayed. Taxa with FDR<0.05 are highlighted in bold. For more information on average relative abundance of the displayed OTUs in the study population and the detailed results of Kruskal-Wallis analyses, see Tables S3 and S5. Samples are color-coded based on microbiota cluster type assignment. Red triangles indicate consumption of each HMO, as summarized in red text.

For all HMO types, there was a general trend in relating consumption efficiency and infant faecal microbiota cluster class. RDA showed that microbial cluster type alone could explain 8.4% of variation in the consumption category and that the cluster effect was statistically significant (FDR<0.05). The lowest efficiency of consumption was linked to microbial cluster type A, with 49.4% of all HMOs consumed at "low" level, 10% at "medium" level and 40.6% at "high" level. Infants classified in microbial cluster type B showed high HMO consumption levels, with 47.3% of all HMO types consumed at "high" level, 21.3% consumed at "medium" level and 31.4% at "low" level. Infants classified in microbial cluster type C, showed "high" consumption for 49.8% of all HMOs, "medium" consumption for 24.4% and "low" consumption for 25.8% of the HMOs recorded. The microbial cluster type consumption efficiency pattern varied for different HMO types.

The Chi-square analysis was used to test the correlation between the proportion of infants in "high", "medium", and "low" consumption categories for each HMO and the infant

microbial cluster groups. Significant ($p<0.05$) differences were detected between clusters with respect to consumption of 2´FL, LNFPIII, LNFPII, DFL and 6´SL. For the aforementioned HMOs, the highest proportion of infants with lowest ability to break down these HMOs was found in cluster type A (Figure 6).

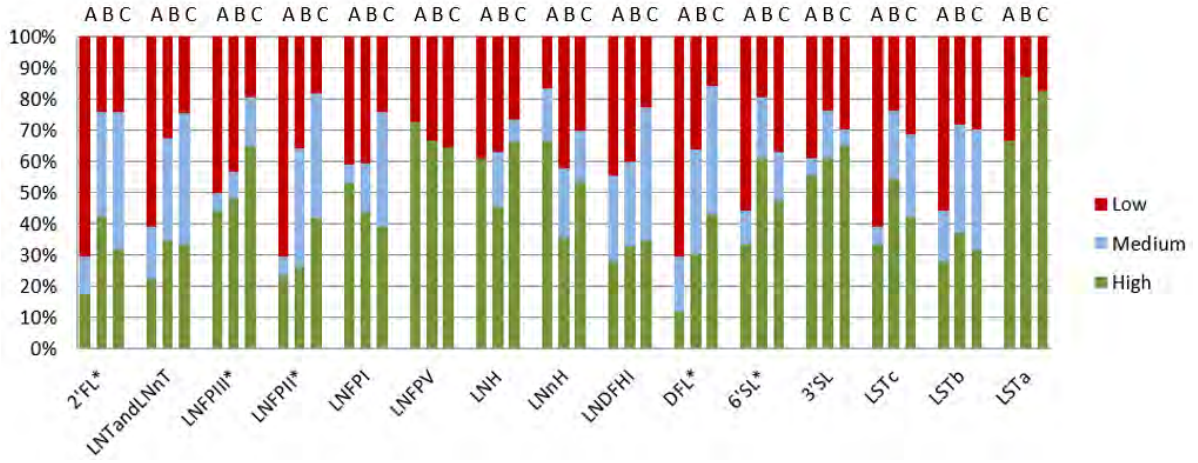


Figure 6. Proportion of infants showing either "high", "medium" or "low" HMO consumption levels, within each microbial cluster class: A, B, C. Significant differences in distribution as determined by Chi-square analysis are indicated with an asterisk.

Kruskal-Wallis analysis was used to compare microbiota composition at the OTU level between infants who were classified as either "high" or "low" consumers for each HMO measured in this study (Figure 7, Table S5). Infants who showed "high" consumption of 2´FL and DFL had significantly higher relative abundance of OTUs *Bifidobacterium* 418 and *Lactobacillus* 744 (FDR<0.05). In addition, "high" DFL consumption was associated with significantly higher relative abundance of *Bifidobacterium* OTUs 406, 643, 423, and 597. Similarly, infants who showed "high" consumption of LNT and LNnT, LNFPIII, LNFPII, LNH had a significantly higher relative abundance of *Bifidobacterium* OTUs 418, 406, 643, 423, 597, 416 (LNFPII only) and *Bifidobacterium* 614 (LNFPII and LNH only). Relative abundance of OTU *Bifidobacterium* 418 was also significantly higher in infants who were efficient degraders of LNnH and LNDFHI. We could not detect statistically significant differences (with FDR<0.05) in the relative abundance of taxa between infants characterised as "high" and "low" degraders of LNFPI, LNFPV, 3´SL, LSTa, LSTb and LSTc.

Kruskal-Wallis test comparing infants who were classified as "low" and "medium" consumers of different HMOs showed statistically significant differences (FDR<0.05) in relative abundance of *Bifidobacterium* 418 for 2´FL and LNFPII, *Bifidobacterium* OTUs 418, 643, 406 and 423 for DFL, *Bifidobacterium* 418 and *Bifidobacterium* 416 for LNFPIII, *Bifidobacterium* OTUs 418, 643, 406, 597 and 423 for LNT and LNnT (data not shown). None of the OTUs differed when the same comparison was done between the "medium" and "high" consumers for any of the HMOs.

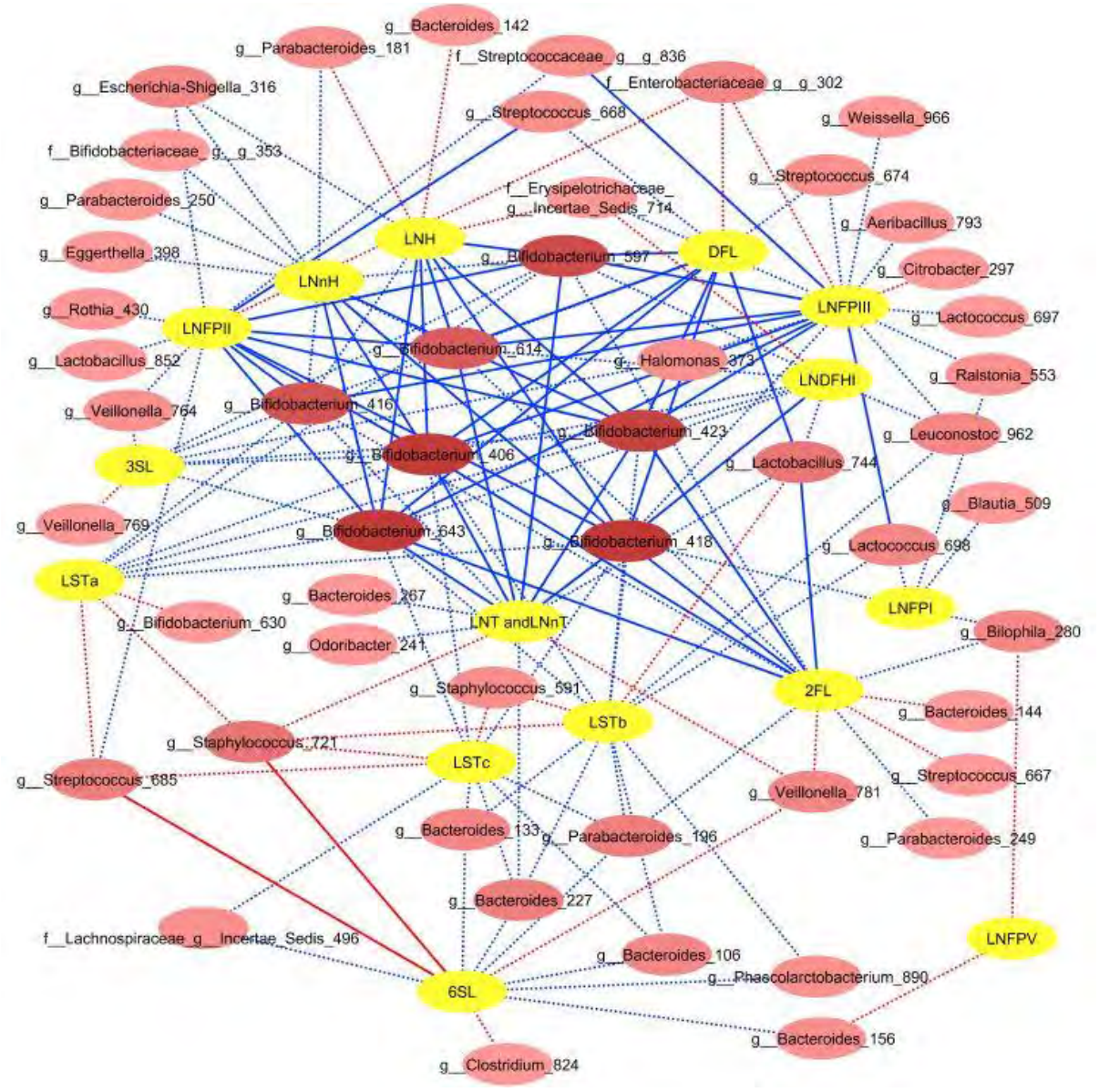


Figure 7. OTUs significantly associated with HMO consumption based on Kruskal-Wallis test including infants classified as high and low consumers for each HMO. Red lines indicate higher OTU relative abundance in relation to low HMO consumption, blue lines indicate higher relative abundance in relation to high consumption. OTU nodes which are connected with the highest number of HMOs are indicated by darker shades of pink. Dotted lines indicate associations with $p<0.05$, solid lines indicate associations with FDR<0.05.

Discussion

The 16S rRNA sequencing analysis described here gave us an opportunity to characterize the faecal microbial ecosystem found in healthy, one month old breastfed infants. Using DMM analysis we showed that infants in this cohort could be categorised into three microbial cluster types based on their faecal microbiota profiles. The occurrence of similar clusters has been reported before in another study [28], although a number of different studies to date also described high variability in infant faecal microbiota composition [29]. In addition,

similar clusters have been defined in the BINGO cohort, as described in more detail in Chapter Three of this thesis.

A number of different factors in early life may affect the dynamics of the developing infant GI microbiota composition. In the first weeks of life the GI microbial ecosystem is not yet fully formed, but instead it undergoes a gradual succession, which leads to the establishment of a stable microbial community that is highly specialized for milk digestion [3]. Our results showed that at about four weeks of age, mode of delivery and gender significantly contributed to explaining the observed variation in microbiota composition ($p<0.05$). The effect of mode of delivery has been also indicated in infants of similar age by another study [30], but there had been no reports in the literature on effect of gender in infants of this age. Interestingly, both, mode of delivery and gender effects have also been detected in the BINGO study described in more detail in Chapters Three and Five of this thesis.

Patterns in maternal breastmilk HMOs have a small effect on faecal microbiota composition in one month old infants

One of the key factors shaping infant microbiota in early life is infant diet. Exclusively breastfed infants show high inter-individual variability, leading to a question of whether the variability can be linked to the breastmilk properties, including the unique HMOs composition of the mother's milk. Breastmilk is a rich source of HMOs and it has been estimated that in healthy breastfed infants about 40% to 97% of the ingested HMOs reach the infant colon [31-33]. These breastmilk HMOs serve as an abundant and diverse carbon source available for bacterial fermentation, and the composition and concentrations of different HMOs are likely among the key factors responsible for shaping the microbiota in the lower intestine of breastfed infants [3]. Using PLS modelling we could detect statistically significant associations between infant faecal microbiota composition and milk LNFPI and 2´FL (Table S4). Both, LNFPI and 2´FL are neutral, fucosylated, unbranched HMOs with α1,2 linkage joining fucose and galactose, and they were on average the third and fourth most abundant HMO measured in our set. The same association was found in another study of three month old breastfed infants, albeit using a much smaller cohort (n=16), which further indicated that LNFPI was positively associated with *Bacteroides* and *Bifidobacterium,* and 2´FL with *Bacteroides* [7]. Based on our RDA analysis (Figure 2), these two HMOs were associated with mixed microbiota cluster type A, that is characterised by low relative abundance of *Bacteroides* and *Bifidobacterium*. In addition, none of the other HMOs showed significant association with microbiota composition, but based on the vector positions (Figure 2), we could argue that higher concentrations of a number of different HMOs in maternal milk could be driving the infant microbiota away from cluster A, but not specifically towards clusters B or C. One of the possible explanations could be that a combined effect of a number of HMO structures may be necessary to guide microbiota development in early life, or that stronger associations develop over a longer period, and that at one month of age the microbial profile of infants in our study was still largely in its transitional phase [34, 35].

In addition, other breastmilk components, such as secretory IgA, lactoferrin, lysozyme, as well as the breastmilk microbiota itself, are all likely to contribute to shaping the structure of microbial communities within infants' GI tract [2, 36-39], possibly concealing the effect the

individual HMOs. Finally, breastmilk also contains its own microbiota, and previous studies showed that *Streptococcus, Staphylococcus*, together with few other genera, (such as *Lactobacillus*, *Serratia, Pseudomonas, Corynebacterium, Ralstonia, Propionibacterium, Sphingomonas,* unclassified Bradyrhizobiaceae*, Herbaspirillum, Rothia, Stenotrophomonas, Acinetobacter, Bacteroides, Halomonas, Veillonella,* and *Delftia*) can be detected in breastmilk [40, 41]. When we used Spearman correlation analyses of the OTU data to identify significant interactions (positive and negative) between members of microbial communities commonly found in infant faeces (Figure 1) we noted that members of the genera *Bifidobacterium, Bacteroides, Parabacteroides, Streptococcus* and *Staphylococcus* seem to be the key players in the networks we identified and that these groups shared mostly positive associations. Thus, based on our findings and the available literature data, we could speculate that the high relative abundance of these taxa and mostly positive associations between them might, at least in part, be due to infant ingesting these bacterial assemblages during breastfeeding.

In our dataset we also detected presence of negative associations between some of the microbial groups. For example, the two most abundant OTUs, *Bifidobacterium* 418 and *Bifidobacterium* 614, were negatively associated with certain *Parabacteroides*, *Citrobacter* and *Clostridium* OTUs (Figure 1). These negative associations might be rooted in the competitive advantage of specific microbial taxa in their ability to effectively utilize HMOs. For example, the negative association between *Bifidobacterium* and clostridia has been documented in earlier studies, specifically in the context of infant feeding mode (breastfeeding vs. formula feeding; Chapter Three of this thesis). Earlier studies showed that clostridia are unable to utilise HMOs, and have also been shown to be much less efficient in consumption of prebiotics such as GOS or FOS in *in vitro* trials [39, 42]. In addition, human milk also contains large amounts of lysozyme, up to 400µg/mL, to which both *B. bifidum* and *B. longum* are resistant, while clostridia, and many other Gram positive and Gram negative bacteria, are highly susceptible [39].

HMO consumption patterns are associated with specific microbial groups

The HMO profiles from the matching breastmilk and faecal samples allowed us to classify infants into five HMO consumption groups. Similar patterns have been reported previously [43], however, possible links between those utilization patterns and the infant faecal microbiota composition have remained largely unknown. Our data showed a strong significant association between “Non-specific” consumption and microbial DMM cluster A, whereas the “Complete” consumption pattern was related with cluster C and “Specific-neutral” with cluster B (Figure 4). Thus, even though the GI tract microbial ecosystem may not yet be completely formed at one month of age, we showed that the degradation of different types of HMOs is carried out by specific bacterial assemblages, which evolved mechanisms for efficient consumption of this abundant food component in milk.

Bifidobacteria are the main group of microorganisms in the infant GI tract, and also the main consumers of HMOs [44]. Our data supports this, as the most abundant and prevalent OTUs in our dataset were of bifidobacterial origin (Table S3). Earlier *in vitro* studies showed highly specific metabolic behaviour of different bifidobacterial species and strains with respect to their ability to utilize different HMOs [6]. The most prevalent OTU in our set was

Bifidobacterium 614*,* with an average relative abundance of 23.3 % and prevalence of 92%. NCBI blast analysis revealed that the OTU sequence (Table S6) matched several different species and strains of *Bifidobacterium,* including various strains of *B. longum (infantis)*, commonly found in the infant GI tract. Our analysis showed that *Bifidobacterium* 614 was associated with high consumption level of various HMOs, specifically 2´FL, DFL, LNDFHI, LNFPII, LNFPIII, LNH, and LNT and LNnT (Table S5). The second most abundant OTU was *Bifidobacterium* 418*,* found in 45% of infants with an average relative abundance of 6.6%. The NCBI blast analysis of OTU 418 returned a 100% match to several different strains of *Bifidobacterium bifidum* (DSM 20456 = ATCC 29521 = JCM 1255**,** NBRC100015, KCTC3202). The presence of *Bifidobacterium* 418 correlated strongly (FDR<0.05) with high consumption levels of 2´FL, LNT and LNnT, LNFPIII, LNFPII, LNH and its isomer LNnH, LNDFHI, and DFL, and with LNFPI, LSTa, LSTb and LSTc ($p<0.05$). *B. bifidum* has been shown to be an efficient HMO degrader in *in vitro* fermentation studies able to secrete glycosidases to degrade HMOs extracellularly, also making it possible for other species/subspecies to access the HMO degradation by-products and metabolites during cross feeding [6]. *In vitro* studies showed that *B. bifidum* DSM 20456 could efficiently degrade LNT, 2´FL, LNnT, LNFPI, LNFPII, LNFPIII, and LNDFHI, though the rate at which it was degrading these HMOs varied [6].

Streptococcus and *Staphylococcus* OTUs showed an increase in relative abundance specifically in relation to high consumption of the fucosylated HMOs - DFL, LNFPII and LNFPIII. However, i*n vitro* studies showed that *Streptococcus* and *Staphylococcus* cannot effectively metabolize breastmilk HMOs [45]. Based on our microbial network analysis (Figure 1), we noted that these groups also showed a strong positive association with *Bifidobacterium* 418 and *Bifidobacterium* 614, suggesting that positive trophic interactions between members of these groups may be present in the infant GI tract. It is also possible that HMOs may enhance growth of certain bacteria without being metabolised. For example, HMOs have been shown to bind and possibly activate a growth-promoting signalling in some strains of breastmilk *Staphylococcus* without being actively metabolised by this strain [45].

Bacteroides and *Parabacteroides* (formerly also *Bacteroides*) are among the first dominant bacterial groups established in the infant GI tract [46]. In general, members of the genus *Bacteroides* can degrade a broad range of simple and complex sugars, oligosaccharides, and polysaccharides, including some HMOs, mucus glycans, and plant derived polysaccharides [46]. Like bifidobacteria, *Bacteroides* spp. can grow on milk glycans as a sole carbon source, yet they cannot degrade a wide range of HMOs, which during infancy, and up to the weaning period, gives bifidobacteria a competitive advantage over members of the genus *Bacteroides* [46]. This might explain why infants with high levels of *Bifidobacterium,* such as those classified in the DMM cluster C tend to have lower levels of *Bacteroides*. On the other hand, *Bacteroides* has been shown to efficiently degrade mucus glycans, and because of the similarity of some HMO structures and mucus glycans, some *Bacteroides* species could also effectively degrade specific HMOs by activating the mucus degrading pathway [46]. These species might be better at competing with bifidobacterial groups that show fewer adaptations for HMO utilisation. Our analysis indicated that infants who were efficient degraders of the sialylated (acidic) HMOs (3´SL, 6´SL, LSTa, LSTb, LSTc) and classified into "Specific neutral" consumption category were also often assigned to the *Bacteroides* dominated DMM cluster B

(Figure 3). This was in agreement with another study which showed that for example *Bacteroides fragilis, Bacteroides vulgatus* and *Bacteroides thetaiotaomicron* were capable of utilizing specific acidic HMOs (3´SL, 6´SL) [47].

Finally, a biologically important microbial group commonly detected in infant faeces are the lactobacilli. Our results show that high levels of degradation of 2´FL, DFL, LNDFHI, LNT and LNnT and LNFPII were significantly correlated with higher relative abundance of this group. Remarkably, the opposite effect was noted for LSTb (Figure 7). Unfortunately, the two interesting lactobacilli OTUs which were identified by the analysis, namely *Lactobacillus* 744 and *Lactobacillus* 852, had sequence reads which returned a hundred percent match to more than a dozen species and strains of lactobacilli in NCBI blast analysis, making it impossible to unequivocally identify these populations to the species level. Several *Lactobacillus* spp. have been frequently isolated from neonate faeces, including: *L. fermentum, L. casei, L. paracasei, L. delbrueckii, L. gasseri, L. rhamnosus* and *L. plantarum* [48]. These lactobacilli were shown to be unable to efficiently ferment HMOs *in vitro* [44, 49], and did not correlate with a decrease in faecal HMOs *in vivo* [7]. However, they have been shown to grow well on HMO metabolites *in vitro* [44]. Thus, via the cross feeding with other bacteria, for example bifidobacteria, it is possible that HMO degradation can be linked with higher relative abundance of lactobacilli, and other community members in the microbial ecosystem within GI tract.

The roles of different breastmilk HMOs in development of infant GI tract microbiota, the occurrence of microbial clusters, and the nutritional and health consequences relating to the existence of different trophic networks that are built upon the degradation of specific HMOs are still largely unknown. Our data reinforced the central role of bifidobacteria in the HMO breakdown, and provided an insight into different microbial assemblages in healthy, one month old infants. Furthermore, carrying out the analyses at the OTU level allowed us to uncover a higher level of detail showing that, for example, bifidobacteria were associated with both clusters B and C (17% and 41%, respectively), but the distribution of specific bifidobacterial OTUs within these clusters was not the same (data not shown). Until now few *in vitro* studies demonstrated that closely related species or strains might exhibit different capacities and be involved in a range of trophic interactions. Future studies should strive to improve our understanding on how bacterial assemblages form *in vivo*, which species or strains are present and interacting with each other, to identify the key species, their role in driving the colonisation, the effects on the host, and finally how this knowledge could be translated into practical applications within infant nutrition and health.

Conclusion

GI tract microbiota composition in one-month old breastfed infants is shaped by multiple factors, including breastmilk HMOs, however, we could not show a strong direct link between specific HMOs in mother's milk and microbial community composition in this cohort. We hypothesized that such link might be formed later in the infancy, or might be a result of the interaction between HMOs and other bioactive components present in breastmilk. We showed that breastmilk HMO degradation patterns differed among infants belonging to different

microbial cluster types. Degradation of specific HMOs could be correlated with statistically significant increase in relative abundance of various phylotypes (OTUs) within the genus *Bifidobacterium*, and to lesser extent within the genera *Bacteroides* and *Lactobacillus*.

Acknowledgements

The authors thank Maria Suarez-Diez (Wageningen University, Laboratory of Systems and Synthetic Biology) for her help with the DMM analysis.

Funding

This research was performed in the public-private partnership CarboHealth coordinated by the Carbohydrate Competence Center (CCC, www.cccresearch.nl) and financed by participating partners and allowances of the TKI Agri&Food program, Ministry of Economic Affairs of the Netherlands.

References

1. Rautava, S., et al., *Microbial contact during pregnancy, intestinal colonization and human disease.* Nat Rev Gastroenterol Hepatol, 2012. **9**(10): p. 565-76.
2. Smilowitz, J.T., et al., *Breast milk oligosaccharides: structure-function relationships in the neonate.* Annu Rev Nutr, 2014. **34**: p. 143-69.
3. Zivkovic, A.M., et al., *Establishment of a Milk- Oriented Microbiota (MOM) in Early Life: How Babies Meet Their MOMs.* Functional Food Reviews, 2013. **5**(1): p. 3–12.
4. Marcobal, A., et al., *Consumption of human milk oligosaccharides by gut-related microbes.* J Agric Food Chem, 2010. **58**(9): p. 5334-40.
5. Bode, L. and E. Jantscher-Krenn, *Structure-function relationships of human milk oligosaccharides.* Adv Nutr, 2012. **3**(3): p. 383S-91S.
6. Asakuma, S., et al., *Physiology of consumption of human milk oligosaccharides by infant gut-associated bifidobacteria.* J Biol Chem, 2011. **286**(40): p. 34583-92.
7. Wang, M., et al., *Fecal microbiota composition of breast-fed infants is correlated with human milk oligosaccharides consumed.* J Pediatr Gastroenterol Nutr, 2015. **60**(6): p. 825-33.
8. Coppa GV1, G.O., Pierani P, Catassi C, Carlucci A, Giorgi PL, *Changes in carbohydrate composition in human milk over 4 months of lactation.* Pediatrics, 1993. **91**(3): p. 637-41.
9. Austin, S., et al., *Temporal Change of the Content of 10 Oligosaccharides in the Milk of Chinese Urban Mothers.* Nutrients, 2016. **8**(6).
10. Sprenger, N., et al., *Longitudinal change of selected human milk oligosaccharides and association to infants' growth, an observatory, single center, longitudinal cohort study.* PLoS One, 2017. **12**(2): p. e0171814.
11. Kummeling, I., et al., *Etiology of atopy in infancy: the KOALA Birth Cohort Study.* Pediatr Allergy Immunol, 2005. **16**(8): p. 679-84.
12. Scheepers, L.E., et al., *The intestinal microbiota composition and weight development in children: the KOALA Birth Cohort Study.* Int J Obes (Lond), 2015. **39**(1): p. 16-25.
13. B. E. P. Snijders, et al., *Cytokines and soluble CD14 in breast milk in relation with atopic manifestations in mother and infant (KOALA Study).* Clinical and Experimental Allergy, 2006. **36**: p. 1609-1615.
14. Penders, J., et al., *Factors Influencing the Composition of the Intestinal Microbiota in Early Infancy.* Pediatrics, 2006. **118**(2): p. 511-521.
15. Ramiro-Garcia, J., et al., *NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes.* F1000Research, 2016. **5**: p. 1791.
16. Albrecht, S., et al., *CE-LIF-MS n profiling of oligosaccharides in human milk and feces of breast-fed babies.* Electrophoresis, 2010. **31**(7): p. 1264-73.
17. Wu, S., et al., *Development of an annotated library of neutral human milk oligosaccharides.* J Proteome Res, 2010. **9**(8): p. 4138-51.
18. Wu Shuai, R.G., J. Bruce German, and Carlito B. Lebrilla, *Annotation and Structural Analysis of Sialylated Human Milk Oligosaccharides.* Journal of Proteome Research, 2011. **10**: p. 856–868.
19. Stephan Thurl, B.M.-W., and Gunther Sawatzki, *Quantification of Individual Oligosaccharide Compounds from Human Milk Using High-pH Anion-Exchange Chromatography.* Analytical Biochemistry, 1996. **235**: p. 202 – 206.
20. Quast, C., et al., *The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.* Nucleic Acids Res, 2013. **41**(Database issue): p. D590-6.
21. Ian Holmes, K.H., Christopher Quince, *Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics.* PLoS ONE, 2012. **7**(2): p. 1-15.
22. Morgan, M., *DirichletMultinomial: Dirichlet-Multinomial Mixture Model Machine Learning for Microbiome Data.* R package version 1.18.0., 2017.
23. Shannon, P., et al., *Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.* Genome Research, 2003. **13**(11): p. 2498-504.
24. Šmilauer, P., & Lepš, J. , *Multivariate Analysis of Ecological Data using CANOCO 5.* Cambridge: Cambridge University Press, 2014.

25. Kuczynski, J., et al., *Using QIIME to analyze 16S rRNA gene sequences from microbial communities.* Curr Protoc Bioinformatics, 2011. **Chapter 10**: p. Unit 10 7.
26. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data.* Nat Methods, 2010. **7**(5): p. 335-6.
27. Bode, L., *Human milk oligosaccharides: every baby needs a sugar mama.* Glycobiology, 2012. **22**(9): p. 1147-62.
28. Fujimura, K.E., et al., *Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation.* Nat Med, 2016. **22**(10): p. 1187-1191.
29. Madan, J.C., et al., *Normal neonatal microbiome variation in relation to environmental factors, infection and allergy.* Curr Opin Pediatr, 2012. **24**(6): p. 753-9.
30. Madan, J.C., et al., *Association of Cesarean Delivery and Formula Supplementation With the Intestinal Microbiome of 6-Week-Old Infants.* JAMA Pediatr, 2016. **170**(3): p. 212-9.
31. P, C., et al., *Survival of human milk oligosaccharides in the intestine of infants.* Adv Exp Med Biol., 2001. **501**: p. 315-23.
32. Mark J. Gnoth, C.K., Evamaria Kinne-Saffran, and Silvia Rudloff, *Human Milk Oligosaccharides Are Minimally Digested In Vitro.* The Journal of Nutrition, 2000. **130**(12): p. 3014-20.
33. Coppa GV, P.P., Zampini L, Bruni S, Carloni I, Gabrielli O., *Characterization of oligosaccharides in milk and feces of breast-fed infants by high-performance anion-exchange chromatography.* Adv Exp Med Bio, 2001. **501**: p. 307-14.
34. JE, K., et al., *Succession of microbial consortia in the developing infant gut microbiome.* Proc Natl Acad Sci U S A, 2011. **108** (Suppl 1): p. 4578-85.
35. Chana Palmer, E.M.B., Daniel B. DiGiulio, David A. Relman, Patrick O. Brown, *Development of the Human Infant Intestinal Microbiota.* PLoS Biology, 2007. **5** |(7): p. 1556-1573.
36. Solis, G., et al., *Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut.* Anaerobe, 2010. **16**(3): p. 307-10.
37. Jost, T., et al., *Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding.* Environ Microbiol, 2014. **16**(9): p. 2891-904.
38. Maga, E.A., B.C. Weimer, and J.D. Murray, *Dissecting the role of milk components on gut microbiota composition.* Gut Microbes, 2013. **4**(2): p. 136-9.
39. Rockova, S., et al., *Growth of bifidobacteria and clostridia on human and cow milk saccharides.* Anaerobe, 2011. **17**(5): p. 223-5.
40. Hunt, K.M., et al., *Characterization of the diversity and temporal stability of bacterial communities in human milk.* PLoS One, 2011. **6**(6): p. e21313.
41. Li, S.-W., et al., *Bacterial Composition and Diversity in Breast Milk Samples from Mothers Living in Taiwan and Mainland China.* Frontiers in Microbiology, 2017. **8**.
42. Rada, V., et al., *Growth of infant faecal bifidobacteria and clostridia on prebiotic oligosaccharides in in vitro conditions.* Anaerobe, 2008. **14**(4): p. 205-8.
43. Albrecht, S., et al., *Occurrence of oligosaccharides in feces of breast-fed babies in their first six months of life and the corresponding breast milk.* Carbohydrate Research, 2011. **346**(16): p. 2540-2550.
44. Schwab, C. and M. Ganzle, *Lactic acid bacteria fermentation of human milk oligosaccharide components, human milk oligosaccharides and galactooligosaccharides.* FEMS Microbiol Lett, 2011. **315**(2): p. 141-8.
45. Hunt, K.M., et al., *Human milk oligosaccharides promote the growth of staphylococci.* Appl Environ Microbiol, 2012. **78**(14): p. 4763-70.
46. Marcobal, A., et al., *Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways.* Cell Host Microbe, 2011. **10**(5): p. 507-14.
47. Yu, Z.-T., C. Chen, and D.S. Newburg, *Utilization of major fucosylated and sialylated human milk oligosaccharides by isolated human gut microbes.* Glycobiology, 2013. **23**(11): p. 1281-1292.
48. Haarman, M. and J. Knol, *Quantitative real-time PCR analysis of fecal Lactobacillus species in infants receiving a prebiotic infant formula.* Appl Environ Microbiol, 2006. **72**(4): p. 2359-65.
49. Ward, R.E., et al., *In vitro fermentation of breast milk oligosaccharides by Bifidobacterium infantis and Lactobacillus gasseri.* Appl Environ Microbiol, 2006. **72**(6): p. 4497-9.

50. Alessio Ceroni, K.M., Hildegard Geyer, Rudolf Geyer, Anne Dell and Stuart M. Haslam, *GlycoWorkbench: A Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans*. Journal of Proteome Research, 2007. **7**: p. 1650–1659.

Supplementary tables and figures

Table S1. HMO categories, names and abbreviations included in this study [50]

Category	Name	Abbreviation
Neutral	2′-Fucosyllactose	2'FL
	3′-Fucosyllactose	3FL
	Lacto-*N*-tetraose	LNT
	Lacto-*N*-neotetraose	LNnT
	Lacto-*N*-fucopentaose I	LNFPI
	Lacto-*N*-fucopentaose II	LNFPII
	Lacto-*N*-fucopentaose III	LNFPIII
	Lacto-*N*-fucopentaose V	LNFPV
	Difucosyllactose	DFL
	Lacto-*N*-difucohexaose I	LNDFHI
	Lacto-*N*-hexaose	LNH
	Lacto-*N*-neo-hexaose	LNnH
Acidic	6'-Sialyllactose	6'SL
	3′-Sialyllactose	3'SL
	Sialyl-lacto-*N*-tetraose a	LSTc
	Sialyl-lacto-*N*-tetraose b	LSTb
	Sialyl-lacto-*N*-tetraose c	LSTa

Table S2. Infant – mother pair demographics

Infant/Mother pairs (n=121)		**n**
Delivery mode	Normal Vaginal	100
	Assisted Vaginal	11
	C-Section	10
Delivery Place	Home	72
	Hospital	49
Gender	Female	59
	Male	62
Gestation (weeks)	Mean ± SEM	40.24 ±1.16
	Minimum	37.14
	Maximum	42.86
Birth Weight (g)	Mean ± SEM	3651 ±43.48
	Minimum	2140
	Maximum	4780
Age at collection (days)	Mean ± SEM	32.56 ±0.5
	Minimum	24
	Maximum	56
Health at collection	Sick	3
	Not Sick	118
Medication Use at time of collection	Antibiotics (Baby)	0
	Antimycotics (Baby)	2
	Antibiotics (Mother)	2

Table S3. Average relative abundance and prevalence of the OTUs shared by at least 5% of infants in this study.

OTU	Average RA (%)	Prevalence (%)
f_Bifidobacteriaceae_64_g_g	0.03	11
f_Bifidobacteriaceae_353_g_g	0.01	7
g_Bifidobacterium_614	23.32	92
g_Bifidobacterium_418	6.64	45
g_Bifidobacterium_622	1.33	50
g_Bifidobacterium_643	0.13	29
g_Bifidobacterium_406	0.13	29
g_Bifidobacterium_423	0.12	27
g_Bifidobacterium_597	0.11	26
g_Bifidobacterium_416	0.03	16
g_Bifidobacterium_630	0.02	9
g_Bifidobacterium_356	0.02	7
Other *g_Bifidobacterium* (n=26)	0.12	26
g_Bacteroides_106	8.61	59
g_Bacteroides_149	4.58	22
g_Bacteroides_144	2.04	24
g_Bacteroides_159	1.94	23
g_Bacteroides_157	0.78	25
g_Bacteroides_125	0.60	26
g_Bacteroides_133	0.49	20
g_Bacteroides_156	0.33	12
g_Bacteroides_126	0.30	20
g_Bacteroides_142	0.20	6
g_Bacteroides_139	0.08	5
g_Bacteroides_88	0.08	5
g_Bacteroides_267	0.05	5
g_Bacteroides_227	0.02	5
Other *g_Bacteroides* (n=50)	1.26	27
g_Parabacteroides_181	3.29	37
g_Parabacteroides_179	2.00	18
g_Parabacteroides_249	0.53	19
g_Parabacteroides_250	0.05	5
g_Parabacteroides_196	0.02	6
Other *g_Parabacteroides* (n=25)	0.79	22
g_Lactobacillus_852	1.03	26
g_Lactobacillus_744	0.78	36
Other g_Lactobacillus (n=22)	0.48	12
f_Streptococcaceae_836_g_g	0.06	18
g_Streptococcus_685	2.46	74
g_Streptococcus_668	0.18	19
g_Streptococcus_674	0.16	36
g_Streptococcus_684	0.15	6
g_Streptococcus_667	0.13	12
Other *g_Streptococcus* (n=14)	0.67	13
Total (n=531)	100.0	100

OTU	Average RA (%)	Prevalence (%)
g_Lactococcus_698	0.35	50
g_Lactococcus_697	0.03	10
g_Lactococcus_696	0.01	6
f_Enterobacteriaceae_302_g_g	3.83	31
g_Escherichia-Shigella_328	11.70	72
g_Escherichia-Shigella_316	0.04	17
g_Staphylococcus_721	2.08	58
g_Staphylococcus_591	0.02	7
g_Veillonella_781	0.83	48
g_Veillonella_769	0.50	20
g_Veillonella_764	0.27	22
g_Veillonella_776	0.09	10
g_Haemophilus_368	0.88	49
g_Clostridium_824	0.85	7
g_Clostridium_885	0.48	9
g_Blautia_509	0.47	7
g_Blautia_471	0.04	5
g_Phascolarctobacterium_890	0.21	7
g_Phascolarctobacterium_892	0.12	5
f_Lachnospiraceae *Incertae_Sedis*_487	0.77	9
f_Lachnospiraceae *Incertae_Sedis*_496	0.05	5
f_Lachnospiraceae *Incertae_Sedis*_941	0.19	7
g_Rothia_430	0.16	22
g_Sutterella_565	0.79	21
g_Leuconostoc_962	0.06	19
g_Enterococcus_842	0.20	18
g_Halomonas_373	0.06	18
g_Haemophilus_371	0.08	17
g_Bilophila_280	0.12	13
g_Flavonifractor_737	0.33	11
g_Aeribacillus_793	0.04	10
g_Collinsella_528	0.10	10
g_Eggerthella_398	0.02	7
g_Ralstonia_553	0.02	7
g_Weissella_966	0.01	7
g_Odoribacter_241	0.10	6
g_Citrobacter_297	0.02	5
g_Negativicoccus_759	0.02	5
g_Varibaculum_438	0.02	5
f_Erysipelotrichaceae *Incertae_Sedis*_714	0.69	7
f_Ruminococcaceae_6_g_g	0.31	6
f_Coriobacteriaceae_929	0.03	8
f_Peptostreptococcaceae *Incertae_Sedis*_394	0.02	9
Remaining Other OTUs (n=312)	6.81	87

Table S4. PLS analysis results showing the association of milk and faecal HMOs with microbiota of 121 infants. p<0.05 highlighted in bold.

	MILK		FAECES	
Compound	R2	p-value	R2	p-value
2′FL	0.449014	**0.01**	0.474618	**0.008**
LNT and LNnT	0.365672	0.321	0.486463	**0.005**
LNFPIII	0.411369	0.1	0.334021	0.489
LNFPII	0.428876	0.059	0.538998	**0.001**
LNFPI	0.453009	**0.006**	0.352368	0.395
LNFPV	0.361059	0.29	0.411458	0.268
LNH	0.296281	0.711	0.419855	0.311
LNnH	0.372147	0.267	0.407214	0.372
LNDFHI	0.348777	0.378	0.425054	**0.02**
DFL	0.375545	0.193	0.484093	**0.001**
6′SL	0.308867	0.679	0.307907	0.651
3′SL	0.330597	0.538	0.335586	0.548
LSTc	0.40498	0.101	0.284052	0.797
LSTb	0.3488	0.387	0.366729	0.303
LSTa	0.37228	0.267	0.303566	0.667

Table S5. OTUs significantly different (p<0.05) in relative abundance between infants classified as high or low consumers of specific HMOs. Differences in relative abundance with FDR<0.05 are indicated in bold.

Taxonomy	**OTU**	**HMO**	**p**	**FDR**	**High consumption group**	**Low consumption group**
g_Bacteroides	106	6′SL	0.008	0.137	0.1021	0.0355
	106	LSTb	0.010	0.192	0.0745	0.0279
	106	LSTc	0.006	0.129	0.1047	0.0288
	133	6′SL	0.005	0.136	0.0079	0.0002
	133	LSTb	0.005	0.131	0.0078	0.0003
	133	LSTc	0.002	0.129	0.0087	0.0002
	142	LNH	0.034	0.276	0.0001	0.0056
	144	2′FL	0.017	0.168	0.0269	0.0271
	156	6′SL	0.021	0.263	0.0020	0.0010
	156	LNFPV	0.040	0.848	0.0005	0.0047
	227	6′SL	0.040	0.296	0.0004	0
	227	LNTandLNnT	0.022	0.176	0.0004	0
	227	LSTb	0.022	0.227	0.0006	0
	227	LSTc	0.030	0.273	0.0004	0
	267	LNTandLNnT	0.041	0.280	0.0011	0
g_Bifidobacterium	**406**	**2′FL**	**0.001**	**0.018**	**0.0015**	**0.0002**
	406	3′SL	0.033	0.410	0.0014	0.0005
	406	**DFL**	**0.000**	**0.001**	**0.0015**	**0**
	406	LNDFHI	0.002	0.059	0.0015	0.0004
	406	**LNFPII**	**0.000**	**0.000**	**0.0024**	**0**
	406	**LNFPIII**	**0.000**	**0.003**	**0.0015**	**0.0001**
	406	**LNH**	**0.000**	**0.000**	**0.0017**	**0.0001**
	406	**LNTandLNnT**	**0.000**	**0.000**	**0.0017**	**0.0001**
	406	LSTa	0.005	0.147	0.0013	0.0001

	406	LSTb	0.022	0.227	0.0012	0.0004
	406	LSTc	0.038	0.279	0.0011	0.0004
	416	2´FL	0.029	0.194	0.0004	0.0001
	416	3´SL	0.048	0.441	0.0003	0.0001
	416	DFL	0.011	0.108	0.0003	0
	416	**LNFPII**	**0.000**	**0.002**	**0.0006**	**0**
	416	**LNFPIII**	**0.003**	**0.030**	**0.0003**	**0**
	416	LNH	0.006	0.072	0.0004	0.0000
	416	LNTandLNnT	0.008	0.091	0.0004	0.0000
	416	LSTa	0.026	0.294	0.0003	0
	418	**2´FL**	**0.000**	**0.004**	**0.0668**	**0.0211**
	418	**DFL**	**0.000**	**0.000**	**0.0742**	**0.0022**
	418	**LNDFHI**	**0.000**	**0.005**	**0.0704**	**0.0265**
	418	LNFPI	0.036	0.588	0.0737	0.0160
	418	**LNFPII**	**0.000**	**0.000**	**0.0967**	**0**
	418	**LNFPIII**	**0.000**	**0.000**	**0.0757**	**0.0031**
	418	**LNH**	**0.000**	**0.000**	**0.0884**	**0.0108**
	418	**LNTandLNnT**	**0.000**	**0.000**	**0.0930**	**0.0105**
	418	LSTa	0.002	0.147	0.0654	0.0254
	418	LSTb	0.001	0.070	0.0717	0.0217
	418	LSTc	0.006	0.129	0.0699	0.0299
	423	**2´FL**	**0.001**	**0.022**	**0.0014**	**0.0001**
	423	3´SL	0.027	0.410	0.0012	0.0005
	423	**DFL**	**0.000**	**0.002**	**0.0012**	**0**
	423	LNDFHI	0.007	0.110	0.0014	0.0004
	423	**LNFPII**	**0.000**	**0.000**	**0.0022**	**0**
	423	**LNFPIII**	**0.000**	**0.005**	**0.0013**	**0.0001**
	423	**LNH**	**0.000**	**0.000**	**0.0016**	**0.0001**
	423	**LNTandLNnT**	**0.000**	**0.000**	**0.0016**	**0.0001**
	423	LSTa	0.008	0.169	0.0012	0.0001
	423	LSTb	0.030	0.254	0.0011	0.0003
	597	2´FL	0.005	0.071	0.0013	0.0001
	597	3´SL	0.016	0.410	0.0012	0.0004
	597	**DFL**	**0.000**	**0.002**	**0.0012**	**0**
	597	LNDFHI	0.019	0.255	0.0013	0.0003
	597	**LNFPII**	**0.000**	**0.000**	**0.0021**	**0**
	597	**LNFPIII**	**0.001**	**0.010**	**0.0013**	**0.0001**
	597	**LNH**	**0.000**	**0.001**	**0.0015**	**0.0001**
	597	**LNTandLNnT**	**0.000**	**0.002**	**0.0015**	**0.0001**
	597	LSTa	0.012	0.190	0.0011	0.0001
	614	2´FL	0.016	0.168	0.2053	0.1097
	614	**DFL**	**0.001**	**0.017**	**0.2532**	**0.1215**
	614	LNDFHI	0.039	0.377	0.2223	0.1482
	614	**LNFPII**	**0.000**	**0.002**	**0.2718**	**0.1180**
	614	LNFPIII	0.008	0.065	0.2290	0.1357

	614	**LNH**	**0.000**	**0.006**	**0.2418**	**0.1266**
	614	**LNTandLNnT**	**0.002**	**0.024**	**0.2272**	**0.1167**
	630	LSTa	0.017	0.231	0.0001	0.0004
	643	**2´FL**	**0.001**	**0.018**	**0.0015**	**0.0002**
	643	3´SL	0.035	0.410	0.0014	0.0005
	643	**DFL**	**0.000**	**0.001**	**0.0014**	**0**
	643	LNDFHI	0.002	0.059	0.0015	0.0004
	643	**LNFPII**	**0.000**	**0.000**	**0.0023**	**0**
	643	**LNFPIII**	**0.000**	**0.003**	**0.0015**	**0.0001**
	643	**LNH**	**0.000**	**0.000**	**0.0017**	**0.0001**
	643	**LNTandLNnT**	**0.000**	**0.000**	**0.0017**	**0.0001**
	643	LSTa	0.005	0.147	0.0013	0.0001
	643	LSTb	0.019	0.227	0.0012	0.0004
	643	LSTc	0.036	0.279	0.0012	0.0005
g_Bilophila	280	2´FL	0.047	0.270	0.0019	0.0002
	280	LNFPI	0.024	0.588	0.0017	0
	280	LNFPV	0.047	0.848	0.0005	0.0016
g_Blautia	509	LNFPI	0.024	0.588	0.0122	0
g_Citrobacter	297	LNFPIII	0.018	0.103	0.0000	0.0004
g_Clostridium	824	6´SL	0.038	0.296	0.0022	0.0193
g_Escherichia-Shigella	316	LNFPII	0.029	0.210	0.0005	0.0002
	316	LNH	0.044	0.327	0.0004	0.0002
g_Halomonas	373	3´SL	0.017	0.410	0.0003	0.0012
	373	LNFPIII	0.008	0.065	0.0009	0.0000
g_Lactobacillus	**744**	**2´FL**	**0.000**	**0.004**	**0.0080**	**0.0013**
	744	**DFL**	**0.000**	**0.001**	**0.0086**	**0.0002**
	744	LNDFHI	0.006	0.110	0.0057	0.0016
	744	LNTandLNnT	0.035	0.255	0.0064	0.0048
	744	LSTb	0.031	0.254	0.0036	0.0049
	852	LNFPII	0.013	0.102	0.0207	0.0019
g_Lactococcus	697	LNFPIII	0.015	0.095	0.0004	0
	698	LNFPI	0.007	0.588	0.0045	0.0010
	698	**LNFPIII**	**0.001**	**0.019**	**0.0044**	**0.0008**
	698	LSTb	0.047	0.273	0.0045	0.0020
g_Leuconostoc	962	LNDFHI	0.042	0.377	0.0005	0.0003
	962	LNFPIII	0.011	0.074	0.0007	0.0001
	962	LSTb	0.040	0.269	0.0008	0.0003
g_Odoribacter	241	LNTandLNnT	0.022	0.176	0.0018	0
g_Parabacteroides	181	LNH	0.012	0.125	0.0130	0.0437
	196	2´FL	0.040	0.251	0.0005	0
	196	6´SL	0.026	0.263	0.0004	0
	196	LSTb	0.042	0.269	0.0003	0
	196	LSTc	0.019	0.244	0.0004	0
	249	2´FL	0.029	0.194	0.0111	0.0002
g_Phascolarctobacterium	890	6´SL	0.026	0.263	0.0036	0

	890	LSTb	0.043	0.269	0.0052	0.0001
g_Ralstonia	553	LNFPI	0.041	0.588	0.0004	0
	553	LNFPIII	0.034	0.162	0.0003	0
g_Rothia	430	LNFPII	0.036	0.226	0.0025	0.0003
g_Staphylococcus	591	LSTb	0.022	0.227	0	0.0003
	591	LSTc	0.018	0.244	0	0.0002
	721	**6´SL**	**0.001**	**0.048**	**0.0087**	**0.0334**
	721	LNTandLNnT	0.047	0.295	0.0099	0.0181
	721	LSTa	0.030	0.294	0.0155	0.0297
	721	LSTb	0.002	0.080	0.0092	0.0239
	721	LSTc	0.005	0.129	0.0078	0.0231
g_Streptococcus	667	2´FL	0.021	0.168	0	0.0005
	668	DFL	0.026	0.194	0.0034	0.0001
	668	**LNFPII**	**0.002**	**0.024**	**0.0036**	**0.0001**
	674	DFL	0.016	0.144	0.0019	0.0003
	674	LNFPIII	0.042	0.181	0.0018	0.0006
	685	**6´SL**	**0.001**	**0.048**	**0.0138**	**0.0289**
	685	LNFPII	0.031	0.210	0.0194	0.0168
	685	LSTa	0.033	0.294	0.0165	0.0360
	685	LSTc	0.021	0.244	0.0150	0.0253
g_Veillonella	764	3´SL	0.035	0.410	0.0018	0.0030
	764	LNFPII	0.043	0.246	0.0038	0.0011
	769	3´SL	0.015	0.410	0.0011	0.0117
	781	2´FL	0.019	0.168	0.0018	0.0097
	781	6´SL	0.008	0.137	0.0022	0.0121
	781	LNTandLNnT	0.010	0.100	0.0023	0.0118
g_Weissella	966	LNFPIII	0.034	0.162	0.0002	0
g_Aeribacillus	793	LNFPIII	0.010	0.074	0.0005	0
f_Enterobacteriaceae_g_g	302	DFL	0.024	0.192	0.0106	0.0782
f_Enterobacteriaceae_g_g	302	LNFPII	0.007	0.067	0.0014	0.0717
f_Enterobacteriaceae_g_g	302	LNFPIII	0.042	0.181	0.0182	0.0565
f_Erysipelotrichaceae g_Incertae_Sedis	714	LNDFHI	0.041	0.377	0.0001	0.0227
f_Erysipelotrichaceae g_Incertae_Sedis	714	LNFPIII	0.023	0.123	0.0121	0
f_Erysipelotrichaceae *Incertae_Sedis*	714	LNH	0.034	0.276	0.0001	0.0200
f_Lachnospiraceae *Incertae_Sedis*	496	6´SL	0.040	0.296	0.0009	0
f_Lachnospiraceae *Incertae_Sedis*	496	LSTc	0.030	0.273	0.0010	0
f_Streptococcaceae_g_g	**836**	**LNFPIII**	**0.002**	**0.019**	**0.0008**	**0.0000**
f_Streptococcaceae_g_g	836	LNFPII	0.046	0.249	0.0009	0.0002

Table S6. DNA sequences of OTUs that differ in relative abundance between high and low consumer groups.

Taxonomy	OTU	Sequences (5'-3')
g_Aeribacillus	793	TACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTT CCTTAAGTCTGATGCCTGTTCGCTCCCCACGCTTTCGCGCCTCAGCGTCAGTTACAGG CCAGAGAGCCGCCTTCGCCACTGGTG
g_Bacteroides	106	TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGATGGAT GTTTAAGTCAGTTGCCTGTTTGATACCCACACTTTCGAGCCTCAATGTCAGTTGCAGC TTAGCAGGCTGCCTTCGCAATCGGAG
	133	TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGCGGAT TGTTAAGTCAGTTGCCTGTTTGATACCCACACTTTCGAGCATCAGCGTCAGTTACACT CCAGTGAGCTGCCTTCGCAATCGGAG
	142	TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGCGGGT TGTTAAGTCAGTTGCCTGTTTGATACCCACACTTTCGAGCATCAGCGTCAGTTACAAT CCAGTAAGCTGCCTTCGCAATCGGAG
	144	TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGTGGAC AGTTAAGTCAGTTGCCTGTTTGATACCCACACTTTCGAGCATCAGTGTCAGTTGCAGT CCAGTGAGCTGCCTTCGCAATCGGAG
	156	TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGTGGAT TGTTAAGTCAGTTGCCTGTTTGATACCCACACTTTCGAGCATCAGTGTCAGTAACAGT CTAGTGAGCTGCCTTCGCAATCGGAG
	227	TACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGATGGAT GTTTAAGTCAGTTGCCTGTTTGATACCCACACTTTCGAGCCTCAATGTCAGTTGCAGC TTAGCAGGCTGCCTTCGCAATCGGAG
	267	TACGGAGGATTCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGATGGAT GTTTAAGTCAGTTGCCTGTTTGATACCCACACTTTCGAGCCTCAATGTCAGTTGCAGC TTAGCAGGCTGCCTTCGCAATCGGAG
g_Bifidobacterium	406	TACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAAAGGGCTCGTAGGCGGTT CGTCGCGTCCGGTGCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTAACGGC CCAGAGACCTGCCTTCGCCATTGGTG
	416	TACGTAGGGCGCAAGCGTTATCCGGATTTATTGGGCGTAAAGGGCTCGTAGGCGGCT CGTCGCGTCCGGTGCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTAACGGC CCAGAGACCTGCCTTCGCCATCGGTG
	418	TACGTAGGGCGCAAGCGTTATCCGGATTTATTGGGCGTAAAGGGCTCGTAGGCGGCT CGTCGCGTCCGGTGCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTGACGGC CCAGAGACCTGCCTTCGCCATCGGTG
	423	TACGTAGGGCGCAAGCGTTATCCGGATTTATTGGGCGTAAAGGGCTCGTAGGCGGTT CGTCGCGTCCGGTGCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTAACGGC CCAGAGACCTGCCTTCGCCATTGGTG
	597	TACGTAGGGTGCAAGCGTTATCCGGAATTATTGGGCGTAAAGGGCTCGTAGGCGGCT CGTCGCGTCCGGTGCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTGACGGC CCAGAGACCTGCCTTCGCCATCGGTG
	614	TACGTAGGGTGCAAGCGTTATCCGGAATTATTGGGCGTAAAGGGCTCGTAGGCGGTT CGTCGCGTCCGGTGCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTAACGGC CCAGAGACCTGCCTTCGCCATTGGTG
	630	TACGTAGGGTGCAAGCGTTATCCGGAATTATTGGGCGTAAAGGGCTCGTAGGCGGTT CGTCGCGTCCGGTGCCTGTTTGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTAACGGC CCAGAGACCTGCCTTCGCCATTGGTG
	643	TACGTAGGGTGCAAGCGTTATCCGGATTTATTGGGCGTAAAGGGCTCGTAGGCGGCT CGTCGCGTCCGGTGCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTGACGGC CCAGAGACCTGCCTTCGCCATCGGTG
g_Bilophila	280	TACGGAGGGTGCAAGCGTTAATCGGAATCACTGGGCGTAAAGCGCACGTAGGCGGC TTGGTAAGTCAGGGGCCTGTTTGCTACCCACGCTTTCGCACCTCAGCGTCAGTTACCG TCCAGGTGGCCGCCTTCGCCACCGGTG
g_Blautia	509	TACGTAGGGGGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGAGCGTAGACGGT GTGGCAAGTCTGATGCCTGTTTGCTCCCCACGCTTTCGAGCCTCAACGTCAGTTACCG TCCAGTAAGCCGCCTTCGCCACTGGTG
g_Citrobacter	297	TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGT CTGTCAAGTCGGATGCCTGTTTGCTCCCCACGCTTTCGCACCTGAGCGTCAGTCTTCG TCCAGGGGGCCGCCTTCGCCACCGGTA
g_Clostridium	824	TACGTAGGTGGCAAGCGTTGTCCGGATTTACTGGGCGTAAAGGGAGCGTAGGCGGA TTTTTAAGTGGGATGCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAG TCCAGAAAGTCGCCTTCGCCACTGGTG
g_Escherichia-Shigella	316	TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGT TTGTTAAGTCAGATGCCTGTTCGCTCCCCACGCTTTCGCACCTGAGCGTCAGTCTTCG TCCAGGGGGCCGCCTTCGCCACCGGTA
g_Halomonas	373	TACGGAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGCGGT CTGATAAGCCGGTTGCCTGTTTGCTACCCACGCTTTCGCACCTCAGCGTCAGTGTCAG TCCAGAAGGCCGCCTTCGCCACTGGTA
g_Lactobacillus	744	TACGTAGGTGGCAAGCGTTATCCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTT TTTTAAGTCTGATGCCTGTTCGCTACCCATGCTTTCGAGCCTCAGCGTCAGTTACAGA CCAGACAGCCGCCTTCGCCACTGGTG
	852	TACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGTGCAGGCGGTT CAATAAGTCTGATGCCTGTTCGCTACCCATGCTTTCGAGCCTCAGCGTCAGTTGCAGA CCAGAGAGCCGCCTTCGCCACTGGTG

g_Lactococcus	697	TACGTAGGTCCCGAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGCGCAGGTGGTT TAATAAGTCTGATGCCTGTTTGCTACCCACGCTTTCGAGCCTCAGTGTCAGTTACAGT CCAGAGAGCCGCTTTCGCCACCGGTG
	698	TACGTAGGTCCCGAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGCGCAGGTGGTT TATTAAGTCTGGTGCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGTGTCAGTTACAGG CCAGAGAGCCGCTTTCGCCACCGGTG
g_Leuconostoc	962	TACGTATGTCCCGAGCGTTATCCGGATTTATTGGGCGTAAAGCGAGCGCAGACGGTT GATTAAGTCTGATGCCTGTTTGCTACCCACACTTTCGAGCCTCAACGTCAGTTGTTGT CCAGTAAGCCGCCTTCGCCACTGGTG
g_Odoribacter	241	TACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGTT TATTAAGTTAGTGGCCTGTTCGCTACCCACGCTCTCGTGCATCAGCGTCAGTTACAGT CTGGTAAGCTGCCTTCGCTATCGGAG
g_Parabacteroides	181	TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGCC TTTTAAGTCAGCGGCCTGTTTGATCCCCACGCTTTCGTGCATCAGCGTCAGTCATGGC TTGGCAGGCTGCCTTCGCAATCGGGG
	196	TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGTGGTG ATTTAAGTCAGCGGCCTGTTTGATCCCCACGCTTTCGTGCTTCAGTGTCAGTTATGGT TTAGTAAGCTGCCTTCGCAATCGGAG
	249	TACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGTGGTG ATTTAAGTCAGCGGCCTGTTTGATCCCCACGCTTTCGTGCTTCAGTGTCAGTTATGGT TTAGTAAGCTGCCTTCGCAATCGGAG
g_Phascolarctobacterium	890	TACGTAGGTGGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCATGTAGGCGGCT TAATAAGTCGAGCGCCCGTTCGCTACCCTGGCTTTCGCATCTCAGCGTCAGACACAG TCCAGAAAGGCGCCTTCGCCACTGGTG
g_Ralstonia	553	TACGTAGGGTCCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCGGTT GTGCAAGACCGATGCCTGTTTGCTCCCCACGCTTTCGTGCATGAGCGTCAGTGTTATC CCAGGGGGCTGCCTTCGCCATCGGTA
g_Rothia	430	TACGTAGGGCGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTTGTAGGCGGTT TGTCGCGTCTGCTGCCTGTTCGCTCCCCATGCTTTCGCTTCTCAGCGTCAGTTACAGC CCAGAGACCTGCCTTCGCCATCGGTG
g_Staphylococcus	591	TACGTAGGGTGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGTAGGCGGTT TTTTAAGTCTGATGCCTGTTTGATCCCCACGCTTTCGCACATCAGCGTCAGTTACAGA CCAGAAAGTCGCCTTCGCCACTGGTG
	721	TACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGTAGGCGGTT TTTTAAGTCTGATGCCTGTTTGATCCCCACGCTTTCGCACATCAGCGTCAGTTACAGA CCAGAAAGTCGCCTTCGCCACTGGTG
g_Streptococcus	667	TACGTAGGTCCCGAGCGTTATCCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTT AGATAAGTCTGAAGCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAA GCCAGAGAGCCGCTTTCGCCACCGGTG
	668	TACGTAGGTCCCGAGCGTTATCCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTT AGATAAGTCTGAAGCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAG ACCAGAGAGCCGCTTTCGCCACCGGTG
	674	TACGTAGGTCCCGAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTT AGATAAGTCTGAAGCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAA GCCAGAGAGCCGCTTTCGCCACCGGTG
	685	TACGTAGGTCCCGAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTT TGATAAGTCTGAAGCCTGTTCGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAG ACCAGAGAGCCGCTTTCGCCACCGGTG
g_Veillonella	764	TACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGA TAGGTCAGTCTGTCTCCCGTTCGCTCCCCTGGCTTTCGCGCCTCAGCGTCAGTTTTCG TCCAGAAAGTCGCCTTCGCCACTGGTG
	769	TACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGA TCAGTCAGTCTGTCTCCCGTTCGCTCCCCTGGCTTTCGCGCCTCAGCGTCAGTTTTCGT CCAGAAAGTCGCCTTCGCCACTGGTG
	781	TACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGA TTGGTCAGTCTGTCTCCCGTTCGCTCCCCTGGCTTTCGCGCCTCAGCGTCAGTTTTCGT CCAGAAAGTCGCCTTCGCCACTGGTG
g_Weissella	966	TACGTATGTTCCAAGCGTTATCCGGATTTATTGGGCGTAAAGCGAGCGCAGACGGTT ATTTAAGTCTGAAGCCTGTTTGCTACCCACACTTTCGAGCCTCAACGTCAGTTACAGT CCAGAAAGCCGCCTTCGCCACTGGTG
f_Enterobacteriaceae	302	TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGT CTGTCAAGTCGGATGCCTGTTTGCTCCCCACGCTTTCGCACCTGAGCGTCAGTCTTTG TCCAGGGGGCCGCCTTCGCCACCGGTA
f_Erysipelotrichaceae *Incertae_Sedis*	714	TACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGAGGGAGCAGGCGGC AGCAAGGGTCTGTGGCCTATTTGCTCCCCACGCTTTCGGGACTGAGCGTCAGTTGCA GGCCAGATCGTCGCCTTCGCCACTGGTG
f_Lachnospiraceae *Incertae_Sedis*	496	TACGTAGGGGGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGAGCGTAGACGGC GAAGCAAGTCTGAAGCCTGTTTGCTCCCCACGCTTTCGAGCCTCAACGTCAGTTATC GTCCAGTAAGCCGCCTTCGCCACTGGTG
f_Streptococcaceae	836	TACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTT CATTAAGTCTGATGCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAGT CCAGAGAGCCGCTTTCGCCTCCGGTG

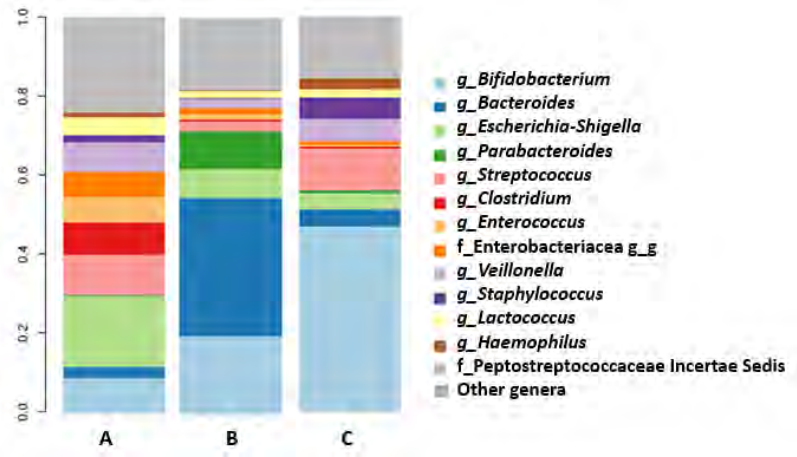
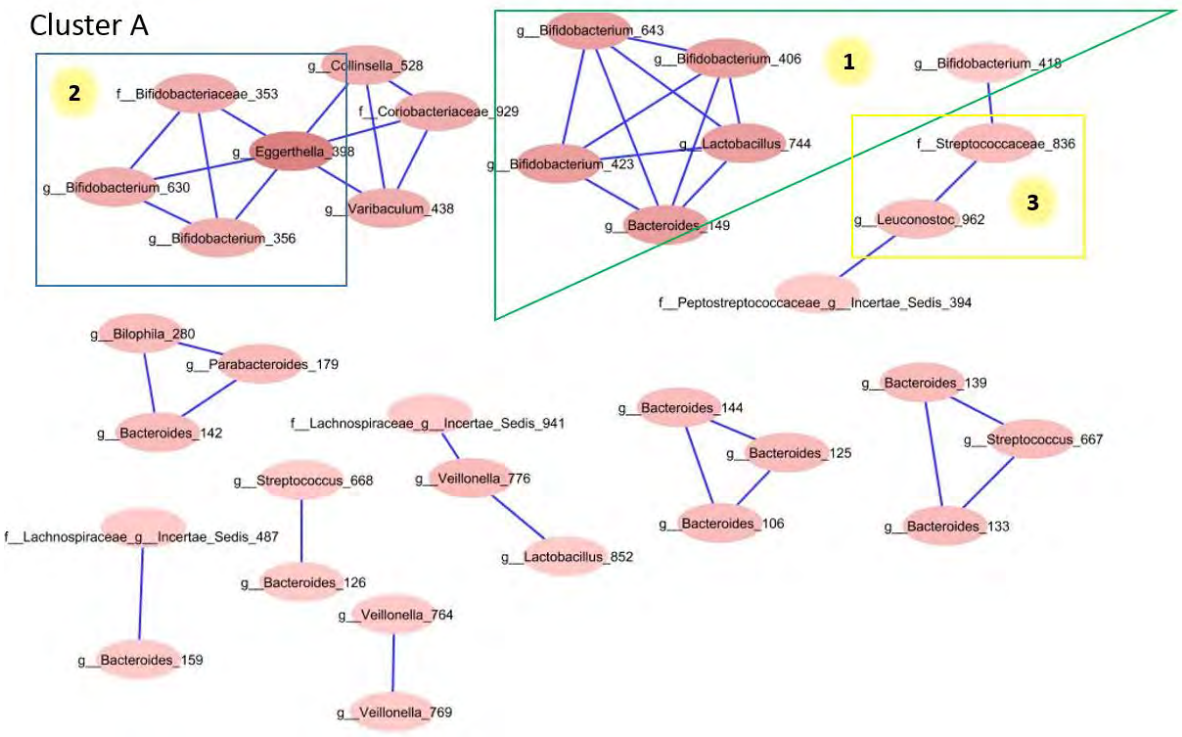
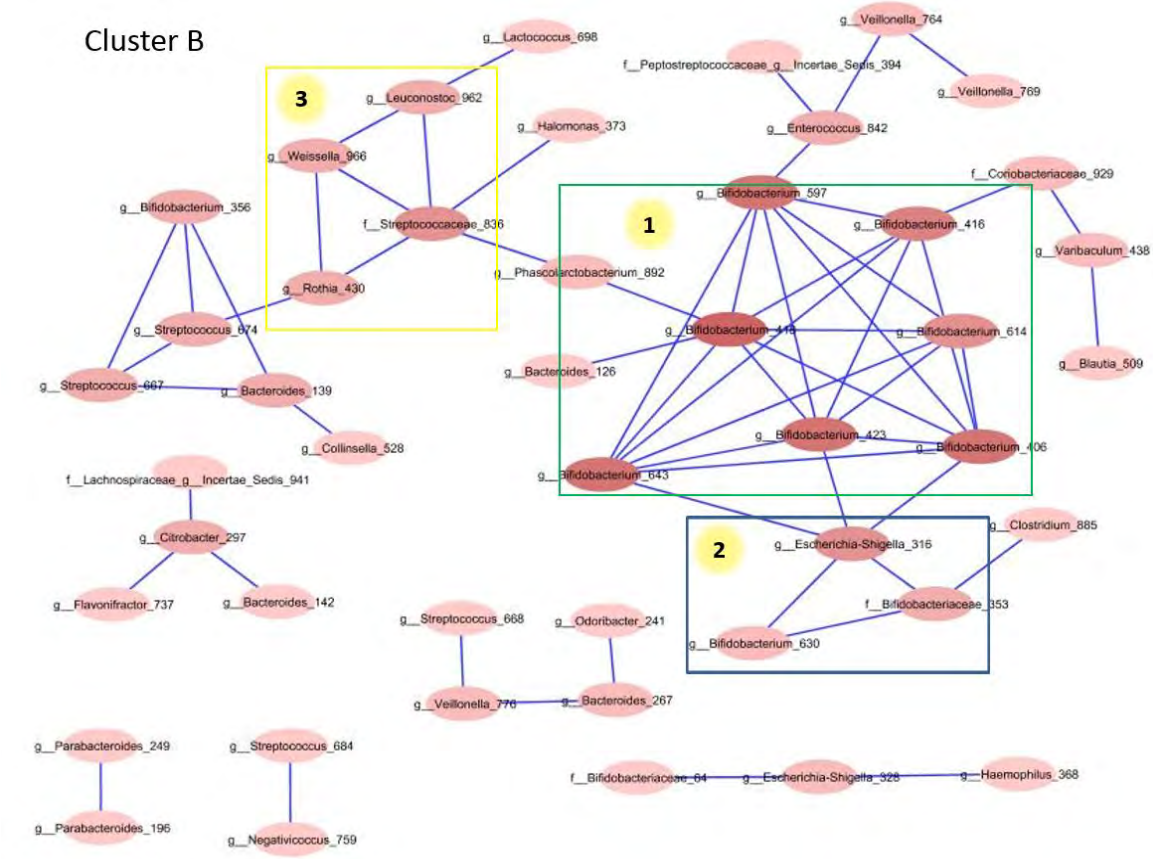


Figure S1. Microbial cluster composition based on DMM modelling of KOALA samples

a.



b.


Cluster B
g__Lactococcus_698
g__Veillonella_764
f__Peptostreptococcaceae_g__Incertae_Sedis_394
g__Veillonella_769
3
g__Leuconostoc_962
g__Halomonas_373
g__Weissella_966
g__Enterococcus_842
f__Coriobacteriaceae_929
g__Bifidobacterium_597
g__Bifidobacterium_356
1
g__Bifidobacterium_416
f__Streptococcaceae_836
g__Varibaculum_438
g__Phascolarctobacterium_892
g__Rothia_430
g__Bifidobacterium_418
g__Bifidobacterium_614
g__Streptococcus_674
g__Blautia_509
g__Bacteroides_126
g__Streptococcus_667
g__Bacteroides_139
g__Bifidobacterium_423
g__Collinsella_528
g__Bifidobacterium_406
g__Bifidobacterium_643
f__Lachnospiraceae_g__Incertae_Sedis_941
g__Clostridium_885
2
g__Escherichia-Shigella_316
g__Citrobacter_297
f__Bifidobacteriaceae_353
g__Flavonifractor_737
g__Bacteroides_142
g__Bifidobacterium_630
g__Streptococcus_668
g__Odoribacter_241
g__Veillonella_776
g__Bacteroides_267
g__Parabacteroides_249
g__Streptococcus_684
f__Bifidobacteriaceae_64
g__Escherichia-Shigella_328
g__Haemophilus_368
g__Parabacteroides_196
g__Negativicoccus_759

c.

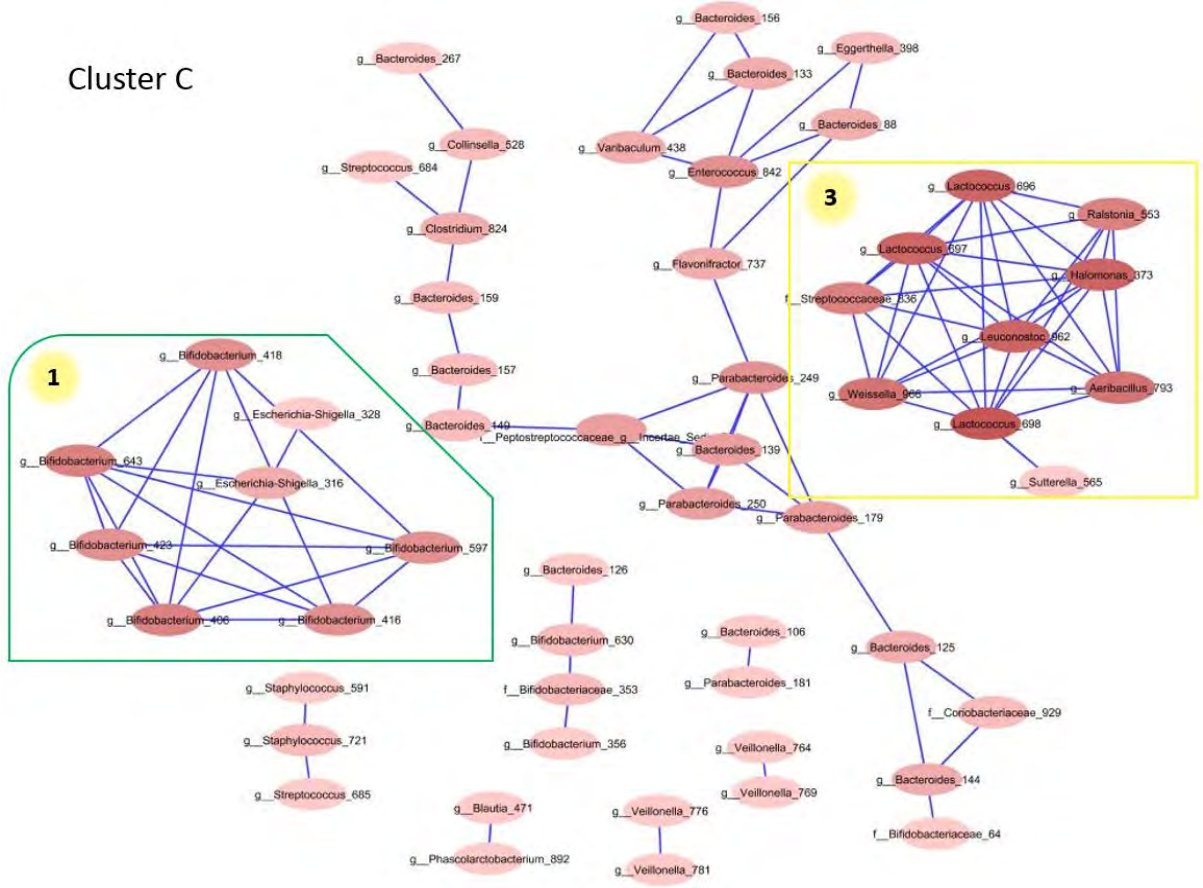

Cluster C
g__Bacteroides_156
g__Eggerthella_398
g__Bacteroides_267
g__Bacteroides_133
g__Bacteroides_88
g__Collinsella_528
g__Varibaculum_438
g__Streptococcus_684
g__Enterococcus_842
3
g__Lactococcus_696
g__Clostridium_824
g__Ralstonia_553
g__Lactococcus_697
g__Flavonifractor_737
g__Halomonas_373
g__Bacteroides_159
f__Streptococcaceae_836
g__Leuconostoc_962
1
g__Bacteroides_157
g__Parabacteroides_249
g__Bifidobacterium_418
g__Weissella_966
g__Aeribacillus_793
g__Escherichia-Shigella_328
g__Bacteroides_149
g__Lactococcus_698
g__Bacteroides_139
g__Bifidobacterium_643
g__Escherichia-Shigella_316
g__Sutterella_565
g__Parabacteroides_250
g__Parabacteroides_179
g__Bifidobacterium_423
g__Bifidobacterium_597
g__Bacteroides_126
g__Bacteroides_106
g__Bacteroides_125
g__Bifidobacterium_406
g__Bifidobacterium_416
g__Bifidobacterium_630
g__Parabacteroides_181
f__Coriobacteriaceae_929
g__Staphylococcus_591
f__Bifidobacteriaceae_353
g__Bacteroides_144
g__Staphylococcus_721
g__Bifidobacterium_356
g__Veillonella_764
f__Bifidobacteriaceae_64
g__Streptococcus_685
g__Veillonella_769
g__Blautia_471
g__Veillonella_776
g__Phascolarctobacterium_892
g__Veillonella_781

Figure S2. OTU network ($p<0.05$, correlation threshold of ±0.03) identified in the faecal microbial communities of infants within each microbial DMM cluster type A (a), B (b) and C (c). Node colour intensity is proportional to the number of connections with other nodes. Blue connecting lines show positive associations, black connecting lines show negative associations. Darker shade and thickness of connecting lines indicates higher Spearman correlation score. The three networks which were identified as significant ($p=0.05$) by the model are indicated with yellow, green and blue boxes and numbered for convenience.

Chapter 5

The association between infant faecal microbiota composition and the degradation of human milk oligosaccharides in healthy breastfed infants at two, six and twelve weeks of age

Klaudyna Borewicz*, Fangjie Gu*, Edoardo Saccenti, Christine Hechler, Roseriet Beijers, Carolina de Weerth, Sander S. van Leeuwen, Henk A. Schols, Hauke Smidt

* These authors contributed equally to this work

Manuscript in preparation

Abstract

In this study we followed 24 mother-infant pairs to investigate the association between selected human milk oligosaccharides (HMOs) in breastmilk and the faecal microbiota composition in healthy, breastfed infants during their first 12 weeks of life. The main objectives were to identify factors that influence faecal microbiota development through the first three months of life, to assess potential links between selected breastmilk HMOs and infant faecal microbiota composition, and to examine associations between microbiota composition and the degradation of milk HMOs in the infant gastrointestinal GI tract. Microbial community composition was analysed using Illumina HiSeq sequencing of PCR-amplified V4 region of the 16S ribosomal RNA gene, and the HMO concentrations in breastmilk and infant faeces were measured using porous graphitized carbon-ultra high performance liquid chromatograph - mass spectrometry (HPLC-MS) and high performance anion exchange chromatography (HPAEC). We confirmed previous findings showing that microbiota composition was influenced by infant age. Furthermore, mode of delivery and Lacto-*N*-fucopentaose III (LNFPIII) concentration were significantly associated with infant faecal microbiota at two weeks of age. At six and 12 weeks of age, significant associations were observed for gender, mode of delivery and 3'Sialyllactose (3´SL), and gender and Lacto-*N*-hexaose (LNH), respectively. Lactation duration had a significant effect on breastmilk HMO content and the HMO content decreased with time, except for 3′-Fucosyllactose (3FL) and LNFPIII. Correlations between breastmilk HMOs and faecal microbiota OTUs were weak and highly variable across different infant ages. We did not detect significant correlations between the most predominant *Bifidobacterium* OTU 1263 and any of the measured HMOs. However, we showed that infant's ability to degrade HMOs increased with infant age and was strongly correlated with an increase in the relative abundance of several OTUs, mainly within the genera *Bifidobacterium, Parabacteroides*, *Escherichia-Shigella, Bacteroides, Actinomyces, Veillonella,* Lachnospiraceae *Incertae Sedis,* and Erysipelotrichaceae *Incertae Sedis*. Members of these taxa might play important roles in the intestinal microbial communities of infants as they were also identified as key groups in the analysis of microbial co-occurrence networks.

Introduction

Microbial colonisation of the infant gastrointestinal (GI) tract begins before or at birth, and in healthy, breastfed infants it progresses towards a microbial community that is dominated by bifidobacteria and metabolically adapted to thrive on human milk [1, 2]. Many host specific and environmental factors have been identified to play a role in the development of human GI tract microbiota [3]. Understanding the impact of these factors and their associated health outcomes has been a growing area of research during recent years [4, 5].

The high concentrations of human milk oligosaccharides (HMOs) in breastmilk are believed to be a main driving force in shaping the bifidobacteria dominated GI ecosystem in breastfed infants, however, only few *in vivo* studies up to date were able to demonstrate this [2, 6]. The composition of breastmilk, including the types and the concentration of different HMOs, varies between mothers, and also across lactation stages [7-9]. Until now, there have been no reports on longitudinal studies investigating the establishment of infant GI microbiota in relation to changes in breastmilk HMO composition. Furthermore, little is known about how the GI microbial community development affects an infant's ability to digest different HMOs, and how this ability changes during early infancy.

In order to fill this knowledge gap, we followed a cohort of 24 healthy breastfed infants and analysed breastmilk and infant faecal samples collected at two, six and 12 weeks post-delivery. The following research questions were addressed: (i) To what extent does the dynamic composition and concentrations of breastmilk HMOs influence faecal microbial composition through the first three months of life in addition to other factors? and (ii) how does microbiota composition affect the degradation of HMOs in the infant GI tract?

Methods

Sample collection

Only infants who were healthy, born at term and did not receive oral antibiotic treatment during the study period were included. The BINGO (Dutch acronym for Biological Influences on Baby's Health and Development) study is an ongoing longitudinal cohort study investigating prenatal predictors of infant health and development. This study was approved by the ethical committee of the Faculty of Social Sciences of the Radboud University [ECSW2014-1003-189]. The study design and infant recruitment criteria can be found at http://www.bingo-onderzoek.nl/deelname/. Both, the infant faecal samples and breastmilk samples were collected within a 48 hour period, by the mothers at home, at two, six and 12 weeks post-partum. Breastmilk samples (approximately 20 ml) were collected into clean, sterile collection cups. Mothers were asked to wash their hands, breasts, and nipples, and collected the first breastmilk in the morning by hand expression, before feeding the infant. Stool samples were collected from infant's diaper using sterile stool collection vials (80×16.5mm; cat#:80.623.022, Sarstedt; Nümbrecht, Germany) with a spoon attached to the lid. Mothers were asked to save all faecal sample up to one-third of the vial. Milk and faecal samples were stored by the participants in

their home freezers until collected by the experimenter within a week after the last collection time point. Subsequently all milk and faecal samples were stored at -80 °C until further processing and analysis. Samples were analysed for breastmilk HMOs, corresponding faecal HMOs, and microbiota composition.

HMO analysis in breastmilk and faeces

Eighteen different HMO types were analysed in milk and infant faeces, including 13 neutral HMOs (2′FL, 3FL, DFL, LNDFHI, LNDFHII, LNFPI, LNFPII, LNFPIII, LNFPV, LNH, LNnH, pLNH, LNTandLNnT) and five acidic HMOs (3´SL, 6´SL, LSTa, LSTb, LSTc) (Table S1). The HMOs were extracted, purified, and quantified by using porous graphitized carbon-ultra high performance liquid chromatography - mass spectrometry (HPLC-MS) [10-12]. 3FL was eluted in a separate fraction and quantified by high performance anion exchange chromatography (HPAEC) [13]. The purification and chromatographic conditions were optimized as described by Gu and co-workers (manuscript in preparation).

DNA extraction, amplification of 16S rRNA genes and sequencing

Total bacterial DNA was extracted using the Maxwell® 16 Total RNA system (Promega; Wisconsin, United States) with Stool Transport and Recovery Buffer (STAR; Roche Diagnostics Corporation, Indianapolis, IN). Briefly, 0.1-0.15g of faecal sample was homogenized in a Precellys 24 homogenizer (Precellyis 24, Bertin Technologies, France) with 0.25 g of sterilized 0.1 mm zirconia beads and three glass beads (2.5 mm) in 350 µL STAR buffer for 3x1 min at 5.5 ms, with 10 s cooling at room temperature between rounds of bead beating. Samples were incubated with shaking at 100 rpm for 15 min at 95 °C and pelleted by 5 min centrifugation at 4 °C and 14000 × g. Supernatant was removed, and the pellets were processed again using 200 µL of fresh STAR buffer. Supernatant was removed, pooled and 250 µL was used for purification with the Maxwell® 16 Tissue LEV Total RNA Purification Kit (AS1220) customized for DNA extraction in combination with the STAR buffer following manufacturer's instructions. DNA was eluted with 50 µL of DNAse and RNAse free water (Qiagen, Hilden, Germany). DNA concentrations were measured spectrophotometrically with a NanoDrop ND-1000 (NanoDrop® Technologies, Wilmington, DE, USA) and adjusted to 20 ng/µL with DNAse and RNAse free water. The V4 region of 16S ribosomal RNA (rRNA) genes was amplified in duplicate PCR reactions, each in a total volume of 50 µL and containing 20 ng of template DNA. Each sample was amplified with 200 nM of uniquely barcoded primers 515F-n (5'-GTGCCAGCMGCCGCGGTAA-) and 806R-n (5'-RGGATTAGATACCC), 10 µl of 5x HF buffer (Finnzymes, Vantaa, Finland), 200 µM dNTP Mix (Roche Diagnostics GmbH, Mannheim, Germany), 1 U Phusion® Hot Start II High Fidelity DNA Polymerase (Finnzymes) and 36.5 µL of DNAse and RNAse free water [14]. The amplification program included a 30 s initial denaturation step at 98 °C, followed by 25 cycles of denaturation at 98 °C for 10 s, annealing at 56 °C for 10 s and elongation at 72 °C for 10 s, and a final extension at 72 °C for 7 min. PCR product presence and size (~290 bp) was confirmed with gel electrophoresis using the Lonza FlashGel® System (Lonza, Cologne, Germany). Seventy unique barcode tags were used in each library [14]. PCR products were purified with HighPrep® PCR kit (MagBio Genomics, Alphen aan den Rijn, Netherlands), and DNA concentrations were measured with

Qubit® dsDNA BR Assay Kit (Life Technologies, Leusden, Netherlands). Hundred nanograms of each barcoded sample was added to an amplicon pool and was then concentrated with HighPrep® PCR kit to 20 µL volume. Concentrations of pools were measured with Qubit® dsDNA BR Assay Kit and adjusted to 100 ng/µL final concentration. The libraries were sent for adapter ligation and Illumina HiSeq2000 sequencing at GATC-Biotech, Konstanz, Germany.

Data analysis

The 16S rRNA sequencing data analysis was carried out using the NG-Tax analysis pipeline [14]. In brief, libraries were filtered to contain only read pairs with perfectly matching barcodes that were subsequently used to separate reads by sample. Operational taxonomic units (OTUs) were assigned using an open reference approach and the SILVA_111_SSU 16S rRNA gene reference database (https://www.arb-silva.de/) [15]. Microbial composition data was expressed as a relative abundance of each OTU obtained with NG-Tax.

Statistical analyses

Measurements were obtained for a total of 127 samples from breastfed (BF) infants, including 24 infants for whom the complete set of data (breastmilk and faecal sample at all time points) was available at two, six and 12 weeks (total 72 samples), and 55 samples from infants for whom data could only be collected at one or two time points due to insufficient sample quantity, or because infants received formula during the time of sample collection. In order to reduce the possible bias due to individual variation when different infants were included in the analyses of different time points, only samples from infants with the complete set of measurements were used in the analyses (n=24).

Milk and faecal HMOs concentrations were measured in µg per mL of milk or µg per gram of faeces. Readout values were normalised for each time point separately around mean using the Probabilistic Quotient Normalization (PQN) method in R (version 3.3.2). Since no data on daily milk intake was collected, we used the average daily volume of ingested breastmilk as indicated in the literature to estimate the amount of each HMO consumed by an infant during a 24 h period [16]. The average estimated amounts of milk consumed were 480 g at week two, 580 g at week six and 630 g at week 12.

Microbial composition data was expressed as relative abundance (RA) of each OTU obtained in the NG-Tax pipeline. OTUs which had a prevalence of less than 5% across all samples were removed and their values were summarized as “Other OTUs”. Alpha diversity indices (Shannon, Chao1, and PD Whole Tree) were calculated using QIIME [17]. Spearman correlations were calculated using R to evaluate associations between OTU members of the faecal microbial community, and between the faecal OTUs and milk or faecal HMO concentrations. Network visualization was done using Cytoscape [18]. Unconstrained (PCA) and constrained (RDA) multivariate analyses were carried out in Canoco5 with the significance assessed using a permutation test [19]. The explanatory variables used in the RDA multivariate analysis included infant age at the time of collection, estimated amounts of the 18 milk HMO ingested during a 24 h period (mg/24 h) or faecal HMOs (µg/g of faeces), gender, place and

mode of delivery, and if an infant was sick at the time of sample collection, as recalled by the mother. Degradation of each breastmilk HMO was calculated as a ratio of HMO concentration in infants' faeces and the concentration of the same HMO measured in mothers' milk. If a given HMO was not detected in milk, the consumption score was not included in the analysis, and if the concentration in faeces exceeded the amount detected in milk, the infant was assigned to the "low" category for that HMO. Resulting values (ratios) were then used to calculate tertiles for each HMO type and to assign infants to either a "low", "medium", or a "high" consumption category for each HMO. The association between faecal microbiota composition and the assignment of each infant to a "low", "medium", or "high" consumption category for each HMO were investigated with RDA analysis in Canoco5, with significance assessed using a permutation test [19]. Bacterial OTUs that differed significantly between infants who were assigned as high and low consumers for each individual HMO were identified with Kruskal-Wallis analysis using QIIME [17].

Nucleotide sequences

BINGO data sets cannot be made publicly available due to the data being part of an ongoing longitudinal study. Parts of the data are available to the research community for scientific collaborations upon request to Prof. dr. C. de Weerth at: Radboud University, Department of Developmental Psychology, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands, e-mail: c.deweerth@psych.ru.nl.

Results

The total number of sequencing reads obtained for the 127 faecal samples of BF infants was 17,899,918 (range: 3,383-721,990 per sample, *M*= 140,944, *SD* =127,715, *SE*= 11,333). The total number of OTUs was 617, and 86 OTUs could be found in more than 5% of all samples (i.e. at least seven samples). For the 24 infants for whom the complete set of data was available at two, six and 12 weeks of age the total number of sequencing reads was 8,550,719 (range: 3,383-421,482 per sample, *M* = 118,760, *SD* = 86,261, *SE* = 10,166). In this subset of 72 samples, 411 OTUs were identified of which 83 OTUs were found in more than 5%, or at least 4 samples (Table S2).

The development of infant faecal microbiota

Alpha diversity indices were calculated for the 24 infants at each time point, and showed no statistically significant differences in richness and diversity at different time points when tested with Kruskal-Wallis and Wilcoxon tests (Figure S1). Unconstrained analysis (PCA) showed that composition of faecal microbiota of individual infants was diverse, but despite the high inter-individual variation, observed changes in microbial profiles were directional and progressed with time towards bifidobacteria dominated communities (Figure 1).

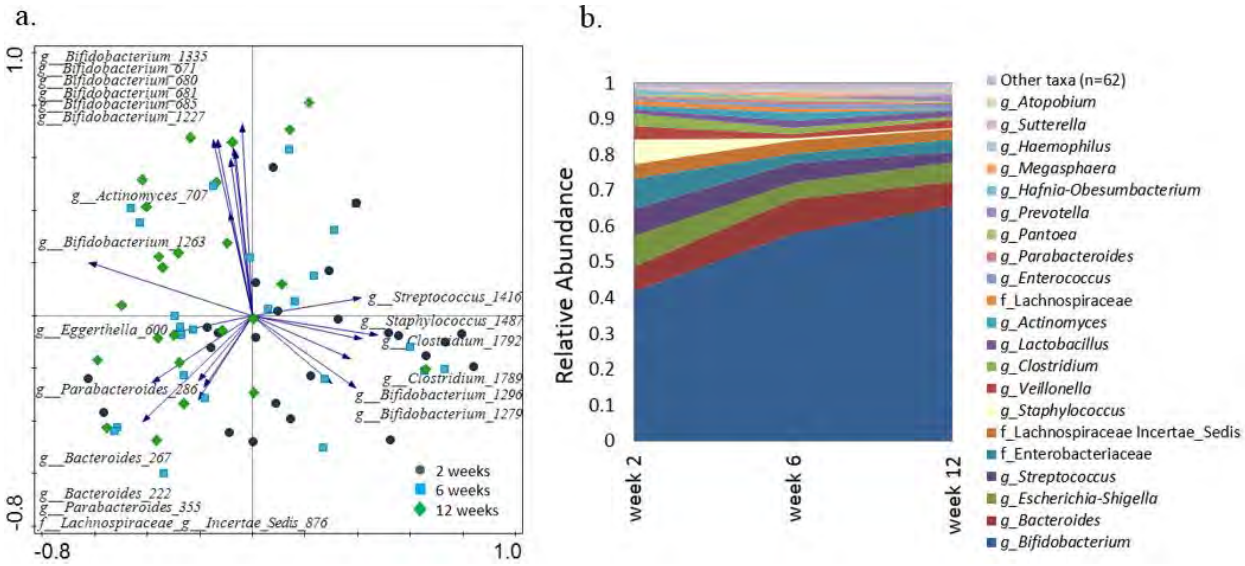
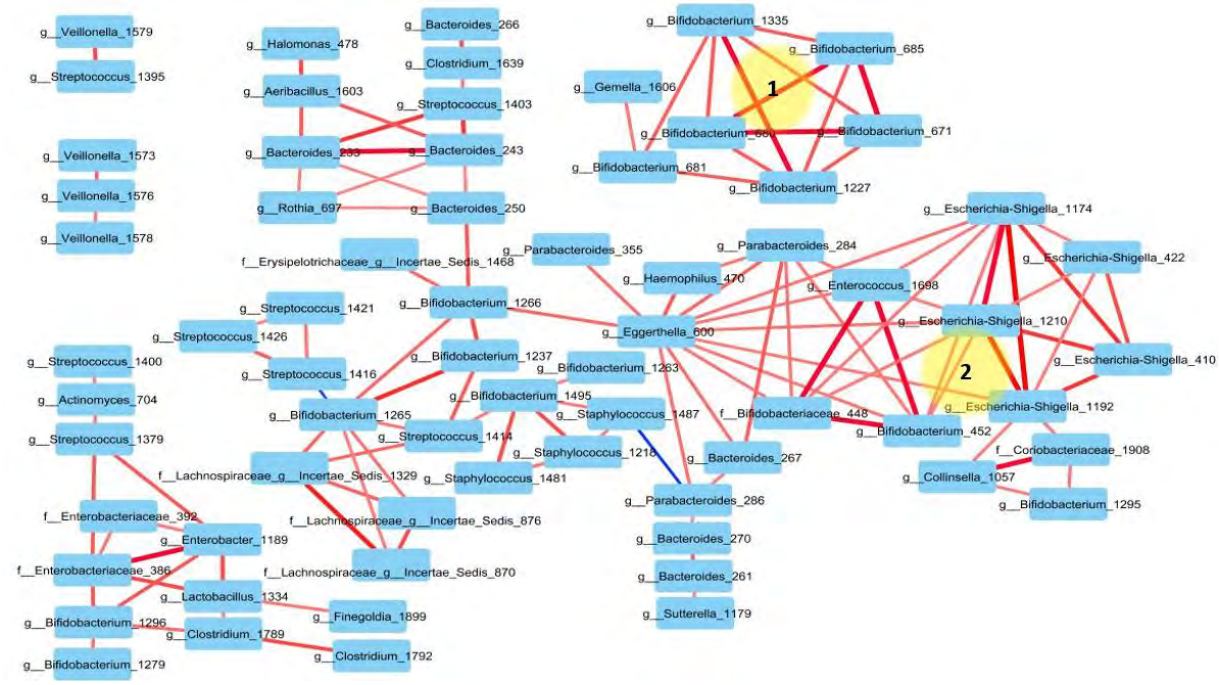


Figure 1. Infant faecal microbiota composition. a) Unconstrained analysis (PCA) showing spatial distribution of samples from 24 infants based on their faecal microbiota composition at two, six and 12 weeks of age and the twenty best fitting OTUs; b) Average relative abundance of genus level taxa in the 24 infants at different time points.

Subsequently, constrained analysis (RDA) was performed on the OTU level faecal microbiota data from 24 infants at all time points combined in order to explore the influence of available explanatory variables (infant age, estimated amounts of the 18 milk HMOs, gender, place and mode of delivery, infant illness). Together these variables explained 56.3% of variation in infant microbiota, but only infant age, gender, place and mode of delivery and the milk HMOs 6´SL, 2′FL, 3FL,3´SL, and LNDFHII significantly ($p<0.05$) contributed to explaining the variation in microbial composition (data not shown). The same analysis was repeated separately for the three individual time points. These analyses showed that at two weeks of age mode of delivery and LNFP III were significantly associated with infant faecal microbiota, at six weeks - gender, mode of delivery and 3´SL, and at 12 weeks - gender and LNH.

In order to explore microbial community assembly during the first three months of life, Spearman correlation analyses were performed to investigate significant ($p<0.05$) correlations between different OTUs at each time point. Correlations which passed the threshold of 0.5 were visualized as OTU networks (Figure 2). The majority of detected correlations were positive as indicated with red lines. As the infants aged, the number of separate, decentralized OTU networks decreased in favour of a more centralized network which was characterized by fewer but stronger connections between its OTU members. The two resulting centres at week 12 were dominated either by members of the genus *Bifidobacterium* (centre "1") or *Escherichia-Shigella-Bifidobacterium-Streptococcus* (centre "2") OTUs. At six weeks only, a third centre was observed and was dominated by members of the genera *Bifidobacterium, Lactobacillus* and Lachnospiraceae *Incertae Sedis* (centre "3").

a.


g__Veillonella_1579
g__Streptococcus_1395
g__Halomonas_478
g__Bacteroides_266
g__Clostridium_1639
g__Aeribacillus_1603
g__Streptococcus_1403
g__Veillonella_1573
g__Bacteroides_233
g__Bacteroides_243
g__Veillonella_1576
g__Rothia_697
g__Bacteroides_250
g__Veillonella_1578
g__Bifidobacterium_1335
g__Bifidobacterium_685
g__Gemella_1606
1
g__Bifidobacterium_671
g__Bifidobacterium_681
g__Bifidobacterium_1227
g__Escherichia-Shigella_1174
f__Erysipelotrichaceae_g__Incertae_Sedis_1468
g__Parabacteroides_355
g__Parabacteroides_284
g__Escherichia-Shigella_422
g__Haemophilus_470
g__Enterococcus_1698
g__Streptococcus_1421
g__Bifidobacterium_1266
g__Escherichia-Shigella_1210
g__Streptococcus_1426
g__Eggerthella_600
g__Streptococcus_1400
g__Bifidobacterium_1237
g__Bifidobacterium_1263
g__Escherichia-Shigella_410
2
g__Streptococcus_1416
g__Actinomyces_704
g__Bifidobacterium_1495
g__Bifidobacterium_1265
g__Staphylococcus_1487
f__Bifidobacteriaceae_448
g__Escherichia-Shigella_1192
g__Streptococcus_1379
g__Streptococcus_1414
g__Bifidobacterium_452
g__Staphylococcus_1218
g__Bacteroides_267
f__Coriobacteriaceae_1908
f__Lachnospiraceae_g__Incertae_Sedis_1329
g__Staphylococcus_1481
g__Collinsella_1057
g__Parabacteroides_286
g__Bifidobacterium_1295
f__Enterobacteriaceae_392
f__Lachnospiraceae_g__Incertae_Sedis_876
g__Enterobacter_1189
g__Bacteroides_270
f__Enterobacteriaceae_386
f__Lachnospiraceae_g__Incertae_Sedis_870
g__Bacteroides_261
g__Lactobacillus_1334
g__Sutterella_1179
g__Finegoldia_1899
g__Bifidobacterium_1296
g__Clostridium_1789
g__Clostridium_1792
g__Bifidobacterium_1279

b.


g__Staphylococcus_1487
g__Flavonifractor_1518
g__Bifidobacterium_1296
g__Bifidobacterium_1263
g__Finegoldia_1899
g__Clostridium_1639
g__Clostridium_1792
g__Eggerthella_600
g__Lactobacillus_1718
g__Parabacteroides_355
g__Clostridium_1789
g__Bacteroides_250
f__Erysipelotrichaceae_g__Incertae_Sedis_1468
g__Bifidobacterium_1237
g__Bacteroides_222
g__Bacteroides_261
g__Bacteroides_266
g__Bifidobacterium_685
g__Parabacteroides_286
g__Bifidobacterium_1335
g__Bifidobacterium_680
1
g__Parabacteroides_284
g__Bacteroides_270
g__Bifidobacterium_681
g__Collinsella_1057
g__Bacteroides_267
g__Sutterella_1179
g__Actinomyces_707
g__Bifidobacterium_1227
g__Escherichia-Shigella_422
g__Bacteroides_233
g__Bifidobacterium_671
Other
f__Coriobacteriaceae_1908
g__Bacteroides_243
g__Enterococcus_1698
g__Escherichia-Shigella_1174
g__Haemophilus_470
g__Varibaculum_708
2
g__Escherichia-Shigella_1192
g__Enterobacter_1189
g__Streptococcus_1400
g__Escherichia-Shigella_410
f__Enterobacteriaceae_392
g__Clostridium_1842
g__Bifidobacterium_452
g__Enterococcus_1702
g__Escherichia-Shigella_1210
g__Veillonella_1579
f__Enterobacteriaceae_386
f__Bifidobacteriaceae_448
g__Bifidobacterium_1295
g__Actinomyces_711
g__Streptococcus_1416
g__Halomonas_478
g__Streptococcus_1421
g__Bifidobacterium_1266
g__Lactobacillus_1334
g__Streptococcus_1379
g__Streptococcus_1426
g__Aeribacillus_1603
g__Actinomyces_695
g__Bifidobacterium_1265
g__Streptococcus_1414
g__Streptococcus_1403
3
g__Veillonella_1573
f__Lachnospiraceae_g__Incertae_Sedis_1329
g__Gemella_1606
g__Staphylococcus_1218
g__Streptococcus_1395
f__Lachnospiraceae_g__Incertae_Sedis_876
g__Veillonella_1576
g__Lactobacillus_1205
f__Lachnospiraceae_g__Incertae_Sedis_870

c.

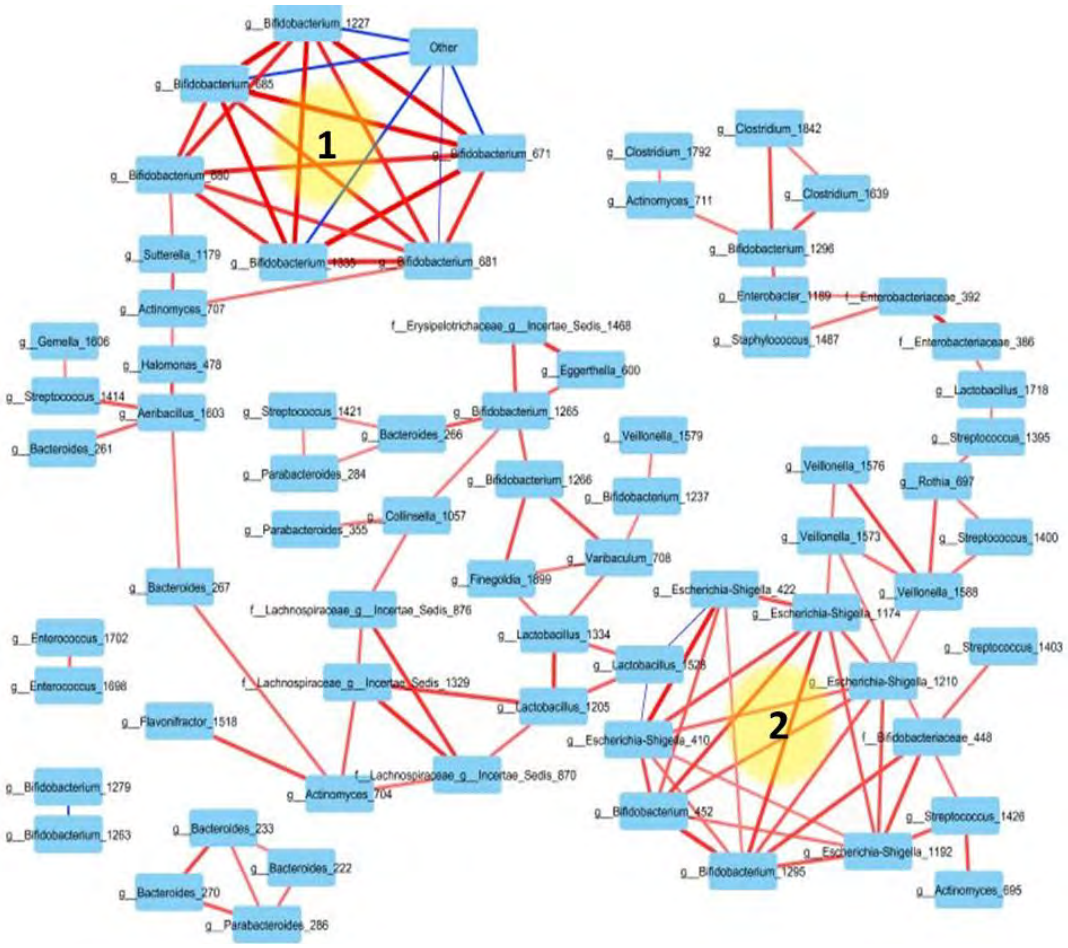


Figure 2. OTU networks showing statistically significant ($p<0.05$) positive and negative associations which passed a Spearman correlation threshold of 0.5 in faecal samples of two (a), six (b) and 12 (c) week old infants. Red lines indicate positive correlations, whereas blue lines represent negative correlations. Wider lines and brighter line colour correspond with higher correlation score indicating stronger associations. Microbial centres of interest are numbered and highlighted in yellow.

Milk HMO content and changes during lactation

Unconstrained analysis (PCA) of milk HMOs showed a clear separation of samples in relation to collection time point postpartum (Figure 3a). This finding was confirmed with RDA analysis which showed that collection time point had a significant effect explaining 16% of variation in the data (data not shown). Neither delivery mode nor maternal stress (measured via saliva cortisol, data not shown) at different time points were significantly associated with the HMO levels. The average concentration (µg/ml) of the HMOs measured in the milk decreased with time of lactation, except for 3FL and LNFPIII, which were secreted at higher amounts at later time points (Figure 3b). Earlier studies showed that the average daily milk intake of breastfed infants increases rapidly in the first three months of life [16]. After adjusting the HMO concentrations for the average estimated amount of milk consumed (480 g, 580 g and 630 g at week two, six and 12, respectively), the average HMO intake seemed to remain stable in the first three months of life (Figure 3c).

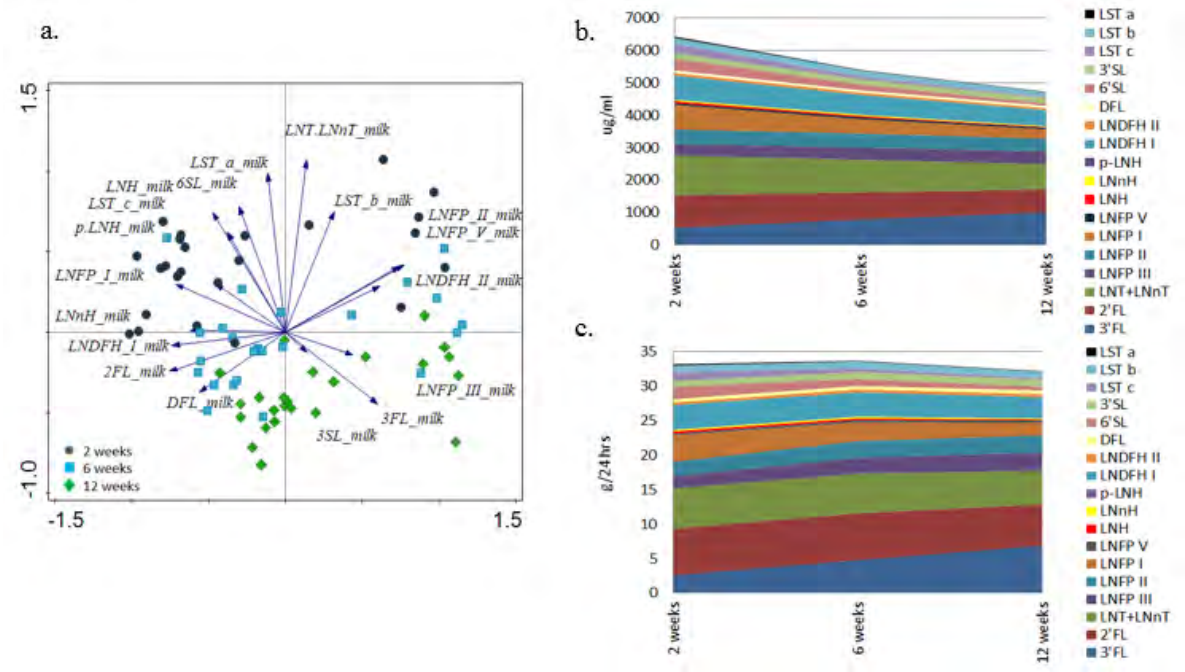


Figure 3. Breastmilk HMO profiling. a) Unconstrained analysis (PCA) showing separate clustering of breastmilk samples of 24 mothers at two, six and 12 weeks of lactation. Both, 3FL and LNFPIII are positively associated with duration of lactation. b) Average concentration of different HMOs found in the breastmilk samples from the 24 mothers included in the study at two, six and 12 weeks of lactation; c) Estimated average daily intake of each HMO in the 24 infants at two, six and 12 weeks of age.

Correlation between milk HMOs and infant faecal microbiota

Average daily milk intake of breastfed infants increases rapidly in the first three months of life [16]. We estimated daily intake (g/day) of each HMO consumed by each infant at two, six and 12 weeks of age and used this data in Spearman correlation analysis with the relative abundance of OTUs for the corresponding samples. The correlations detected were weak and did not exceed correlation values of ± 0.6 (Figure 4). Our analyses showed that the most predominant *Bifidobacterium* OTU 1263 was not significantly correlated with any of the measured HMOs. Few low abundance *Bifidobacterium* OTUs correlated negatively with 6´SL, LSTc, LNH, LNnH, LNFPI, III and V. The strongest positive correlations were detected between few of the *Staphylococcus* OTUs and 6´SL, LSTc, LSTa, LNH, LNnH, and LNFPI. *Streptococcus* OTUs were positively correlated with 3FL, LNFPIII, and pLNH. Positive correlations were also found between OTUs within the genus *Actinomyces* and 3´SL, and *Enterococcus* OTUs and 3FL, LNFPII, III, V and LNDFHII (Figure 4).

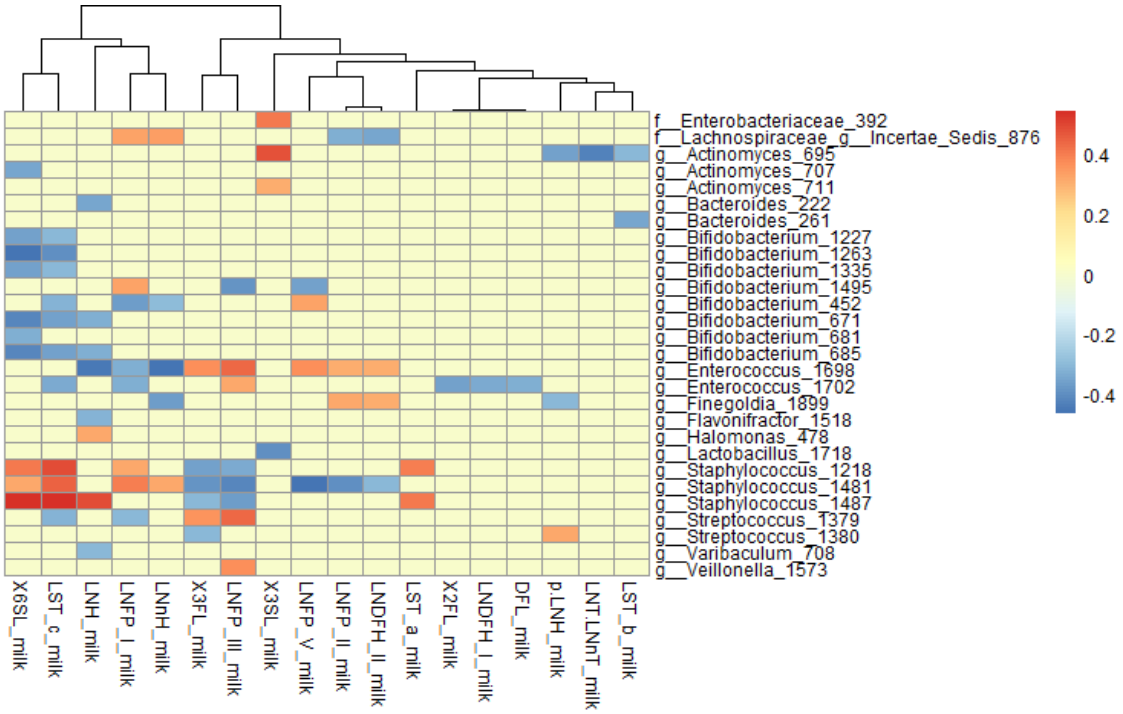


Figure 4. Statistically significant ($p<0.05$) Spearman correlations (correlation threshold ±0.3) between estimated daily intake of different HMOs and faecal microbiota composition at OTU level of 24 infants across the study duration. Positive associations are indicated in red, negative in blue, and correlations that did not pass significance or correlation threshold are left in yellow.

We then performed Spearman correlation analysis for each time point separately in order to address potential milk HMO-OTU associations only present at single time points rather than the entire 12 weeks (Figure S2a,b,c). Again, the majority of correlations detected were weak and did not exceed correlation values of ± 0.6. At the two week time point three bifidobacterial OTUs were positively associated with LNH, pLNH and LNFPII, and the main *Bifidobacterium* OTU 1263 was negatively associated with LNFPII and III. At six weeks three different low abundance bifidobacterial OTUs were positively associated with LNFPV, LNnH, LNH, LSTb and c, whereas at 12 weeks LNnH, LNFPII, and LNDFHII showed a low positive correlation with three low abundance bifidobacterial OTUs. Overall there was no consistency in the type (positive or negative), or strength (passing the correlation threshold of ± 0.3, $p<0.05$) of associations between specific OTU-HMO pairs when all ages were combined, or when individual time points were analysed separately.

Correlation between infant faecal microbiota and HMOs excreted in infant faeces.

Undigested milk HMOs were secreted in infant faeces and their concentrations varied between infants and time points. Our hypothesis was that aside of being directly dependent on the amounts of the HMOs ingested with the milk, the faecal HMO concentrations could serve as an indicator of the efficiency of the degradation (consumption) of different HMOs by infant GI tract microbiota. The RDA analysis showed that infant age had a significant effect on faecal HMO concentrations (Figure 5a), and the concentrations of all HMOs in faeces decreased with infant age, with the exception of 3FL (Figure 5b).

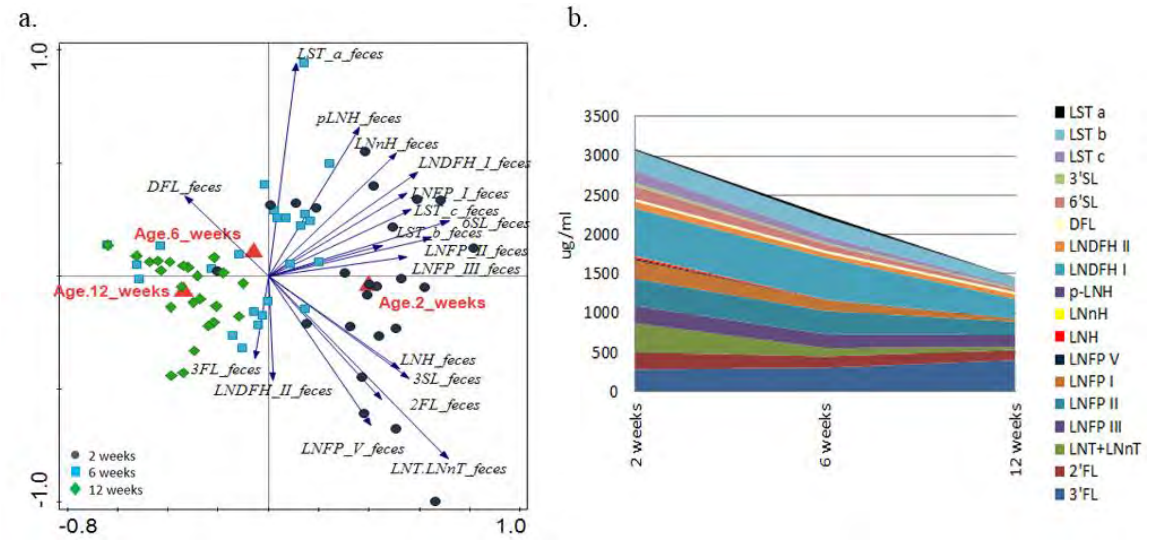


Figure 5. Faecal HMO profiling. a) Constrained analysis (RDA) showing spatial distribution of faecal samples of 24 infants at two, six and 12 weeks of age based on the concentrations of residual HMOs detected in infant faeces; b) Average concentration of different HMOs (µg/ml of faeces) found in infant faeces decreases with infant age, except for 3FL, which for most infants increased with age.

Spearman correlation analysis between the faecal HMO concentrations and faecal microbiota of the 24 infants at all three time points showed that the majority of statistically significant negative associations were present between bifidobacterial OTUs and thirteen different HMOs (Figure 6). The main *Bifidobacterium* OTU 1263 was negatively correlated with nine different HMOs, of which LNH, LNT+LNnT and LNFPV showed strongest correlations. Two of the bifidobacterial OTUs (1296 and1495) showed only positive correlations with faecal HMOs. Positive associations were also observed between various OTUs of *Streptococcus, Staphylococcus, Clostridium* and *Escherichia-Shigella* and thirteen different HMOs.

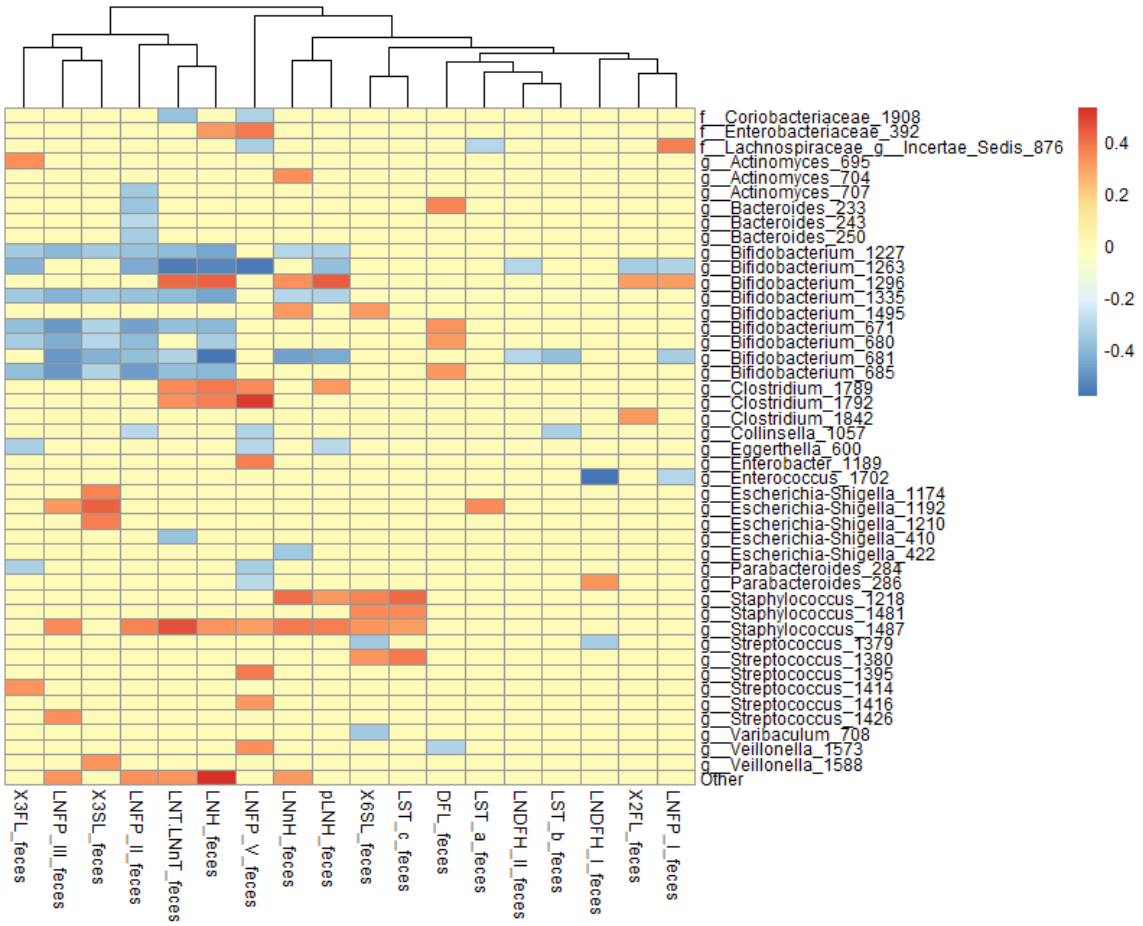


Figure 6. Statistically significant ($p<0.05$) Spearman correlations (correlation threshold ±0.3) between HMOs detected in infant faeces and faecal microbiota composition at OTU level of 24 infants across the study duration. Positive associations are indicated in red, negative in blue, and correlations that did not pass significance or correlation threshold are marked in yellow.

Spearman correlation analysis between bacterial OTUs and faecal HMO concentrations was repeated on data from each separate time point. Correlation scores reached ± 0.6. At two weeks of age the strongest negative correlations were found between various HMOs and highly abundant *Bifidobacterium* OTU 1263 and 681 (Figure S3a). At week six, an additional six lower abundance bifidobacterial OTUs showed negative correlations with pLNH, LNH, LNFPII and III. At both time points *Bifidobacterium* OTU 1263 was the only bifidobacterial population correlating negatively with 3FL possibly highlighting the unique link between this HMO and the major *Bifidobacterium* OTU during the initial stages of the development of GI tract microbiota (Figure S3b). At 12 weeks of age, the HMO concentrations in faeces were very low or no longer within the detectable range. The correlations between bifidobacteria and faecal HMOs were less clear, especially between the various HMOs and the bifidobacterial OTUs. The strongest negative correlations and the highest number of associations for this bacterial group were identified between *Bifidobacterium* OTU 1263 and 2′FL, LNDFHII and LNFPV (Figure S3c).

In order to account for the initial availability of different milk HMO, we calculated ratios between each faecal and milk HMO, for each mother-infant pair at each time point. The

resulting ratios were then used to estimate the consumption level for each HMO as either "high", "medium" or "low" based on tertiles. The RDA analysis showed that high consumption was associated with older infant age and higher levels of bifidobacteria. Low consumption of pLNH, 2′FL, LNH, LNnH, LNTandLNnT, LNFPI and V, 3´SL and LNDFHI, medium consumption of LNDFHII, LNFPII and 6´SL, and high consumption of pLNH, LNH, LNnH, LNFPI and V, LSTa, LNTandLNnT, and 3´SL were significantly associated with the microbial composition (FDR<0.05) (Figure 7). Overall high consumption was detected in association with various bifidobacterial OTUs including the most predominant *Bifidobacterium* OTU 1263, as well as several OTUs within genera *Parabacteroides*, *Escherichia-Shigella, Bacteroides, Actinomyces, Velionela,* and Erysipelotrichaceae *Incertae Sedis* (Figure 6 and 7).

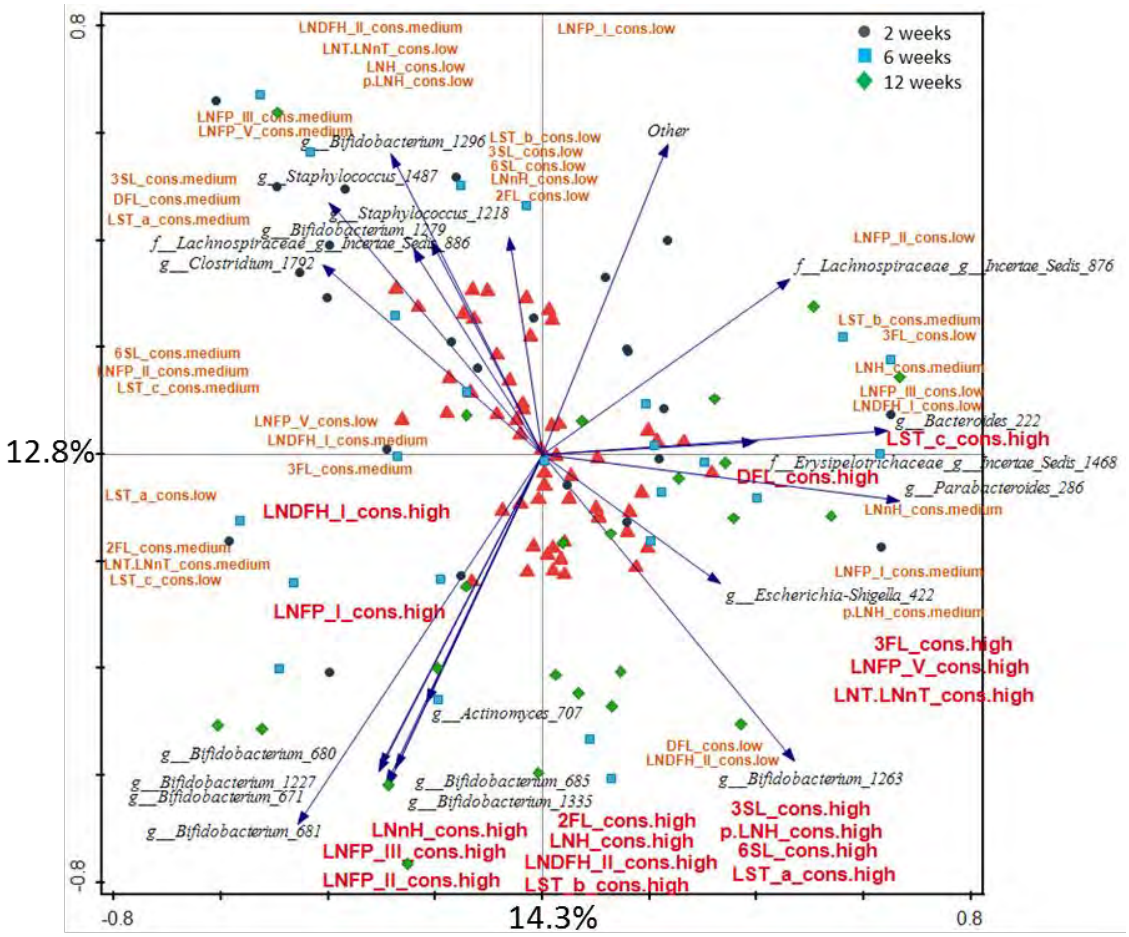


Figure 7. Constrained analysis (RDA) showing spatial distribution of faecal samples of 24 infants at two, six and 12 weeks of age based on their OTU composition and using the estimated level of consumption (low, medium and high) for each HMO as an explanatory variable.

Finally, we compared relative abundance of OTUs between infants assigned into a low and high consumption groups for each HMO (Table S3). As expected, infants who were classified in the high consumption category had a significantly higher relative abundance of various bifidobacterial OTUs and in most cases the most predominant OTU *Bifidobacterium* 1263. Only the high consumption of 6´SL, DFL, LSTa and LSTc was not significantly linked

with the presence of bifidobacteria, but instead was linked with higher relative abundances of other bacterial groups (Table S3).

Discussion

Faecal microbiota composition of young infants varies between individuals, is highly dynamic, and depends on multiple factors, of which some seem to be age-related [20-23]. In our study we followed a cohort of 24 mother-infant pairs and collected breastmilk and infant faecal samples at two, six and 12 weeks post-partum. Our goal was to find a link between breastmilk HMO composition, infant faecal microbiota, and the ability of an infant to utilize different HMOs. During the study period we noted directional changes in both, the milk HMO concentrations and the infant faecal microbiota composition. Microbial community analyses revealed that in the first three months of life factors such as infant age, gender, place and mode of delivery and certain milk HMOs, namely 6´SL, 2′FL, 3FL, 3´SL, LNDFHII, could explain 35% of the variation in microbiota composition, and that these factors had a significant ($p<0.05$) effect on microbiota structure. Separate analyses at each time point revealed that the effect of different factors also varied with age. When infants were two weeks old, mode of delivery and LNFPIII showed a significant association with the faecal microbiota composition. The important role of the mode of delivery in the initial seeding of the GI tract has been previously reported [24, 25] and has been linked to various health outcomes, both in infancy and beyond [26, 27]. At six weeks the significant effect of the mode of delivery could still be detected, but also gender and milk 3´SL seemed to play a role. A recent study using animal models showed that gender specific, microbiota-independent differences in immunity may lead to the selection of a gender-specific GI microbiota in adult germ free mice [28]. Gender-related differences in faecal microbiota had been also reported in adults [29], and in pre-term infants [30], but up to date, there are no published reports on timing and the possible mechanisms that might underlie the gender differentiation of GI microbiota in healthy full term infants. One of the simplest possible explanations might be the fact that male infants tend to have a higher daily milk intake as compared to female infants of the same age, leading to a higher transfer of microbial and HMO components from the mother's milk [31]. However, it is likely that other factors could also play a role. At 12 weeks of age, both gender and LNH were significantly associated with microbiota, however, the significant association between microbiota composition and mode of delivery was no longer present. With the accumulating evidence linking mode of delivery with various health effects later in life [26, 27, 32], it is likely that the microbiota related programming of the host happens soon after birth, or during specific "windows of opportunity", yet even when microbiota recovers to its normal state, the long term health effects of such disturbance might persist throughout life [32].

One of the crucial factors shaping the development of GI microbial community during infancy is the type of feeding that an infant receives. Multiple studies comparing faecal microbiota of healthy, full term infants receiving formula vs. breastmilk show different colonisation patterns with regard to feeding mode [33-37]. Human milk is a highly complex biofluid which evolved not only to fulfil nutritional needs of a growing baby, but also to guide

the early life development of its GI microbial community [2, 38]. One of the main components of milk supporting microbial colonisation in early life are HMOs [6, 39, 40]. Earlier research showed that the breastmilk HMO content is highest in colostrum, and the concentrations of HMOs decrease in mature milk [8, 41]. Furthermore, the HMO content of breastmilk is genetically predetermined and varies between mothers [9]. Our data is in agreement with these findings showing that concentrations of the 16 HMOs measured per mL of milk varied between mothers, and decreased between two and 12 weeks of lactation, except for 3FL and LNFPIII, which increased in concentration as lactation progressed. Earlier studies on infant feeding and nutrition showed that the daily intake of milk by an infant increases in the first months life [16]. Unfortunately, we did not collect data on daily milk intake in our study. To address this limitation, we used literature reported feeding volumes and calculated the approximate daily amounts of ingested HMOs at each study time point. Based on this data, we noted that the increase in volumes consumed at six and 12 weeks of age could compensate for the decreased HMO concentrations, and on average the daily amounts of ingested HMOs remained relatively stable during the whole study period. The exception was in the intake of 3FL and LNFPIII which gradually increased in time. Interestingly, this increase in the 3FL and LNFPIII intake corresponded with the increase in the faecal concentrations of these two HMOs, suggesting that, on average, their supply likely exceeded the ability of the infant GI tract microbiota to utilize these two HMOs.

In order to better understand the relationship between the supply of different breastmilk HMOs and their effect on individual bacterial OTUs (species/strains), we searched for positive correlations between separate HMOs measured in mother's milk samples and the microbial OTUs detected in infant faeces in each of the mother–infant pairs. Unfortunately, we did not measure the absolute abundance of bacteria in infant faeces during this study and are aware of the limitation of using relative abundance data. However, culture-based studies showed that the total faecal bacterial load, as well as *Bifidobacterium* counts tend to increase in the first month of life [34]. Thus, the observed increase in relative abundance of bifidobacteria in our data should reflect the actual increase in the abundance of this group, rather than being a simple artefact of a decrease in the abundance of other taxa. Our hypothesis was that quantity of HMOs might selectively promote growth of either the primary or secondary HMO degraders, leading to increase in their abundance within the microbial community. One of the signature bacterial groups found in faeces of breastfed infants is the genus *Bifidobacterium* [2]. *In vitro* studies showed that *Bifidobacterium bifidum* JCM1254 [40], *Bifidobacterium longum* subsp. *infantis*, *Bacteroides fragilis* and *Bacteroides vulgatus* [39] grow well on HMOs as sole carbon source. Thus, we expected to find positive correlations between certain HMOs and the bifidobacterial OTUs. However, when using data from the three time points combined, we saw an opposite trend - as the predicted daily intake of most HMOs was stable in time, the relative abundance of bifidobacteria was increasing, and the Spearman correlation analysis returned no, or negative associations. The same analyses repeated for individual time points showed very few positive correlations between bifidobacterial OTUs and HMOs, but the results varied between the time points and no two identical milk HMO- bifidobacterial OTU associations passed the correlation threshold and significance cut off for any of the three different age groups (Figure S2 a, b, c). The overall low number of positive associations between HMOs in milk and bifidobacterial

OTUs could be due to the fact that, in addition to HMOs promoting growth of selected microbes capable to utilize this carbon source, there may be other mechanisms controlling the microbial community structure. For example, presence of other breastmilk components, such as lysozyme, secretory IgA and other endogenous factors can supress growth of certain members of the community and thus, indirectly allow other bacterial species to dominate the infant GI tract ecosystem [34]. We observed positive associations at all time points for breastmilk 3´SL and unidentified OTUs within family Enterobacteriaceae and *Actinomyces* 695. 6´SL was positively associated with clostridia – *Clostridium* 1789 at week two and *Clostridium* 1639 at week 12. Finally, LNFPIII was positively associated with *Enterococcus* 1698 at 6 and 12 weeks. At two and at six weeks Lachnospiraceae *Incertae Sedis* 876 was negatively associated with LNDFHII, *Lactobacillus* 1718 was negatively associated with 3´SL, and *Bifidobacterium* 1295 with LNFPI.

In our analyses we hoped to find a direct link between milk HMOs and infant faecal microbiota. However, it is likely that the effect of HMOs on microbiota may already start in the breastmilk itself. A recent study on human milk investigated the associations between the HMO content and microbiota composition in colostrum and reported strong positive correlations between different HMOs and various microbial groups, including streptococci, staphylococci, enterococci and bifidobacteria, in particular *Bifidobacterium breve* and LNFPIII [42]. As described earlier, our RDA analysis showed a significant association of milk LNFPIII with infant faecal microbiota in all time points combined. There was a strong positive association of milk LNFPIII with OTUs belonging to the genera *Veillonella*, *Enterococcus* and *Streptococcus* (Figure S2). Thus, it is likely that some of the correlations we detected in our data were due to a combined effect of the HMOs modulating both, the microbiota of the mother's milk and the infant GI tract, as well as due to the transfer of bacteria during breastfeeding. Studies on mature breastmilk microbiota and the microbial transfer of microbiota from mother to infant show that breastmilk contains a distinct microbial community and that breastfed infants receive on average nearly 30% of the bacteria from breastmilk and 10% from areolar skin in the first 30 days of life [43]. The study also concluded that the association was lower in older infants, and it was proportional to the frequency of breastfeeding that an infant received [43]. In fact, preliminary analyses of the microbiota in the milk samples in our study detected a significant positive correlation between *Staphylococcus* in breastmilk and infant faeces at two, six and 12 weeks of age combined, but the significant association was not found for *Streptococcus* (data not shown).

Infant GI microbiota plays an important role in energy metabolism via utilising otherwise indigestible HMOs. Our data showed that the average concentrations of faecal HMOs decreased with age, likely indicating that microbiota of older infants is more adapted and efficient in degrading these compounds (Figure 5b). Furthermore, we noted that the increase in efficiency was correlated with the increase in the relative abundance of several bifidobacterial OTUs, *Parabacteroides*, *Escherichia-Shigella, Bacteroides, Actinomyces, Veillonella,* and Erysipelotrichaceae *Incertae Sedis* (Figure 7). We hypothesized that strong negative Spearman correlations would be detected when higher relative abundance of bacterial groups that were involved in the HMO metabolism would lead to higher HMO consumption and thus, lower

faecal HMO content. As expected, overall 17 of the 18 faecal HMOs showed a significant negative correlation with 23 different OTUs for all time points combined (Figure 6). Furthermore, 13 faecal HMOs were negatively correlating with nine different *Bifidobacterium* OTUs, whereas in the analyses correlating milk HMOs and microbiota (Figure 4) only six milk HMOs and eight bifidobacteria OTUs were negatively correlated, with the only overlapping result detected for LNH and *Bifidobacterium* 671 and *Bifidobacterium* 685. This implies, that the majority of detected negative correlations for *Bifidobacterium* were the result of the HMO degradation, rather than the associations already present between OTUs and milk HMOs. Interestingly, two of the *Bifidobacterium* OTUs (1296 and 1495) showed only positive correlations with faecal HMOs (Figure 6). The highly abundant *Bifidobacterium* OTU 1263 was negatively correlated with nine different HMOs in faeces, especially LNH, LNT and LNnT and LNFPV.

Negative associations were also observed for LNFPII and *Bacteroides,* and for LNFPV and *Parabacteroides* suggesting the role of these bacteria in the HMO degradation. The fact that *Bacteroides* and *Parabacteroides* (formerly also *Bacteroides*) were identified in our analysis is in line with earlier studies showing that *Bacteroides* spp. can grow on selected milk glycans as a sole carbon source by activating the mucus degradation pathway [44]. Finally, LNDFHI in both, milk and faeces was negatively associated with *Enterococcus* OTU 1702, but the association was stronger in faeces. Even though *in vitro* studies showed that in a monoculture *Enterococcus* was not able to grow on milk HMOs [39], another study showed that this group was found in breastmilk [42], that its abundance in infant faeces could be predicted from the maternal HMO profile and that it was positively correlated with the abundance of *Bifidobacterium, Streptococcus* and *Veillonella* [6]. One of the suggested explanations was that *Enterococcus* can cross feed on HMO fermentation products or HMO breakdown by-products that are released in the ecosystem by HMO degrading bifidobacteria or *Bacteroides* spp. [6].

The correlation analysis of infant faecal HMOs and infant faecal OTUs for all time points combined also detected numerous significant positive associations between various HMOs and *Streptococcus, Staphylococcus, Escherichia-Shigella*, and *Clostridium* OTUs (Figure 6). In both, milk and faeces LSTc, 6´SL and LNnH showed strong significant correlation with staphylococci, while LNFPIII and 3FL were positively correlated with streptococci. Earlier studies showed that neither *Streptococcus, Staphylococcus, Escherichia-Shigella* [6, 39], nor *Clostridium* [39] could effectively utilize and grow on milk HMOs. However, all these bacterial groups are members of the microbiota of breastmilk and areolar skin [43, 45, 46], and even if they are not involved in HMO degradation, the HMO presence may confer a protective effect on these groups. Finally, the positive link might be due to breastfeeding practises or frequency, for example when infants feed more often, they likely ingest more of both, the bacteria and the HMOs, and if the HMOs are not well digested, the positive associations may still persist in the faeces.

By comparing the amounts of breastmilk HMOs and the HMOs detected in infant faeces, we classified infants into low, medium or high consumption categories and compared microbial profiles of infants who were classified as high and low consumers for each HMO

(Table S3). Overall, those who were good degraders of 2′FL, 3FL, 3´SL, LNDFHI, II, LNFPI, II, III, V, LNH, LNnH, pLNH, LSTb, and LNT and LNnT had on average a significantly higher relative abundance of one or more *Bifidobacterium* OTUs, confirming the important role of this bacterial group in the HMO degradation (Figure 7). The highly abundant *Bifidobacterium* 1263 was associated with the degradation of all these HMOs except for LSTb, which was degraded in the presence of *Bifidobacterium* 681- the third most abundant *Bifidobacterium* OTU in our data set. Aside of bifidobacteria, members of the genus *Bacteroides* were significantly more abundant in infants who were good degraders of 2′FL, LNFPI, II, V, and pLNH, and *Parabacteroides* in the high degraders of 3FL, LNFP V, LNH, LNT and LNnT, indicating that these microbial groups might have a mutualistic or symbiotic relationship degrading those compounds. In addition, *Halomonas*, *Enterococcus*, *Lactobacillus,* St*aphylococcus, Suterella*, *Varibaculum*, *Veillonella*, *Streptococcus, Actinomyces*, Lachnospiraceae *Incertae Sedis* were also associated with degradation of the same HMOs as bifidobacteria. Interestingly, four of the tested HMOs, namely 6´SL, DFL, LSTa and LSTc showed no significant increase of any of the bifidobacterial OTUs in relation to high degradation, but instead were associated with various OTUs belonging to *Bacteroides*, *Streptococcus*, *Varibacullum* (6´SL), *Actinomyces*, *Clostridium*, *Collinsella* and *Streptococcus* (LSTc), and *Haemophilus*, *Veillonella* (DFL), and Lachnospiraceae *Incertae Sedis*, and *Halomonas* (LSTa).

In this study we showed that selected breastmilk HMOs have a limited influence on shaping the microbiota community structure in faeces of breastfed infants. However, we found a strong link between degradation levels of various HMOs and specific microbial groups, in particular different members of *Bifidobacterium*. Earlier studies showed that different bifidobacterial species vary in their ability to break down HMOs, and some species can degrade HMOs without experiencing a detectable population growth [47, 48]. Thus, including metatranscriptome or metaproteome analyses in this set would have been very helpful in understanding the community dynamics in regard of HMO metabolism in the infant GI tract. Our findings could provide the basis for assembling simple synthetic communities to study microbial interactions and community structure changes which are centred around degradation of different HMO structures. *In vitro* fermentation studies incorporating purified compounds would also allow to eliminate confounders, such as presence of milk's own microbiota and presence of milk components, which have a regulatory effect on microbiota in both, milk and in the infant GI tract.

Conclusion

Our study showed that the faecal microbiota of breastfed infants during the first 12 weeks of life is highly diverse, dynamic and influenced by age and other factors. The effect of mode of delivery disappeared after six weeks of age, whereas the effect of gender became detectable. Overall, microbiota development in this cohort followed a normal colonization pattern resulting in faecal microbial communities dominated by *Bifidobacterium,* in particular the most predominant *Bifidobacterium* OTU 1263. Breastmilk HMO analyses showed that the composition of the 18 HMOs that were measured varied between mothers and throughout the

duration of lactation. In our analysis we did not observe strong and consistent positive correlations between the HMOs in maternal breastmilk and specific microbial OTUs including bifidobacteria in infants' faeces. Thus, we believe that HMO composition is only one of many factors regulating colonization and structure of the infant GI microbial community. However, our study confirmed the key role of bifidobacteria in the infants' ability to degrade most of the measured HMOs, in addition to indicating the role of other microbial taxa in the degradation or metabolism of specific HMOs.

Acknowledgements

The authors thank Maria Suarez-Diez (Wageningen University, Laboratory of Systems and Synthetic Biology) for her help with the DMM analysis.

Funding

This research was performed in the public-private partnership CarboHealth coordinated by the Carbohydrate Competence Center (CCC, www.cccresearch.nl) and financed by participating partners and allowances of the TKI Agri&Food program, Ministry of Economic Affairs of the Netherlands.

References

1. JE, K., et al., *Succession of microbial consortia in the developing infant gut microbiome.* Proc Natl Acad Sci U S A, 2011. **108** (Suppl 1): p. 4578-85.
2. Zivkovic, A.M., et al., *Establishment of a Milk- Oriented Microbiota (MOM) in Early Life: How Babies Meet Their MOMs.* Functional Food Reviews, 2013. **5**(1): p. 3–12.
3. Chu, D.M., et al., *Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery.* Nat Med, 2017. **23**(3): p. 314-326.
4. Kummeling, I., et al., *Etiology of atopy in infancy: the KOALA Birth Cohort Study.* Pediatr Allergy Immunol, 2005. **16**(8): p. 679-84.
5. Scheepers, L.E., et al., *The intestinal microbiota composition and weight development in children: the KOALA Birth Cohort Study.* Int J Obes (Lond), 2015. **39**(1): p. 16-25.
6. Wang, M., et al., *Fecal microbiota composition of breast-fed infants is correlated with human milk oligosaccharides consumed.* J Pediatr Gastroenterol Nutr, 2015. **60**(6): p. 825-33.
7. Coppa GV1, G.O., Pierani P, Catassi C, Carlucci A, Giorgi PL, *Changes in carbohydrate composition in human milk over 4 months of lactation.* Pediatrics, 1993. **91**(3): p. 637-41.
8. Austin, S., et al., *Temporal Change of the Content of 10 Oligosaccharides in the Milk of Chinese Urban Mothers.* Nutrients, 2016. **8**(6): p. 22.
9. Sprenger, N., et al., *Longitudinal change of selected human milk oligosaccharides and association to infants' growth, an observatory, single center, longitudinal cohort study.* PLoS One, 2017. **12**(2): p. e0171814.
10. Albrecht, S., et al., *CE-LIF-MS n profiling of oligosaccharides in human milk and feces of breast-fed babies.* Electrophoresis, 2010. **31**(7): p. 1264-73.
11. Wu, S., et al., *Development of an annotated library of neutral human milk oligosaccharides.* J Proteome Res, 2010. **9**(8): p. 4138-51.
12. Wu Shuai, R.G., J. Bruce German, and Carlito B. Lebrilla, *Annotation and Structural Analysis of Sialylated Human Milk Oligosaccharides.* Journal of Proteome Research, 2011. **10**: p. 856–868.
13. Stephan Thurl, B.M.-W., and Gunther Sawatzki, *Quantification of Individual Oligosaccharide Compounds from Human Milk Using High-pH Anion-Exchange Chromatography.* Analytical Biochemistry, 1996. **235**: p. 202 – 206.
14. Ramiro-Garcia, J., et al., *NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes.* F1000Research, 2016. **5**: p. 1791-1809.
15. Quast, C., et al., *The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.* Nucleic Acids Res, 2013. **41**(Database issue): p. D590-6.
16. (DC), I.o.M.U.C.o.N.S.D.P.a.L.W., *Nutrition During Lactation. 5, Milk Volume.* National Academies Press, 1991(0-309-53767-3): p. 1-326.
17. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data.* Nature Methods, 2010. doi:10.1038/nmeth.f.303.
18. Shannon, P., et al., *Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.* Genome Research, 2003. **13**(11): p. 2498-504.
19. Šmilauer, P., & Lepš, J. , *Multivariate Analysis of Ecological Data using CANOCO 5.* Cambridge: Cambridge University Press, 2014.
20. Yatsunenko, T., et al., *Human gut microbiome viewed across age and geography.* Nature, 2012. **486**(7402): p. 222-227.
21. Scholtens, P.A.M.J., et al., *The Early Settlers: Intestinal Microbiology in Early Life.* Annual Review of Food Science and Technology, 2012. **3**(1): p. 425-447.
22. Timmerman, H.M., et al., *Intestinal colonisation patterns in breastfed and formula-fed infants during the first 12 weeks of life reveal sequential microbiota signatures.* Scientific Reports, 2017. **7**(1): p. 8327.
23. Bäckhed, F., et al., *Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life.* Cell Host & Microbe, 2015. **17**(5): p. 690-703.

24. Biasucci, G., et al., *Mode of delivery affects the bacterial community in the newborn gut.* Early Hum Dev, 2010. **86** Suppl **1**: p. 13-5.
25. Rutayisire, E., et al., *The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review.* BMC Gastroenterol, 2016. **16**(1): p. 86.
26. Black, M., et al., *Planned Cesarean Delivery at Term and Adverse Outcomes in Childhood Health.* JAMA, 2015. **314**(21): p. 2271-9.
27. Kuhle, S., O.S. Tong, and C.G. Woolcott, *Association between caesarean section and childhood obesity: a systematic review and meta-analysis.* Obes Rev, 2015. **16**(4): p. 295-303.
28. Fransen, F., et al., *The Impact of Gut Microbiota on Gender-Specific Differences in Immunity.* Frontiers in Immunology, 2017. **8**.
29. Mueller, S., et al., *Differences in fecal microbiota in different European study populations in relation to age, gender, and country: a cross-sectional study.* Appl Environ Microbiol, 2006. **72**(2): p. 1027-33.
30. Cong, X., et al., *Gut Microbiome Developmental Patterns in Early Life of Preterm Infants: Impacts of Feeding and Gender.* PLoS One, 2016. **11**(4): p. e0152751.
31. Nancy F. Butte, et al., *Infant Feeding Mode Affects Early Growth and Body Composition.* Pediatrics, 2000. **106**(6): p. 1355-1366.
32. Abrahamsson, T.R., et al., *Low gut microbiota diversity in early infancy precedes asthma at school age.* Clin Exp Allergy, 2014. **44**(6): p. 842-50.
33. Harmsen, H.J.M., et al., *Analysis of Intestinal Flora Development in Breast-Fed and Formula-Fed Infants by Using Molecular Identification and Detection Methods.* Journal of Pediatric Gastroenterology and Nutrition, 2000. **30**(1): p. 61-67.
34. Kleessen, B., et al., *Influence of two infant formulas and human milk on the development of the faecal flora in newborn infants.* Acta Pediatrica, 1995(84): p. 1347-1356.
35. Kunz, C. and S. Rudloff, *Biological functions of oligosaccharides in human milk.* Acta Paediatr, 1993. **82**: p. 903-12.
36. Stark, P. and A. Lee, *The microbial ecology of the large bowel of breast-fed and formula-fed infants during the first year of life.* J Med Microbiol, 1982. **15**(2): p. 189-203.
37. Balmer, S. and B. Wharton, *Diet and faecal flora in the newborn: breast milk and infant formula.* Archives of Disease in Childhood, 1989. **64**(12): p. 1672-1677.
38. Smilowitz, J.T., et al., *Breast milk oligosaccharides: structure-function relationships in the neonate.* Annu Rev Nutr, 2014. **34**: p. 143-69.
39. Marcobal, A., et al., *Consumption of human milk oligosaccharides by gut-related microbes.* J Agric Food Chem, 2010. **58**(9): p. 5334-40.
40. Asakuma, S., et al., *Physiology of consumption of human milk oligosaccharides by infant gut-associated bifidobacteria.* J Biol Chem, 2011. **286**(40): p. 34583-92.
41. Coppa, G., et al., *Changes in carbohydrate composition in human milk over 4 months of lactation.* Pediatrics, 1993. **91**(3): p. 637-41.
42. Aakko, J., et al., *Human milk oligosaccharide categories define the microbiota composition in human colostrum.* Benef Microbes, 2017. **8**(4): p. 563-567.
43. Pannaraj, P.S., et al., *Association Between Breast Milk Bacterial Communities and Establishment and Development of the Infant Gut Microbiome.* JAMA Pediatr, 2017. **171**(7): p. 647-654.
44. Marcobal, A., et al., *Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways.* Cell Host Microbe, 2011. **10**(5): p. 507-14.
45. Hunt, K.M., et al., *Characterization of the diversity and temporal stability of bacterial communities in human milk.* PLoS One, 2011. **6**(6): p. e21313.
46. Jost, T., et al., *Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding.* Environ Microbiol, 2014. **16**(9): p. 2891-904.
47. Kiyohara, M., et al., *An exo-alpha-sialidase from bifidobacteria involved in the degradation of sialyloligosaccharides in human milk and intestinal glycoconjugates.* Glycobiology, 2011. **21**(4): p. 437-47.
48. Ward, R.E., et al., *In vitro fermentability of human milk oligosaccharides by several strains of bifidobacteria.* Mol Nutr Food Res, 2007. **51**(11): p. 1398-405.

Supplementary Figures and Tables

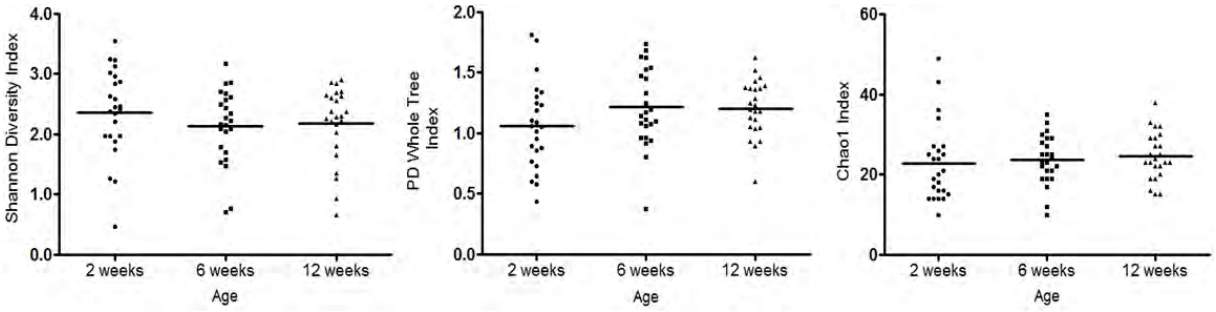
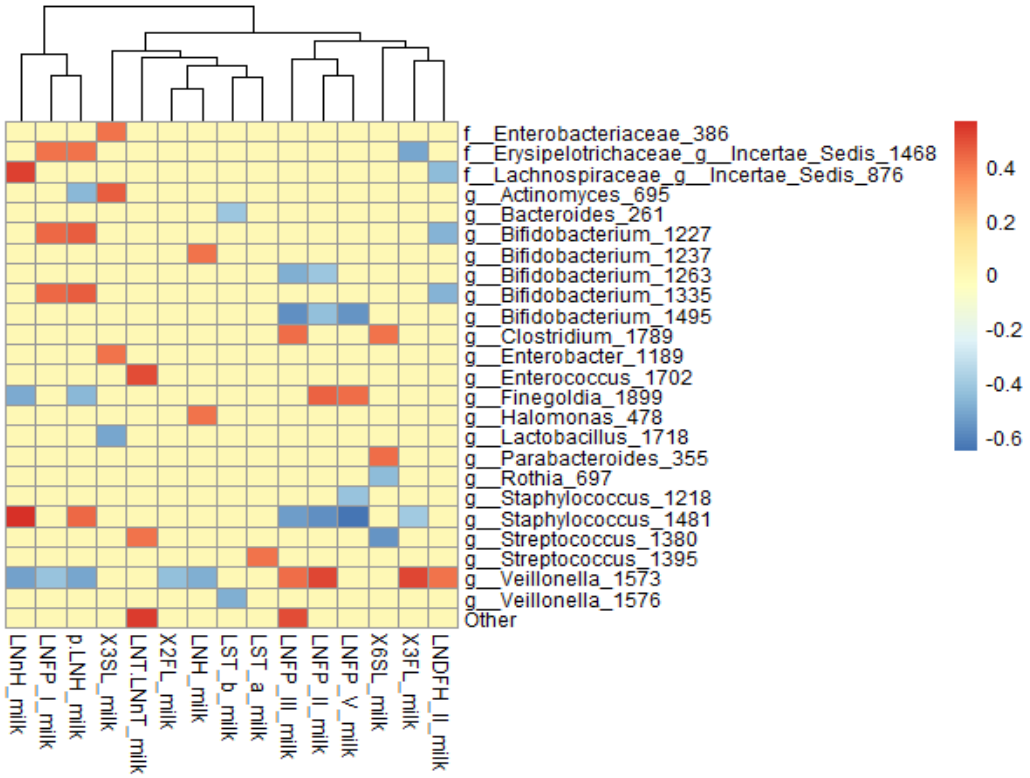
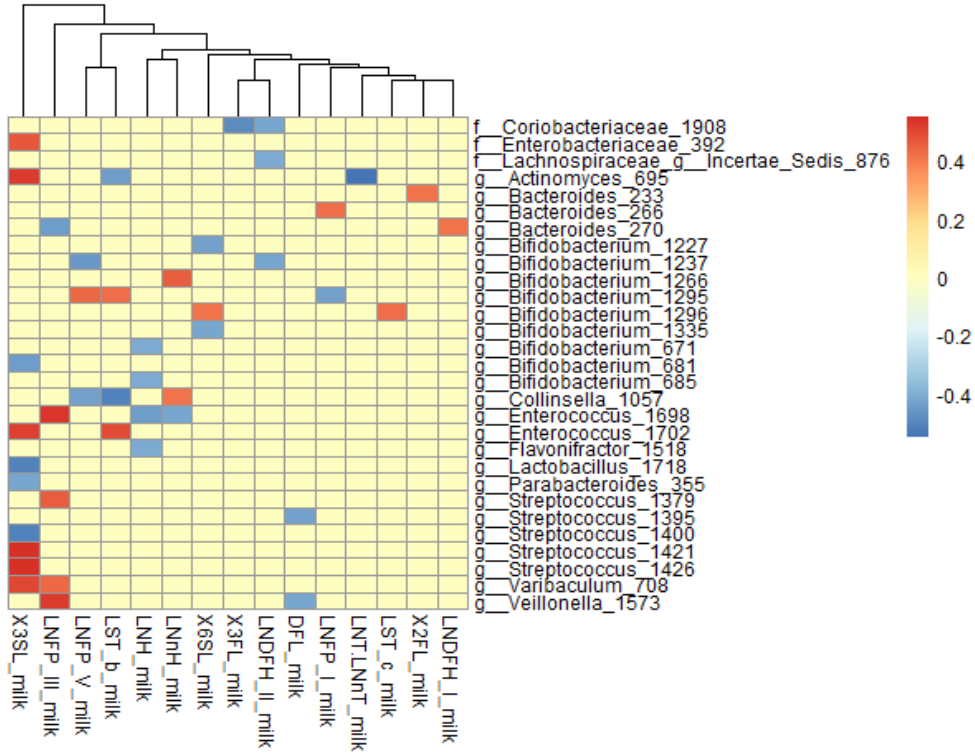


Figure S1. Alpha diversity indexes at different time points. Diversity was estimated using Shannon diversity index, which evaluates both the number of species and the evenness of their distribution. Richness, which estimates the number of different species present in each sample, was measured with the Chao1 Index and with PD Whole Tree, the latter of which takes into account phylogenetic differences between species (OTUs). No significant differences were found between any of the age groups.

a.



b.



c.

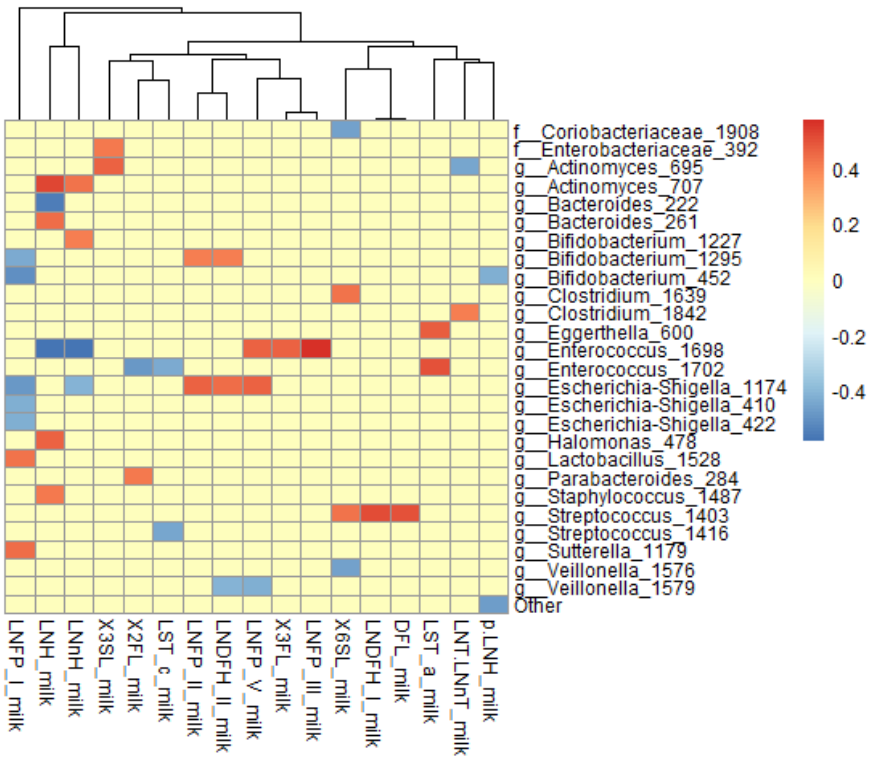
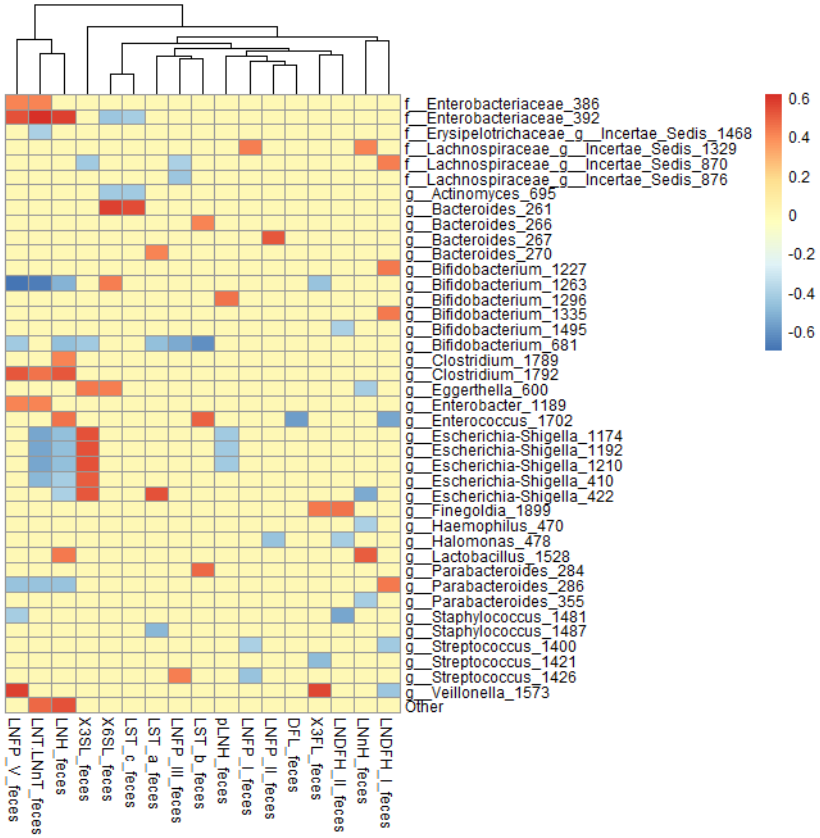
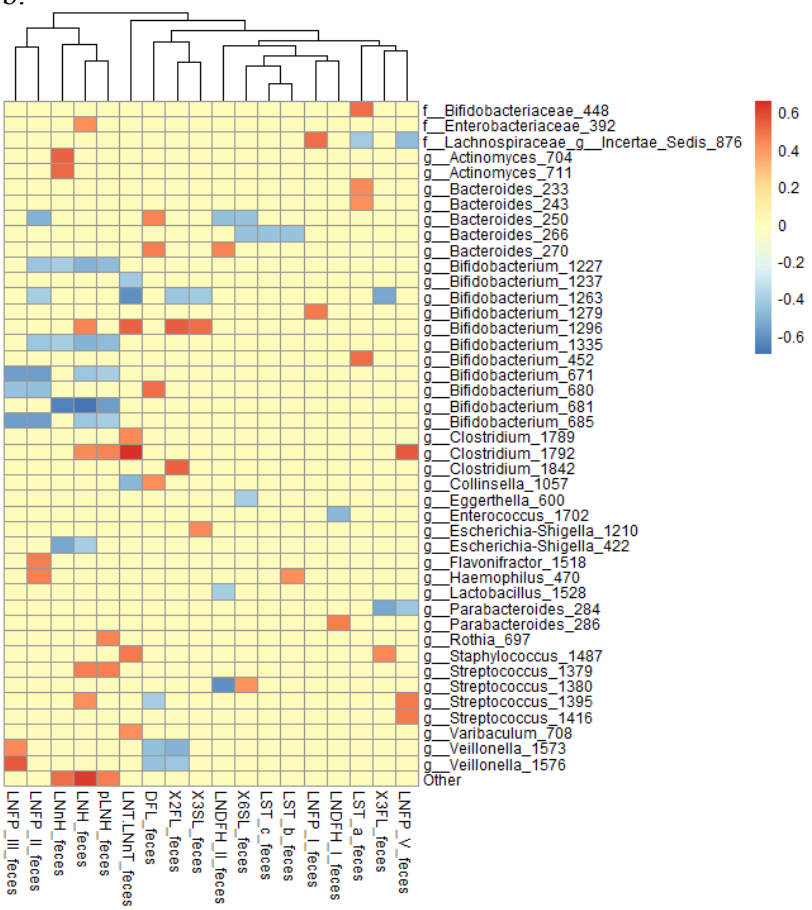


Figure S2. Statistically significant ($p<0.05$) Spearman correlations (correlation threshold ±0.3) between an estimated daily intake of different breastmilk HMOs and faecal microbiota composition at OTU level of 24 infants at two (a), six (b) and 12 (c) weeks after birth. Positive associations are indicated in red, negative in blue, yellow denotes correlations that did not pass the significance or the correlation thresholds.

a.


f__Enterobacteriaceae_386
f__Enterobacteriaceae_392
f__Erysipelotrichaceae_g__Incertae_Sedis_1468
f__Lachnospiraceae_g__Incertae_Sedis_1329
f__Lachnospiraceae_g__Incertae_Sedis_870
f__Lachnospiraceae_g__Incertae_Sedis_876
g__Actinomyces_695
g__Bacteroides_261
g__Bacteroides_266
g__Bacteroides_267
g__Bacteroides_270
g__Bifidobacterium_1227
g__Bifidobacterium_1263
g__Bifidobacterium_1296
g__Bifidobacterium_1335
g__Bifidobacterium_1495
g__Bifidobacterium_681
g__Clostridium_1789
g__Clostridium_1792
g__Eggerthella_600
g__Enterobacter_1189
g__Enterococcus_1702
g__Escherichia-Shigella_1174
g__Escherichia-Shigella_1192
g__Escherichia-Shigella_1210
g__Escherichia-Shigella_410
g__Escherichia-Shigella_422
g__Finegoldia_1899
g__Haemophilus_470
g__Halomonas_478
g__Lactobacillus_1528
g__Parabacteroides_284
g__Parabacteroides_286
g__Parabacteroides_355
g__Staphylococcus_1481
g__Staphylococcus_1487
g__Streptococcus_1400
g__Streptococcus_1421
g__Streptococcus_1426
g__Veillonella_1573
Other
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
LNFP_V_feces
LNT.LNnT_feces
LNH_feces
X3SL_feces
X6SL_feces
LST_c_feces
LST_a_feces
LNFP_III_feces
LST_b_feces
pLNH_feces
LNFP_I_feces
LNFP_II_feces
DFL_feces
X3FL_feces
LNDFH_II_feces
LNnH_feces
LNDFH_I_feces

b.


f__Bifidobacteriaceae_448
f__Enterobacteriaceae_392
f__Lachnospiraceae_g__Incertae_Sedis_876
g__Actinomyces_704
g__Actinomyces_711
g__Bacteroides_233
g__Bacteroides_243
g__Bacteroides_250
g__Bacteroides_266
g__Bacteroides_270
g__Bifidobacterium_1227
g__Bifidobacterium_1237
g__Bifidobacterium_1263
g__Bifidobacterium_1279
g__Bifidobacterium_1296
g__Bifidobacterium_1335
g__Bifidobacterium_452
g__Bifidobacterium_671
g__Bifidobacterium_680
g__Bifidobacterium_681
g__Bifidobacterium_685
g__Clostridium_1789
g__Clostridium_1792
g__Clostridium_1842
g__Collinsella_1057
g__Eggerthella_600
g__Enterococcus_1702
g__Escherichia-Shigella_1210
g__Escherichia-Shigella_422
g__Flavonifractor_1518
g__Haemophilus_470
g__Lactobacillus_1528
g__Parabacteroides_284
g__Parabacteroides_286
g__Rothia_697
g__Staphylococcus_1487
g__Streptococcus_1379
g__Streptococcus_1380
g__Streptococcus_1395
g__Streptococcus_1416
g__Varibaculum_708
g__Veillonella_1573
g__Veillonella_1576
Other
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
LNFP_III_feces
LNFP_II_feces
LNnH_feces
LNH_feces
pLNH_feces
LNT.LNnT_feces
DFL_feces
X2FL_feces
X3SL_feces
LNDFH_II_feces
X6SL_feces
LST_c_feces
LST_b_feces
LNFP_I_feces
LNDFH_I_feces
LST_a_feces
X3FL_feces
LNFP_V_feces

c.

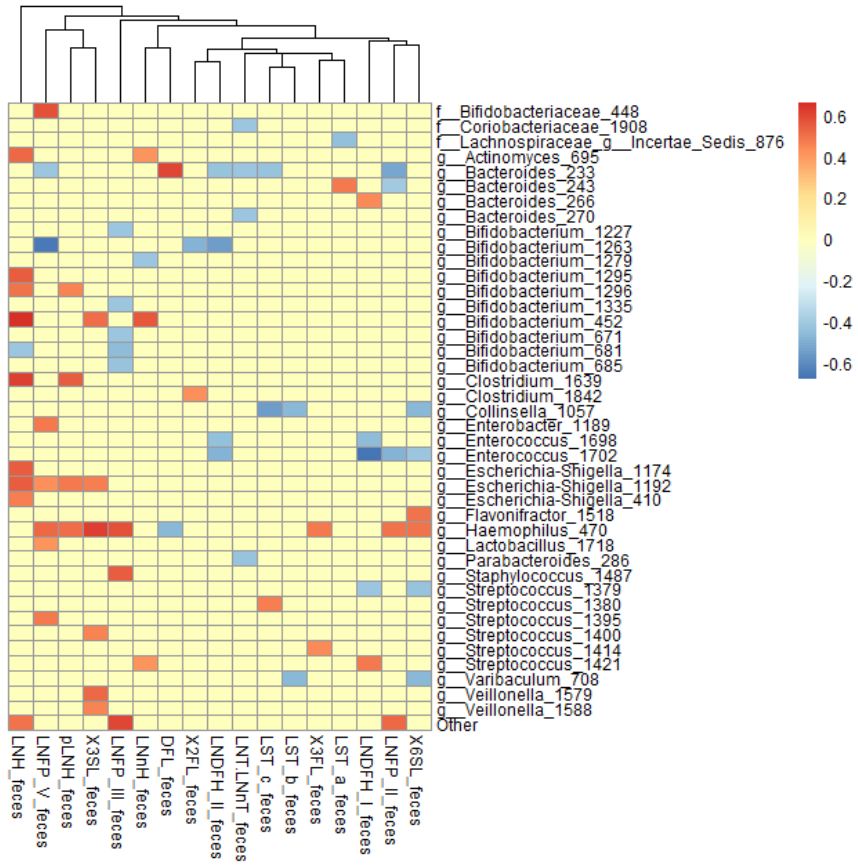


Figure S3. Statistically significant ($p<0.05$) Spearman correlations (correlation threshold ±0.3) between different HMOs detected in infant faeces and faecal microbiota composition at OTU level of 24 infants at two (a), six (b) and 12 (c) weeks after birth. Positive associations are indicated in red, negative in blue, and correlations that did not pass the significance or the correlation thresholds are marked in yellow.

Table S1. HMO categories, names and abbreviations included in this study

Category	Name	Abbreviation
Neutral	2′-Fucosyllactose	2′FL
	3′-Fucosyllactose	3FL
	Lacto-*N*-tetraose	LNT
	Lacto-*N*-neotetraose	LNnT
	Lacto-*N*-fucopentaose I	LNFPI
	Lacto-*N*-fucopentaose II	LNFPII
	Lacto-*N*-fucopentaose III	LNFPIII
	Lacto-*N*-fucopentaose V	LNFPV
	Difucosyllactose	DFL
	Lacto-*N*-difucohexaose I	LNDFHI
	Lacto-*N*-hexaose	LNH
	Lacto-*N*-neo-hexaose	LNnH
Acidic	6'-Sialyllactose	6'SL
	3′-Sialyllactose	3'SL
	Sialyl-lacto-*N*-tetraose a	LSTc
	Sialyl-lacto-*N*-tetraose b	LSTb
	Sialyl-lacto-*N*-tetraose c	LSTa

Table S2. Average relative abundance (Avg RA), standard error of abundance means (SEM), and OTU prevalence of OTUs (n=411) at two, six and 12 weeks of age. Only OTUs which were present in at least 5% of all samples are shown

Genus	OTU #	Week two (n=24)			Week six (n=24)			Week 12 (n=24)			Total (n=72)		
		Avg RA (%)	± SEM	Prevalence (%)	Avg RA (%)	± SEM	Prevalence (%)	Avg RA (%)	± SEM	Prevalence (%)	Avg RA (%)	± SEM	Prevalence (%)
Actinomyces	695	0.89	0.88	8.33	2.17	2.16	12.5	0.17	0.16	8.33	1.08	0.77	9.72
	704	0.3	0.28	12.5	0.06	0.04	12.5	0.01	0.01	4.17	0.12	0.09	9.72
	707	0	0	0	0.01	0.01	4.17	0.07	0.03	20.83	0.03	0.01	8.33
	711	0	0	0	0.07	0.05	8.33	0.12	0.08	16.67	0.06	0.03	8.33
Aeribacillus	1603	0.01	0.01	4.17	0.02	0.01	12.5	0.01	0.01	4.17	0.01	0	6.94
Bacteroides	222	2.67	1.17	33.33	7.97	3.24	50	5.48	2.16	50	5.37	1.36	44.44
	233	0.09	0.08	8.33	0.04	0.02	16.67	0.1	0.04	25	0.07	0.03	16.67
	243	0.03	0.03	8.33	0.07	0.05	8.33	0.39	0.2	20.83	0.16	0.07	12.5
	250	0.05	0.04	8.33	0.07	0.05	12.5	0.23	0.12	16.67	0.12	0.04	12.5
	261	1.31	1.29	12.5	0.07	0.07	4.17	0.04	0.02	12.5	0.47	0.43	9.72
	266	0.36	0.28	8.33	0.08	0.05	12.5	0.02	0.01	8.33	0.15	0.1	9.72
	267	0.5	0.4	12.5	0.27	0.19	16.67	0.1	0.08	12.5	0.29	0.15	13.89
	270	0.32	0.29	12.5	0.08	0.04	16.67	0.07	0.04	12.5	0.16	0.1	13.89
Bifidobacterium	1227	0.11	0.08	8.33	0.14	0.07	20.83	0.28	0.1	37.5	0.17	0.05	22.22
	1237	0.13	0.09	8.33	0.17	0.1	16.67	0.21	0.14	25	0.17	0.06	16.67
	1263	22.23	5.39	83.33	35.1	6.23	91.67	51.26	4.96	100	36.2	3.46	91.67
	1265	0.04	0.02	12.5	0.01	0.01	4.17	0.01	0.01	4.17	0.02	0.01	6.94

	1266	0.01	0.01	4.17	0.02	0.01	8.33	0.03	0.02	8.33	0.02	0.01	6.94
	1279	11.92	3.18	58.33	11.93	3.82	79.17	6.01	2.86	79.17	9.95	1.91	72.22
	1295	0.05	0.03	16.67	0.06	0.03	20.83	0.07	0.03	20.83	0.06	0.02	19.44
	1296	0.07	0.04	16.67	0.05	0.03	12.5	0.02	0.02	4.17	0.05	0.02	11.11
	1335	0.09	0.06	8.33	0.12	0.06	20.83	0.24	0.08	37.5	0.15	0.04	22.22
	1495	0.06	0.03	16.67	0	0	0	0	0	0	0.02	0.01	5.56
	452	0.02	0.02	4.17	0.02	0.01	8.33	0.05	0.02	20.83	0.03	0.01	11.11
	671	0.03	0.03	4.17	0.08	0.04	16.67	0.24	0.08	37.5	0.12	0.03	19.44
	680	0.01	0.01	4.17	0.02	0.01	12.5	0.09	0.04	25	0.04	0.01	13.89
	681	5.92	3.51	16.67	9.46	4.41	41.67	6.48	2.38	58.33	7.29	2.02	38.89
	685	0.03	0.03	4.17	0.08	0.04	16.67	0.28	0.1	37.5	0.13	0.04	19.44
Clostridium	1629	0.02	0.02	4.17	0.52	0.52	4.17	0.18	0.18	4.17	0.24	0.18	4.17
	1639	0.03	0.03	4.17	0.04	0.04	4.17	0.28	0.2	8.33	0.12	0.07	5.56
	1789	0.06	0.03	12.5	0.02	0.01	8.33	0	0	0	0.03	0.01	6.94
	1792	1.87	0.84	20.83	0.79	0.37	45.83	0.32	0.15	25	0.99	0.31	30.56
	1842	0	0	0	0.08	0.07	12.5	0.08	0.07	8.33	0.06	0.03	6.94
Collinsella	1057	0.05	0.05	4.17	0.09	0.06	12.5	0.12	0.07	16.67	0.09	0.03	11.11
Eggerthella	600	0.09	0.05	12.5	0.14	0.08	20.83	0.09	0.06	12.5	0.1	0.04	15.28
Enterobacter	1189	0.02	0.02	8.33	0.01	0.01	4.17	0.02	0.01	8.33	0.02	0.01	6.94
Enterococcus	1698	0.01	0.01	4.17	0.15	0.1	12.5	0.18	0.1	20.83	0.11	0.05	12.5
	1702	0.48	0.25	25	0.77	0.35	37.5	0.84	0.3	54.17	0.7	0.17	38.89
Escherichia-Shigella	1174	0.04	0.02	16.67	0.02	0.01	12.5	0.06	0.03	20.83	0.04	0.01	16.67
	1192	0.04	0.02	16.67	0.01	0.01	4.17	0.02	0.02	8.33	0.02	0.01	9.72
	1210	0.04	0.02	16.67	0.03	0.02	16.67	0.05	0.02	16.67	0.04	0.01	16.67
	410	0.12	0.05	29.17	0.08	0.03	29.17	0.17	0.05	50	0.12	0.02	36.11
	422	8.45	3.68	66.67	4.42	1.58	58.33	5.09	1.46	75	5.99	1.42	66.67
Finegoldia	1899	0.06	0.05	12.5	0.06	0.06	4.17	0.01	0.01	4.17	0.04	0.03	6.94
Flavonifractor	1518	0	0	0	0.02	0.01	8.33	0.04	0.03	8.33	0.02	0.01	5.56
Gemella	1606	0.01	0.01	4.17	0.01	0.01	4.17	0.02	0.01	8.33	0.01	0.01	5.56
Haemophilus	470	0.18	0.08	25	0.32	0.15	37.5	0.7	0.41	41.67	0.4	0.15	34.72
Halomonas	478	0.01	0.01	8.33	0.03	0.02	12.5	0.01	0.01	8.33	0.02	0.01	9.72
Lactobacillus	1205	0	0	0	0.01	0.01	4.17	0.04	0.02	16.67	0.01	0.01	6.94
	1334	0.01	0.01	4.17	0.01	0.01	4.17	0.03	0.02	12.5	0.02	0.01	6.94
	1528	0.55	0.31	20.83	1.28	0.41	41.67	1.7	0.51	50	1.18	0.25	37.5
	1718	0.46	0.32	12.5	0.71	0.34	25	0.04	0.03	8.33	0.4	0.16	15.28
Parabacteroides	284	0.19	0.14	12.5	0.12	0.09	16.67	0.04	0.03	8.33	0.12	0.06	12.5
	286	0.34	0.17	29.17	0.72	0.31	29.17	0.47	0.28	29.17	0.51	0.15	29.17
	355	0.07	0.05	12.5	0.08	0.06	8.33	0.14	0.08	12.5	0.1	0.04	11.11
Rothia	697	0.02	0.01	8.33	0.04	0.02	20.83	0.06	0.03	20.83	0.04	0.01	16.67
Staphylococcus	1218	0.28	0.08	45.83	0.01	0.01	4.17	0	0	0	0.1	0.03	16.67
	1481	0.09	0.03	33.33	0	0	0	0	0	0	0.03	0.01	11.11
	1487	6.45	1.08	91.67	0.54	0.13	70.83	0.32	0.13	54.17	2.44	0.49	72.22
Streptococcus	1379	0.05	0.05	4.17	0.02	0.02	8.33	0.22	0.09	25	0.1	0.04	12.5
	1380	0.33	0.1	50	0.13	0.05	33.33	0.25	0.11	37.5	0.24	0.05	40.28

	1395	1.18	1.13	8.33	0.7	0.57	8.33	0.03	0.02	8.33	0.64	0.42	8.33
	1400	0.11	0.04	33.33	0.17	0.06	45.83	0.35	0.08	62.5	0.21	0.04	47.22
	1403	0.22	0.17	12.5	0.05	0.05	4.17	0.02	0.01	12.5	0.1	0.06	9.72
	1414	0.12	0.12	4.17	0.24	0.18	8.33	0.2	0.17	8.33	0.19	0.09	6.94
	1416	5.22	1.4	87.5	4.18	1.65	83.33	1.82	0.61	62.5	3.74	0.76	77.78
	1421	0.03	0.01	16.67	0.03	0.01	12.5	0.01	0.01	8.33	0.02	0.01	12.5
	1426	0.07	0.02	25	0.03	0.02	12.5	0.01	0.01	4.17	0.03	0.01	13.89
Sutterella	1179	0.12	0.12	4.17	0.08	0.05	12.5	0.41	0.35	8.33	0.2	0.12	8.33
Varibaculum	708	0	0	0	0.18	0.13	16.67	0.12	0.06	16.67	0.1	0.05	11.11
Veillonella	1573	0.61	0.44	25	0.18	0.12	20.83	0.81	0.35	45.83	0.53	0.19	30.56
	1576	1.04	0.61	37.5	0.12	0.06	16.67	0.22	0.11	29.17	0.46	0.21	27.78
	1578	0.11	0.1	16.67	0.01	0.01	4.17	0	0	0	0.04	0.03	6.94
	1579	0.04	0.04	4.17	0	0	4.17	0.03	0.01	16.67	0.03	0.02	8.33
	1588	1.74	0.7	54.17	0.65	0.36	54.17	0.95	0.48	41.67	1.11	0.31	50
f_Bifidobacteriaceae	448	0.01	0.01	4.17	0.02	0.01	8.33	0.02	0.01	12.5	0.02	0.01	8.33
f_Coriobacteriaceae	1908	0.02	0.02	4.17	0.08	0.04	12.5	0.13	0.07	16.67	0.08	0.03	11.11
f_Enterobacteriaceae	386	0.03	0.02	8.33	0.04	0.03	8.33	0.09	0.04	29.17	0.05	0.02	15.28
	392	6.02	3.97	29.17	2.45	1.6	37.5	3.07	1.59	37.5	3.85	1.51	34.72
f_Erysipelotrichaceae *Incertae_Sedis*	1468	0.13	0.1	12.5	0.08	0.08	4.17	0.06	0.04	8.33	0.09	0.05	8.33
f_Lachnospiraceae *Incertae_Sedis*	1329	0.05	0.03	12.5	0.03	0.02	12.5	0.02	0.02	8.33	0.03	0.01	11.11
	870	0.15	0.08	16.67	0.07	0.04	12.5	0.03	0.02	12.5	0.08	0.03	13.89
	876	3.21	1.33	33.33	2.02	0.9	25	1.25	0.64	20.83	2.16	0.58	26.39
Other OTUs (n=328)		12.07	3.94	95.83	9.09	2.44	95.83	6.71	2.18	95.83	9.29	1.7	95.83

Table S3. Differentially abundant OTUs in faecal samples of 24 infants at two, six and 12 weeks of age associated with high and low level of HMO consumption (Kruskal-Wallis; $p<0.05$, FDR<0.05 where indicated with*). The higher values of the average relative abundances (RA) are marked in bold.

HMO	OTU	Test-Statistic	p	FDR	High consumption group RA	Low consumption group RA
2′FL	*g_Bacteroides_261*	5.60	0.018	0.223	**0.002**	0
	*g_Bifidobacterium_1263**	11.16	0.001	0.036	**0.549**	0.198
	*g_Bifidobacterium_681**	16.59	0.000	0.004	**0.094**	0
	g_Bifidobacterium_1227	5.60	0.018	0.223	**0.001**	0
	g_Bifidobacterium_1335	5.60	0.018	0.223	**0.001**	0
	g_Bifidobacterium_671	4.36	0.037	0.247	**0.001**	0
	g_Bifidobacterium_685	4.36	0.037	0.247	**0.001**	0
	g_Bifidobacterium_1296	5.60	0.018	0.223	0	**0.001**
	g_Clostridium_1842	4.36	0.037	0.247	0	**0.001**
	g_Escherichia-Shigella_1210	4.63	0.031	0.247	0.000	**0.001**
	g_Escherichia-Shigella_410	4.06	0.044	0.255	0.001	**0.002**
	g_Halomonas_478	5.60	0.018	0.223	**0.000**	0

	g_Streptococcus_1416	4.09	0.043	0.255	0.016	**0.047**
	g_Veillonella_1588	4.84	0.028	0.247	0.006	**0.022**
	Other	4.63	0.031	0.247	0.045	**0.094**
3FL	*g_Bifidobacterium_1263**	12.04	0.001	0.023	**0.529**	0.235
	g_Bifidobacterium_685	7.81	0.005	0.124	**0.002**	0.000
	g_Bifidobacterium_1227	4.94	0.026	0.208	**0.002**	0.001
	g_Bifidobacterium_671	7.65	0.006	0.124	**0.002**	0.000
	g_Bifidobacterium_1335	5.07	0.024	0.208	**0.002**	0.001
	g_Bifidobacterium_680	6.41	0.011	0.165	**0.001**	0
	g_Bifidobacterium_452	4.09	0.043	0.215	**0.001**	0
	g_Bifidobacterium_1296	4.45	0.035	0.215	0	**0.001**
	g_Enterococcus_1698	5.23	0.022	0.208	**0.002**	0
	*g_Enterococcus_1702**	12.10	0.001	0.023	**0.017**	0.002
	g_Finegoldia_1899	4.45	0.035	0.215	0	**0.001**
	g_Lactobacillus_1528	5.45	0.020	0.208	**0.018**	0.006
	g_Parabacteroides_284	4.11	0.043	0.215	**0.003**	0.000
	g_Staphylococcus_1218	4.04	0.044	0.215	0.000	**0.001**
	g_Staphylococcus_1487	6.41	0.011	0.165	0.010	**0.034**
	g_Sutterella_1179	4.09	0.043	0.215	**0.006**	0
	g_Varibaculum_708	4.09	0.043	0.215	**0.001**	0
	g_Veillonella_1588	5.60	0.018	0.208	0.003	**0.016**
3′SL	f_Lachnospiraceae *Incertae_Sedis_1329*	4.85	0.028	0.151	**0.001**	0
	g_Bifidobacterium_1263	4.93	0.026	0.151	**0.418**	0.267
	*g_Bifidobacterium_681**	10.87	0.001	0.043	**0.104**	0.028
	g_Bifidobacterium_1227	7.87	0.005	0.055	**0.003**	0.000
	g_Bifidobacterium_1335	7.87	0.005	0.055	**0.002**	0.000
	g_Bifidobacterium_685	6.33	0.012	0.084	**0.002**	0.000
	g_Bifidobacterium_671	6.24	0.013	0.084	**0.002**	0.000
	g_Bifidobacterium_680	6.25	0.012	0.084	**0.001**	0
	g_Bifidobacterium_452	4.28	0.039	0.198	0.000	**0.001**
	g_Escherichia-Shigella_422	8.67	0.003	0.055	0.021	**0.129**
	g_Escherichia-Shigella_410	6.98	0.008	0.080	0.001	**0.002**
	g_Escherichia-Shigella_1174	9.13	0.003	0.055	0.000	**0.001**
	g_Escherichia-Shigella_1210	8.19	0.004	0.055	0.000	**0.001**
	*g_Escherichia-Shigella_1192**	11.85	0.001	0.043	0	**0.001**
	g_Escherichia-Shigella_382	5.67	0.017	0.107	0	**0.000**
	g_Staphylococcus_1487	3.95	0.047	0.214	**0.026**	0.025
	g_Varibaculum_708	4.18	0.041	0.198	**0.001**	0
	g_Veillonella_1579	6.47	0.011	0.084	0.00	**0.001**
	g_Veillonella_1588	8.10	0.004	0.055	0.004	**0.022**
6′SL	*g_Bacteroides_250*	4.20	0.040	0.440	**0.002**	0.000
	g_Bifidobacterium_1495	4.26	0.039	0.440	0	**0.001**
	g_Collinsella_1057	3.79	0.051	0.497	**0.002**	0.001
	g_Haemophilus_470	4.81	0.028	0.410	0.001	**0.007**
	g_Staphylococcus_1218	5.45	0.020	0.410	0.000	**0.002**

	g_Staphylococcus_1481	5.16	0.023	0.410	0.000	**0.001**
	g_Streptococcus_1379	5.01	0.025	0.410	**0.002**	0.000
	g_Streptococcus_1380	7.72	0.005	0.410	0.001	**0.003**
	g_Varibaculum_708	6.68	0.010	0.410	**0.002**	0
DFL	f_Coriobacteriaceae*_1908*	7.14	0.008	0.425	0	**0.002**
	g_Bacteroides_233	3.85	0.050	0.425	0.000	**0.001**
	g_Bacteroides_250	3.96	0.047	0.425	0.001	**0.002**
	g_Bacteroides_266	4.12	0.042	0.425	0.001	**0.004**
	g_Collinsella_1057	4.96	0.026	0.425	0.0	**0.003**
	g_Eggerthella_600	4.45	0.035	0.425	0.000	**0.002**
	g_Haemophilus_470	4.22	0.040	0.425	**0.008**	0.001
	g_Varibaculum_708	5.80	0.016	0.425	0	**0.001**
	g_Veillonella_1573	4.45	0.035	0.425	**0.012**	0.001
LNDFHI	f_Enterobacteriaceae*_386*	4.26	0.039	0.339	**0.001**	0
	g_Actinomyces_707	4.26	0.039	0.339	**0.001**	0
	g_Bifidobacterium_671	4.77	0.029	0.339	**0.002**	0.001
	g_Bifidobacterium_685	4.89	0.027	0.339	**0.002**	0.001
	g_Enterococcus_1702	10.79	0.001	0.089	**0.015**	0.003
	g_Flavonifractor_1518	4.26	0.039	0.339	0	**0.001**
	g_Parabacteroides_286	6.41	0.011	0.328	0.001	**0.011**
	g_Streptococcus_1416	5.76	0.016	0.339	0.012	**0.049**
	g_Streptococcus_1421	6.68	0.010	0.328	0	**0.000**
	g_Streptococcus_1426	5.44	0.020	0.339	0	**0.001**
LNDFHII	*g_Actinomyces_695*	4.27	0.039	0.476	0	**0.009**
	g_Bifidobacterium_1263	6.33	0.012	0.258	**0.516**	0.318
	g_Enterococcus_1698	6.71	0.010	0.258	**0.003**	0
	g_Enterococcus_1702	4.40	0.036	0.476	**0.016**	0.003
	g_Halomonas_478	3.81	0.051	0.476	**0.000**	0.000
	g_Lactobacillus_1528	9.53	0.002	0.088	**0.019**	0.005
	g_Streptococcus_1400	9.66	0.002	0.088	**0.003**	0.001
LNFPI	*g_Actinomyces_707*	4.27	0.039	0.259	**0.001**	0
	g_Bacteroides_233	4.75	0.029	0.232	**0.001**	0.001
	g_Bifidobacterium_1263	7.02	0.008	0.117	**0.450**	0.221
	*g_Bifidobacterium_681**	21.93	0.000	0.000	**0.162**	0.004
	g_Bifidobacterium_1227	9.06	0.003	0.064	**0.003**	0.000
	g_Bifidobacterium_1335	9.06	0.003	0.064	**0.003**	0.000
	g_Bifidobacterium_685	6.20	0.013	0.124	**0.002**	0.000
	g_Bifidobacterium_671	6.20	0.013	0.124	**0.002**	0.000
	g_Bifidobacterium_1296	6.71	0.010	0.119	0	**0.001**
	g_Clostridium_1792	4.28	0.039	0.259	0.002	**0.027**
	g_Streptococcus_1400	8.84	0.003	0.064	**0.003**	0.001
	g_Sutterella_1179	5.46	0.019	0.169	**0.006**	0
	Other	7.60	0.006	0.101	0.073	**0.120**
LNFPII	f_Enterobacteriaceae*_386*	3.93	0.048	0.180	**0.001**	0.000
	f_Lachnospiraceae *Incertae_Sedis_870*	4.12	0.042	0.174	0.000	**0.002**

	g_Actinomyces_707	6.68	0.010	0.083	**0.001**	0
	g_Bacteroides_233	5.01	0.025	0.146	**0.001**	0.000
	g_Bacteroides_243	6.68	0.010	0.083	**0.003**	0
	g_Bacteroides_250	6.68	0.010	0.083	**0.002**	0
	g_Bifidobacterium_1227	5.16	0.023	0.143	**0.003**	0.002
	g_Bifidobacterium_1263	9.57	0.002	0.033	**0.546**	0.274
	g_Bifidobacterium_1335	5.16	0.023	0.143	**0.002**	0.001
	g_Bifidobacterium_671	11.03	0.001	0.026	**0.003**	0.000
	g_Bifidobacterium_680	9.30	0.002	0.033	**0.001**	0
	g_Bifidobacterium_681	6.55	0.011	0.083	0.084	**0.085**
	g_Bifidobacterium_685	11.21	0.001	0.026	**0.003**	0.000
	g_Collinsella_1057	4.26	0.039	0.169	**0.001**	0
	g_Enterococcus_1698	4.48	0.034	0.169	**0.003**	0.000
	g_Enterococcus_1702	4.28	0.039	0.169	**0.013**	0.005
	g_Haemophilus_470	7.18	0.007	0.083	0.000	**0.008**
	g_Parabacteroides_284	4.06	0.044	0.174	0.000	**0.002**
	g_Staphylococcus_1218	9.30	0.002	0.033	0	**0.002**
	g_Staphylococcus_1481	5.44	0.020	0.142	0	**0.001**
	g_Staphylococcus_1487	11.51	0.001	0.026	0.011	**0.032**
	g_Varibaculum_708	4.48	0.034	0.169	**0.002**	0.000
	Other	4.34	0.037	0.169	0.034	**0.093**
LSTc	*g_Actinomyces_695*	6.68	0.010	0.170	**0.032**	0
	g_Clostridium_1792	6.79	0.009	0.170	**0.011**	0.001
	g_Collinsella_1057	6.68	0.010	0.170	**0.002**	0
	g_Staphylococcus_1218	6.68	0.010	0.170	0	**0.002**
	g_Streptococcus_1380	11.90	0.001	0.049	0.000	**0.003**
	g_Streptococcus_1395	4.26	0.039	0.445	**0.008**	0
	g_Streptococcus_1403	4.26	0.039	0.445	0	**0.001**
LNFPIII	*g_Actinomyces_707*	5.44	0.020	0.124	**0.001**	0
	g_Bifidobacterium_1263	4.97	0.026	0.150	**0.504**	0.313
	*g_Bifidobacterium_681**	10.72	0.001	0.023	**0.120**	0.034
	*g_Bifidobacterium_685**	12.84	0.000	0.016	**0.003**	0.000
	*g_Bifidobacterium_1227**	9.18	0.002	0.030	**0.003**	0.001
	*g_Bifidobacterium_671**	12.65	0.000	0.016	**0.003**	0.000
	*g_Bifidobacterium_1335**	9.34	0.002	0.030	**0.003**	0.001
	*g_Bifidobacterium_680**	10.69	0.001	0.023	**0.001**	0
	g_Enterococcus_1702	5.41	0.020	0.124	**0.015**	0.005
	g_Escherichia-Shigella_1192	5.44	0.020	0.124	0	**0.001**
	g_Escherichia-Shigella_422	6.04	0.014	0.124	0.022	**0.116**
	g_Halomonas_478	4.26	0.039	0.212	**0.000**	0
	g_Parabacteroides_355	5.60	0.018	0.124	0.000	**0.003**
	*g_Staphylococcus_1487**	10.07	0.002	0.026	0.015	**0.028**
	g_Streptococcus_1426	5.75	0.016	0.124	0.000	**0.001**
	g_Varibaculum_708	5.44	0.020	0.124	**0.001**	0
LNFPV	f_Enterobacteriaceae_392	4.43	0.035	0.171	0.008	**0.101**

	f_Lachnospiraceae *Incertae_Sedis_876*	7.31	0.007	0.090	**0.034**	0.006
	g_Bacteroides_222	8.57	0.003	0.074	**0.054**	0.007
	g_Bacteroides_261	4.53	0.033	0.170	**0.011**	0
	g_Bacteroides_267	4.14	0.042	0.183	**0.006**	0.000
	*g_Bifidobacterium_1263**	18.02	0.000	0.001	**0.475**	0.156
	g_Bifidobacterium_1296	5.17	0.023	0.143	0	**0.001**
	g_Clostridium_1639	4.53	0.033	0.170	0	**0.002**
	g_Clostridium_1789	7.84	0.005	0.089	0	**0.001**
	*g_Clostridium_1792**	18.33	0.000	0.001	0.001	**0.028**
	g_Collinsella_1057	5.39	0.020	0.135	**0.002**	0
	g_Eggerthella_600	4.93	0.026	0.153	**0.002**	0.000
	g_Enterobacter_1189	6.16	0.013	0.104	0	**0.001**
	g_Parabacteroides_284	7.21	0.007	0.090	**0.003**	0
	g_Parabacteroides_286	6.61	0.010	0.098	**0.009**	0.001
	g_Staphylococcus_1487	9.09	0.003	0.074	0.013	**0.032**
	g_Streptococcus_1395	6.16	0.013	0.104	0	**0.002**
	g_Streptococcus_1416	6.63	0.010	0.098	0.031	**0.055**
	g_Veillonella_1573	4.30	0.038	0.175	0.002	**0.014**
	g_Veillonella_1588	5.71	0.017	0.122	0.004	**0.024**
LNH	f_Coriobacteriaceae*_1908*	3.90	0.048	0.175	**0.001**	0
	f_Enterobacteriaceae*_392*	4.67	0.031	0.145	0.025	**0.069**
	g_Actinomyces_695	6.99	0.008	0.061	0	**0.032**
	g_Actinomyces_704	5.89	0.015	0.095	0.000	**0.003**
	g_Actinomyces_707	3.90	0.048	0.175	**0.000**	0
	g_Bacteroides_222	3.90	0.048	0.175	**0.055**	0.044
	g_Bacteroides_261	4.62	0.032	0.145	**0.009**	0
	*g_Bifidobacterium_1263**	20.45	0.000	0.000	**0.480**	0.161
	*g_Bifidobacterium_681**	21.88	0.000	0.000	**0.116**	0.000
	*g_Bifidobacterium_1227**	12.26	0.000	0.007	**0.003**	0
	*g_Bifidobacterium_1335**	12.26	0.000	0.007	**0.003**	0
	*g_Bifidobacterium_685**	10.37	0.001	0.014	**0.002**	0
	*g_Bifidobacterium_671**	10.37	0.001	0.014	**0.002**	0
	g_Bifidobacterium_680	6.93	0.008	0.061	**0.001**	0
	*g_Bifidobacterium_1296**	12.84	0.000	0.007	0	**0.001**
	g_Clostridium_1639	5.16	0.023	0.134	0	**0.002**
	g_Clostridium_1789	8.88	0.003	0.025	0	**0.001**
	g_Clostridium_1792	9.23	0.002	0.023	0.002	**0.027**
	g_Collinsella_1057	3.90	0.048	0.175	**0.001**	0
	g_Parabacteroides_355	4.62	0.032	0.145	**0.002**	0
	g_Staphylococcus_1487	4.79	0.029	0.145	0.021	**0.037**
	g_Streptococcus_1395	6.02	0.014	0.094	0.000	**0.019**
	g_Veillonella_1588	3.94	0.047	0.175	0.006	**0.022**
	Other*	16.51	0.000	0.001	0.042	**0.165**
LNnH	*g_Actinomyces_704*	8.75	0.003	0.054	0.000	**0.004**
	g_Bifidobacterium_1263	6.51	0.011	0.085	**0.417**	0.241

	*g_Bifidobacterium_681**	15.41	0.000	0.008	**0.114**	0.005
	g_Bifidobacterium_1227	6.97	0.008	0.080	**0.003**	0.000
	g_Bifidobacterium_1335	7.06	0.008	0.080	**0.002**	0.000
	g_Bifidobacterium_685	5.43	0.020	0.115	**0.002**	0.000
	g_Bifidobacterium_671	5.52	0.019	0.115	**0.002**	0.000
	*g_Bifidobacterium_1296**	10.65	0.001	0.026	0.000	**0.001**
	g_Bifidobacterium_1495	7.84	0.005	0.069	0	**0.001**
	g_Clostridium_1629	5.79	0.016	0.115	0	**0.007**
	g_Clostridium_1789	5.03	0.025	0.136	0.000	**0.001**
	g_Escherichia-Shigella_410	4.75	0.029	0.150	**0.002**	0.001
	g_Escherichia-Shigella_422	7.69	0.006	0.069	**0.070**	0.039
	*g_Staphylococcus_1218**	13.02	0.000	0.013	0.000	**0.003**
	g_Staphylococcus_1481	6.67	0.010	0.085	0.000	**0.001**
	*g_Staphylococcus_1487**	10.50	0.001	0.026	0.013	**0.048**
	g_Streptococcus_1400	3.89	0.049	0.235	**0.003**	0.001
	Other	5.67	0.017	0.115	0.057	**0.146**
LSTa	f_Bifidobacteriaceae_*448*	6.65	0.010	0.189	0	**0.000**
	f_Lachnospiraceae *Incertae_Sedis_1329*	4.88	0.027	0.215	**0.001**	0
	f_Lachnospiraceae *Incertae_Sedis_870*	6.49	0.011	0.189	**0.001**	0
	f_Lachnospiraceae *Incertae_Sedis_876*	6.69	0.010	0.189	**0.037**	0.001
	g_Bacteroides_243	4.00	0.046	0.331	0.000	**0.002**
	g_Escherichia-Shigella_1174	5.46	0.020	0.215	0.000	**0.001**
	g_Escherichia-Shigella_1192	8.45	0.004	0.185	0	**0.001**
	g_Escherichia-Shigella_1210	5.25	0.022	0.215	0.000	**0.001**
	g_Escherichia-Shigella_382	4.91	0.027	0.215	0	**0.000**
	g_Escherichia-Shigella_422	5.42	0.020	0.215	0.022	**0.119**
	g_Halomonas_478	4.88	0.027	0.215	**0.000**	0
	g_Veillonella_1588	8.17	0.004	0.185	0.004	**0.018**
LSTb	*g_Actinomyces_707*	4.26	0.039	0.622	**0.001**	0
	g_Bifidobacterium_1296	5.44	0.020	0.427	0	**0.001**
	g_Bifidobacterium_681	5.84	0.016	0.427	**0.149**	0.045
	g_Collinsella_1057	6.68	0.010	0.427	**0.002**	0
	g_Staphylococcus_1218	6.47	0.011	0.427	0.000	**0.002**
pLNH	*g_Bacteroides_261*	4.49	0.034	0.198	0.009	0
	*g_Bifidobacterium_1263**	15.91	0.000	0.002	**0.457**	0.173
	*g_Bifidobacterium_681**	16.21	0.000	0.002	**0.118**	0.028
	*g_Bifidobacterium_1227**	10.16	0.001	0.022	**0.003**	0
	*g_Bifidobacterium_1335**	10.16	0.001	0.022	**0.003**	0
	*g_Bifidobacterium_685**	8.40	0.004	0.041	**0.002**	0
	*g_Bifidobacterium_671**	8.40	0.004	0.041	**0.002**	0
	g_Bifidobacterium_680	5.98	0.015	0.105	**0.001**	0
	*g_Bifidobacterium_1296**	15.57	0.000	0.002	0	**0.002**
	g_Clostridium_1639	7.27	0.007	0.068	0	**0.004**
	g_Clostridium_1789	5.36	0.021	0.128	0	**0.001**
	g_Clostridium_1792	6.68	0.010	0.077	0.003	**0.022**

	g_Eggerthella_600	6.76	0.009	0.077	**0.002**	0
	g_Rothia_697	5.64	0.018	0.117	0.000	**0.001**
	g_Staphylococcus_1487	3.98	0.046	0.225	0.020	**0.032**
	g_Veillonella_1588	3.97	0.046	0.225	0.006	**0.022**
	Other*	10.08	0.001	0.022	0.063	**0.136**
LNT and LNnT	f_Bifidobacteriaceae_*448*	4.26	0.039	0.154	**0.000**	0
	f_Coriobacteriaceae_*1908*	6.68	0.010	0.065	**0.002**	0
	*g_Bifidobacterium_1263**	18.97	0.000	0.001	**0.561**	0.162
	g_Bifidobacterium_681	5.82	0.016	0.081	0.054	**0.060**
	*g_Bifidobacterium_1296**	7.96	0.005	0.046	0	**0.001**
	*g_Bifidobacterium_1227**	10.69	0.001	0.016	**0.002**	0
	*g_Bifidobacterium_1335**	10.69	0.001	0.016	**0.002**	0
	g_Bifidobacterium_452	6.68	0.010	0.065	**0.001**	0
	g_Bifidobacterium_671	10.69	0.001	0.016	**0.002**	0
	g_Bifidobacterium_680	6.68	0.010	0.065	**0.001**	0
	g_Bifidobacterium_685	10.69	0.001	0.016	**0.002**	0
	g_Clostridium_1789	5.44	0.020	0.081	0	**0.001**
	*g_Clostridium_1792**	8.03	0.005	0.046	0.002	**0.027**
	g_Eggerthella_600	6.02	0.014	0.081	**0.002**	0.000
	*g_Escherichia-Shigella_1174**	9.30	0.002	0.029	**0.001**	0
	g_Escherichia-Shigella_1192	5.44	0.020	0.081	**0.001**	0
	g_Escherichia-Shigella_1210	5.75	0.016	0.081	**0.001**	0.000
	*g_Escherichia-Shigella_410**	12.93	0.000	0.014	**0.002**	0.000
	g_Escherichia-Shigella_422	5.77	0.016	0.081	0.068	**0.072**
	g_Parabacteroides_284	5.44	0.020	0.081	**0.002**	0
	*g_Staphylococcus_1487**	7.64	0.006	0.050	0.010	0.035
	Other	5.72	0.017	0.081	0.053	0.153
*FDR<0.05						

Chapter 6

In vitro fermentation behaviour of isomalto/malto-polysaccharides using human faecal inoculum indicates prebiotic potential

Fangjie Gu*, Klaudyna Borewicz*, Bernadette Richter, Pieter H. van der Zaal, Hauke Smidt, Pieter L. Buwalda, Henk A. Schols

* These authors contributed equally to this work

Mol Nutr Food Res. April 2018

Abstract

Scope: This study characterized intestinal fermentation of isomalto/malto-polysaccharides (IMMPs), by monitoring degradation of IMMPs, production of short chain fatty acids (SCFAs), lactic acid and succinic acid as well as enzyme activity and microbiota composition in time.
Methods and results: IMMP-94 (94% α-(1→6) glycosidic linkages), IMMP-96, IMMP-27 and IMMP-dig27 (after removal of digestible starch segments from IMMP-27) were fermented batchwise *in vitro* using human faecal inoculum. Fermentation digesta samples were taken for analysis in time up till 48 h. The fermentation of α-(1→6) glycosidic linkages in IMMP-94, IMMP-96 and IMMP-dig27 started after 12 h and finished within 48 h. IMMP-27 fermentation started directly after inoculation utilising α-(1→4) glycosidic linkages, however, the utilization of α-(1→6) linked glucoses was delayed and started only after depletion of α-(1→4) linked glucose moieties. SCFAs were produced in high amounts with acetic acid and succinic acid being the major products next to propionic acid and butyric acid. The polysaccharide fraction was degraded into isomalto-oligosaccharides (IMOs) mainly by extracellular enzymes. The smaller IMOs were further degraded by cell-associated enzymes. Overall microbial diversity and the relative abundance of *Bifidobacterium* and *Lactobacillus*, significantly increased during fermentation of IMMPs.
Conclusion: IMMPs containing segments of α-(1→6) linked glucose units are slowly-fermentable fibres with prebiotic potential.

Introduction

Prebiotics and their health benefits are of growing research interest nowadays. A dietary prebiotic is defined as "a substrate that is selectively utilized by host microorganisms conferring a health benefit" [1]. Well-documented prebiotics include lactulose, inulin, fructo-oligosaccharides (FOS) and galacto-oligosaccharides (GOS). These substrates have been shown to selectively stimulate the growth and activity of bifidobacteria, lactic acid bacteria and other health beneficial bacteria [2, 3]. Fermentation of prebiotics in the colon by these and other bacterial groups leads to production of short chain fatty acids (SCFAs) that are beneficial for gut health [4]. The present study focuses on a novel type of undigestible α-glucans, the isomalto/malto-polysaccharides (IMMPs).

IMMPs are produced from starch with the use of a 4,6-α-glucanotransferase (GTFB) enzyme from *Lactobacillus reuteri* 121 [5, 6]. The GTFB enzyme transfers a glucose moiety from the non-reducing end of α-(1→4) linked glucose chains, as present in starch and starch-derived maltodextrins to the non-reducing end of other glucose chain generating α-(1→6) linkages between glucose units in a stepwise manner, which results in the formation of IMMP containing linear chains of α-(1→6) linked glucose residues [6]. The conversion rate to α-(1→6) glycosidic linkages is positively correlated with the amylose content of the substrates, and negatively correlated with the original level of α-(1→4,6) linked branches present in amylopectin [6]. For this reason, the joint action of GTFB and debranching enzymes, e.g. isoamylase or pullulanase, leads to higher conversion rates from α-(1→4) to α-(1→6) glycosidic linkages [6]. The percentage of α-(1→6) glycosidic linkages can reach more than 90%, depending on the origin of the starch used as the substrate and the involvement of debranching enzymes [6].

IMMPs have been suggested to have potential health-beneficial effects because the α-(1→6) rich segments can escape digestion in the upper gastrointestinal tract, and be utilized as carbon source by microbiota in the large intestine [6]. This has been reported for compounds such as isomalto-oligosaccharides (IMOs) and dextran, which have similarities in structure when compared to IMMPs. IMOs are gluco-oligosaccharides consisting of predominantly α-(1→6) glycosidic linkages, with the degree of polymerization (DP) ranging from 2 to 10 [7]. IMOs have been shown to promote growth of lactobacilli and bifidobacteria in both *in vitro* fermentation and *in vivo* rat models [8–10]. Dextran, another well-known glucose homopolysaccharide with consecutive α-(1→6) glycosidic linkages, has been reported to stimulate bifidobacteria and lactobacilli during *in vitro* fermentation with human faecal microbiota, and to lead to increased production of butyrate [11]. Therefore, based on the structural similarity between IMMPs, IMOs and dextran, we expected that IMMPs would bear prebiotic potential as well.

The starch origin and involvement of debranching enzymes during synthesis of IMMPs lead to structural differences, which in turn, may influence the IMMPs' fate during fermentation in the colon. The difference can be in the proportion of α-(1→6) and α-(1→4) glycosidic linkages. It has been shown by NMR spectroscopy that the relative amount of α-(1→6) linkages can be very different, ranging from 7% to over 90% [6], with the remaining linkages being α-(1→4). Although α-(1→4) glycosidic linkages are in general readily digested by human digestive enzymes, introduction of α-(1→6) linkages may help neighbouring α-(1→4) linked

units to escape digestion and to enter the colon. Such starches that have been chemically or enzymatically modified to resist digestion, are considered to be Resistant Starch type IV [12]. It remains unclear to what extent the α-(1→4) linked glucose segments of IMMPs would end up in the colon and have an influence on fermentation of α-(1→6) linked glucosyl residues. Based on earlier studies on the fate of retrograded tapioca starch, it can be speculated that the presence of resistant starch could influence the fermentation of other fibres [13]. IMMPs with similar percentages of α-(1→6) linkages could differ in the distribution of molecular chain length, depending on the side-chain length distribution of the parental starch. It remains unknown whether such differences in molecular chain length would influence the fermentation behaviour of IMMPs.

Leemhuis et al. [6] showed preliminary results of *in vitro* fermentation of IMMPs, including an increase in microbial biomass, as monitored by optical density, and an increase in concentrations of acetic acid and propionic acid. However, the influence of additional factors on IMMPs fermentation, including molecular weight and the presence of α-(1→4) glycosidic linkages, still needs to be determined. Furthermore, previous research showed that the production of enzymes by faecal microbiota varies depending on substrate properties, including sugar composition, linkage type and chain length [13–15]. The prebiotic potential of IMMPs is still unknown since detailed effects on microbiota composition have yet to be established [6].

Therefore, to evaluate the prebiotic potential of IMMPs, a comprehensive *in vitro* batch fermentation of selected types of IMMPs with a standardized human faecal inoculum was performed in the present study. The fermentation behaviour of IMMPs at a molecular level and production of individual organic acids were studied, and a link to microbiota composition was made. In addition, bacterial enzyme activities involved in the IMMP degradation were studied in order to help explaining the mechanism of bacterial utilization of IMMPs.

Materials and Methods

Materials

Three different types of IMMPs were used in this study. In order to facilitate the comparison of results, the IMMPs in this study were named after their percentages of total α-(1→6) glucosyl linkages. The total α-(1→6) linked glucosyl content, consisting of both α-(1→6) and α-(1→4,6) glycosidic linkages was determined by hydrogen-1 nuclear magnetic resonance (^{1}H NMR) spectroscopy, with the methodology and results already published previously [16]. IMMP-94 (94% α-(1→6) linkages) was potato starch (AVEBE, Veendam, the Netherlands) modified with *Lactobacillus reuteri* 121 GTFB 4,6-α-glucanotransferase [6] and pullulanase (Promozyme D2) (Novozymes, Bagsvaerd, Denmark) and was kindly provided by Dr. Hans Leemhuis (AVEBE). IMMP-27 (27%) and IMMP-96 (96%) were synthesized from potato starch and Etenia^{TM} 457 starch (AVEBE), respectively, as published by van der Zaal et al. [16] and described below:

Synthesis of IMMP-27

Potato Starch was suspended at 2.5% (w/v) in 20 mM sodium acetate buffer, pH=4.9, containing 5 mM $CaCl_2$. The suspension was autoclaved at 121 °C for 15 min and cooled to 37 °C. IMMP synthesis was carried out by adding 0.3 mg GTFB-ΔN /g substrate and incubating the reaction mixture at 37 °C for 24 h. GTFB-ΔN is GTFB with N-terminal truncation [17], and the synthesis of GTFB-ΔN is described elsewhere [16]. GTFB-ΔN was inactivated in a water bath at 95 °C for 15 min. The solution was cooled to 50 °C, Amberlite® MB-20 resin (Dow, Midland, MI, USA) was added to remove salts and then incubated at 50 °C for 2 h. The resin was sieved out and the IMMP solution was freeze-dried.

Synthesis of IMMP-96

Amylomaltase treated potato starch (Etenia™ 457) was used as substrate, and the treatment was similar to that of IMMP-27. In addition GTFB-ΔN, pullulanase (Promozyme D2) was also added at an amount of 2 μL/g substrate, and the incubation time was extended to 41 h. The other conditions of the synthesis was kept the same as described in the synthesis of IMMP-27.

Experimental set-up and removal of digestible starch segments

A schematic overview of the present study is shown in Figure 1.

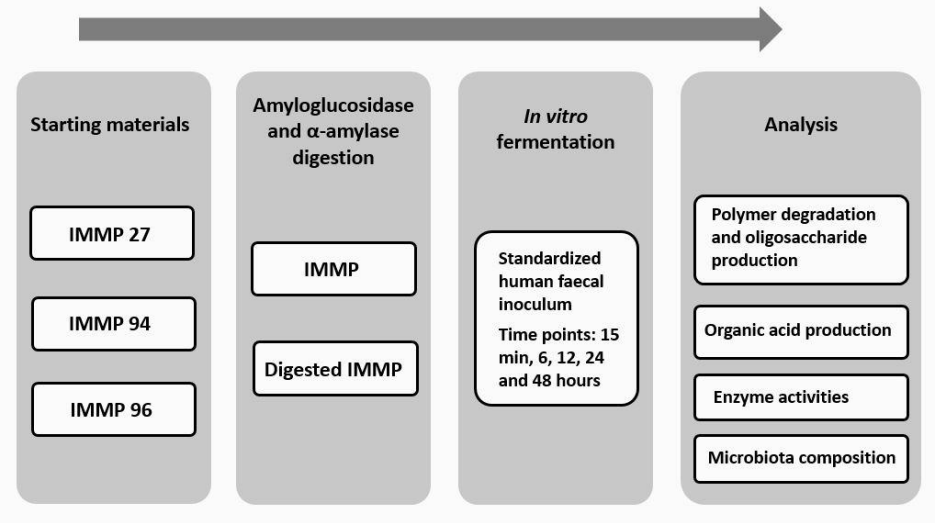


Figure 1. Experimental set-up of the *in vitro* fermentation of IMMPs

Three types of IMMPs, namely IMMP-27, IMMP-94 and IMMP-96 were used, representing extremes with respect to the percentage of total α-(1→6) linked glucosyl residues. Each of the three IMMPs were split and parts were either left untreated or treated with pancreatic α-amylase and amyloglucosidase (Resistant Starch assay kit, Megazyme, Bray, Ireland) to remove α-(1→4) linked glucosyl residues in order to obtain the resistant fibre. Hereto, IMMP-27, IMMP-94 and IMMP-96 were treated with two starch digesting enzymes. The concentrations of both enzymes and the incubation conditions were according to Megazyme protocols. After inactivating the enzymes at 100 ºC for 5 min, IMMPs were

recovered by ethanol precipitation with a final ethanol concentration of 70%. The supernatant containing glucose and small maltodextrins was removed by decanting after centrifugation at 10,000 g for 15 min at room temperature. The ethanol precipitation step was repeated twice. Afterwards, the pellet was washed once with pure ethanol and air-dried at 30 ^{o}C. The sample obtained after the removal of digestible α-(1→4) linked segments from IMMP-27 was named IMMP-dig27.

In order to determine the level of removal of α-(1→4) linked glucosyl residues in IMMPs after enzymatic digestion, the sugar content of the supernatant collected from the ethanol precipitation step was measured colorimetrically by a phenol-sulphuric acid assay, using D-glucose as standard for calibration [18, 19]. Since less than 2% of glucose moieties were removed from both IMMP-94 and IMMP-96, only the parental IMMP-94 and IMMP-96 were included in the following *in vitro* fermentation, whereas both untreated IMMP-27 and IMMP-dig27 were used.

In vitro fermentation

An *in vitro* fermentation was performed to simulate the fermentation of IMMPs in the human colon according to the procedure described by Rösch et al. [14], with the modification that carbohydrates (pectin, xylan, arabinogalactan, amylopectin and starch) and Tween 80 were left out of the standard ileal efflux medium (SIEM), in order to reduce as much as possible background fermentation from the medium components. The modified SIEM medium contained 40% (v/v) Bacto Peptone-casein-ox bile (BCO) medium, 1.6% (v/v) salt solution, 0.8% (v/v) $MgSO_4$ (50 g/L), 0.4% (v/v) cysteine hydrochloride (40 g/L), 0.08% (v/v) vitamin solution, 10% (v/v) MES buffer (1 M, pH 6.0) and the rest is water. The BCO medium contained (g/L): Bacto Peptone, 60.0; casein, 60.0; Ox bile, 1. The salt solution contained (g/L): $K_2HPO_4 \cdot 3H_2O$, 156.25; NaCl, 281.25; $CaCl_2 \cdot 2H_2O$, 28.13; $FeSO_4 \cdot 7H_2O$, 0.31; hemin, 0.63. The vitamin solution contained (mg/L): menadion, 1.0; biotine, 2.0; vitamin B12, 0.5; pantothenate, 10.0; nicotinamide, 5.0; *p*-aminobenzoic acid, 5.0; thiamine, 4.0. The ingredients used to make SIEM were purchased from Tritium Microbiologie (Veldhoven, the Netherlands).

A standard human faecal inoculum was prepared by TNO (Zeist, the Netherlands), and was kindly provided by Prof. Dr. K. Venema. The faecal inoculum was pooled from seven healthy volunteers (male: n=3, average age=46.3 y (range: 26-57), BMI= 24.1 ± 2.42 kg/m2; female: n=4, average age=37.7 y (27-52), BMI= 24.2 ± 1.91 kg/m2). The pooling procedure was described and validated previously [20, 21].

Each *in vitro* fermentation took place in 20 mL serum bottles sealed with a butyl rubber stopper and with the final volume of the fermentation liquid being 10 mL. Bottles and medium were flushed with gas mixture containing 81% N_2, 15% CO_2 and 4% H_2 to remove oxygen. The final concentration of IMMP was 10 mg/mL, and faecal inoculum was added to a final concentration of 1% (v/v). Negative control incubations were included and did not receive any faecal inoculum (inoculum blanks) or IMMP substrate (IMMP blanks). A baseline sample (defined as 0 h) was taken within the first 15 min after addition of the inoculum, after which bottles were incubated at 37 ^{o}C and shaking at 140 rpm. For sampling, sterile syringes and needles were used to take aliquots (2 – 3.5 mL) at time points 15 min, 6 h, 12 h, 24 h and 48 h.

Assessment of IMMP degradation by HPSEC-RI

Part of the fermentation digest was heated at 100 °C for 5 min and then centrifuged at 18,600 g for 10 min at room temperature. The supernatant was diluted four times with water to be used for high performance size exclusion chromatography (HPSEC). A set of four TSK-Gel® SuperAW columns (Tosoh Bioscience, Tokyo, Japan) were used on an Ultimate 3000 HPLC (Dionex, Sunnyvale, CA, USA) system: a guard column (SuperAW-L, 3.5 cm x 4.6 mm ID) and three analytical columns (SuperAW 4000, 3000 and 2500; 15 cm x 6.0 mm ID). Ten µL of sample was injected and eluted at 0.6 mL/min 0.2 M $NaNO_3$ isocratically. The column temperature was 55 °C. Eluted components were monitored by an refractive index (RI) detector (Shodex RI-101, Showa Denko K.K., Kawasaki, Japan). Molecular weights of IMMPs were estimated using a pullulan (Polymer Laboratories, Palo Alto, CA, USA) calibration curve. Chromeleon™ 7.1 software (Dionex) was used to process data from HPSEC.

Analysis of oligosaccharide production by HPAEC-PAD

The supernatant obtained after centrifugation of the fermentation digest was tenfold diluted before analysis using high performance anion exchange chromatography in combination with pulsed amperometric detection (HPAEC-PAD). The oligosaccharide peaks were annotated using dextranase-treated IMMP-94 as a standard. The dextranase-treated IMMP-94 was prepared as follows: 0.25 unit of dextranase from *Chaetomium erraticum* (Sigma-Aldrich, St. Louis, MO, USA) was added to 5 mg IMMP-94 in 1 mL of 0.1 M sodium maleate buffer containing 5 mM $CaCl_2$ at pH 6, and incubated at 37 °C for 30 min. The reaction was stopped by heating at 99 °C for 5 min, and the supernatant was diluted five times for HPAEC analysis after centrifuging at 18,600 g for 10 min at room temperature.
Ten µL of sample was injected to a Dionex ICS 5000 system (Dionex) with a CarboPac PA-1 column (250 mm x 2 mm ID) and a CarboPac PA guard column (25 mm x 2 mm ID). The column temperature was 20 °C. The flow rate of the two mobile phases (A) 0.1 M NaOH and (B) 1 M NaOAc in 0.1 M NaOH was set to 0.3 mL/min. The gradient elution was applied as follows: 0 – 40 min, 0 – 40% B; 40 – 40.1 min, 40 – 100% B; 40.1 – 45 min, 100% B; 45 – 45.1 min, 100 – 0% B; 45.1 – 60 min, 0% B. The elution was monitored by a PAD (Dionex ISC-5000 ED). Chromeleon™ 7.1 software (Dionex) was used to process data from HPAEC.

Extraction and activity of bacterial enzymes

Part of the fermentation digest (0.4 mL) was snap frozen in liquid nitrogen and stored at -80 °C before enzyme extraction. The following fermentation digests were selected according to HPAEC results (see section Results and Discussion for further details): IMMP blank, IMMP-27 at 12 and 48 h, IMMP-dig27 at 12 and 24 h, and IMMP-94 at 12 and 24 h. Protein extraction was performed as described elsewhere [14] with some modifications. To obtain the fraction of extracellular enzymes (EE), the fermentation digest was first centrifuged (21,000 g, 4 °C, 10 min), and the supernatant was applied on a 10kDa centrifugal filter (VWR®, Amsterdam, the Netherlands) at 4 °C and 18,600 g to remove any mono- and oligosaccharides produced during fermentation. A volume of 0.4 mL 25 mM MES buffer pH 5.8 containing 1 mM phenylmethylsulfonyl fluoride and 1 mM dithiothreitol was used to reconstitute the retentate

(EE). The pellet from the first centrifugation step was washed once with 1.5 mL buffer, centrifuged again and then suspended in 0.4 mL of the same MES buffer. The suspension was sonicated at 30% amplitude for 30 s and repeated three times with 40 s break in between [13]. The supernatant after centrifugation was used as cell-associated enzymes (CE).

The enzyme activity of EE and CE towards PNP-glucose substrates and starch was determined using a colour reaction, as described previously [14] with some modifications. In the glycosidase assay, only PNP-α-D-glucopyranoside and PNP-β-D-glucopyranoside were included as substrates. In the polysaccharide assay, soluble potato starch (Sigma-Aldrich) and IMMP-94 were used as substrates. Potato starch was incubated at 99 ºC until solubilized. The substrate (3.125 mg/mL) was mixed with enzyme extracts in a 4:1 ratio, yielding a final substrate concentration of 2.5 mg/mL. The amount of reducing sugar released after 1 h incubation was determined by 4-hydroxybenzoic acid hydrazide (PAHBAH) assay using glucose as a standard [13]. Enzyme activities were expressed in mU (nmol-reduced-end-formed* mL-digest^{-1} min^{-1}).

Analysis of SCFAs and other organic acids by GC-FID and HPLC-RI

Determination of SCFAs (acetic acid, propionic acid and butyric acid) by gas chromatography (GC) and of lactic acid and succinic acid by high performance liquid chromatography (HPLC) was done as described previously [22] with some modifications. For GC, 70 µL of twofold diluted supernatant of the fermentation digest was mixed with 70 µL 0.15 M oxalic acid and allowed to stand at room temperature for 30 min. Then 199 µL water and 1 µL of 5 mg/mL 2-ethylbutyric acid was added. The temperature profile during GC analysis was as follows: from 100 ºC to 165 ºC at 5 ºC min^{-1}, then held at 165 ºC for 1 min. Chromeleon™ 7.1 software (Dionex) was used to process data from HPLC. Xcalibur TM software (Thermo Scientific, Breda, the Netherlands) was used to process data from GC.

DNA extraction, 16S ribosomal RNA gene sequencing and microbial composition analysis

The pellets obtained from centrifugation of fermentation digest were snap frozen in liquid nitrogen, stored at -80 ºC and used for microbial composition analysis. Total bacterial DNA was extracted using the Maxwell® 16 Total RNA system (Promega, Wisconsin, USA) with Stool Transport and Recovery Buffer (STAR; Roche Diagnostics Corporation, Indianapolis, IN). Briefly, bacterial pellets were homogenized with 0.25 g of sterilized 0.1 mm zirconia beads and three glass beads (2.5 mm) in 300 µL STAR buffer for 3 × 1 min at 5.5 m/s using a bead beater (Precellys 24, Bertin Technologies), with cooling at room temperature for 10 s in between. Samples were incubated with shaking at 100 rpm for 15 min at 95 ºC and pelleted by 5 min centrifugation at 4 ºC and 14,000 g. Supernatant was removed and the pellets were processed again using 200 µL fresh STAR buffer. Samples were incubated at 95 ºC and centrifuged as before. Supernatant was removed, pooled with the first supernatant and 250 µL was used for purification with Maxwell® 16 Tissue LEV Total RNA Purification Kit (AS1220) customized for DNA extraction in combination with the STAR buffer following manufacturer's instructions (Promega). DNA was eluted with 50 µL of DNAse and RNAse free water (Qiagen, Hilden, Germany). DNA concentrations were measured with a NanoDrop ND-1000 spectrophotometer (NanoDrop® Technologies, Wilmington, DE, USA) and adjusted to 20

ng/µL with DNAse and RNAse free water. The V4 region of 16S ribosomal RNA (rRNA) genes was amplified. PCR reactions were done in duplicates, each in a total volume of 50 µL and containing 20 ng of template DNA. Each sample was amplified with a unique barcoded primer pair 515F-n (5'-GTGCCAGCMGCCGCGGTAA-) and 806R-n (5'-RGGATTAGATACCC) (200 nM each [23]), 1x HF buffer (Finnzymes, Vantaa, Finland), 200 µM dNTP Mix (Roche Diagnostics GmbH, Mannheim, Germany), 1 U Phusion® Hot Start II High Fidelity DNA Polymerase (Finnzymes) and 36.5 µL of DNAse and RNAse free water. The amplification program included 30 s initial denaturation step at 98 ºC, following by 25 cycles of denaturation at 98 ºC for 10 s, annealing at 56 ºC for 10 s and elongation at 72 ºC for 10 s, and a final extension at 72 ºC for 7 min. The PCR product presence and size (~290 bp) was confirmed with gel electrophoresis using the Lonza FlashGel® System (Lonza, Cologne, Germany). Seventy unique barcode tags were used in each library and artificial control (Mock) communities representative of human intestinal microbiota were included [23]. PCR products were purified with HighPrep® PCR kit (MagBio Genomics, Alphen aan den Rijn, the Netherlands), and DNA concentrations were measured with Qubit® dsDNA BR Assay Kit (Life Technologies, Leusden, the Netherlands). 100 ng of each barcoded sample was added to an amplicon pool that was subsequently concentrated with HighPrep® PCR kit to 20 µL volume. The concentration was measured with Qubit® dsDNA BR Assay Kit and adjusted to 100 ng/µL final concentration. The libraries were sent for adapter ligation and HiSeq sequencing (GATC-Biotech, Konstanz, Germany). Data processing and analysis was carried out using NG-Tax [23]. Alpha diversity analyses were carried out in QIIME [24, 25]. Relative abundance at genus level was used for calculating pairwise Pearson correlation scores between biological replicates, and the values for the different taxa were averaged for each replicate pair.

Results and Discussion

Between batch similarity

The *in vitro* fermentation experiments were run in two separate batches, the first one using IMMP-27 and IMMP-94, and the second batch with IMMP-dig27 and IMMP-96. Between batch similarity was estimated based on Pearson correlation scores of genus-level microbiota composition data for the IMMP blank samples from different batches, at time 0, 24 and 48 h and were 0.98, 0.94 and 0.88 respectively. A high reproducibility for the results between the two batches was found, validating the between batch comparisons to be carried out when necessary. Pearson correlation scores also showed high levels of similarity between the biological duplicates at genus level (average Pearson score of 0.97, *SD* ± 0.03 for IMMP treatment groups and 0.90, *SD* ± 0.22 for IMMP blank groups).

Physicochemical characterization of IMMP-94 and IMMP-96

IMMP-94 and IMMP-96 contained high percentages of α-(1→6) linked glycosidic chains, as a result of including the debranching enzyme pullulanase during the synthesis by GTFB. The two IMMPs were derived from different starches, namely normal potato starch and amylomaltase treated potato starch (Etenia™). Amylomaltase treatment results in a

disappearance of the amylose fraction and a broader chain length distribution of the amylopectin fraction, due to the disproportionating effect of the amylomaltase enzyme [26]. In order to verify potential differences in the molecular weight distribution, both IMMPs were compared by HPSEC using samples prior to fermentation (Figure 2, Aa and Ba). IMMP-94 showed a broad molecular weight distribution with populations being eluted between 10 – 12.9 min (1.7 – 65 kDa). IMMP-96 showed a slightly clearer bimodal distribution, with higher RI response towards both ends of the same elution window (10 – 12.9 min), indicating that IMMP-96 contained both shorter and longer chains and fewer medium length chains, as compared to IMMP-94.

Influence of IMMP molecular weight distribution on its *in vitro* fermentation

The degradation of IMMPs during *in vitro* fermentation was monitored by HPSEC up to 48 h (Figure 2).

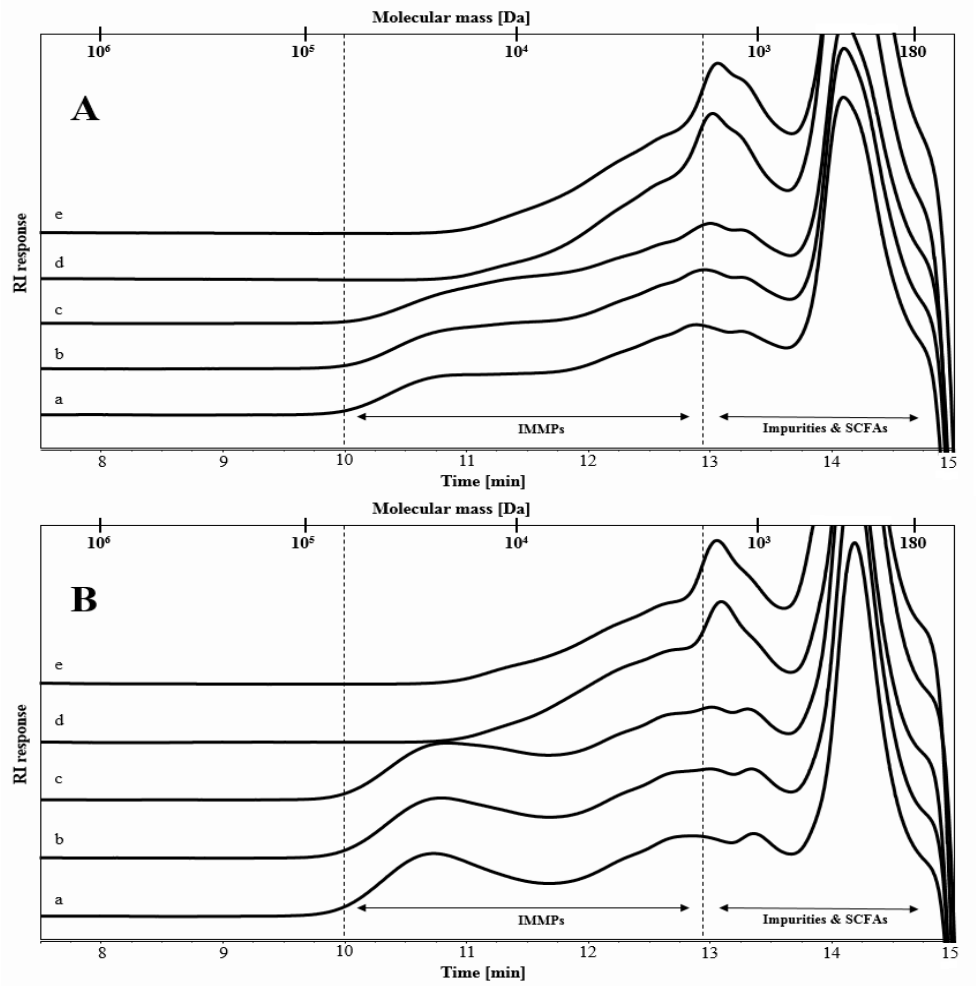
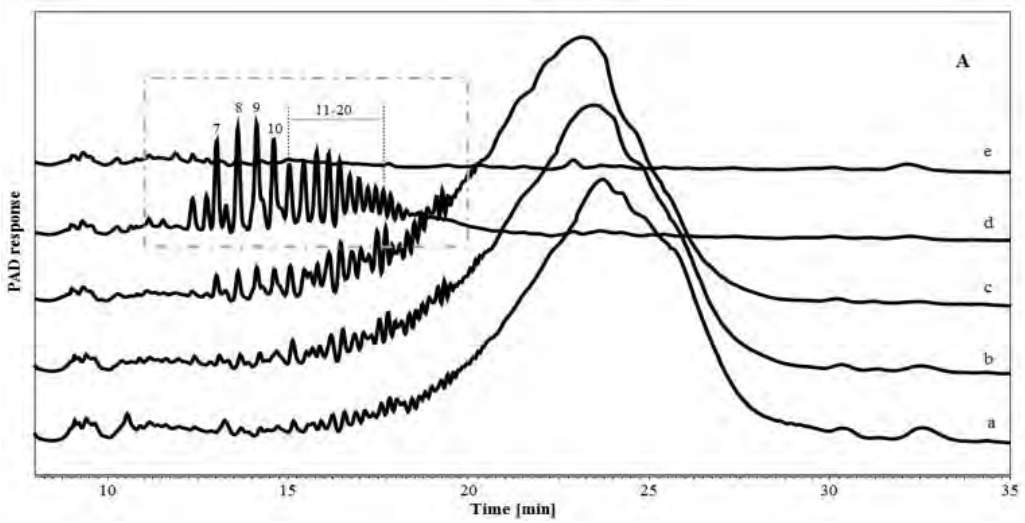


Figure 2. High performance size exclusion chromatography (HPSEC) elution patterns of (A) IMMP-94 originating from potato starch, and (B) IMMP-96 originating from EteniaTM 457, before (a) and after in vitro fermentation for 6 h (b), 12 h (c), 24 h (d) and 48 h (e). Calibration of the system with pullulan standards is indicated

The fermentation behaviour of IMMP-94 and IMMP-96 was similar, and the same types of oligomeric dextran fragments were formed and utilized in time. For both IMMPs, the HPSEC elution patterns remained the same during the first 12 h, followed by a shift in molecular size from larger to smaller molecules from 12 h to 24 h. No further difference was noted between 24 h and 48 h of incubation suggesting that the degradation of the polysaccharides fraction of IMMP-94 and IMMP-96 mainly took place between 12 and 24 h of fermentation. To have a better overview of smaller size molecules being formed during fermentation, HPAEC was performed (Figure 3). For both IMMPs, a broad peak being eluted between 20 and 25 min was seen during the first 12 h of fermentation. This peak included a wide range of not well-separated IMMP molecules, which partly corresponded to the 10 – 65 kDa population in the HPSEC chromatograms (Figure 2, A&B, lines a, b, c). At 24 h of fermentation, these polymers had disappeared, and a series of well-separated oligosaccharide peaks which eluted between 11 and 20 min could be observed. The oligosaccharide peaks were annotated according to the HPAEC elution pattern of IMMP-94 treated with a pure dextranase from *Chaetomium erraticum* (results not shown). The oligosaccharide fraction of the fermentation digest comprised α-(1→6) linked IMOs with a DP of 7 to over 20. IMOs with DP < 7 were absent at 24 h, which could be due to instant consumption of smaller oligosaccharides by the microbiota during fermentation, indicating a preference of the microbiota for the utilization of smaller molecules. At 48 h of incubation, the oligosaccharide fraction had disappeared, and no carbohydrate peaks were present in the chromatogram (Figure 3). Overall, the HPAEC results of IMMP-94 and IMMP-96 were in accordance with HPSEC results. Despite similarities in the fermentation behaviour of IMMP-94 and IMMP-96, the overall rate of fermentation of IMMP-96 was slower (Figure 3), as indicated by the presence of polymeric material being eluted between 18 and 20 min after 24 h of incubation. This difference in fermentation rate could be due to the difference in chain length distributions of the two IMMPs. This finding agreed with a previous study which reported that IMOs of different chain length led to different utilization and fermentation rate when using human faecal microbiota [27].



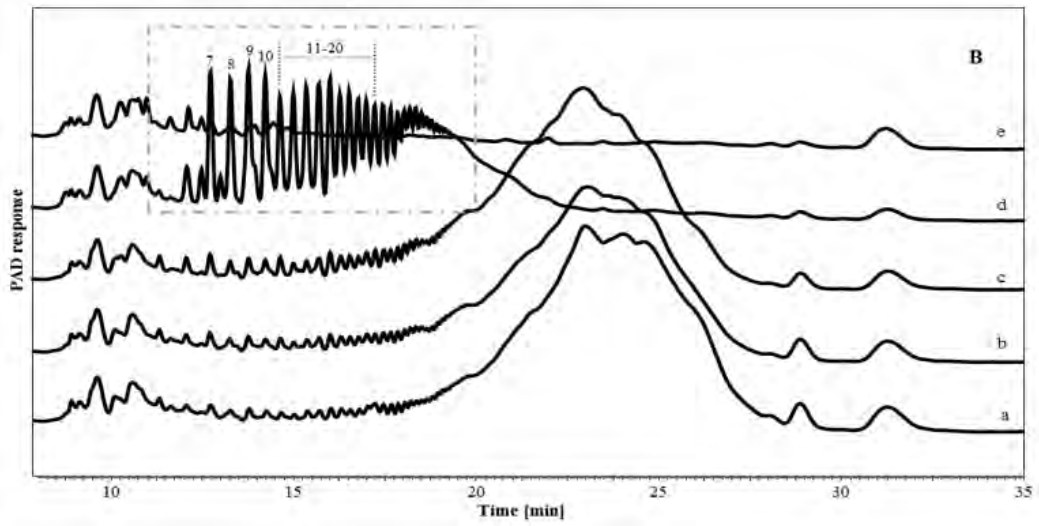


Figure 3. High performance anion exchange chromatography (HPAEC) elution patterns of (A) IMMP-94 originating from potato starch, and (B) IMMP-96 originating from Etenia^{TM} 457, before (a) and after *in vitro* fermentation for 6 h (b), 12 h (c), 24 h (d) and 48 h (e). Isomalto-oligosaccharides are annotated in a box, with the number indicating their degree of polymerization (DP).

pH and production of organic acids upon *in vitro* fermentation of IMMP-94 and IMMP-96

Analysis of the pH of fermentation digesta and organic acid production at different time points confirmed that the fermentation of IMMP-94 and IMMP-96 started after 12 h of incubation (Figure 4). For both IMMP-94 and IMMP-96, the pH remained stable at around pH 6.2 during the first 12 h, followed by a decrease to around pH 5.2 at 24 h, and a slight further decrease at 48 h (Figure 4). It is noteworthy that the drop of the pH to 5.0 at 48 h was larger as compared to a drop of pH to 6.0 at 48 h previously observed for resistant gluco-dextrin fermentation in a comparable set-up [14]. The pH decreased as a result of organic acid production. In line with the change of pH, the largest increase in the concentration of SCFAs was observed from 12 h to 24 h, followed by a further increase from 24 h to 48 h (Figure 4). Acetic acid, propionic acid and butyric acid are in general the three main SCFAs produced during *in vitro* fermentation of carbohydrates. Lactic acid and succinic acid should also be taken into consideration, since they are intermediates in SCFA production during fermentation [28].

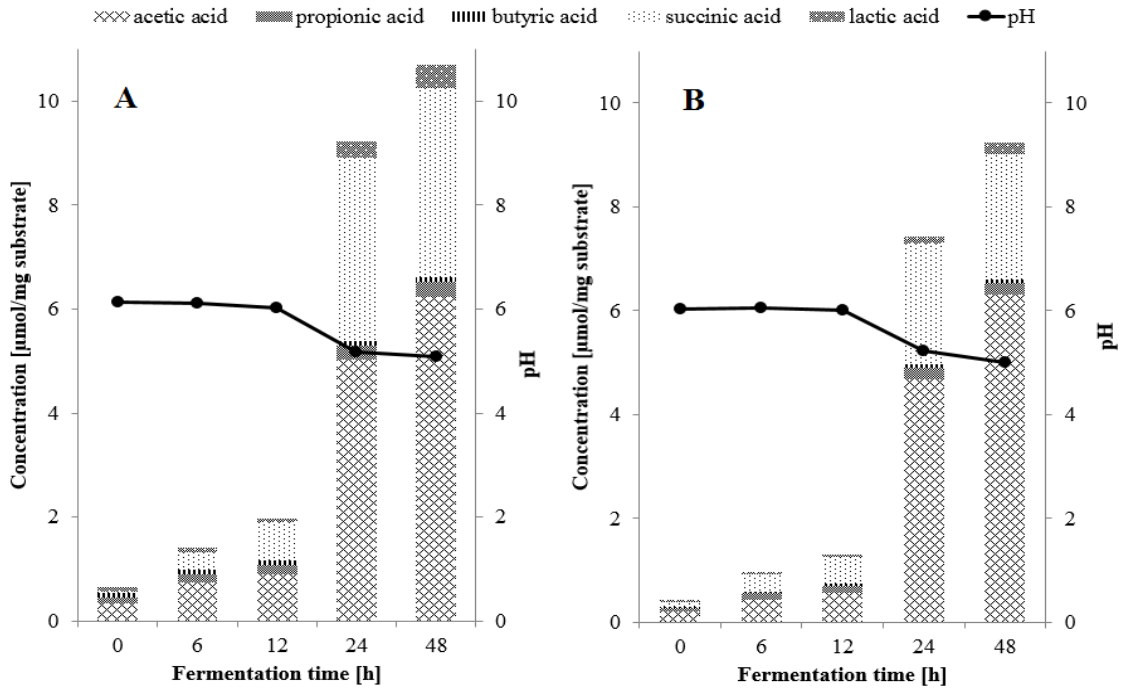


Figure 4. Concentrations of short chain fatty acids, lactic acid and succinic acid and the pH (-●-) during *in vitro* fermentation with human faecal inoculum of (A) IMMP-94 and (B) IMMP-96.

For both IMMPs, the most predominant SCFA produced was acetic acid, with minor amounts of propionic acid and butyric acid (Figure 4). Unexpectedly, the second most produced organic acid for both IMMPs was succinic acid. Succinic acid is an intermediate of intestinal SCFA production, and is utilized by members of the phylum Bacteroidetes and the family Veillonellaceae to form propionic acid [29–31]. In the current study, however, succinic acid accumulated during the incubation without further conversion. This accumulation of succinic acid can also explain the lower pH at the end of fermentation, since succinic acid has a pK_{a1} of 4.2 [32], which is lower than that of other SCFAs (approximately 4.8, [4]).

The drop of pH and the predominant production of acetic acid are in line with a previous report on *in vitro* fermentation of IMMPs [6]. Formation of succinic acid was not reported in that study, however, it should be noted that only acetic acid and propionic acid were measured. Information about succinic acid was also not presented for studies where IMOs or dextrans were fermented [8, 10, 11], but was reported for fermentation studies where other prebiotics were used as substrate, e.g. lactulose and inulin [33–35]. It has been reported previously that *Bacteroides fragilis* produced acetate and succinate mainly in the presence of sufficient carbon source, whereas it converted succinate to propionate when carbon sources were limited [28]. An *in vivo* study was performed in collaboration with the University Medical Centre Groningen (the Netherlands), where IMMPs were fed to mice (unpublished results). Also in that study, significant amounts of succinic acid were found in the faeces, providing additional evidence that the production and accumulation of succinic acid during IMMPs fermentation was not an artefact of the *in vitro* fermentation set up.

The degradation of IMMPs started later and continued over a longer time than that of other commonly studied prebiotics. In a comparable *in vitro* fermentation set-up, utilization of

FOS started at around 2 h after faecal inoculation and was completed within 9 h [36]. Therefore, IMMPs can be considered to be a slowly fermentable fibre, although it should be noted that a direct comparison between the substrates might be necessary to unequivocally confirm observations described here. Slowly fermentable fibres are of great interest, because most colonic diseases occur distally, where proteolytic fermentation may take place when carbohydrates are lacking [2, 37]. The slow fermentability of IMMPs makes them beneficial to gut health by increasing the delivery of SCFAs to the distal colon. Besides, given the fact that IMOs of DP <10 were shown to be bifidogenic [9, 38, 39], results presented here indicate that IMMPs are a good fibre source to make these IMOs available for the fermentation by the colonic microbiota.

Physicochemical characterization of IMMP-27 and IMMP-dig27

Starch, due to its high content of α-(1→4) linked glucosyl residues, is mostly digested in the human small intestine. In contrast, when mixed with α-(1→6) linked glucosyl moieties such as in IMMPs, it is possible that part of the α-(1→4) linked glucoses could escape digestion and enter the large intestine. To investigate the influence of α-(1→4) linked glucosyl residues on the fermentation of α-(1→6) linked glucose segments by colonic microbiota, the *in vitro* fermentation of IMMP-27 and IMMP-dig27 was compared. IMMP-27 contains 27% α-(1→6) linked glucosyl residues, whereas IMMP-dig27 is the α-(1→6) glucan enriched fraction of IMMP-27, after being treated with an excess of α-amylase and amyloglucosidase that removed >70% of glucose moieties.

The molecular size distribution of IMMP-27 and IMMP-dig27 was determined by HPSEC (lines a, Figure 5A, B). The overall molecular size of IMMP-27 was larger than that of IMMP-dig27. Molecules that eluted at 8 – 10 min (65 – 850 kDa) in IMMP-27 were not observed in IMMP-dig27, indicating that this fraction of molecules was digested to smaller fragments due to the removal of α-(1→4) linked glucose moieties by α-amylase and amyloglucosidase.

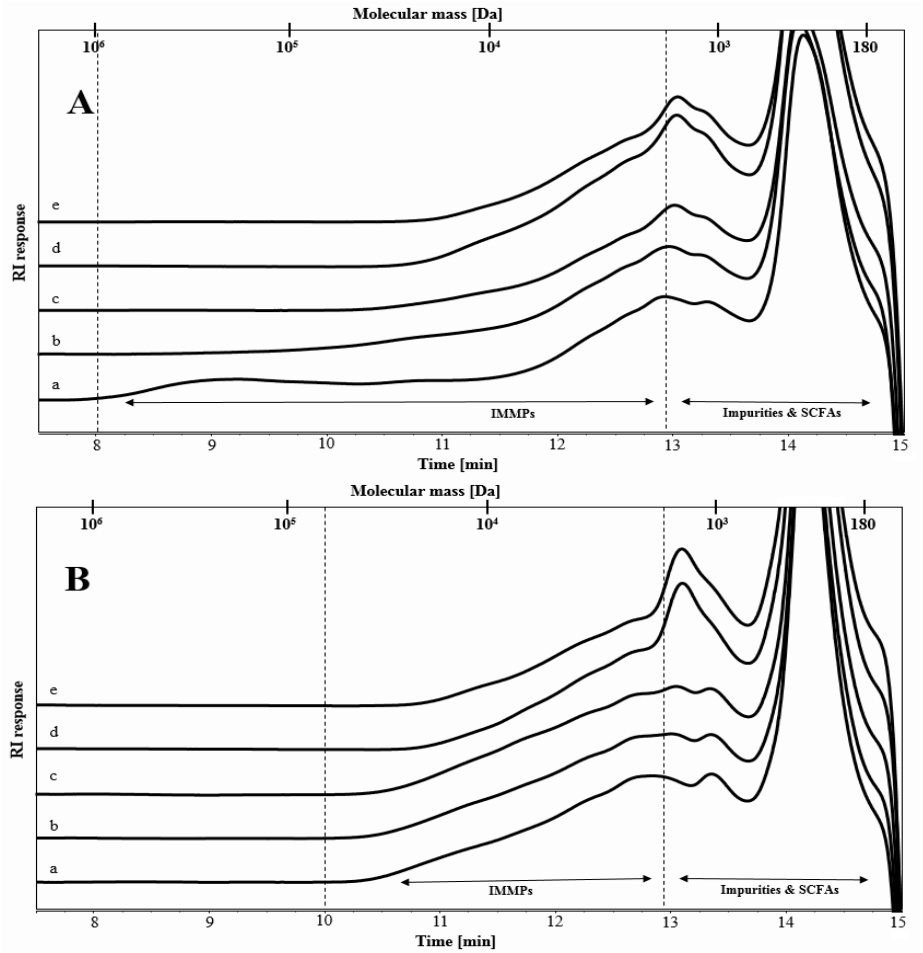


Figure 5. High performance size exclusion chromatography (HPSEC) elution patterns of (A) IMMP-27 and (B) IMMP-dig27, before (a) and after *in vitro* fermentation for 6 h (b), 12 h (c), 24 h (d) and 48 h (e). Calibration of the system with pullulan standards is indicated.

Influence of α-(1→4) glycosidic linkages on bacterial utilization of α-(1→6) linked glucose during *in vitro* fermentation of IMMPs

The change in molecular size distribution of IMMP-27 and IMMP-dig27 during *in vitro* fermentation was monitored using HPSEC (Figure 5). For IMMP-27, HPSEC chromatograms showed differences between 0 h and 6 h, with molecules ranging in size between 65 – 850 kDa being degraded within 6 h of fermentation. After 6 h, the chromatograms of IMMP-27 did not show any further increase in the proportion of the smaller molecules which eluted at 8 – 10 min (65 – 850 kDa). In contrast, the chromatograms of IMMP-dig27 remained the same in the first 12 h, and there was a shift in molecular size distribution to smaller molecules between 12 h and 24 h. No changes in the elution patterns were observed from 24 h to 48 h, indicating that the degradation of IMMP polymers was completed. The oligomer profiles of IMMP-27 and IMMP-dig27 during fermentation obtained by HPAEC showed that for IMMP-27, α-(1→4) linked maltodextrin peaks were already present at 15 min, and were still present at 6 h (Figure 6). At 12 h, these maltodextrin peaks were hardly present, whereas new peaks, probably representing oligosaccharides consisting of both α-(1→4) and α-(1→6) linkages, became more apparent

(Figure 6). At 24 h, a series of well-separated α-(1→6) linked IMO peaks which eluted between 11 and 20 min appeared, and a broad fraction eluting between 20 and 24 min representing unseparated dextran oligomers of higher DPs was clearly seen. The peaks of IMOs (11 – 20 min) were still present at 48 h of fermentation, whereas the unseparated fraction (20 – 24 min) disappeared. For IMMP-dig27, with hardly any α-(1→4) linkages present in the substrate, the IMMP molecules remained intact during the first 12 h of fermentation. However, no carbohydrates were detected at 24 h of fermentation, indicating that a very quick and complete fermentation took place between 12 h and 24 h.

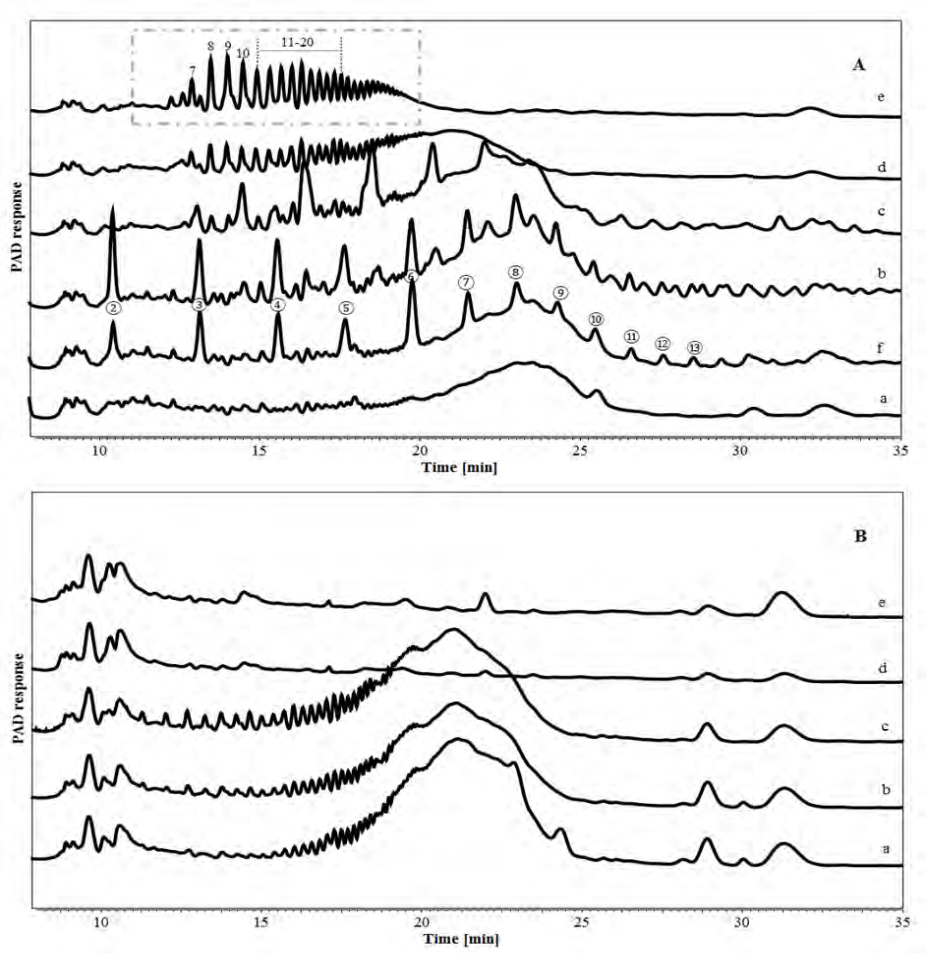


Figure 6. High performance anion exchange chromatography (HPAEC) elution patterns of (A) IMMP-27 and (B) IMMP-dig27, before (a) and after *in vitro* fermentation for 15 min (f), 6 h (b), 12 h (c), 24 h (d) and 48 h (e).) Isomalto-oligosaccharide peaks (7-20) in a box and maltodextrin peaks (②-⑬) are annotated, with the number indicating the DP.

The different degradation patterns of IMMP-27 and IMMP-dig27 suggest that in the *in vitro* fermentation model used here, human faecal microbiota could utilize the α-(1→4) linkages directly, whereas α-(1→6) linkages were utilized only after the α-(1→4) linkages were depleted. Different enzymes are required to digest α-(1→4,6) linkages, and bacteria present in the faecal inoculum could be induced to produce corresponding hydrolytic enzymes by the presence of specific substrates in the colon [40]. However, when mixtures of compounds are present, the

availability of one substrate could delay the fermentation of another, possibly less favourable substrate.

Our results suggest that the presence of α-(1→4) linked glucosyl residues could postpone the utilization of α-(1→6) linked glucosyl residues *in vitro* and that fermentation of IMMPs with high levels of α-(1→6) glycosidic linkages may require colonic microbiota to undergo an adaptation period. Furthermore, this adaptation period might relate to the molecular size of the α-(1→6) glucan chains. The fermentation behaviour of IMMP-dig27 resembled that of IMMP-94 and IMMP-96, in line with the fact that all three substrates are rich in α-(1→6) linked- and depleted in α-(1→4) linked glucose residues.

The complete degradation of IMMP-dig27, however, was faster than that of the other two IMMPs. This could be explained by the smaller molecular sizes of IMMP-dig27 'dextran'-segments as compared to IMMP-94 and IMMP-96, indicating that fermentation of α-(1→6) linkages is quicker for smaller IMMP molecules. Therefore, the fermentation of IMMPs depends not only on the presence of α-(1→4) glycosidic linkages, but also on molecular length distribution of IMMPs, although it would be necessary to further investigate whether α-(1→4) linked glucosyl residues still present within the IMMPs would escape digestion and enter the colon *in vivo*.

pH and production of organic acids upon fermentation of IMMP-27 and IMMP-dig27

For IMMP-27, the pH dropped continuously from the beginning of the fermentation until 24 h, which agrees with the steadily increasing level of SCFAs, lactic acid and succinic acid produced during the first 24 h (Figure 7).

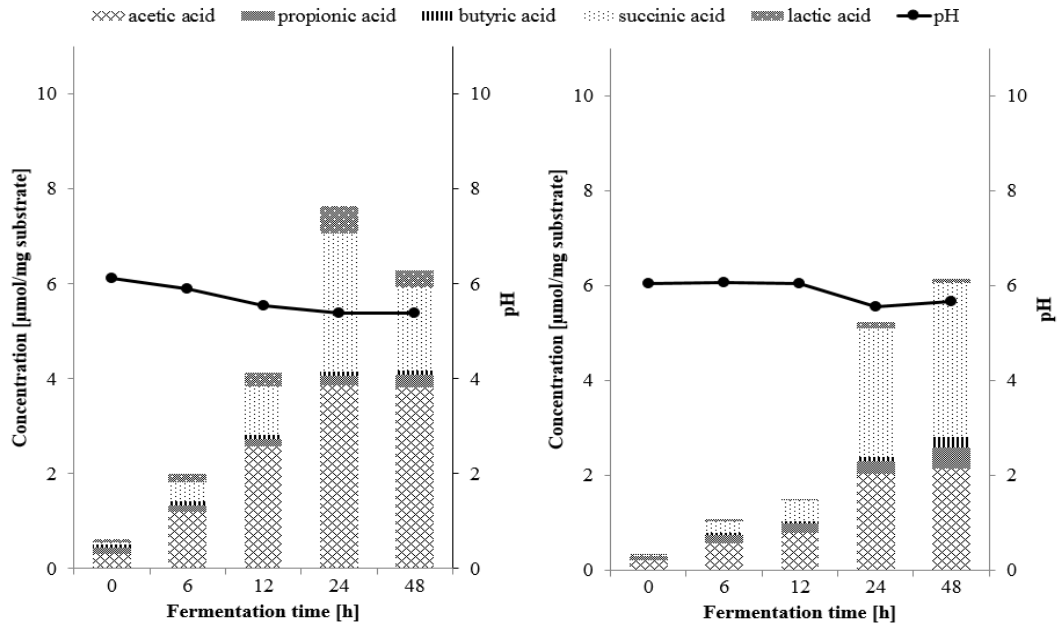


Figure 7. Concentrations of short chain fatty acids, lactic acid and succinic acid present and the pH (-●-) during *in vitro* fermentation with human faecal inoculum of (A) IMMP-27 and (B) IMMP-dig27.

From 24 h to 48 h, the pH remained stable and the concentrations of lactic acid and succinic acid decreased. During the fermentation of IMMP-dig27, the pH dropped between 12 and 24 h, concomitant with the most pronounced increase in the level of total organic acids, resembling the results for IMMP-94 and IMMP-96. The pH profiles and SCFAs production of IMMP-27 and IMMP-dig27 fermentation further confirmed that the human faecal microbiota used here readily utilized α-(1→4) glucan chains, whereas utilization of the α-(1→6) linked chains was delayed. A slight increase of pH (from 5.5 to 5.7) was observed between 24 and 48 h when fermenting IMMP-dig27. This could be an indication of proteolytic fermentation, of which one of the end-products is ammonia ([41], not measured in this study). The onset of proteolytic fermentation was possibly a result of carbohydrate depletion of IMMP-dig27 after 24 h fermentation.

Furthermore, acetic acid and succinic acid were the two major products for IMMP-27 and IMMP-dig27, as reported above for IMMP-94 and IMMP-96. Overall, the production of SCFAs with IMMP-dig27 resembled that with IMMP-94 and IMMP-96, except that the production of acetic acid was much lower in final concentration for IMMP-dig27 between 24 and 48 h (Figure 4, Figure 7). The lower production of acetic acid explained the slightly higher pH at 48 h in fermentation of IMMP-dig27 (5.7) compared to that of IMMP-94 (5.0) and IMMP-96 (5.0). Furthermore, within the first 12 h of fermentation of IMMP-27, where mainly the α-(1→4) linked glucose was utilized by faecal bacteria, succinic acid was already produced in large quantity (Figure 7). Therefore, the production of succinic acid was not specific to fermentation of the α-(1→6) linked glucosyl residues of IMMPs.

Enzyme activity upon fermentation of IMMPs

During *in vitro* fermentation, IMMP molecules with higher degree of polymerization (DP) were degraded into molecules with lower DP by enzymes that were produced by faecal microbiota, followed by further degradation into glucose, which was then utilised by bacteria present. To investigate which enzymes were produced during IMMP fermentation, proteins were extracted from fermentation digests at selected time points, chosen based on the HPAEC patterns: IMMP-94 (12 h & 24 h), IMMP-27 (12 h & 48 h) and IMMP-dig27 (12 h & 24 h). These time points indicated the time before the α-(1→6) linked glucan chains started to be degraded (all three IMMPs), the time when IMOs of DP 7 – 20 were predominantly present (IMMP-27 and IMMP-94) or even fully utilized (IMMP-dig27). IMMP-94 was used to represent IMMPs that were rich in the α-(1→6) linkages. Besides, the IMMP blank which contained inoculum with no IMMPs at time 0 h was included as the baseline of enzyme activity. From all time points, two types of enzyme extracts were obtained: extracellular enzyme extract (EE) and cell-associated enzyme extract (CE). Four substrates, PNP-α-D-glucopyranoside and PNP-β-D-glucopyranoside, potato starch and IMMP-94, were tested to determine the presence and activity of α- and β-1,4-glucosidases, starch-degrading enzymes and dextran-degrading enzymes (Table 1).

Table 1. Enzyme activity in mU (nmol mL-digest^{-1} min^{-1}) of enzyme extracts (EE: extracellular enzymes; CE: cell-associated enzymes) from *in vitro* fermentation samples.

***In vitro* fermentation substrates**	**Fermentation time [h]**	**Enzyme extract**	**Enzyme assay substrates**			
			PNP-α-D-glucopyranoside	**PNP-β-D-glucopyranoside**	**Soluble potato starch**	**IMMP 94**
Inoculum blank	0	EE	31	27	-	-
		CE	0.5 [I]	0.2	-	-
IMMP 94	12	EE	32	47	43	46
		CE	71 [II]	2.5	1	0.4
	24	EE	7 [II]	2	20	69
		CE	448	46	6	29
IMMP 27	12	EE	45	25	14	21
		CE	127 [II]	10 [I]	1	-
	48	EE	61	4	60	20
		CE	101 [I]	13	28	8
IMMP dig27	12	EE	64 [I]	N.A.	36	31
		CE	1 [I]	0.03	1	7
	24	EE	25	6	17	32
		CE	76	7	4	74

[I] Results given by single test; [II] results given by duplicates; all other results given by triplicates.
-Not detectable.
N.A. Not analysed.

At baseline (IMMP blank-0 h), all enzyme activities measured were negligible, especially the starch/dextran-degrading enzymes, as neither CE nor EE showed detectable activity towards soluble potato starch or IMMP-94 (Table 1). When IMMPs were present during the fermentation, the enzyme activities towards PNP-α-D-glucopyranoside increased at 12 h, especially in CE of IMMP-27-12 h (127 mU compared to 0.5 mU in IMMP blank). For both IMMP-94 and IMMP-dig27, the enzyme activities of CE towards PNP-α-D-glucopyranoside at 24 h were much higher than those at 12 h. In general, the enzyme activity towards PNP-α-D-glucopyranoside was much higher than the activity towards PNP-β-D-glucopyranoside for all enzyme extracts, suggesting that the microbiota was induced to produce enzymes to degrade the α-glucans used in this study. Enzyme activities towards soluble potato starch and IMMP-94 were also higher when IMMPs were present as substrates in the fermentation, and the enzyme activities of EE were much higher than that of CE. The EE enzyme extracts of IMMP-94 showed an increasing activity towards IMMP-94 from 12 h (46 mU) to 24 h (69 mU), whereas a declining activity towards soluble potato starch from 12 h (43 mU) to 24 h (20 mU) was observed. This confirms that the production of α-(1→6) linked glucose hydrolytic enzymes was induced by the presence of α-(1→6) linkage-rich substrates after the disappearance of α-(1→4) linkages. The decrease in activity of α-(1→4) linked glucose endo-acting enzyme was most probably due to the absence of starch, and the α-(1→4) linked glucose hydrolytic enzyme that

was found active at the beginning was no longer produced during later stages of the fermentation.

The overall distribution of the four enzyme activities in CE and EE followed a certain tendency: activities towards soluble potato starch and IMMP-94, i.e. α-amylase and dextranase, were higher in EE than in CE, whereas activities towards PNP-α/β-D-glucopyranoside were higher in CE than in EE. This suggests that α-(1→4) and α-(1→6) linked glucose polysaccharide-degrading enzymes, which comprise mainly of endo-acting enzymes [14], were excreted by microbes to cleave IMMP polysaccharides into smaller oligosaccharides. These smaller oligosaccharides could then be taken up by microbial cells to be further degraded by glucosidases, which are exo-acting enzymes. This agrees with previous findings that exo-acting enzymes were mostly cell-bound whereas end-acting enzymes were mostly extracellular [14]. Also the absence of IMOs of DP lower than 7 in the well-separated IMO fraction in the HPAEC chromatograms (Figure 3, Figure 6) seems to match this theory, because bacterial cells, together with the smaller oligosaccharides that had already passed the cell membrane, were removed from fermentation digest by centrifugation before HPAEC analysis.

According to HPAEC (Figure 6), degradation of IMMP-27 was mainly targeting α- α-(1→4) linkages in the first 12 h, and switched to α-(1→6) linked glucose residues afterwards. Furthermore, at 48 h, IMOs of DP 7-20, which were products of degradation of IMMP polysaccharides by endo-acting enzymes, were present. This means that the α-(1→4) linkage degrading enzymes were active during the first 12 h of fermentation, whereas afterwards, glycanase activity was taken over by the α-(1→6) linkage degrading enzymes. However, this did not agree with the enzyme activities measured during fermentation of IMMP-27: the combined CE and EE enzyme activities towards soluble potato starch were higher at 48 h (88 mU) than at 12 h (15 mU). In addition, the enzyme extracts of IMMP-27-48 h showed higher combined CE and EE enzyme activities towards soluble potato starch (88 mU) than towards IMMP-94 (28 mU). This suggests that production of α-(1→4) linkage degrading enzymes was not suppressed after the substrates were depleted.

As to IMMP-dig27, the combined CE and EE enzyme activity towards soluble potato starch declined from 12 h (37 mU) to 24 h (21 mU), whereas the activity towards IMMP-94 increased from 12 h (38 mU) to 24 h (106 mU). This observation suggested that the microbial enzyme production of IMMP-dig27 fermentation resembled that of IMMP-94 fermentation. This agreed with the results of molecular degradation patterns and SCFA production, as discussed above.

Microbiota composition during fermentation of IMMPs

The microbiota composition during fermentation of IMMPs was analysed to evaluate the prebiotic potential of IMMPs, and to make a link with the structural changes of IMMPs and production of SCFAs, lactic acid and succinic acid. Multivariate analysis of bacterial community dynamics over time in the different *in vitro* fermentations, using weighted Unifrac distances as a measure for differences in microbial composition, showed a directional shift in community composition in relation to incubation time and the type of IMMPs used. A strong segregation of samples with IMMPs present after 24/48 h of incubation could be observed

(Figure 8). This indicates that both the duration of incubation and the presence of different IMMP substrates played an important role in shaping the microbial communities *in vitro*.

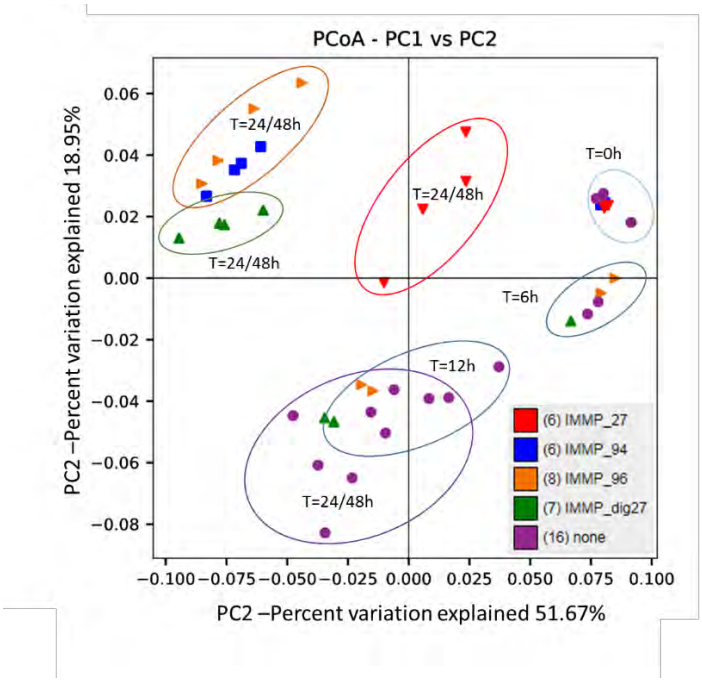


Figure 8. PCoA based on weighted (relative abundance) Unifrac distances between observed microbial communities for *in vitro* fermentation of IMMPs with human faecal inoculum at different time points.

A similar segregation of samples was also found with unweighted analyses that only take presence and absence of microbial groups into account (data not shown). The microbial alpha diversity, as determined based on Shannon's diversity index, changed as the fermentation progressed and decreased in the blank, but increased in digesta with the IMMPs present (Figure 9).

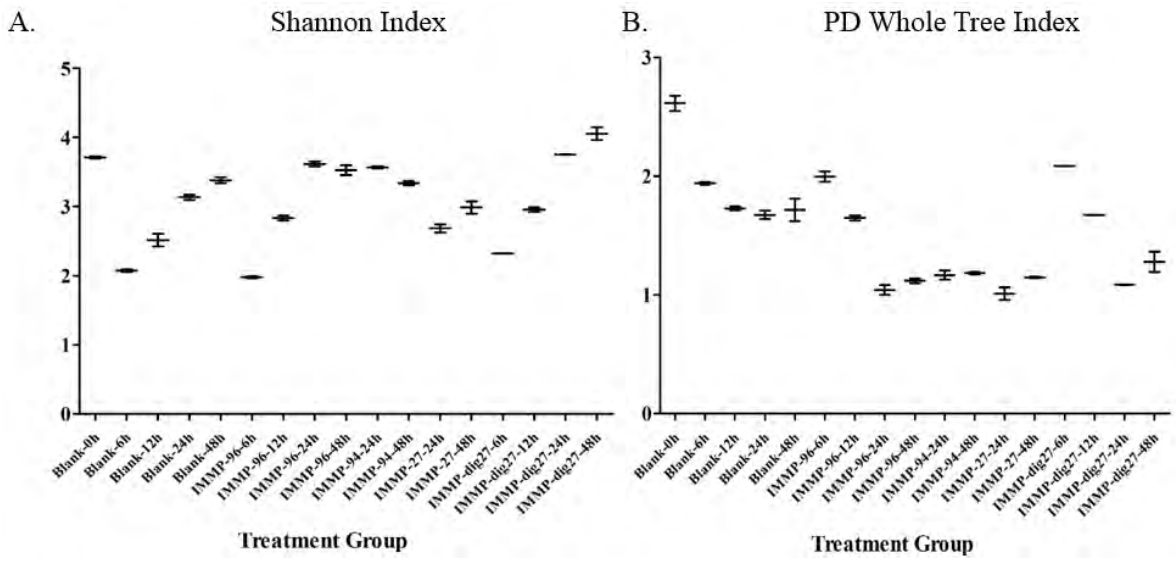
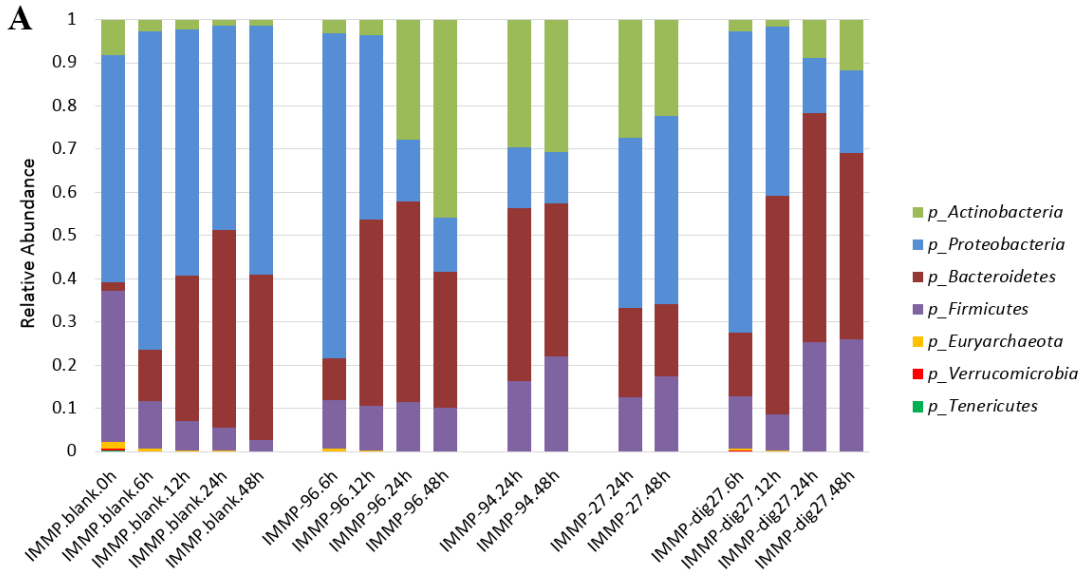


Figure 9. Microbial alpha-diversity estimates, including (A) Shannon diversity index, (B) Phylogenetic Diversity Whole Tree for *in vitro* fermentation of IMMPs with human faecal inoculum at different time points.

Shannon's diversity index accounts for both abundance and evenness of the species present. There was a high predominance of members of the *Escherichia-Shigella* group at the beginning of the fermentation, possibly due to the presence of residual amounts of oxygen during initial inoculum activation. As fermentation progressed, the presence of IMMPs and depletion of oxygen enabled growth of other bacterial groups leading to an increase in the evenness of the community. Although the microbiota composition at the start of fermentation was different from that normally found in faeces of healthy adults, it is interesting to note that such a dysbiotic community was "normalized" by IMMPs towards a more typical colonic microbiota. It is tempting to speculate that this "normalization" effect might also occur *in vivo* and could facilitate ecosystem recovery following states of dysbiosis (e.g. after diarrhoea). In the IMMP blank sample, the ecosystem was starved, thus the growth of other bacterial groups was much slower. Phylogenetically weighted species richness, as measured by the PD (Phylogenetic Diversity) Whole Tree index, decreased in all treatment groups in the first hours of incubation, whereas it remained relatively stable after 24 h. Despite high structural similarities between IMMP-94 and IMMP-96, Pearson correlation scores at genus level were 0.77, 0.89 and 0.34 at times 0, 24 and 48 h, respectively, suggesting different microbial response patterns towards these two substrates. In line with this observation, the average relative abundance of different phyla changed with time, and was influenced by the type of IMMP being fermented (Figure 10A). Levels of Proteobacteria decreased in all groups until 24 h of incubation and remained stable or slightly increased at 48 h. This was accompanied by a gradual increase in Bacteroidetes up to 24 h, followed by decrease at 48 h. Firmicutes showed a rapid decrease in abundance at 6 h and gradual increase at later time points, except for IMMP blank where their relative abundance continued to decline. The levels of Actinobacteria were very low, and decreased to 4.7% in the IMMP blank at 48 h. Their relative abundance was higher in the IMMP digesta as compared to the IMMP blank.



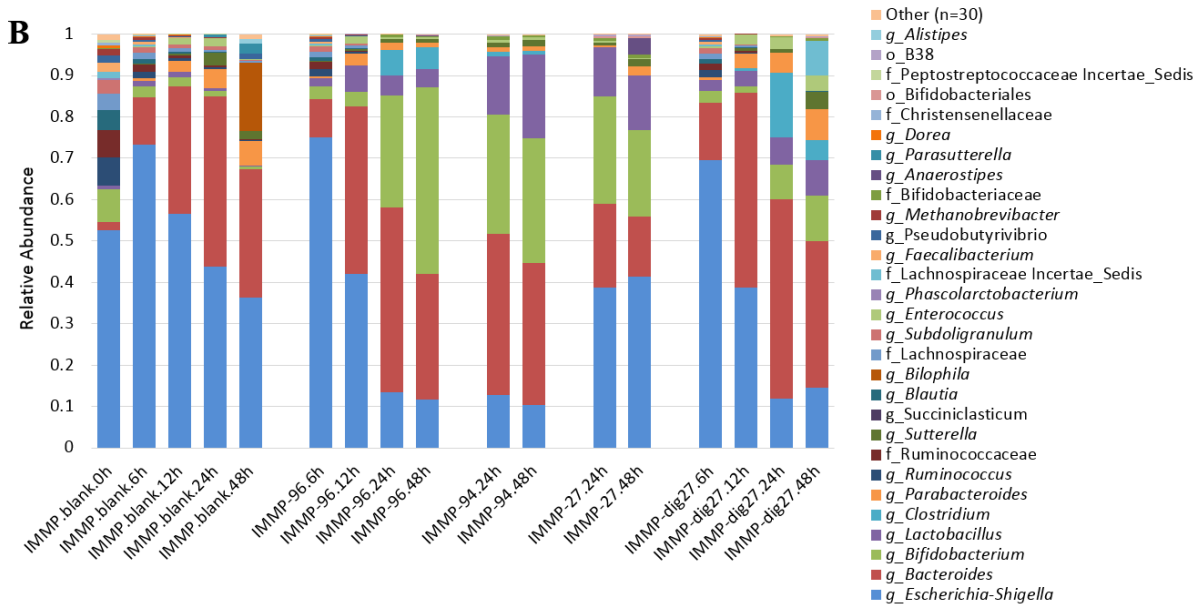


Figure 10. Relative abundance of taxa detected during *in vitro* fermentation of IMMPs with human faecal inoculum at different time points, considering phylum (A) and genus (B) levels.

At the genus level, four bacterial taxa, namely *Escherichia-Shigella, Bacteroides, Bifidobacterium* and *Lactobacillus* were predominant in all IMMP digesta, with their combined relative abundance ranging from 41% to 97% of all detected reads (Figure 10B). The detailed relative abundance of taxa during IMMPs fermentation at genus level is given in Supporting Table S1. The duration of *in vitro* fermentation was positively correlated with the increase of the relative abundance of *Bacteroides* and a corresponding decrease in *Escherichia-Shigella*. The presence of IMMPs correlated with high (up to 50%) relative abundance of genera *Bifidobacterium* and *Lactobacillus*, as compared to the IMMP blank group, in which less than 5% of all reads belonged to these taxa. This prebiotic effect was especially strong in IMMP-27, IMMP-94 and IMMP-96 after 24 h of incubation, at which time the fermentation of α-(1→6) linked glucosyl residues was predominant. The increase in relative abundance of *Bifidobacterium* and *Lactobacillus* was specific to the presence of IMMPs and was not observed in the IMMP blank, indicating that fermentation of IMMPs promoted the growth of *Bifidobacterium* and *Lactobacillus* (Table 2).

Table 2. Genus level taxa with significantly different relative abundance in combined IMMP groups at 24 h and 48 h of incubation as compared to IMMP blank groups at 24 h and 48 h using Kruskal-Wallis analysis.

Taxon	p value	FDR	Bonferroni	Average RA	
				IMMP	Blank
g_Bifidobacterium	0.00001	0.00006	0.00029	**0.26264**	0.00803
g_Escherichia-Shigella	0.01045	0.02438	0.43878	0.18368	**0.31904**
g_Lactobacillus	0.00001	0.00006	0.00029	**0.09625**	0.00460
g_Parabacteroides	0.00922	0.02278	0.38733	0.02309	**0.04232**
g_Sutterella	0.00023	0.00109	0.00981	0.01420	**0.03009**
f_Bifidobacteriaceae_g_g	0.00000	0.00006	0.00017	**0.00583**	0.00000
o_Bifidobacteriales_g_g	0.00479	0.01341	0.20118	**0.00188**	0.00000
g_Parasutterella	0.00000	0.00005	0.00005	0.00018	0.01251
f_Lachnospiraceae_g_g	0.00001	0.00006	0.00024	0.00000	**0.00285**
g_Ruminococcus	0.00003	0.00016	0.00111	0.00000	**0.00265**
g_Subdoligranulum	0.00003	0.00016	0.00111	0.00000	**0.00230**
g_Bilophila	0.00011	0.00060	0.00476	0.00000	**0.07763**
g_Eggerthella	0.00045	0.00171	0.01879	0.00000	**0.00314**
g_Butyricimonas	0.00045	0.00171	0.01879	0.00000	**0.00190**
g_Alistipes	0.00164	0.00492	0.06895	0.00000	**0.00452**
g_Pseudobutyrivibrio	0.00164	0.00492	0.06895	0.00000	**0.00490**
f_Ruminococcaceae_g_g	0.00164	0.00492	0.06895	0.00000	**0.00130**
g_Methanobrevibacter	0.00565	0.01484	0.23749	0.00000	**0.00075**

For IMMP-96 and IMMP-dig27, the relative abundance of *Bifidobacterium* remained very low in the first 12 h (2 – 5%), then increased rapidly to a high level at 24 h (27% for IMMP-96; 9% for IMMP-dig27). From 24 h to 48 h, *Bifidobacterium* relative abundance continued to largely increase for IMMP-96 (44%), whereas it only slightly increased for IMMP-dig27 (11%). The growth pattern of bifidobacteria was in line with the degradation pattern of IMMPs which consisted mostly of α-(1→6) linkages, as both started only after 12 h of fermentation. The highest increase in relative abundance of bifidobacteria occurred from 12 h to 24 h, where IMMP polysaccharides were degraded into α-(1→6) linked IMOs with DP of 7 to over 20. The growth pattern of bifidobacteria also agreed with the formation of SCFAs, as shown previously. Formation of SCFAs during fermentation contributes to acidification of the colonic lumen [4]. A lower pH in the colon is favourable for bifidobacteria and lactic acid bacteria, while impeding the overgrowth of more pH-sensitive pathogenic bacteria [2, 42] For IMMP-94 and IMMP-27, microbiota composition at 6 h and 12 h was not analysed, due to a scarcity of the fermentation digest. Both IMMPs showed high levels of *Bifidobacterium* at 24 h (29% for IMMP-94; 25% for IMMP-27), and at 48 h the relative abundance of this genus remained almost the same for IMMP-94 (30%) but slightly decreased for IMMP-27 (21%).
The observed changes in relative abundance of *Lactobacillus* differed among different IMMPs, with the strongest increase observed for IMMP-94 and IMMP-27 at 24 h and 48 h, whereas the increase in relative abundance was weaker with IMMP-96 and IMMP-dig27. In the presence of IMMP-96 there was a rapid increase in the relative abundance at 6 h, followed by a gradual decline at later time points, whereas with IMMP-dig27 the relative abundance of this genus showed a steady increase with time. The more pronounced increase in relative abundance of *Lactobacillus* during the fermentation of IMMP-94 and IMMP-27 as compared to the other two

substrates was in line with the higher level of lactic acid produced during fermentation of IMMP-94 and IMMP-27 (Figure 4 and Figure 7). In addition, there was an increase in the relative abundance of genera *Enterococcus* and *Parabacteroides* during the fermentation of IMMP-dig27 and with the IMMP blank, but not for the other IMMPs (Figure 10B), a result which we cannot explain in a straightforward way. However, there is growing evidence suggesting that metabolic webs and complex polysaccharide utilization networks exist between different members of intestinal microbiota, with different species specializing to utilize different polysaccharides, expanding the number and types of glycoside hydrolase produced in the presence of a competitor, or acting as producers or recipients of the polysaccharide breakdown products [43, 44].

We observed a high accumulation of succinate during the *in vitro* fermentation of all IMMPs. This might be due to activity of *Bacteroides* which in the gut can use CO_2 to reduce formate to succinate to generate ATP in a primitive electron transport chain [43]. Succinate is then excreted as an end product and can be utilised by secondary fermenters, or it can be further converted by *Bacteroides* to propionate if the CO_2 is limiting. In fact, the ability to convert succinate to propionate has been described for both Bacteroidetes and Veillonellaceae [31]. In auxotrophic *Bacteroides* spp. this conversion of succinate to propionate is modulated by the availability of vitamin B12, which in the gut is produced by certain members of Firmicutes and Actinobacteria [45]. Thus, the accumulation of succinic acid in our experiment could be, among others, a result of high CO_2 levels, vitamin B12 limitation, or might be linked to the absence of members of the family Veillonellaceae, which were not detected in our microbiota composition analysis (Figure 10B).

Concluding remarks

The IMMPs in this study showed delayed and slow-fermenting behaviour compared to other prebiotics during their *in vitro* fermentation by a human faecal inoculum. Measurable production of enzymes targeting α-(1→6) linked glucose was detected after 12 h of incubation. The presence of α-(1→4) linked glucosyl linkages in the IMMPs further postponed the bacterial utilization of α-(1→6) linked glucosyl residues, suggesting that when available, the α-(1→4) linked glucosyl residues are preferentially used by the faecal microbiota. We also found that α-(1→6) linked glucose oligomers with lower DP were preferentially used, as compared to those with higher DP. Organic acids were produced at high total amounts during IMMPs fermentation, with acetic acid and succinic acid being the predominant metabolites in all incubations. The HPAEC chromatograms and enzyme production analysis showed that the polysaccharide fraction of IMMPs was degraded mainly by extracellular enzymes into α-(1→6) linked IMOs, among which the IMOs with DP lower than 7 might be transported into microbial cells and further degraded by cytoplasmic enzymes. Fermentation of IMMPs led to the increase in diversity and evenness of bacterial communities, and promoted the increase in relative abundance especially of genera *Bifidobacterium* and *Lactobacillus*, lending a strong support for the prebiotic potential of these fibres.

References

1. Gibson GR, Hutkins R, Sanders ME, Prescott SL, Reimer RA, Salminen SJ, Scott K, Stanton C, Swanson KS, Cani PD, Verbeke K, Reid G. 2017. *Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics*. Nat Rev Gastroenterol Hepatol **14**:491–502.
2. Gibson GR. 2004. Fibre and effects on probiotics (the prebiotic concept). Clin Nutr Suppl **1**:25–31.
3. Scott KP, Martin JC, Duncan SH, Flint HJ. 2014. *Prebiotic stimulation of human colonic butyrate-producing bacteria and bifidobacteria, in vitro*. FEMS Microbiol Ecol 87:30–40.
4. Verspreet J, Damen B, Broekaert WF, Verbeke K, Delcour JA, Courtin CM. 2016. *A Critical Look at Prebiotics Within the Dietary Fiber Concept*. Annu Rev Food Sci Technol 7:annurev-food-081315-032749.
5. Dobruchowska JM, Gerwig GJ, Kralj S, Grijpstra P, Leemhuis H, Dijkhuizen L, Kamerling JP. 2012. *Structural characterization of linear isomalto-/malto-oligomer products synthesized by the novel GTFB 4, 6-α-glucanotransferase enzyme from Lactobacillus reuteri 121*. Glycobiology **22**:517–528.
6. Leemhuis H, Dobruchowska JM, Ebbelaar M, Faber F, Buwalda PL, van der Maarel MJEC, Kamerling JP, Dijkhuizen L. 2014. *Isomalto/malto-polysaccharide, a novel soluble dietary fiber made via enzymatic conversion of starch.* J Agric Food Chem **62**:12034–12044.
7. Goffin D, Delzenne N, Blecker C, Hanon E, Deroanne C, Paquot M. 2011. *Will isomalto-oligosaccharides, a well-established functional food in Asia, break through the European and American market? The status of knowledge on these prebiotics.* Crit Rev Food Sci Nutr **51**:394–409.
8. Kaulpiboon J, Rudeekulthamrong P, Watanasatitarpa S, Ito K, Pongsawasdi P. 2015. *Synthesis of long-chain isomaltooligosaccharides from tapioca starch and an in vitro investigation of their prebiotic properties.* J Mol Catal B Enzym **120**:127–135.
9. Ketabi A, Dieleman LA, Gänzle MG. 2011. *Influence of isomalto ‐ oligosaccharides on intestinal microbiota in rats.* J Appl Microbiol **110**:1297–1306.
10. Rycroft CE, Jones MR, Gibson GR, Rastall RA. 2001. *A comparative in vitro evaluation of the fermentation properties of prebiotic oligosaccharides.* J Appl Microbiol **91**:878–887.
11. Olano-Martin E, Mountzouris KC, Gibson GR, Rastall RA. 2000. *In vitro fermentability of dextran, oligodextran and maltodextrin by human gut bacteria.* Br J Nutr **83**:247–255.
12. Ashwar BA, Gani A, Shah A, Wani IA, Masoodi FA. 2015. *Preparation, health benefits and applications of resistant starch-a review.* Starch/Staerke 287–301.
13. Jonathan MC, Haenen D, Souza Da Silva C, Bosch G, Schols HA, Gruppen H. 2013. *Influence of a diet rich in resistant starch on the degradation of non-starch polysaccharides in the large intestine of pigs*. Carbohydr Polym **93**:232–239.
14. Rösch C, Venema K, Gruppen H, Schols HA. 2015. *Characterisation and in vitro fermentation of resistant maltodextrins using human faecal inoculum and analysis of bacterial enzymes present.* Bioact Carbohydrates Diet Fibre **6**:46–53.
15. Ramasamy US, Schols HA, Gruppen H. 2014. *Characteristics of bacterial enzymes present during in vitro fermentation of chicory root pulp by human faecal microbiota.* Bioact Carbohydrates Diet Fibre **4**:115–124.
16. van der Zaal PH, Schols HA, Bitter JH, Buwalda PL. 2017. *Isomalto/malto-polysaccharide structure in relation to the structural properties of starch substrates.* Carbohydr Polym 0–1.
17. Bai Y, van der Kaaij RM, Leemhuis H, Pijning T, Leeuwen SS van, Jin Z, Dijkhuizen L. 2015. *Biochemical characterization of the Lactobacillus reuteri glycoside hydrolase family 70 GTFB type of 4,6-α-glucanotransferase enzymes that synthesize soluble dietary starch fibers.* Appl Environ Microbiol **81**:7223–7232.
18. Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F. 1956. *Colorimetric Method for Determination of Sugars and Related Substances.* Anal Chem **28**:350–356.
19. Saha AK, Brewer CF. 1994. Determination of the concentrations of oligosaccharides, complex

type carbohydrates, and glycoproteins using the phenol-sulfuric acid method. Carbohydr Res **254**:157–167.
20. Aguirre M, Ramiro-Garcia J, Koenen ME, Venema K. 2014. *To pool or not to pool? Impact of the use of individual and pooled fecal samples for in vitro fermentation studies.* J Microbiol Methods **107**:1–7.
21. Minekus M, Smeets-Peeters M, Bernalier A, Marol-Bonnin S, Havenaar R, Marteau P, Alric M, Fonty G, Huis in't Veld JHJ. 1999. *A computer-controlled system to simulate conditions of the large intestine with peristaltic mixing, water absorption and absorption of fermentation products.* Appl Microbiol Biotechnol **53**:108–114.
22. Ladirat SE, Schols HA, Nauta A, Schoterman MHC, Schuren FHJ, Gruppen H. 2014. *In vitro fermentation of galacto-oligosaccharides and its specific size-fractions using non-treated and amoxicillin-treated human inoculum.* Bioact Carbohydrates Diet Fibre **3**:59–70.
23. Ramiro-Garcia J, Hermes GDA, Giatsis C, Sipkema D, Zoetendal EG, Schaap PJ, Smidt H. 2016. *NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes.* F1000Research **5**:1791.
24. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. *QIIME allows analysis of high-throughput community sequencing data.* Nat Methods **7**:335–6.
25. Kuczynski J, Stombaugh J, Walters WA, Gonz??lez A, Caporaso JG, Knight R. 2012. *Using QIIME to analyze 16s rRNA gene sequences from microbial communities.* Curr Protoc Microbiol 1–30.
26. Van Der Maarel MJEC, Capron I, Euverink GJW, Bos HT, Kaper T, Binnema DJ, Steeneken PAM. 2005. *A novel thermoreversible gelling product made by enzymatic modification of starch.* Starch/Staerke **57**:465–472.
27. Wu Q, Pi X, Liu W, Chen H, Yin Y, Yu HD, Wang X, Zhu L. 2017. *Fermentation properties of isomaltooligosaccharides are affected by human fecal enterotypes.* Anaerobe **48**:206–214.
28. Macfarlane S, Macfarlane GT. 2003. *Regulation of short-chain fatty acid production*. Proc Nutr Soc **62**:67–72.
29. Watanabe Y, Nagai F, Morotomi M. 2012. *Characterization of Phascolarctobacterium succinatutens sp. Nov., an asaccharolytic, succinate-utilizing bacterium isolated from human feces.* Appl Environ Microbiol **78**:511–518.
30. Reichardt N, Duncan SH, Young P, Belenguer A, McWilliam Leitch C, Scott KP, Flint HJ, Louis P. 2014. *Phylogenetic distribution of three pathways for propionate production within the human gut microbiota.* ISME J.
31. Flint HJ, Scott KP, Louis P, Duncan SH. 2012. *The role of the gut microbiota in nutrition and health*. Nat Rev Gastroenterol Hepatol **9**:577–589.
32. Biswas A, Shogren RL, Kim S, Willett JL. 2006. *Rapid preparation of starch maleate half-esters.* Carbohydr Polym **64**:484–487.
33. Sahota SS, Bramley PM, Menzies IS. 1982. *The fermentation of lactulose by colonic bacteria.* J Gen Microbiol **128**:319–325.
34. Demigné C, Jacobs H, Moundras C, Davicco M-JJ, Horcajada M-NN, Bernalier A, Coxam V. 2008. *Comparison of native or reformulated chicory fructans, or non-purified chicory, on rat cecal fermentation and mineral metabolism*. Eur J Nutr **47**:366–374.
35. Dimitrovski D, Velickova E, Dimitrovska M, Langerholc T, Winkelhausen E. 2016. *Synbiotic functional drink from Jerusalem artichoke juice fermented by probiotic Lactobacillus plantarum PCS26.* J Food Sci Technol **53**:766–74.
36. Leijdekkers AGM, Aguirre M, Venema K, Bosch G, Gruppen H, Schols HA. 2014. *In vitro fermentability of sugar beet pulp derived oligosaccharides using human and pig fecal inocula.* J Agric Food Chem **62**:1079–1087.
37. Wong JMW, de Souza R, Kendall CWC, Emam A, Jenkins DJA. 2006. *Colonic health: fermentation and short chain fatty acids.* J Clin Gastroenterol **40**:235–243.
38. Kohmoto T, Fukui F, Takaku H, Machida Y, Arai M, Mitsuoka T. 1988. *Effect of Isomalto-oligosaccharides on Human Fecal Flora.* Bifidobact Microflora **7**:61–69.

39. Kohmoto T, Fukui F, Takaku H, Mitsuoka T. 1991. *Dose-response Test of Isomaltooligosaccharides for Increasing Fecal Bifidobacteria.* Agric Biol Chem **55**:2157–2159.
40. Macfarlane GT, Englyst HN. 1986. *Starch utilization by the human large intestinal microflora.* J Appl Bacteriol **60**:195–201.
41. Manning TS, Gibson GR. 2004. *Prebiotics.* Best Pract Res Clin Gastroenterol 18:287–298.
42. Topping DL, Clifton PM. *Short-Chain Fatty Acids and Human Colonic Function . Roles of resistant starch and nonstarch polysaccharides.* Physiol Rev **81**: 1031-1064.
43. Fischbach MA, Sonnenburg JL. 2011. *Eating for two: How metabolism establishes interspecies interactions in the gut.* Cell Host Microbe **10**:336–347.
44. Datta M Sen, Gore J. 2014. *Evolution: "Snowed" in with the enemy.* Curr Biol 24:R33–R35.
45. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI. 2009. *Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat.* Cell Host Microbe **6**:279–289.

Chapter 7

Isomalto/malto-polysaccharides maintain normal gut functioning while promoting growth and activity of beneficial bacteria

Klaudyna Borewicz*, Bastian Hornung*, Fangjie Gu, Pieter H. van der Zaal, Henk Schols, Peter J. Schaap, Hauke Smidt

* These authors contributed equally to this work

Manuscript submitted

Abstract

Isomalto/malto-polysaccharides (IMMPs) are a novel type of soluble dietary fibres with a prebiotic potential capable of promoting growth of beneficial microbes in the gut. However, the mode of action of IMMPs remains unknown. Previous studies on IMMPs showed an increase in total bacteria, especially lactobacilli, and a higher production of short chain fatty acids (SCFA) when IMMPs were fed to rats or used as a carbon source during *in vitro* fermentation. In this study, we investigated with metatranscriptomics how IMMPs with different amounts of α-(1→6) glycosidic linkages affected microbial function during incubation with human faecal inoculum. We showed that microbial community dynamics during fermentation varied depending on the type of IMMP used and that the observed changes were reflected in the community gene expression profiles based on metatranscriptome analysis. Members of *Bacteroides*, *Lactobacillus* and *Bifidobacterium* were the predominant degraders of IMMPs, and the increased activity of these bacteria correlated with high amounts of α-(1→6) glycosidic linkages. We also noted an increase in relative abundance of these bacteria and an activation of pathways involved in SCFA synthesis. Our findings could provide a baseline for more targeted approaches in designing and engineering prebiotics for specific bacteria and to achieve more controlled modulation of microbial composition and activity towards desired health outcomes.

Introduction

The human gut is a home to a diverse ecosystem inhabited by bacteria, archaea, viruses and eukaryotes, which play an important role in their host's health and well-being [1-3]. These organisms interact with each other and with the host via a complex network of relations, and knowing the mechanisms of these interactions and how to influence them might provide a useful tool for refining the function of this ecosystem to promote homeostasis and to strengthen host's immunity against infections [4]. Currently there are only a few ways to manipulate the composition and function of the gut microbiota. These range from extreme measures, such as the use of antibiotics [5] or faecal transplantations [6], to milder ones, such as the implementation of various dietary regimes and the use of dietary supplements, especially pro- and prebiotics [7]. Prebiotics are complex carbohydrates, often soluble dietary fibres, that cannot be digested by human enzymes but are readily used by the colonic microbiota and provide a health benefit for the host [8]. A range of different prebiotics may preferably stimulate growth and activity of specific microbial groups (e.g. butyrogenic bacteria [9]), leading to the production of different metabolites with health-supporting effects. However, the exact mode of action of most prebiotics remains unknown and their specific impact on microbial interactive networks needs to be investigated.

Isomalto/malto-polysaccharides (IMMPs) comprise a novel class of soluble dietary fibres with prebiotic potential. These fibres are synthetized from starch by enzymatic conversion of α-(1→4) glycosidic linkages into α-(1→6) glycosidic linkages by 4,6-α-glucanotransferase (GTFB) from *Lactobacillus reuteri* 121 [10]. The resulting α-(1→6) linkages present in IMMPs make these fibres resistant to digestion by human digestive enzymes in the small intestine. As such, this modified starch can pass undigested into the large intestine where it is fermented by the resident microbes capable of breaking down the α-(1→6) glycosidic linkages. This property of the IMMPs makes them potentially interesting as a prebiotic food ingredient capable of modulating the intestinal microbiota and exerting health promoting effects onto the host. A previous study has reported an increased production of short chain fatty acids (SCFA), especially acetate and propionate, when IMMPs were used as a carbon source for microbial *in vitro* fermentation with human faecal inoculum as the microbial source [10]. In this study, we investigated the effects of three different IMMPs on microbial composition and function during *in vitro* batch fermentations with faecal inoculum from healthy human adults. Here we show that specific changes of the microbiota, such as growth of *Bifidobacterium* and *Lactobacillus* can be attributed to the IMMPs, and that these changes are also reflected at the transcriptomic level, i.e. upregulation of specific gene groups, as well as in enzymatic activity and an increase in production of SCFAs.

Materials and Methods

In vitro fermentation; design and sampling

The faecal inoculum stock was prepared at TNO (Zeist, The Netherlands) from fresh faeces of seven healthy adult donors. The stock was mixed, aliquoted and stored anaerobically

at -80 °C [11]. Sterile 20 mL anaerobic serum bottles were filled with 10 mL of the Standard Ileal Efflux Medium (SIEM; Tritium Microbiology, Eindhoven, The Netherlands). The SIEM was prepared following the manufacturer's guidelines, but omitting the carbohydrate source and Tween 80. Before inoculation, a faeces stock aliquot was mixed with SIEM at 1:10 v/v and incubated overnight at 37 °C. The activated inoculum was then added to the fermentation bottles at 1% (v/v) final concentration. Three different IMMP fibres were tested, with 27% (IMMP-27), 94% (IMMP-94) and 96% (IMMP-96) of total α-(1→6) glycosidic linkages. In addition, a pre-treated IMMP-27 (IMMP-dig27) sample was included after it had been digested with α-amylase and amyloglucosidase to imitate passage through the small intestine (Chapter Six of this thesis). Samples were prepared and processed in duplicate with fibres added to individual fermentation bottles at a final concentration of 10 mg/mL. Flasks were incubated at 37 °C, and 0.5 to 2 mL of each culture was removed at different sampling time points, depending on the experiment.

In experiment A, cultures supplied with two different prebiotic fibres (IMMP-27 and IMMP-94) and one control culture without any substrate (IMMP blank) were monitored over 48 hours (in duplicate), and aliquots were removed at time points 0 (up to 15 min after addition of the prebiotic), 24 h and 48 h. In experiment B, cultures were supplied with two other prebiotics, IMMP-dig27 and IMMP-96, and one culture was left with no prebiotic (IMMP blank). Experiment B was monitored for 48 hours, and samples were taken at 6 h, 12 h, 24 h and 48 h (in duplicate). An aliquot of the activated blank inoculum was taken at time point 0, just before the addition of the IMMP. All samples (18) from experiment A were subjected to metatranscriptomic sequencing. In experiment B the metatranscriptomics sequencing was done for the activated inoculum at time point 0 h and for the treatment groups at all time points (17). Samples for metatranscriptomics were harvested and immediately stabilized in RNAprotect (Qiagen, Hilden, Germany) following the manufacturer's instructions, and bacterial pellets were stored at -80 °C for up to three weeks before processing.

RNA extraction and Illumina sequencing

Total RNA was extracted by using the beat beating - TRIzol - column method modified from Kang et al. [12]. Briefly, bacterial pellets were re-suspended in 100 µL TE buffer (30 mM Tris-HCl, 1 mM EDTA, pH = 8.0) containing 15 mg/mL Lysozyme, 10 U/mL of Mutanolysin and 100 µg/mL of Proteinase K. Samples were vortexed for 10 s and incubated at room temperature for 10 min, and 400 µL of RLT buffer (Qiagen) containing 4 µL of β-mercaptoethanol was added. Samples were then vortexed, mixed with 500 µL of TRIzol Max reagent (Invitrogen, Carlsbad, CA, USA) and homogenized with 0.8 g of sterilized 0.1 mm zirconia beads for three min (3 × 1 min with cooling in between) at 5.5 ms using a bead beater (Precellys 24, Bertin Technologies). Following the beating step, samples were cooled on ice, gently mixed by inverting the tube with 200 µL of ice cold chloroform for 15 s and centrifuged for 15 min at 4 °C at 12,000 × g. The aqueous phase containing total RNA was transferred to fresh tubes and mixed with an equal volume of 70% ethanol. The mixture was placed on a Qiagen RNeasy mini column (RNeasy Mini Kit, Qiagen) and centrifuged at 8,000 × g for 15 s to bind RNA into the column. Filtrates were discarded, and the RNA binding step was repeated until the complete sample was filtered through the column.

The columns were rinsed with 350 µL of RW1 buffer (RNeasy Mini Kit, Qiagen), and 80 µL DNAse I solution (Roche, Manheim, Germany) was applied to the column and incubated for 15 min at RT to digest DNA. The columns were rinsed twice with 350 µL RW1 buffer, and twice with 700 µL of RPE buffer (RNeasy Mini Kit, Qiagen), following with a final wash with 80% ethanol. Columns were dried by a 2 min centrifugation at maximum speed, and total RNA was eluted with 30 µL of DNAse/RNAse free water. The total RNA concentrations were measured spectrophotometrically with an ND-1000 spectrophotometer (NanoDrop® Technologies, Wilmington, DE, USA), and residual DNA concentrations were measured with Qubit® dsDNA BR Assay Kit (Life Technologies, Leusden, the Netherlands). Samples which contained over 10 ng/µL DNA contamination were treated with Turbo DNAfree® Kit (Ambion, Bleiswijk, Netherlands) following manufacturer's instructions and purified using RNeasy Mini Kit. Total RNA quality was evaluated using Experion RNA StdSens kit (Biorad Laboratories INC, USA), total RNA concentrations were measured with NanoDrop® and DNA contamination concentrations were measured with Qubit®dsDNA BR Assay Kit (Life Technologies). Between 3-5 µg of total RNA from each sample was used for mRNA enrichment with RiboZero Bacterial rRNA Removal Kit (Illumina, San Diego, CA, USA), and the quality and quantity of enriched mRNA was assessed as described above for total RNA. Between 200-500 ng of enriched mRNA was used for cDNA production using ScriptSeq®v2RNA-Seq Library Preparation Kit (Epicentre, Madison, WI, USA), FailSafe®PCR Enzyme Mix (Epicentre) and ScriptSeq®Index PCR Primers (Epicentre) for amplification and barcoding of di-tagged cDNA. The PCR product presence was confirmed with gel electrophoresis using the FlashGel® System (Lonza, Rockland, ME, USA). PCR products were then purified with HighPrep® PCR kit (MagBio Genomics, Gaithersburg, MD, USA) and concentrations of indexed cDNA were measured using Qubit®dsDNA BR Assay Kit (Invitrogen). Approximately 28 ng of DNA from each sample was added to a pool, and final volume of each library was adjusted to 25 µL using HighPrep® PCR kit. Two libraries were prepared containing either 17 or 18 samples, with final concentrations of 20 ng/µL in each library. Libraries were sent for single end 150 bp Illumina HiSeq2000 sequencing (GATC, Konstanz, Germany).

Bioinformatic processing, read assembly and annotation

The bioinformatics workflow was adapted from Davids et *al.* [13]. SortMeRNA v1.9 [14] software was used to screen the metatranscriptome data against all databases deployed with the program and to remove rRNA reads. Adapters were trimmed with cutadapt v1.2.1 [15] using default settings. Quality trimming was performed with PRINSEQ Lite v0.20.0 [16] with a minimum sequence length of 40 bp and a minimum quality of 30 on both ends of the read, and as mean quality. All reads containing more than three Ns or non-IUPAC characters were discarded.

Reads from experiment A (Suppl. Figure S1) were pooled and assembled with IDBA_UD version 1.1.1 [17] using two rounds of assembly; firstly, with the options –min_count 200 and – min_support 5, and secondly, the reads, which could not be mapped to this assembly (with bowtie2 v2.0.6 [18], standard parameters) were extracted, and assembled with standard options, but with the output from the previous run provided as long reads. Contigs

with an A/T content of >80% were removed from the final assembly. Because both experiments A and B were performed with aliquots from the same inoculum, we did not include reads from experiment B in the assembly, but rather mapped reads to the assembly generated from reads obtained from experiment A as described below. Prodigal v2.5 was used for prediction of protein coding DNA sequences with the option for meta samples [19]. Protein sequences were annotated with InterProScan 5.4-47.0 [20] on the Dutch science grid (offered by the Dutch National Grid Initiative via SurfSara), and enriched by adding EC numbers using PRIAM version March 06, 2013 [21]. Carbohydrate active enzymes were predicted with dbCAN release 3.0 [22]. Further enrichment for EC numbers was obtained by matching all InterProScan derived domain names against the BRENDA database (download 13.06.13) [23] and using a text mining algorithm that included removal of the non-alphanumerical characters (colons, commas, brackets, etc.), partial and generic terms (type, terminal, subunit, domain, enzyme, like, etc.), as well as other smaller modifications. Details are provided in Supplementary Materials and Methods.

Read counts from experiment A and B (Figure S1) were obtained with Bowtie2 v2.0.6 [18] using default settings. BAM files were converted with SAMtools v0.1.18 [24], and gene coverage was calculated with subread version 1.4.6 [25]. Read mappings to the RNA-assemblies were inspected with Tablet [26].

Taxonomic assignments

RNA sequences from the metatranscriptome assembly were compared with Blast 2.2.29 [27] against the NCBI NT database (download 22.01.2014) using standard parameters, besides an E-value of 0.0001, to the human microbiome (download 08.05.2014), NCBI bacterial draft genomes (download 23.01.2014), NCBI protozoa genomes (download 08.05.2014), and the human genome (download 30.12.2013, release 08.08.2013, NCBI Homo sapiens annotation release 105). Taxonomy was estimated with a custom version of the LCA algorithm as implemented in MEGAN [28], but with the following changes: only hits, which exceeded a bit-score of 50 were considered, and of these, only hits with a length of more than 100 nucleotides and which did not deviate more than 10% from the longest hit were accepted.

From all sequences from the assembly, which did not have a match in any of the former blast analyses, another run with the – blastn option was performed against the same databases, and in case this did not yield any results, a blastp of the predicted proteins was performed against a custom version of the KEGG Orthology database (http://www.genome.jp/kegg/ko.html, download 25.04.2014). Taxonomic assignment was again performed with the LCA algorithm, and for the blastp run only hits which did not deviate by more than 10% from the hit with the maximum identity were considered.

Differential expression

Differential expression analysis was performed at genus level in R version 3.1.1 [29] with the TCC package release 1.6.5 [30], with 36 iterations and the combination of tmm normalization and edgeR, with an FDR=0.1. Only genes with a q-value (multitest corrected p-

value) of less than 0.01 in any of the relevant comparisons were considered to be significantly differentially expressed, unless otherwise mentioned.

Metabolic mapping

Two rounds of clustering were performed to detect patterns in the expressed genes (Figure S2). All genera, which either had an average read count of >=10 per gene, or which exceeded 1% of all reads in any given condition, were clustered into groups based on relative counts per group using the k-means algorithm in Scipy version 1.6.1 [31]. To determine the stability of the clustering, 50 iterations with a clustering between 1 and 20 clusters were performed, with the option "iter" set to 100.000. Afterwards the average cluster support per amount of clusters over all the iterations was computed, and additionally, the clustering was investigated with a custom Python implementation of clustergrams [32]. Within the clustered genera, genes with similar expression patterns were identified with the DBSCAN algorithm [33]. Clustering on expression patterns was performed with ELKI 0.7.0~20150828 [34], the –minpts parameter was fixed to 3 and the epsilon parameter was varied in percentages. Final clustering was evaluated using the Tau index as implemented in ELKI, and the clustering result with the best Tau was chosen, unless a lower Tau led to better cluster separation.

Only genes which were differentially expressed in at least one sampling time point in any of the incubations (i.e. Ino.BL, IMMP-27, IMMP-94, IMMP-96, IMMP-dig27), were considered in the clustering analysis. Genes were normalized per row before the clustering. All derived EC numbers were mapped with custom scripts onto the KEGG database [35] and visualized with Python Scipy version 1.6.1 and NumPy version 0.9.0 [31]. Correlations were calculated with the mentioned versions of Scipy/NumPy. Differentially expressed genes were mapped separately for groups of interest, and changed functions were derived from visual inspections. Cofactor requirements were investigated with the Expasy database [36].

Data accessibility:

The raw data has been uploaded to the EBI under project number PRJEB13209.

Results

We performed two *in vitro* batch fermentation experiments to investigate the influence of different IMMPs on human faecal microbiota. Our aim was to understand how the IMMPs containing different amounts of α-(1→6) glycosidic linkages were broken down by bacteria over time, and how the chemical structure of these compounds affected the functional dynamics of the microbial community during fermentation. Experiment A included fermentation of IMMPs of varying percentage of α-(1→6) glycosidic linkages (27%, IMMP-27; 94%, IMMP-94) at three different time points. This was complemented by experiment B that was performed with IMMP with 96% α-(1→6) linkages (IMMP-96) and IMMP-27 after treatment with α-amylase and amyloglucosidase (IMMP-dig27). Furthermore, in experiment B an additional set of time points was evaluated to provide a more detailed understanding of microbial community dynamics. In both experiments a control blank that did not receive any

IMMP substrate was included. We then performed metatranscriptome sequencing of all these samples, and assembled the resulting data into one reference metatranscriptome. Afterwards, machine learning techniques were applied to identify groups of similarly behaving bacteria and to discover consistent dynamic patterns in gene expression.

Quality control and statistics

The metatranscriptome was sequenced and subjected to a quality control process before the data was further analysed (Figure S1). As a result, 320 million reads (89% of the raw reads and 54% of all bases) passed the quality check and were used for assembly into contigs. In experiment A, the assembly yielded over 140,000 contigs, with more than 200,000 protein coding genes, and contained, on average, 81% of the input reads (range 71% - 85%) per sample. Read counts for experiment B were acquired by mapping to the same assembly obtained from experiment A (Table S1), and showed the same average mapping rate (81%, range 71% - 89%). After mapping, the biological replicates within each experiment showed a spearman correlation of on average 86% (range 78% - 93%), indicating good reproducibility within the sets of samples from the same treatment group.

Community structure and activity patterns

Taxonomic classification to at least the superkingdom of bacteria was assigned to 190,000 of the 200,000 genes obtained from the RNA-assemblies. Less than 3,000 genes were assigned to eukaryotes and less than 2,000 to Archaea. Of the bacterial groups, most genes were assigned to the orders Bacteroidales (>67,000), Clostridiales (>40,000), Lactobacillales (27,000) and Enterobacteriales (>14,000). The genus with the highest number of assigned genes was the genus *Bacteroides* (>54.000; Figure 1).

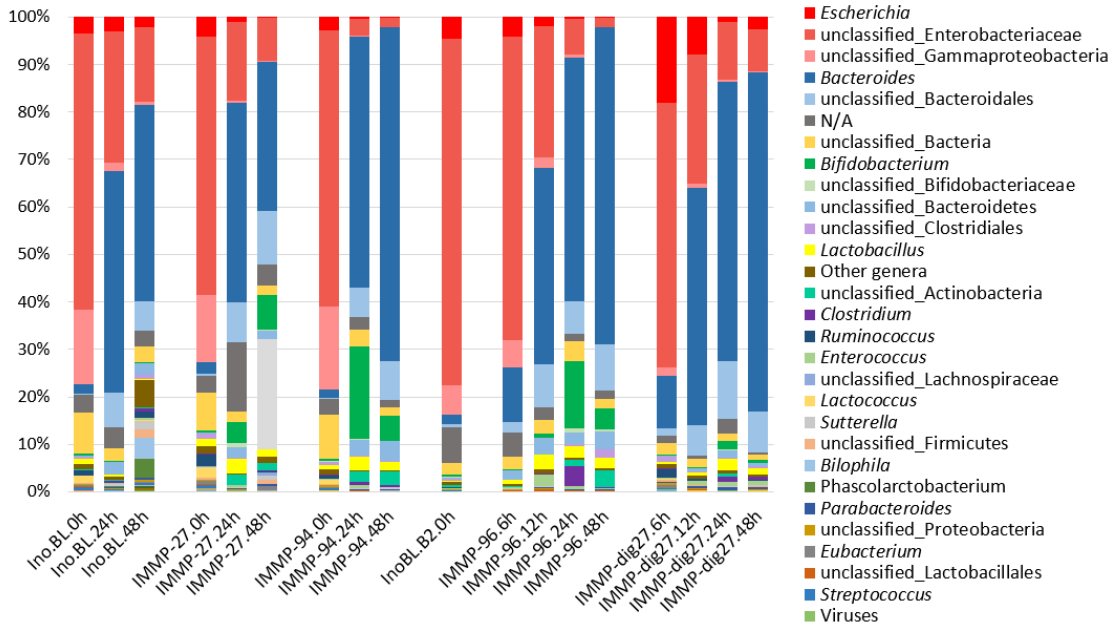


Figure 1. Average relative transcript expression of different genus level taxa in incubations sampled at time points 0 h, 6 h, 12 h, 24 h and 48 h. When the taxonomic assignment could not be made at genus level, the lowest classifiable taxonomy assignment was used for display. Low abundance genera are summarised as “Other genera” for display purposes.

To identify bacterial activity patterns, we focused on RNA reads for which a KEGG Orthology (KO) or EC identifiers could be assigned. The percentage of reads with defined KO or EC ranged from 42% to 83% for different samples. Most of the expression data with assigned KO or EC identifiers came from 22 bacterial groups, of which 12 could be assigned to a known genus, and only a small number of genes was assigned to minor groups (3%), unclassifiable sequences (3%), and sequences not classifiable beyond the superkingdom bacteria (3.5%). In the activated inoculum at the start of the incubation (0 h), unclassified Enterobacteriaceae were the most active group (Figure 1). However, once the incubation had started, the relative activity of *Bacteroides* increased in all treatment groups. In all samples combined across all treatment groups and time points, 39% of all expression data came from the genus *Bacteroides* and 27% from unclassified Enterobacteriaceae. Overall, the relative abundance of different bacterial groups based on the metatranscriptome data corresponded to the pattern in the relative abundance of different taxa based on the 16S rRNA gene analysis described previously by Gu *et al.* (Chapter Six of this thesis) (Figure 2).

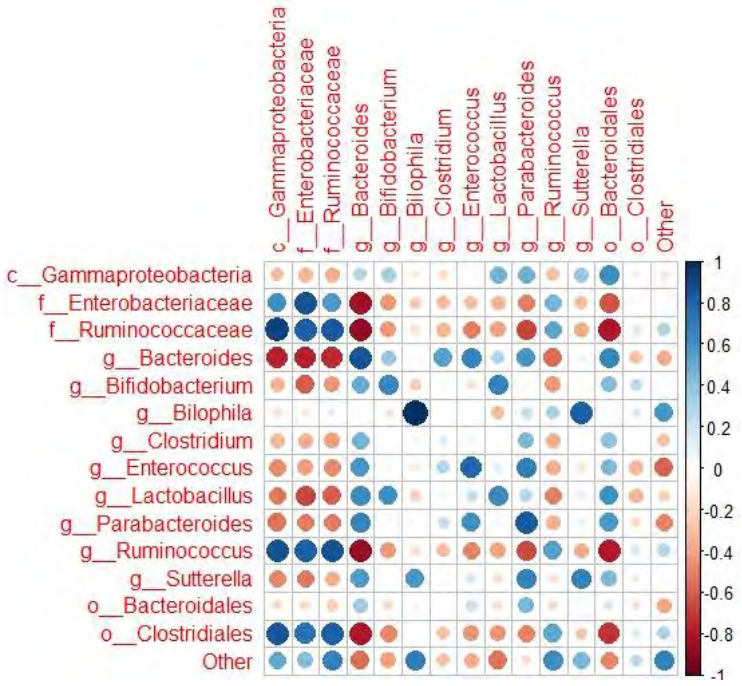


Figure 2. Correlation between the relative activity of the main bacterial groups based on metatranscriptome data, and their relative abundance based on 16S rRNA gene sequencing data (see Chapter Six for details). In case when genus level assignment was ambiguous, unclassified fraction within the next higher taxonomic level was used.

Global and IMMP specific co-occurrence of taxa

It is known that in microbial ecosystems bacterial taxa occupy different niches and co-exist forming a complex network of co-dependencies. We wanted to assess whether, based on the metatranscriptome data, we could identify bacterial groups which co-occurred in our samples and in relation to specific IMMPs. We performed clustering analysis based on mRNA reads from all samples in our dataset to test for global co-occurrence patterns. We showed that clustering into nine groups was most stable. An overview of organism assignment per cluster,

with number of assigned genes and differentially expressed genes is provided in Table S2. One of these clusters was present in all 0 h samples, but decreased or was absent at all other time points. This cluster consisted of mostly anaerobic gut inhabitants and specialized degraders (*Ruminococcus* and *Lactococcus*) as well as reads that could be largely classified as contamination from the sampling (e.g. *Homo, Mus, Bos*, unclassified Mammalia). The second and third cluster consisted mainly of genera that also include many probiotic organisms, e.g. *Bifidobacterium, Lactobacillus, Enterococcus* and sequences, which could not be classified beyond a related higher order (e.g. unclassified Bifidobacteriaceae, unclassified Lactobacillaceae). These clusters also contained a related phage group (*Myoviridae,* mainly *Lactobacillus* phages), and an unrelated genus (*Fusobacterium*). The identified genera in cluster two and three showed an increasing pattern in terms of relative transcript abundance in all the cultures which were supplied with IMMP substrates, whereas relative transcript abundance was decreased or undetected in the control cultures without IMMPs. The fourth cluster was dominated by *E. coli* and related higher order classifications (e.g. unclassified Enterobacteriaceae), together with other enterobacteria such as *Enterobacter, Citrobacter* and *Klebsiella,* and the unrelated genus *Eubacterium*. This cluster was mainly present in the samples without prebiotics, and declined in the samples with prebiotics. The fifth cluster was dominated by *Bacteroides*, and showed an increase with time in all incubations. This cluster also included *Parabacteroides, Prevotella, Flavobacterium*, and *Desulfosporosinus*. The sixth cluster consisted only of *Clostridium*/unclassified Clostridia, which showed some increase with time in all incubations. The seventh cluster contained *Anaerostipes* and related higher order classifications (unclassified Clostridiales, unclassified Lachnospiraceae) and showed a similar pattern as cluster six. No clear pattern was seen for the eighth cluster consisting of *Corynebacterium, Ethanoligenes, Odoribacter*, and *Sutterella*. Finally, the ninth group consisted of different bacterial genera, some of which also containing known pathogens (*Bilophila, Phascolarctobacterium*), some related to non-carbohydrate metabolizing bacteria (*Acidaminococcus*), and some known gut symbionts like *Veillonella* and *Megasphaera*. This group was common in samples of incubations without any prebiotics at 48 h, and was nearly absent in all the other samples.

Detection of specific gene expression patterns

Besides the co-occurrence of bacterial groups, the specific gene expression patterns within these groups were investigated as well, based on the optimal gene clustering for all bacterial groups using DBSCAN. The clustering with the optimal tau was chosen for all bacterial groups, except for the genus *Enterococcus*, for which a suboptimal tau lead to better cluster separation. As a result, the DBSCAN gene clustering analysis revealed the presence of three main patterns in the expression in nearly all observed bacterial groups (Figure S3). These three patterns comprised in all cases at least 80% of all investigated genes, which were not considered noise. The first pattern was present in all incubations, and was characterized by genes which were expressed only at 0 h, and not expressed at any later time points. The second pattern was found only in the control group and only at 48 h. The third, and the most common pattern found in all experimental groups included genes that were not expressed at 0 h, but showed upregulation at the later time points during incubation. This pattern was characteristic

for genes assigned to the genera *Enterococcus* and *Bacteroides*, which showed big gene clusters increasingly expressed over time in all treatment groups including the control group. *Bifidobacterium/Lactobacillus* and *Clostridium* also showed the same pattern, but only in the groups where IMMPs were present. *Eubacterium hallii,* showed the same gene expression pattern, but only in the group supplemented with IMMP-27 (Figure S3).

The expression levels of genes assigned to a specific bacterial group indicates its contribution to utilising the specified substrate, or its by-products. The high overall relative activity of bifidobacteria (and unclassified Bifidobacteriaceae), lactobacilli, enterococci, and unclassified Actinobacteria was positively correlated with the presence of IMMPs (Figure 1). Contrary, the activity of unclassified Proteobacteria, *Prevotella*, *Sutterella*, *Acinetobacter*, *Eggerthella*, *Acidaminococcus*, *Streptococcus*, *Phascolarctobacterium*, and *Bilophila* was negatively associated with the presence of IMMPs, as compared to the control group.

General metabolic effects of IMMP

We wanted to further investigate the activity of the bacterial groups associated with the fermentation of different IMMPs. Our analysis of the metabolic clusters revealed that five bacterial groups found in the faecal inoculum, namely *Bifidobacterium/Lactobacillus, Enterococcus, Bacteroides, Clostridium,* and *Eubacterium hallii,* showed a considerable upregulation of general metabolic pathways like glycolysis, nucleic acid or fatty acid biosynthesis, as compared to the gene expression at 0 h. When we compared metabolic patterns between different bacterial groups, we did not find large differences at a pathway level, but the groups exhibited overall different metabolic patterns. Members of the genus *Bacteroides* active in our incubations showed at first a unique partial upregulation of Vitamin B12 metabolism. An investigation of the cofactor requirements showed that Vitamin B12 in *Bacteroides* is essential for methionine synthase and methylmalonyl-CoA mutase, the latter of which produces methylmalonyl-CoA from succinyl-CoA (Figure 3).

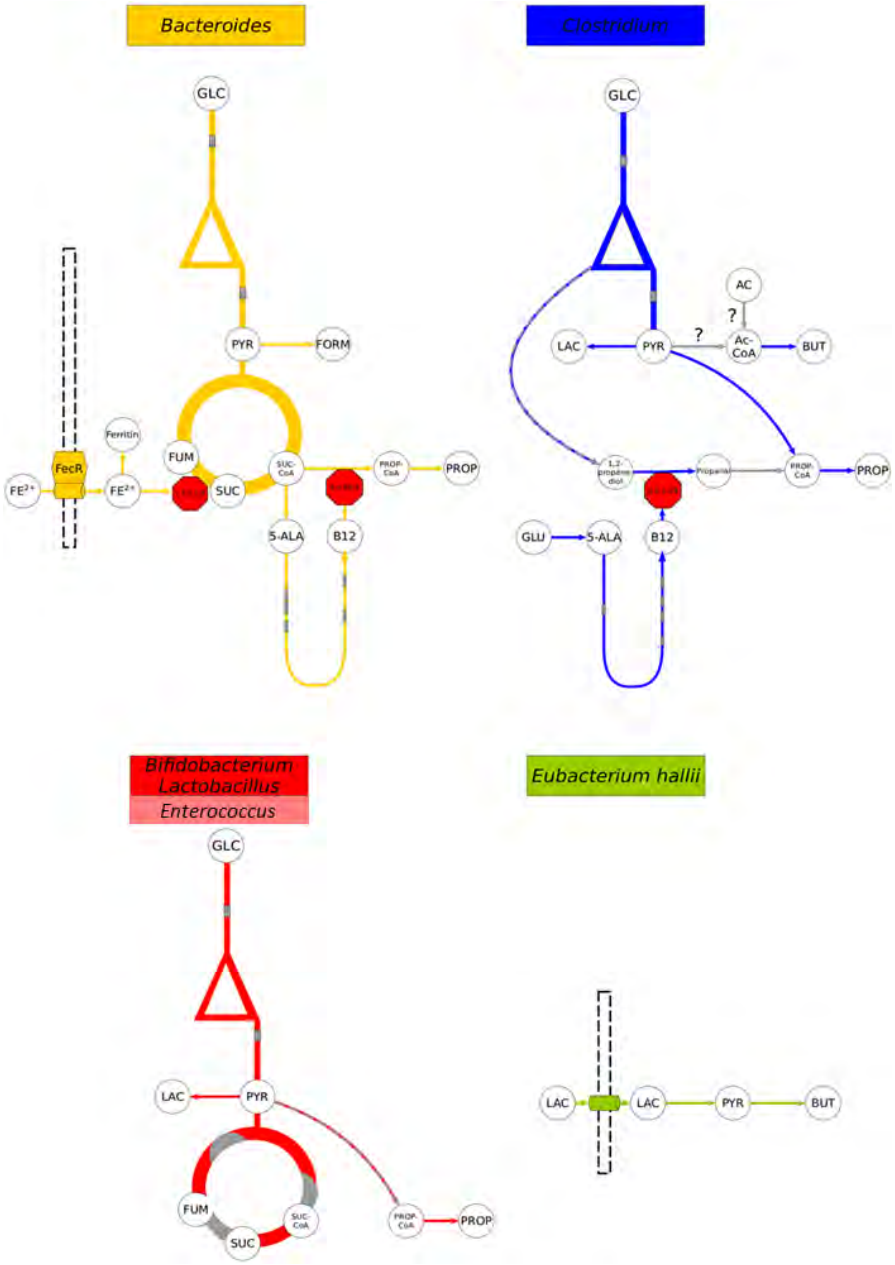


Figure 3. Overview of the metabolism of specific microbial groups observed in the samples taken during *in vitro* fermentation of different IMMPs by human faecal inoculum All samples show in general the same patterns for all organisms, besides for *Eubacterium hallii*, which only showed expression in the samples with IMMP-dig27. The genus *Enterococcus* showed the same pattern as *Bifidobacterium/Lactobacillus*, but at lower relative transcript abundance. Grey indicates that certain genes were not differentially expressed within a pathway. 5-ALA = 5-Aminolevulinate, AC = Acetate, Ac-CoA = Acetyl-CoA, BUT = Butyrate, FORM = Formate, FUM = Fumarate , GLC = Glucose, LAC = Lactate, PROP = Propionate, PROP-CoA = Propanoyl-CoA, PYR = Pyruvate, SUC = Succinate, SUC-CoA = Succinyl-CoA

Methylmalonyl-CoA mutase is involved in propionate biosynthesis, and our data showed that the whole pathway for propionate biosynthesis was, in fact, upregulated. The data further showed that many genes coding for proteins involved in iron scavenging were also

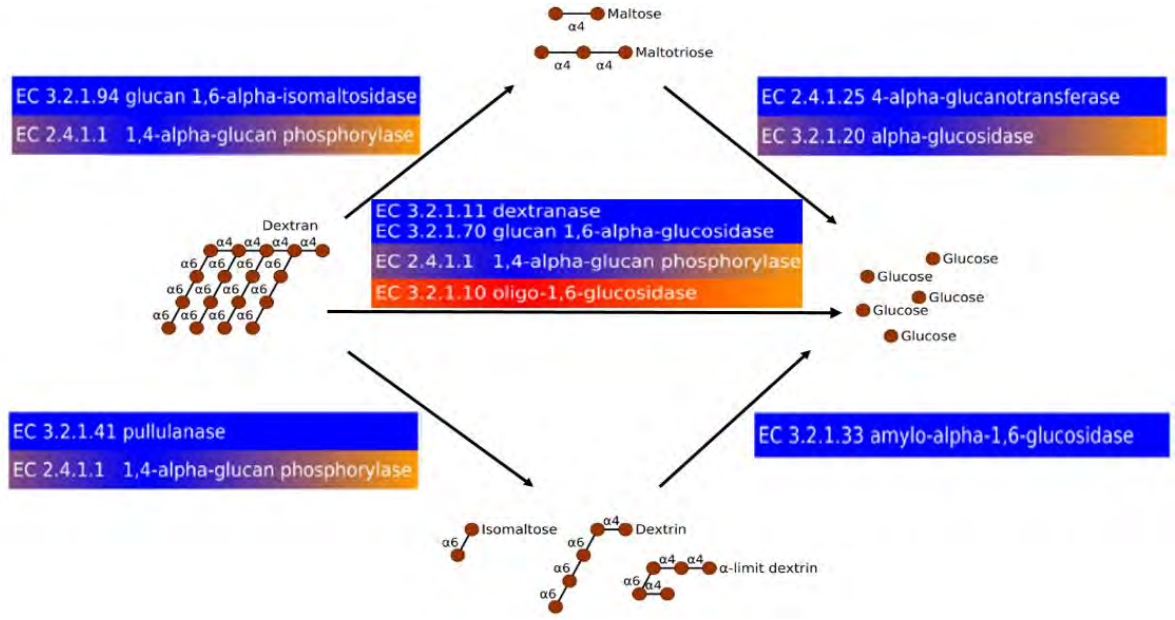
upregulated (e.g. FecR). One of the genes coding for an enzyme with iron requirements was the succinate dehydrogenase gene, which converts succinate into fumarate. This function, as well as all others in the TCA cycle, showed upregulation in all samples tested. The genus *Clostridium* also showed an upregulation of genes involved in Vitamin B12 production, but the biosynthesis occurred via glutamate, whereas in the *Bacteroides* group it was produced via succinate. The genes in the pathway for propionate production were overall upregulated (production via acetyl-CoA, not succinyl-CoA), similar to the genes in lactate and butyrate production pathways. The only other enzyme requiring Vitamin B12 in the microbiome was a multimer of propanediol dehydratase or glycerol dehydratase (ambiguous taxonomic assignment), which are both involved in the breakdown of glycerol/glycerone phosphate to propanol/propionate/1,3-propanediol. However, a full upregulation of either pathway was not observed. The *Bifidobacterium*/*Lactobacillus* group and the *Enterococcus* group showed upregulation of genes related to production of lactate from pyruvate, and *Bifidobacterium*/*Lactobacillus* group also showed upregulation of genes encoding proteins involved in butyrate production, but it is unclear if butyrate would be directly produced from pyruvate, or derived from external acetate. *Eubacterium hallii*, on the other hand, showed high activity related to converting lactate into butyrate, as also shown previously [37]. In addition, our data indicated that formate was produced by the *Enterococcus* and *Bacteroides* populations.

Microbial groups directly involved in the degradation of the IMMPs

In order to gain insight into which bacterial groups are directly involved in degradation of different IMMPs, we used the KEGG reference pathway for starch and sucrose metabolism. We surveyed our data for the expression of the genes encoding enzymes that are known to be involved in sucrose and starch metabolism. More specifically we focused on genes encoding enzymes from glycoside hydrolase family 13 (http://www.cazy.org/GH13_bacteria.html), as this family includes a number of bacterial proteins shown to be essential in degradation of similar compounds, such as isomaltooligosaccharides (IMO) [38]. The majority of genes listed in the KEGG starch and sucrose metabolism pathway were detected in our transcriptome data (Figure S4), as well as some additional genes in glycoside hydrolase family 13 (EC 3.2.1.135, 3.2.1.68 and 3.2.1.11), which were not listed in the KEGG pathway, but which are known to be activated during the degradation of pullulan and dextran [39-42]. It is interesting to note that the relative contribution of these starch and sucrose metabolism genes to the total number of genes from each sample did not correlate with the presence or absence of IMMPs in the samples. The only exception was incubation with pre-treated IMMP-27, in which starch and sucrose metabolism genes reached 10% at 12 h and about 12% at 48 h, whereas in other groups they ranged between 4 to 5% (Figure S5). Despite of the similarities in the overall expression of the starch and sucrose metabolism genes in all samples, we could see differences in the relative abundance of genes coding for specific enzymes depending on the IMMP used, and the duration of the fermentation (Figure S6).

One of the aims of this study was to better understand the functional dynamics of the bacterial communities during IMMP degradation. Previously reported HPAEC and HPSEC analyses (Chapter Six of this thesis) showed that the degradation of IMMP-94 and IMMP-96

occurred between 12 h and 24 h of the incubation. At 24 h and 48 h we noted an increase in the expression of genes coding for enzymes that might be directly involved in the hydrolysis of α-(1→6) glycosidic linkages, namely EC 3.2.1.10 – oligo-1,6-glucosidase, EC 3.2.1.11 – dextranase, and EC 3.2.1.33 – amylo-α-1,6- glucosidase (Figure S7a,b). There was also an increase in the expression of genes coding for enzymes that can hydrolyse α-(1→4) glycosidic linkages, mainly the EC 3.2.1.1 – α-amylase, EC 3.2.1.20 – α–glucosidase 4-α-glucanotransferase, and EC 2.4.1.25 – 4-α-glucanotransferase. Since IMMP-27 contains lower amounts of α-(1→6) linkages, its degradation also involves the activation of the same genes, however, the expression levels of the genes encoding enzymes which hydrolyse α-(1→6) linkages were much lower (Figure S7a,b). Bacterial groups that contributed the most to the primary degradation of IMMP's α-(1→6) linkages were *Lactobacillus*, *Bifidobacterium* and *Bacteroides*, all expressing the genes encoding EC 3.2.1.10 oligo-1,6-glucosidase and EC 3.2.1.11 dextranase. On the other hand, we predicted that α-(1→4) linkages were hydrolysed mainly by *Bacteroides*, unclassified Bacteroidales, unclassified Enterobacteriaceae, *Lactobacillus* and *Bifidobacterium* via EC 3.2.1.1 alpha-amylase and EC 2.4.1.1 glycogen/amylophosphorylase (Figure S7c). Based on the transcript data, *Bifidobacterium* and *Lactobacillus* were mainly active in the degradation of IMMP-94 and IMMP-96 at 24 h (Figure 4, and Figure S7c). These genera were also active in degradation of IMMP-27 and the pre-treated IMMP-27, but their relative contributions were much lower (Figure 4, and Figure S7c). The breakdown of IMMPs at 24 h and 48 h was otherwise dominated by *Bacteroides*, with the exception of pre-treated IMMP-27 at 48 h, which showed a high level of expression of genes assigned to unclassified Enterobacteriaceae. Figure 4 summarises our model of IMMP degradation and confirms the specialised role of lactobacilli and bifidobacteria in hydrolysis of α-(1→6) linkages. It also reveals the important contribution of *Bacteroides* as both, primary and secondary degraders of IMMPs and their by-products.



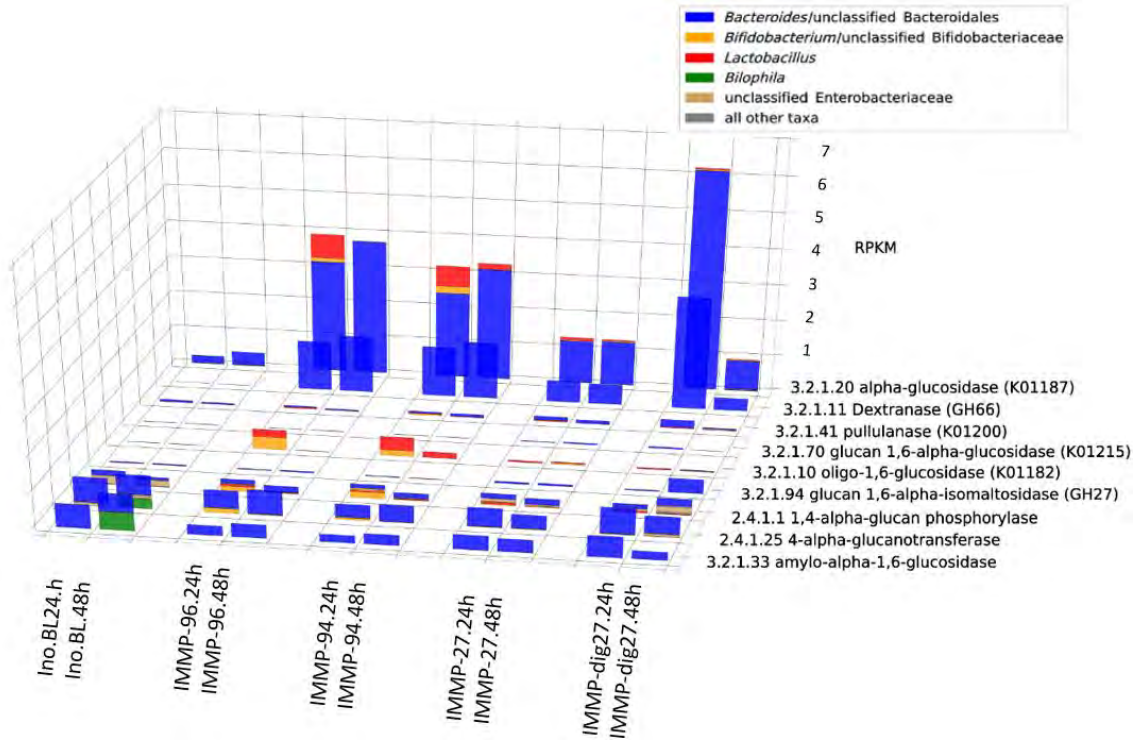


Figure 4. Overview of the main degradation pathways starting from dextran. Colours in the top panel indicate the main contributors to a reaction. The bottom panel shows the overall expression in reads per kilobase per million (RPKM) per organism at time points 24 and 48h for all conditions.

Discussion

Prebiotic food components should be resistant to host's gastric enzymes, fermentable by the host's intestinal microbiota and capable of promoting growth and activity of bacterial groups associated with health [43]. The IMMPs seem to fulfil all these criteria [8, 44, 45]. Earlier studies demonstrated that hydrogenated and high DP IMMPs are not or little-digestible by rat gastric enzymes [45], and that diets containing IMMPs are associated with higher numbers of lactobacilli, and an overall increase in the number of intestinal bacteria [46]. Moreover, a recent study with human inoculum reported that IMMPs can be fermented by human large intestinal microbiota and that SCFAs, in particular acetate and propionate, are produced, indicating that IMMPs may stimulate activity of probiotic groups [10]. This is in accordance with earlier findings from a small human trial that showed an increased level of bifidobacteria in subjects who received IMOs in their diets [44].

In our study, we confirmed the prebiotic character of the IMMPs and showed that the specific effect of different IMMPs on human faecal microbiota composition and activity varied during *in vitro* fermentation, depending on the relative amount of α-(1→6) glycosidic linkages present in the substrate. When IMMP-94 and IMMP-96 were used as a carbon source, we observed a strong upregulation of genes in the probiotic cluster, specifically genes assigned to bifidobacteria and lactobacilli. Furthermore, high relative activity of these bacteria

corresponded with an increase in their relative abundance as estimated by the 16S rRNA gene sequencing (Chapter Six of this thesis). In contrast, when the pre-treated IMMP-27 was used as a substrate, the relative activity of bifidobacteria and lactobacilli was lower. These bacteria were also less active in the control, and their activity peak in the presence of IMMP-27 was delayed to 48 h. Interestingly, all IMMP treatment groups showed a time lag between the maximum relative activity, and the increase in the corresponding bacterial relative abundance as measured by rRNA gene-targeted community analysis (Chapter Six of this thesis). For example, the maximum activity of bifidobacteria was observed at 24 h of incubation when IMMP-94 and IMMP-96 were used as substrates. Yet, bifidobacteria reached their highest relative abundance only at 48 h when their relative activity had already decreased. The relative activity of lactobacilli followed a pattern similar to that of bifidobacteria in all treatment groups, except for incubations with IMMP-94 where lactobacilli showed maximum relative activity at 12 h, whereas bifidobacteria activity peaked at 24 h. Relative activity of *Bacteroides* was very high in all groups, regardless of the incubation time, presence and type of the IMMP that was used as a carbon source. *Bacteroides* spp. are known to be generalists that are able to break down a wide array of carbon sources [47]. Bifidobacteria and lactobacilli are often very specialised and can grow on substrates that are chemically not accessible to other bacteria in the microbial ecosystem. This may be the reason that these groups show delayed activity in relation to *Bacteroides*, as only after the depletion of the easily accessible IMMP fractions containing the α-(1→4) glycosidic linkages the bacterial groups capable of utilising the α-(1→6) glycosidic linkages gained a competitive advantage. It is known that bacteria can sense specific polysaccharides and produce specific sets of enzymes according to their individual nutrient prioritization schemes [48]. The patterns of activity and growth in the presence of IMMP-27, pre-treated IMMP-27 and in the control group may confirm this hypothesis, as we observed increased relative abundance of *Parabacteroides*, *Sutterella, Parasutterella*, *Enterococcus*, unclassified Lachnospiraceae *Incertae Sedis*, *Eggerthella* and few other groups (Chapter Six of this thesis). High relative abundance of these groups could be explained by the presence of residual α-(1→4) glycosidic linkages, or the presence of products generated during the enzymatic conversion of the α-(1→4) glycosidic linkages into α-(1→6) glycosidic linkages during the IMMP pre-treatment process, or by more efficient scavenging on other bacteria or their metabolites.

While some of the beneficial bacteria increased in relative abundance and activity with the presence of IMMPs, we also noted that the exclusive use of these prebiotics put a selective pressure on other beneficial microbes. For example, *Lactococcus lactis* and *Ruminococcus bromii* - two specialized beneficial degraders, did not show any survival in our samples (Chapter Six of this thesis). This can be explained by the lack of suitable substrate for both species, given that neither any simple mono- or disaccharides (for *Lactococcus* [49]) nor type II or III resistant starch (for *Ruminococcus* [50]) were present in this experiment. Although both organisms can be considered a probiotic, they were not stimulated in the particular prebiotic environment tested here, enforcing the notion that prebiotics can selectively stimulate activity and growth of specific groups, whereas in general, a diverse diet may be necessary to comprehensively support a stable community of commensal microbes.

In our study we also observed a clear effect of having no carbohydrate source in the control samples. With the absence of the prebiotics, there was a switch of the community from processing carbohydrates to utilising amino acids [51], as indicated by the increase of relative abundance of *Acidaminococcus* (Chapter Six of this thesis). In addition, there was an increase of *Bilophila* in the control samples, which is an organism previously associated with gut dysbiosis [52].

IMMP Degradation Model

A total of 130 families of glycoside hydrolases, 22 families of polysaccharide lyases, and 16 families of carbohydrate esterases have been described, and many of these enzymes are encoded only by microbes (www.cazy.org) [53]. We surveyed our data for the presence of genes encoding the enzymes that are known to be involved in sucrose and starch metabolism, mostly genes from glycoside hydrolase family 13. Few studies up to date looked at the genetics and enzymology of degradation of IMMPs mainly in lactobacilli [40, 54, 55], bifidobacteria [56] and *Bacteroides*. However, microbial species in the gut do not act in isolation, but rather interact with each other through a network of syntrophic interactions often making the utilization of the substrate more effective [57]. Metabolic potential and fermentation efficiency vary between different species, and complete IMMPs degradation in the gut is a result of different bacterial groups working together in a complementary fashion, likely leading to the formation of microbial food chains [57, 58]. Certain bacterial groups may show a higher activity at specific degradation steps, as measured by the expression of specific genes coding for enzymes required to catalyse given reactions. This is also visible in our experiments. The expression of oligo-1-6-glucosidase encoding genes was dominated by lactobacilli and bifidobacteria when IMMP-94 and IMMP-96 were used as a substrate, whereas *Bacteroides* and unclassified Bacteroidales were also highly active in the presence of IMMP-27 or IMMP-dig27. Similar patterns could be observed in expression of other genes that code for enzymes involved in sucrose and starch metabolism (Suppl. Figure S7c). While some of the carbohydrate breakdown steps were dominated by known probiotic genera, many of the primary and secondary degradation processes were also performed by members of *Bacteroides*. Our data showed that once the fermentation started, one of the very specialized enzymes, dextranase, was produced only by *Bacteroides*. Other processes were found reliant on multiple genera as based on the gene expression data. For example, the breakdown of the IMMPs to maltose and maltotriose by α – amylases was dominated by *Bacteroides*, whereas the further metabolisation was performed also by bifidobacteria and lactobacilli. Furthermore, other groups such as enterobacteria or *Parabacteroides* were not involved in most of these breakdown processes, but still constituted viable populations in the communities. Their functional role in the community is, however, not clear.

Metabolites of fermentation

Experimental results showed that the administration of IMMPs lead to an increased production of different SCFAs, mainly acetate and succinate (Chapter Six of this thesis). While succinate normally does not accumulate in this medium [59], the excess of substrate [60], high CO_2 levels, and the upregulation of all the necessary steps [61] in our metabolic mapping,

including the necessity for iron, could explain such accumulation. In addition, previous studies showed that succinate accumulation was associated with oversupply of complex substrates, such as prebiotics, or in our case IMMPs [62] or when further metabolisation of succinate is unnecessary [59]. It is also possible that lack of Vitamin B12, which is necessary for propionate production [61], and for which an upregulation could be observed, resulted in the accumulation of succinate instead of propionate. However, we are unable to conclude the exact reason based on our data. One of the other propionate production pathways, the acrylate pathway [63], could not be detected in the data. However, it is tempting to speculate that the production of propionate proceeded via the direct fermentation of pyruvate via 3-hydroxypropionate and Acryloyl-CoA in the current study. This pathway has not been described before, but it is potentially visible in the data, with just a few reactions missing. Furthermore, the potential of producing propionate via 1,2-propanediol directly through methylglyoxal is indicated in the data. Unfortunately, no definite conclusions can be drawn due to missing steps in the metabolism of the involved populations (*Bifidobacterium/Lactobacillus* for the former, and *Clostridium* for both), however, the possibility of these alternative pathways should be investigated. Besides succinate, propionate and acetate, also lactate and butyrate were observed as metabolites (Chapter Six of this thesis).

Dietary fibres, including modified starches such as IMMPs offer a promising, non-invasive way to intentionally manipulate gut microbiota composition. Investigations of whole bacterial communities and understanding of the mechanisms by which microorganisms interact to degrade different dietary carbohydrates are essential for our ability to manipulate gut microbiota to benefit our health. We showed how IMMPs can increase the relative abundance and activity of beneficial bacteria, making these novel prebiotics potentially useful in improving host's health from the aspect of nutrition, to achieve prevention or even alleviation of diseases.

Funding

This research was performed in the public-private partnership CarboHealth coordinated by the Carbohydrate Competence Center (CCC, www.cccresearch.nl) and financed by participating partners and allowances of the TKI Agri&Food program, Ministry of Economic Affairs of the Netherlands. This work was furthermore supported by Wageningen University and the Wageningen Institute for Environment and Climate Research (WIMEK) to BH through the IP/OP program Systems Biology [project KB-17-003.02-023].

Acknowledgements

The authors thank Jasper Koehorst (Wageningen University, Laboratory of Systems and Synthetic Biology) for his help with the transcriptome annotation, and Maria Suarez-Diez and Edoardo Saccenti (Wageningen University, Laboratory of Systems and Synthetic Biology) for helpful discussions. This work was carried out on the Dutch national e-infrastructure with the

support of SURF Foundation. The authors want to thank B. van den Bogert (Wageningen University, Laboratory of Microbiology) for the design of Figure S1.

Conflict of Interest

None declared.

References

1. Dethlefsen, L., M. McFall-Ngai, and D.A. Relman, *An ecological and evolutionary perspective on human-microbe mutualism and disease.* Nature, 2007. **449**(7164): p. 811-8.
2. Flint, H.J., et al., *Interactions and competition within the microbial community of the human colon: links between diet and health.* Environ Microbiol, 2007. **9**(5): p. 1101-11.
3. Cani, P.D. and A. Everard, *Talking microbes: When gut bacteria interact with diet and host organs.* Mol Nutr Food Res, 2016. **60**(1): p. 58-66.
4. Garrett, W.S., J.I. Gordon, and L.H. Glimcher, *Homeostasis and inflammation in the intestine.* Cell, 2010. **140**(6): p. 859-70.
5. Becattini, S., Y. Taur, and E.G. Pamer, *Antibiotic-Induced Changes in the Intestinal Microbiota and Disease.* Trends Mol Med, 2016. **22**(6): p. 458-78.
6. Bowman, K.A., E.K. Broussard, and C.M. Surawicz, *Fecal microbiota transplantation: current clinical efficacy and future prospects.* Clin Exp Gastroenterol, 2015. **8**: p. 285-91.
7. Hood, L., R. Balling, and C. Auffray, *Revolutionizing medicine in the 21st century through systems approaches.* Biotechnol J, 2012. **7**(8): p. 992-1001.
8. Roberfroid, M., *Prebiotics: The Concept Revisisted.* J Nutr, 2007. **137**.
9. Scott, K.P., et al., *Prebiotic stimulation of human colonic butyrate-producing bacteria and bifidobacteria, in vitro.* FEMS Microbiol Ecol, 2014. **87**(1): p. 30-40.
10. Leemhuis, H., et al., *Isomalto/malto-polysaccharide, a novel soluble dietary fiber made via enzymatic conversion of starch.* J Agric Food Chem, 2014. **62**(49): p. 12034-44.
11. Aguirre, M., et al., *To pool or not to pool? Impact of the use of individual and pooled fecal samples for in vitro fermentation studies.* J Microbiol Methods, 2014. **107**: p. 1-7.
12. Kang, S., et al., *An efficient RNA extraction method for estimating gut microbial diversity by polymerase chain reaction.* Curr Microbiol, 2009. **58**(5): p. 464-71.
13. Davids, M., et al., *Functional Profiling of Unfamiliar Microbial Communities Using a Validated De Novo Assembly Metatranscriptome Pipeline.* PLOS ONE, 2016. **11**(1): p. e0146423.
14. Kopylova, E., L. Noe, and H. Touzet, *SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data.* Bioinformatics, 2012. **28**(24): p. 3211-7.
15. Martin, M., *Cutadapt removes adapter sequences from high-througput sequencing reads.* EMBnet.journal, 2011. **17**: p. 10-12.
16. Schmieder, R. and R. Edwards, *Quality control and preprocessing of metagenomic datasets.* Bioinformatics, 2011. **27**(6): p. 863-4.
17. Peng, Y., et al., *IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth.* Bioinformatics, 2012. **28**(11): p. 1420-1428.
18. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
19. Hyatt, D., et al., *Prodigal: prokaryotic gene recognition and translation initiation site identification.* BMC Bioinformatics, 2010. **11**: p. 119.
20. Hunter, S., et al., *InterPro in 2011: new developments in the family and domain prediction database.* Nucleic Acids Res, 2012. **40**(Database issue): p. D306-12.
21. Claudel-Renard, C., et al., *Enzyme-specific profiles for genome annotation: PRIAM.* Nucleic Acids Res, 2003. **31**(22): p. 6633-9.
22. Yin, Y., et al., *dbCAN: a web resource for automated carbohydrate-active enzyme annotation.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W445-51.
23. Chang, A., et al., *BRENDA in 2015: exciting developments in its 25th year of existence.* Nucleic Acids Res, 2015. **43**(Database issue): p. D439-46.
24. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
25. Liao, Y., G.K. Smyth, and W. Shi, *The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote.* Nucleic Acids Res, 2013. **41**(10).
26. Milne, I., et al., *Using Tablet for visual exploration of second-generation sequencing data.* Brief Bioinform, 2013. **14**(2): p. 193-202.

27. Altschul, S.F., et al., *Basic Local Alignment Search Tool.* Journal of Molecular Biology, 1990. **215**(3): p. 403-410.
28. Huson, D.H., et al., *Integrative analysis of environmental sequences using MEGAN4.* Genome Res, 2011. **21**(9): p. 1552-60.
29. R Core Team, *R: A language and environment for statistical computing.* 2012, Vienna: R Foundation for Statistical Computing.
30. Sun, J., et al., *TCC: an R package for comparing tag count data with robust normalization strategies.* BMC Bioinformatics, 2013. **14**.
31. van der Walt, S., C. Colbert, and G. Varoquaux, *The NumPy Array: A structure for Efficient Numerical Computation.* Computing In Science & Engineering, 2011. **13**: p. 22-30.
32. Schonlau, M., *Visualizing non-hierarchical and hierarchical cluster analyses with clustergrams.* Computational Statistics, 2004. **19**: p. 95-111.
33. Ester, M., et al., *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.* Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996.
34. Schubert, E., et al., *A Framework for Clustering Uncertain Data.* Proceedings of the VLDB Endowment, 2015. **18**(12).
35. Kanehisa, M., et al., *KEGG for integration and interpretation of large-scale molecular data sets.* Nucleic Acids Res, 2012. **40**(Database issue): p. D109-14.
36. Bairoch, A., *The ENZYME database in 2000.* Nucleic Acids Res, 2000. **28**(1): p. 304-305.
37. Duncan, S.H., P. Louis, and H.J. Flint, *Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product.* Appl Environ Microbiol, 2004. **70**(10): p. 5810-7.
38. Abou Hachem, M., et al., *A Snapshot into the Metabolism of Isomalto-oligosaccharides in Probiotic Bacteria.* Journal of Applied Glycoscience, 2013. **60**(2): p. 95-100.
39. Kuchtová, A. and Š. Janeček, *Domain evolution in enzymes of the neopullulanase subfamily.* Microbiology, 2016. **162**(12): p. 2099-2115.
40. Ganzle, M.G. and R. Follador, *Metabolism of oligosaccharides and starch in lactobacilli: a review.* Front Microbiol, 2012. **3**: p. 340.
41. Bailey, R.W. and R.T.J. Clarke, *A bacterial dextranase.* Biochemical Journal, 1959. **72**(1): p. 49-54.
42. Khalikova, E., P. Susi, and T. Korpela, *Microbial Dextran-Hydrolyzing Enzymes: Fundamentals and Applications.* Microbiology and Molecular Biology Reviews, 2005. **69**(2): p. 306-325.
43. Gibson, G.R., et al., *Dietary modulation of the human colonic microbiota: updating the concept of prebiotics.* Nutr Res Rev, 2004. **17**(2): p. 259-75.
44. Kohmoto, T., et al., *Effect of Isomalto-oligosaccharides on Human Fecal Flora.* Bifidobacteria and Microflora, 1988. **7**(2): p. 61-69.
45. Kaneko, T., A. Yokoyama, and M. Suzuki, *Digestibility Characteristics of Isomaltooligosaccharides in Comparison with Several Saccharides Using the Rat Jejunum Loop Method.* Bioscience, Biotechnology, and Biochemistry, 1995. **59**(7): p. 1190-1194.
46. Ketabi, A., L.A. Dieleman, and M.G. Ganzle, *Influence of isomalto-oligosaccharides on intestinal microbiota in rats.* J Appl Microbiol, 2011. **110**(5): p. 1297-306.
47. Cockburn, D.W. and N.M. Koropatkin, *Polysaccharide Degradation by the Intestinal Microbiota and Its Influence on Human Health and Disease.* Journal of Molecular Biology, 2016. **428**(16): p. 3230-3252.
48. Flint, H.J., et al., *Microbial degradation of complex carbohydrates in the gut.* Gut Microbes, 2012. **3**(4): p. 289-306.
49. De Vos, W.M., *Metabolic engineering of sugar catabolism in lactic acid bacteria.* Antonie van Leeuwenhoek, 1996. **70**: p. 223-242.
50. Ze, X., et al., *Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon.* ISME J, 2012. **6**(8): p. 1535-43.
51. Rogosa, M., *Acidaminococcus gen. n., Acidaminococcus fermentans sp. n., Anaerobic Gram-negative Diplococci Using Amino Acids as the Sole Energy Source for Growth.* J Bacteriol, 1969. **92**(2).

52. Baron, E.J., *Bilophila wadsworthia: a Unique Gram-negative Anaerobic Rod.* Anaerobe, 1997. **3**(2): p. 83-86.
53. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013.* Nucleic Acids Res, 2014. **42**(Database issue): p. D490-5.
54. Moller, M.S., et al., *Enzymology and structure of the GH13_31 glucan 1,6-alpha-glucosidase that confers isomaltooligosaccharide utilization in the probiotic Lactobacillus acidophilus NCFM.* J Bacteriol, 2012. **194**(16): p. 4249-59.
55. Hu, Y., et al., *Metabolism of isomalto-oligosaccharides by Lactobacillus reuteri and bifidobacteria.* Lett Appl Microbiol, 2013. **57**(2): p. 108-14.
56. Liu, S., et al., *Starch and starch hydrolysates are favorable carbon sources for bifidobacteria in the human gut.* BMC Microbiol, 2015. **15**: p. 54.
57. Koropatkin, N.M., E.A. Cameron, and E.C. Martens, *How glycan metabolism shapes the human gut microbiota.* Nat Rev Microbiol, 2012. **10**(5): p. 323-35.
58. Fischbach, M.A. and J.L. Sonnenburg, *Eating for two: how metabolism establishes interspecies interactions in the gut.* Cell Host Microbe, 2011. **10**(4): p. 336-47.
59. Gibson, G.R., J.H. Cummings, and G.T. MacFarlane, *Use of a Three-Stage Continuous Culture System to Study the Effect of Mucin on Dissimilatory Sulfate Reduction and Methanogenesis by Mixed Populations of Human Gut Bacteria.* Appl Environ Microbiol, 1988. **54**(11): p. 2750-2755.
60. Macfarlane, S. and G.T. Macfarlane, *Regulation of short-chain fatty acid production.* Proc Nutr Soc, 2003. **62**(1): p. 67-72.
61. Macy, J.M. and I. Probst, *The biology of gastrointestinal bacteroides.* Annual Reviews in Microbiology, 1979. **33**: p. 561-591.
62. Sakata, T., et al., *Influences of probiotic bacteria on organic acid production by pig caecal bacteria in vitro.* Proc Nutr Soc, 2003. **62**(1): p. 73-80.
63. Hosseini, E., et al., *Propionate as a health-promoting microbial metabolite in the human gut.* Nutr Rev, 2011. **69**(5): p. 245-58.

Supplementary Tables and Figures

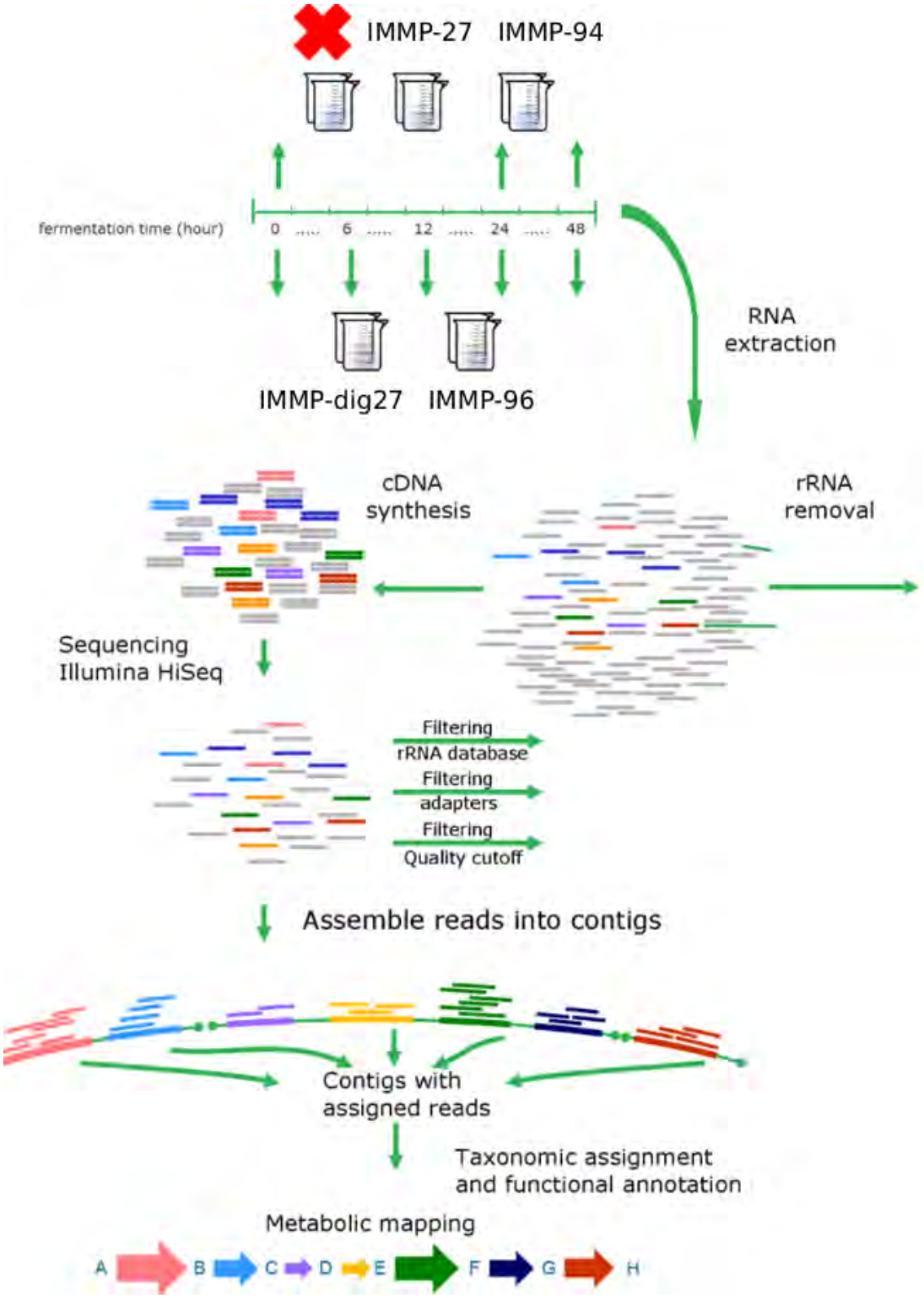


Figure S1: Experimental design.

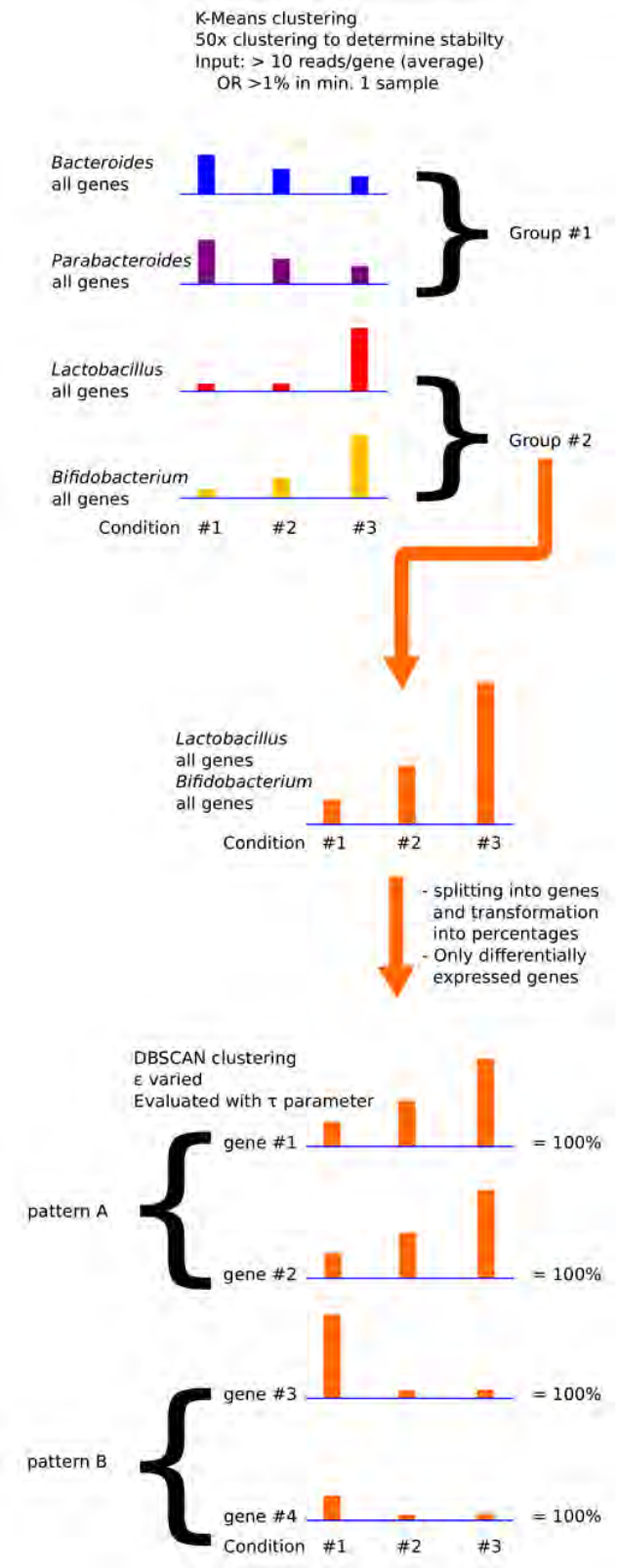


Figure S2: Overview of the clustering procedure. First, expression was lumped at the genus level. On the accumulated expression data k-means clustering was performed, until a stable clustering was achieved. The genes of the grouped genera were afterwards subjected to DBSCAN clustering. The stability of the clustering was evaluated with the Tau-parameter. Only genes, which were at least once differentially expressed, were used in the clustering process to reduce the noise.

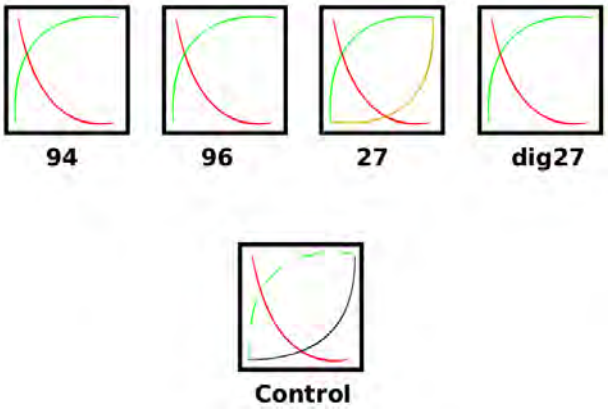


Figure S3. Overview of the main gene expression patterns. All groups (*Bacteroides, E.coli, Lactobacillus/Bifidobacterium, Enterococcus*; besides *Eubacterium hallii*) showed in all prebiotic conditions increase in relative transcript abundance in roughly the same proportion (green). Some groups (*Bacteroides, Escherichia*) also showed comparable increase in expression in the control condition (dotted green line). Furthermore, all groups showed a downregulation of certain genes in all conditions (red), and an upregulation of a group of genes in the control condition (black). *Eubacterium hallii* showed only increase in transcript abundance at the last time point with the prebiotic IMMP-27 (yellow).

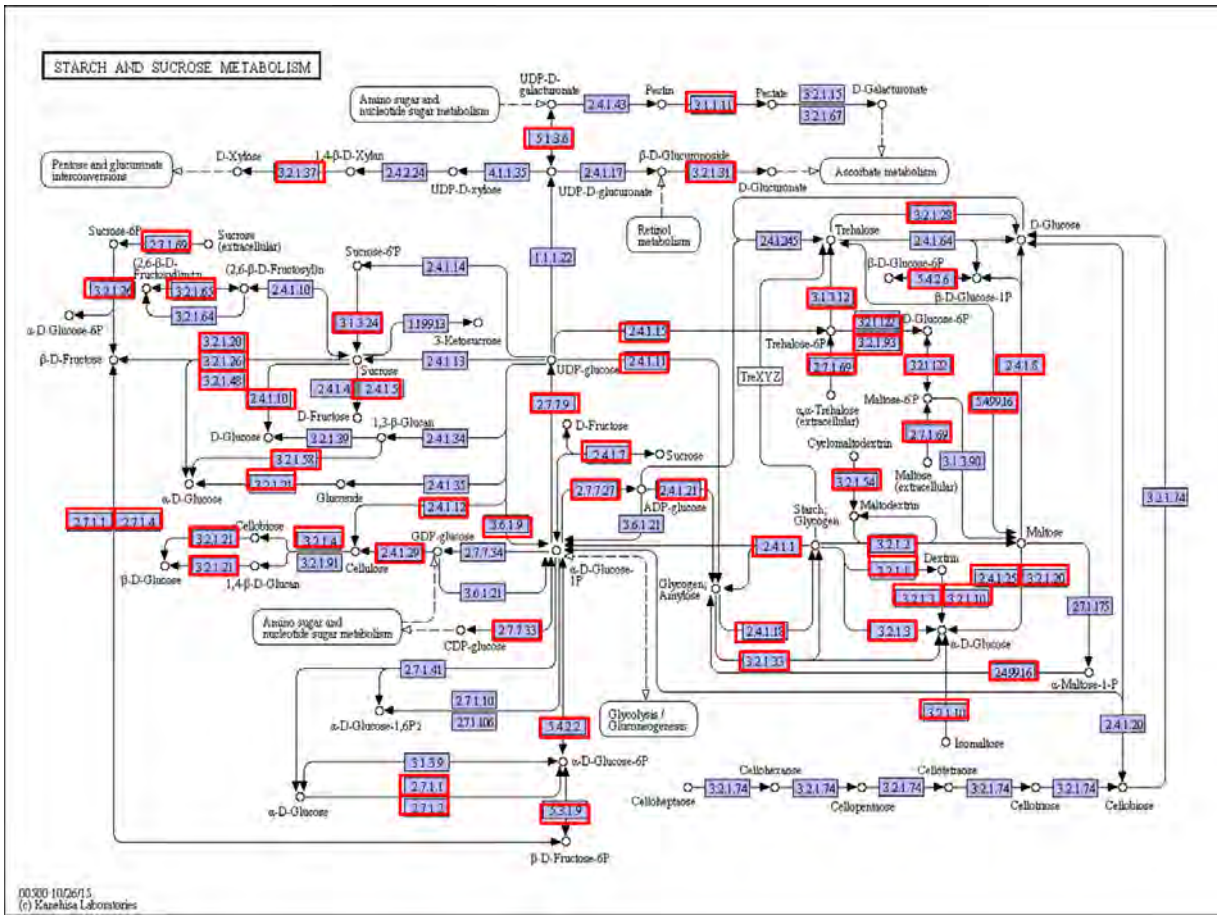


Figure S4. Starch and sucrose metabolism enzymes detected in the data.

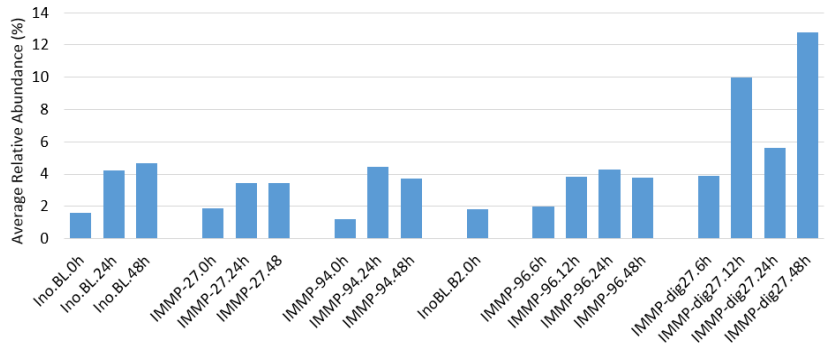


Figure S5. Relative abundance (percentage) of starch and sucrose metabolism enzyme encoding genes detected in the metatranscriptome data.

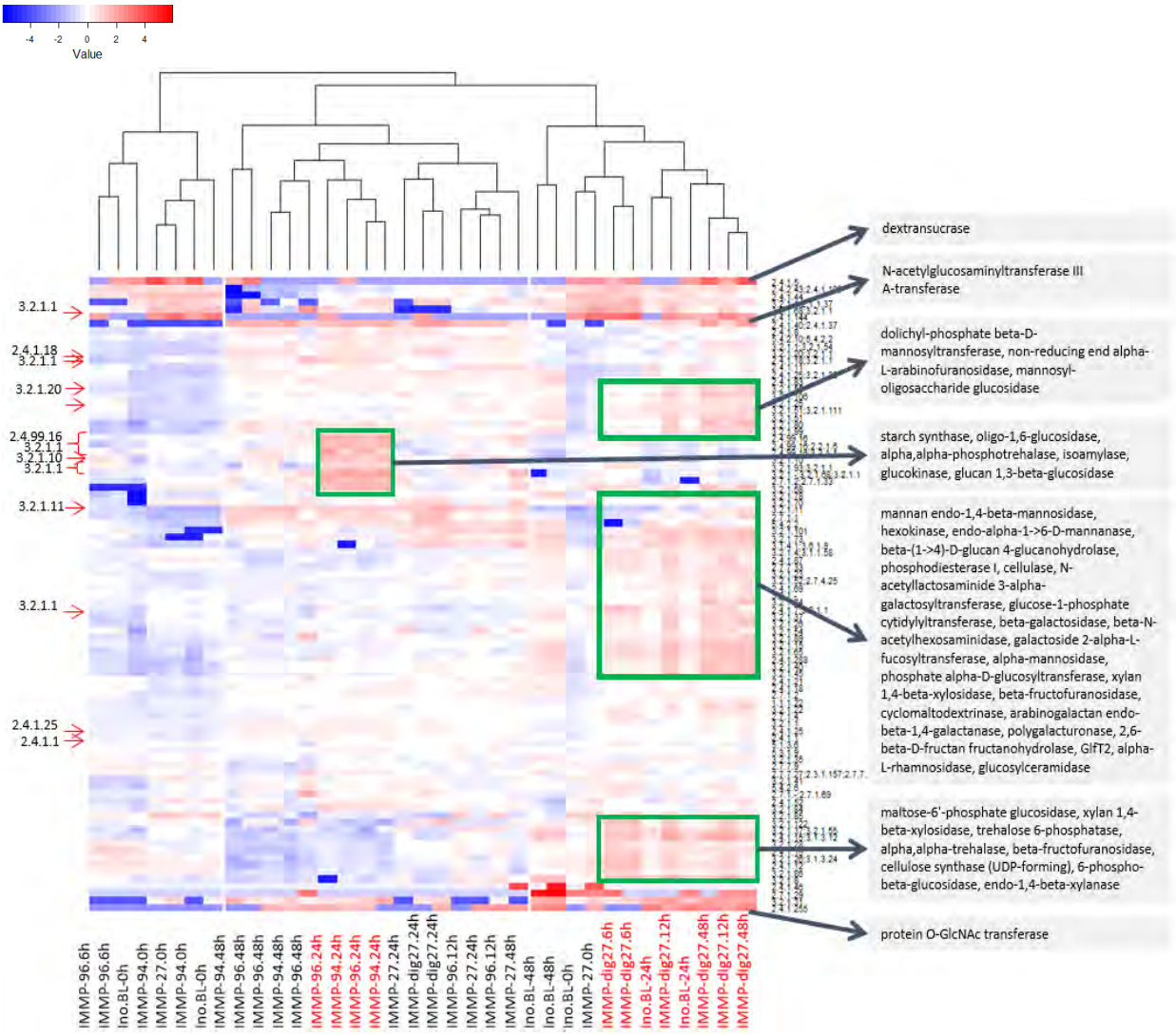
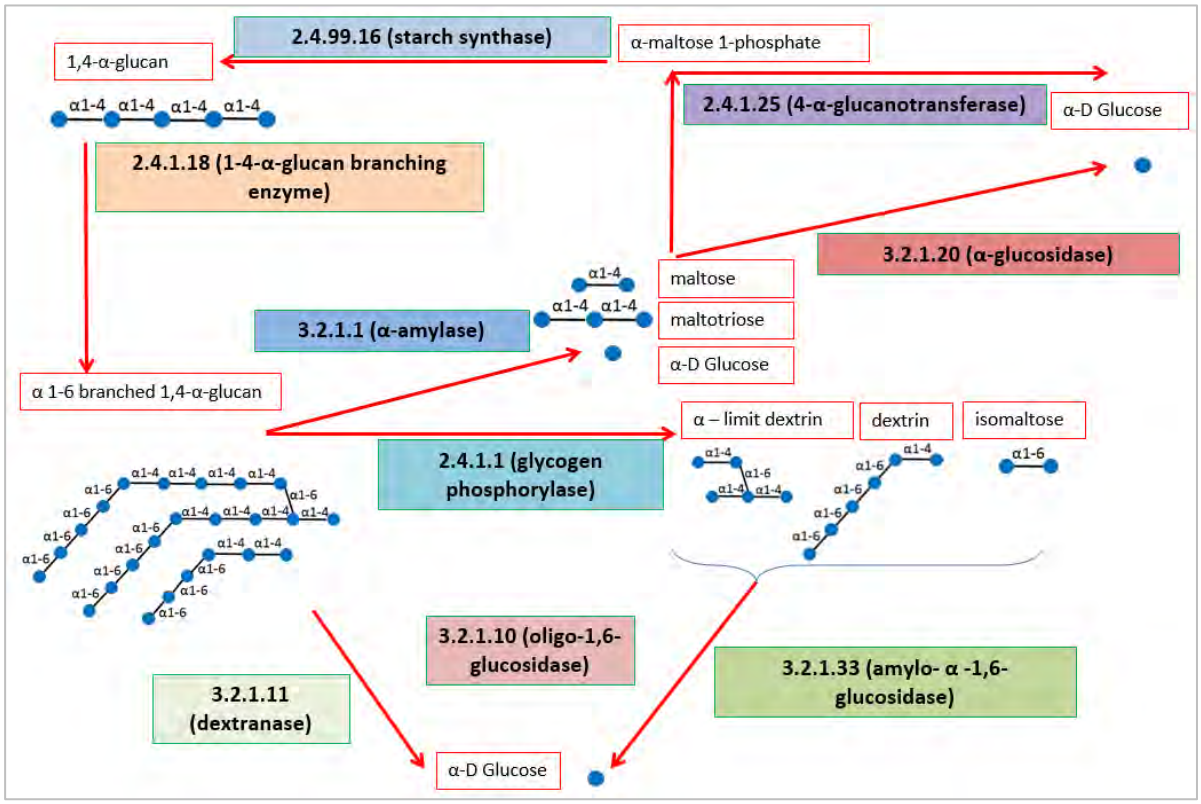
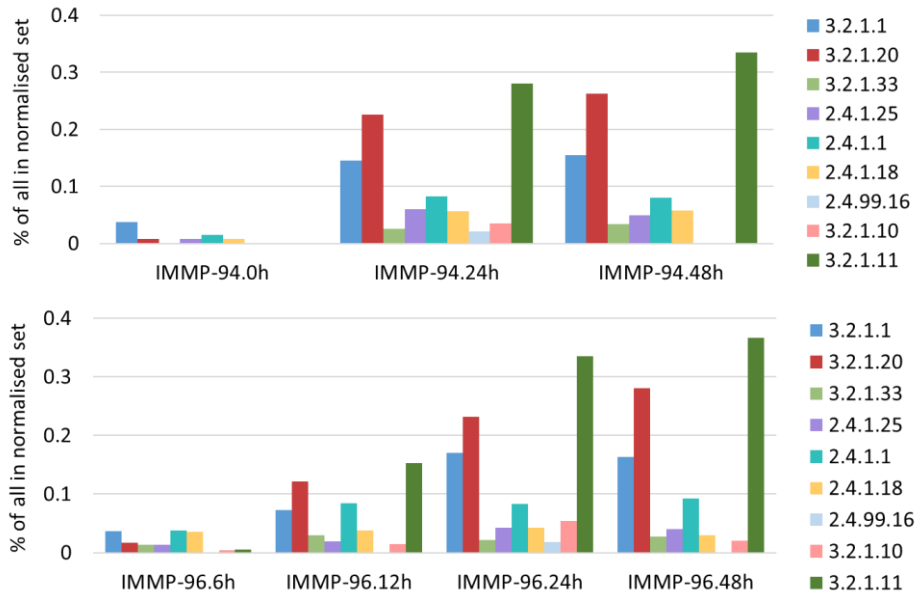


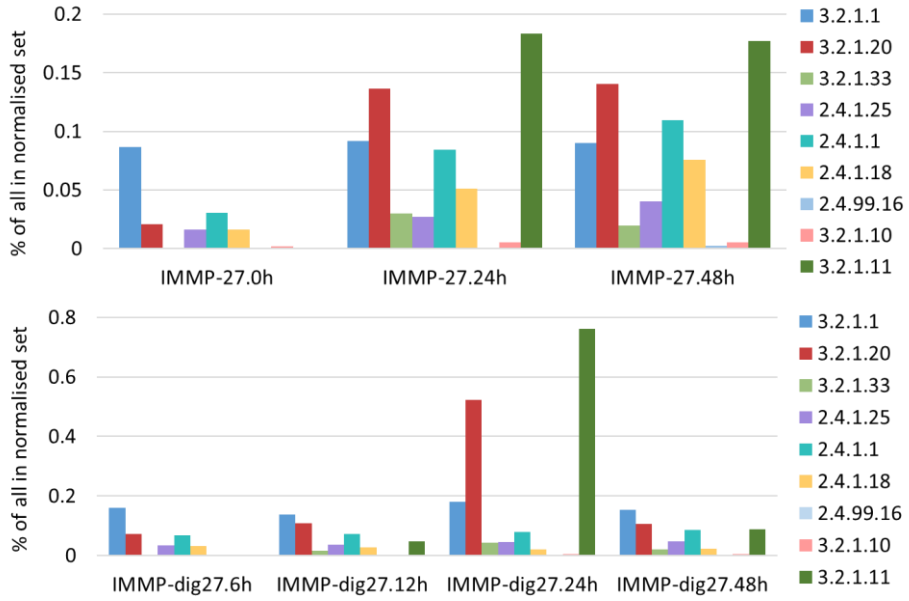
Figure S6. Heatmap of log10 transformed relative abundances of expressed genes detected in our data coding for starch and sucrose metabolism enzymes. Samples clustered based on the similarities between the up and down regulated genes. The red arrows indicate selected genes that code for enzymes described in our IMMP degradation model. Green boxes highlight the gene upregulation patterns for different IMMPs at various incubation times.

a.

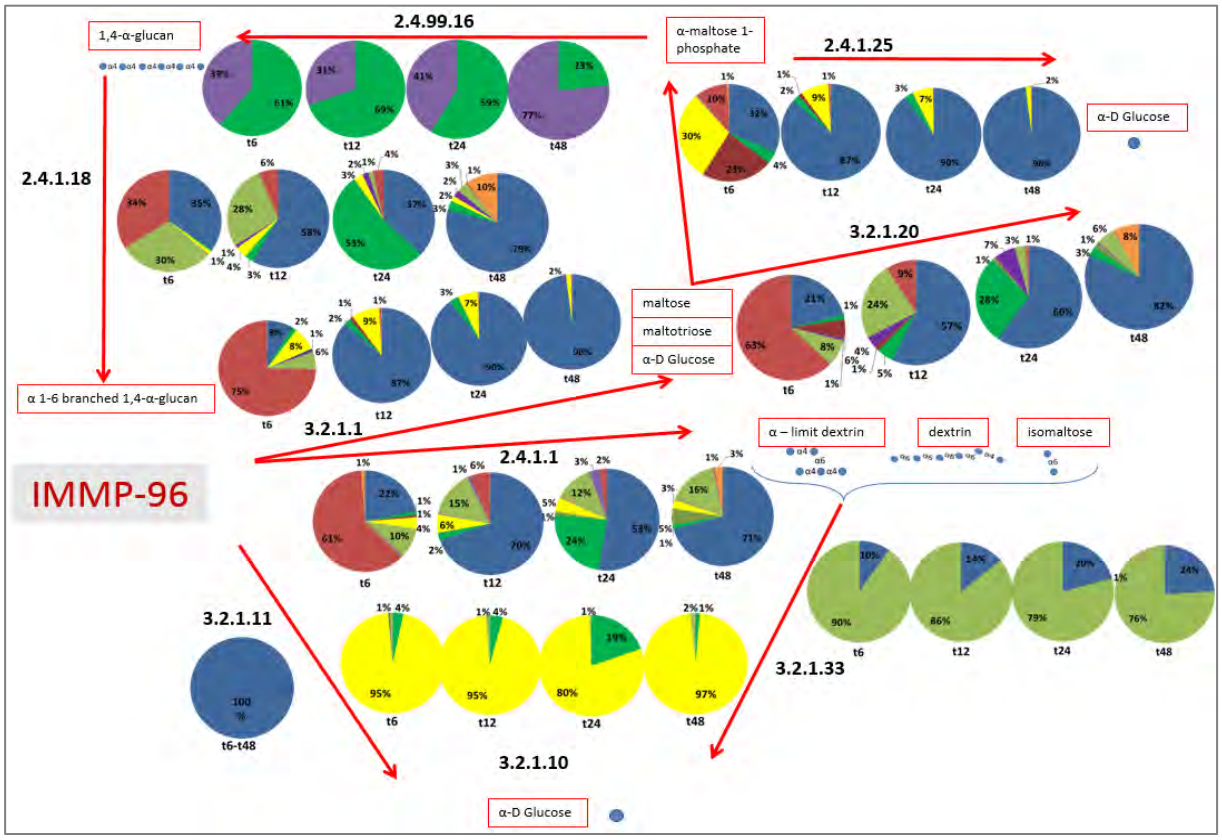

2.4.99.16 (starch synthase)
α-maltose 1-phosphate
1,4-α-glucan
α1-4 α1-4 α1-4 α1-4
2.4.1.25 (4-α-glucanotransferase)
α-D Glucose
2.4.1.18 (1-4-α-glucan branching enzyme)
3.2.1.20 (α-glucosidase)
maltose
maltotriose
α-D Glucose
3.2.1.1 (α-amylase)
α 1-6 branched 1,4-α-glucan
α – limit dextrin
dextrin
isomaltose
2.4.1.1 (glycogen phosphorylase)
3.2.1.10 (oligo-1,6-glucosidase)
3.2.1.33 (amylo- α -1,6-glucosidase)
3.2.1.11 (dextranase)
α-D Glucose

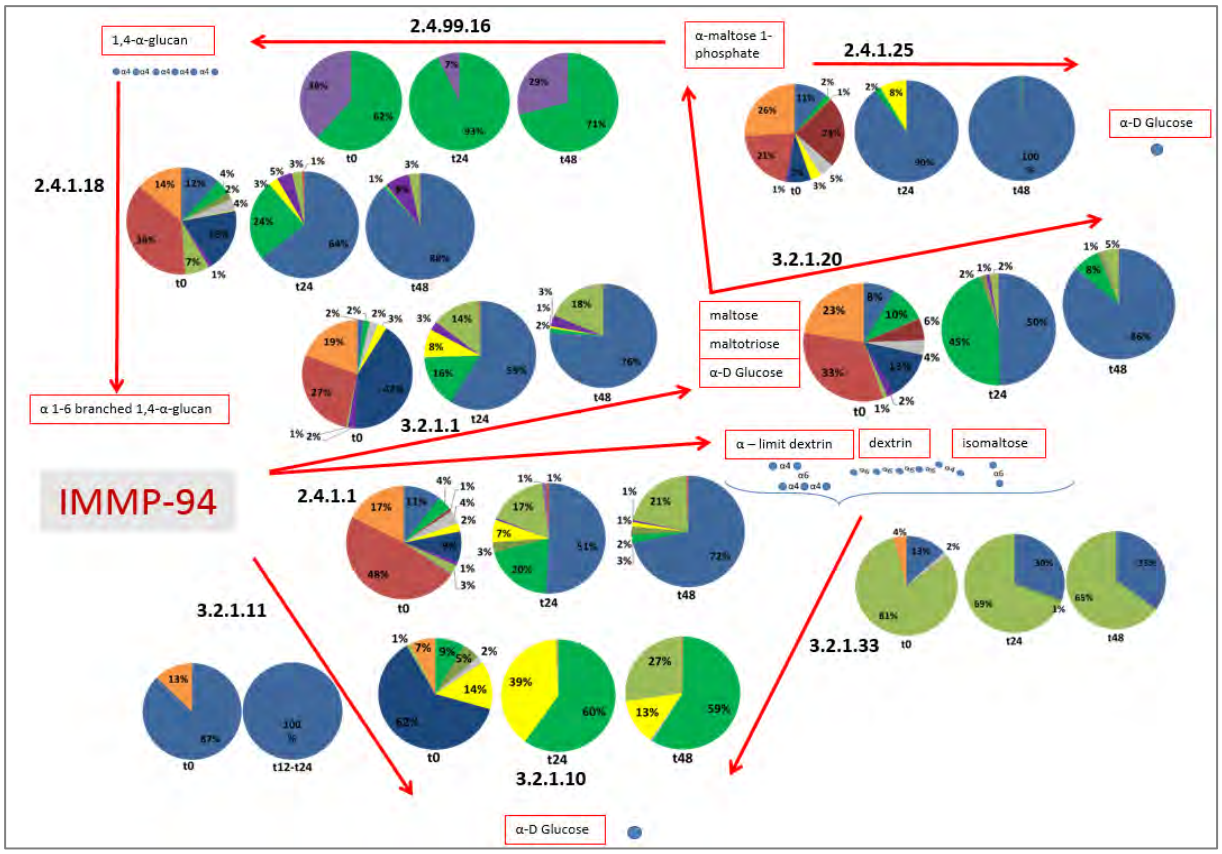
b.

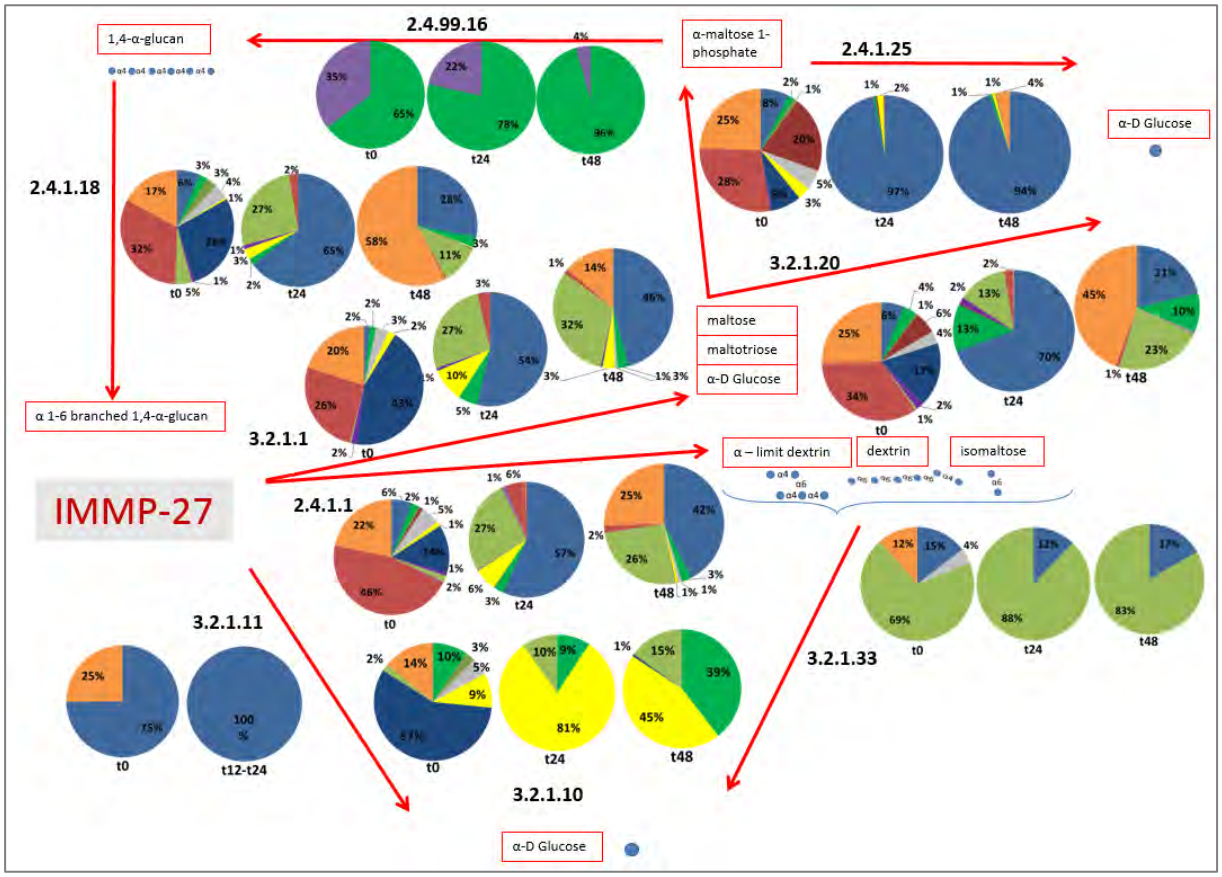

% of all in normalised set
0.4
0.3
0.2
0.1
0
IMMP-94.0h
IMMP-94.24h
IMMP-94.48h
IMMP-96.6h
IMMP-96.12h
IMMP-96.24h
IMMP-96.48h
3.2.1.1
3.2.1.20
3.2.1.33
2.4.1.25
2.4.1.1
2.4.1.18
2.4.99.16
3.2.1.10
3.2.1.11


% of all in normalised set
0.2
0.15
0.1
0.05
0
IMMP-27.0h
IMMP-27.24h
IMMP-27.48h
3.2.1.1
3.2.1.20
3.2.1.33
2.4.1.25
2.4.1.1
2.4.1.18
2.4.99.16
3.2.1.10
3.2.1.11
% of all in normalised set
0.8
0.6
0.4
0.2
0
IMMP-dig27.6h
IMMP-dig27.12h
IMMP-dig27.24h
IMMP-dig27.48h

c.


1,4-α-glucan
2.4.99.16
α-maltose 1-phosphate
2.4.1.25
α-D Glucose
2.4.1.18
3.2.1.20
maltose
maltotriose
α-D Glucose
α 1-6 branched 1,4-α-glucan
3.2.1.1
α – limit dextrin
dextrin
isomaltose
IMMP-96
2.4.1.1
3.2.1.11
3.2.1.33
3.2.1.10
α-D Glucose
t6
t12
t24
t48
t6-t48


1,4-α-glucan
2.4.99.16
α-maltose 1-phosphate
2.4.1.25
α-D Glucose
2.4.1.18
3.2.1.20
maltose
maltotriose
α-D Glucose
α 1-6 branched 1,4-α-glucan
3.2.1.1
α – limit dextrin
dextrin
isomaltose
IMMP-94
2.4.1.1
3.2.1.11
3.2.1.33
t0
t24
t48
t12-t24
3.2.1.10
α-D Glucose


1,4-α-glucan
2.4.99.16
α-maltose 1-phosphate
2.4.1.25
α-D Glucose
2.4.1.18
3.2.1.20
maltose
maltotriose
α-D Glucose
α 1-6 branched 1,4-α-glucan
3.2.1.1
α – limit dextrin
dextrin
isomaltose
IMMP-27
2.4.1.1
3.2.1.11
3.2.1.33
t0
t24
t48
t12-t24
3.2.1.10
α-D Glucose

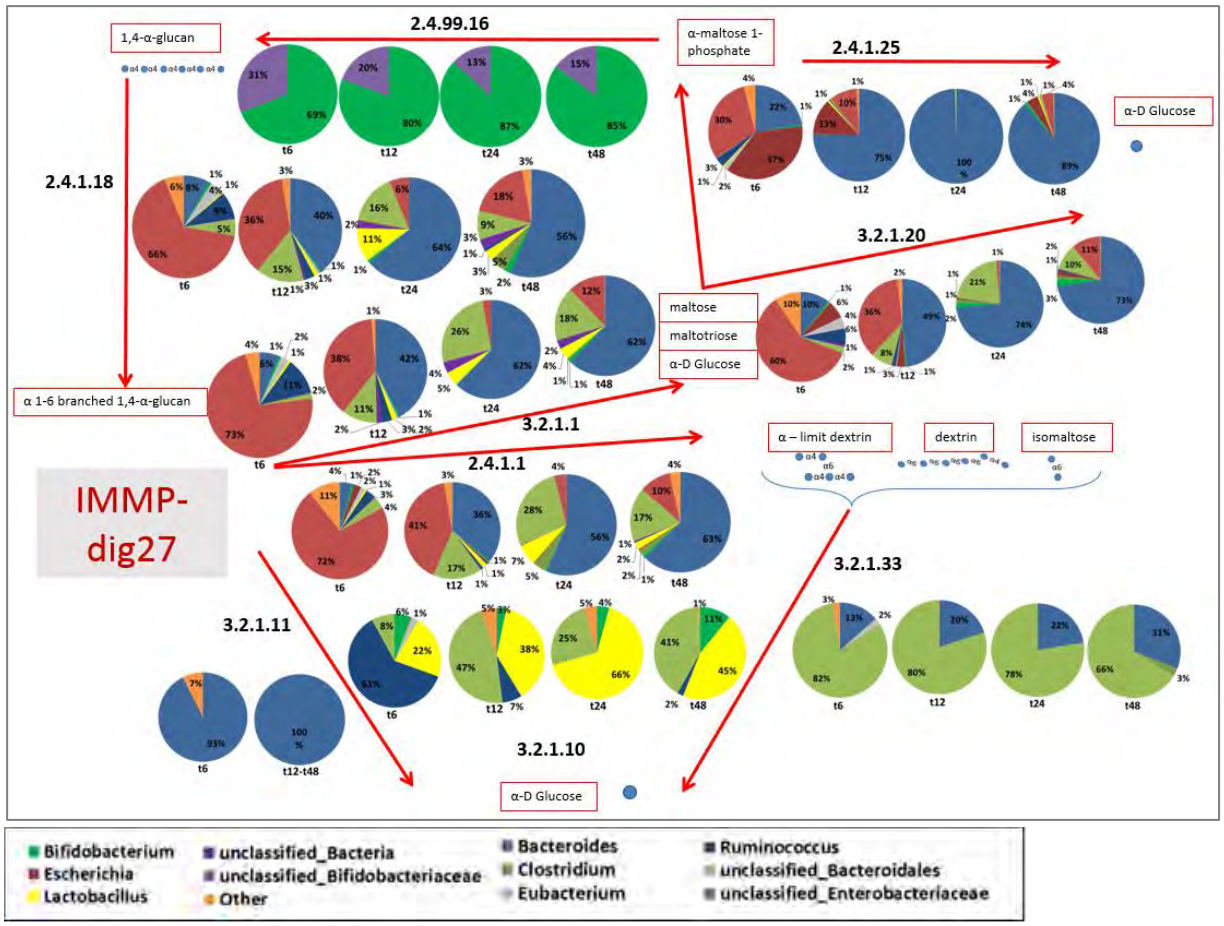


Figure S7. IMMP degradation model. a. main enzymes involved in the pathway, b. relative abundance of transcripts of genes coding for enzymes needed for IMMP degradation, c. relative contribution of different bacterial groups to expression of genes coding for the enzymes in the IMMP degradation pathway.

Table S1. Overview of the RNA-seq metrics

Condition	Total reads	rRNA	% rRNA	Non-rRNA	Trimmed bases due to adapters	Bases trimmed due to adapters (%)	Sequences passing prinseq quality filtering	% passing prinseq quality filtering	Mean length	Total % of bases passing ALL filtering steps	Mapping rate in % to assembly
Blank, repl. 1, 0 h	17182356	388147	2,26	16794209	713631983	28,14	15120013	90,03	108,34	63,56	79,06
Blank, repl. 2, 0 h	12843968	234407	1,83	12609561	620076640	32,57	11339128	89,92	101,95	60,00	75,99
IMMP-27, repl. 1, 0 h	18661592	402676	2,16	18258916	844672334	30,64	16454635	90,12	104,71	61,55	71,29
IMMP-27, repl. 2, 0 h	21152485	438493	2,07	20713992	881059209	28,17	18728000	90,41	108,06	63,78	74,42
IMMP-94, repl. 1, 0 h	30405866	555904	1,83	29849962	1087733238	24,13	26908763	90,15	113,98	67,25	79,06
IMMP-94, repl. 2, 0 h	19354896	777385	4,02	18577511	588584413	20,98	16575459	89,22	119,53	68,24	80,79
Blank, repl. 1, 24h	27195831	243329	0,89	26952502	917731871	22,55	24835747	92,15	114,96	69,99	81,28
Blank, repl. 2, 24h	26279510	263802	1,00	26015708	1199020108	30,52	23724640	91,19	103,99	62,59	81,43
IMMP-27, repl. 1, 24h	12540536	1832716	14,61	10707820	587094636	36,31	9467522	88,42	93,59	47,10	85,32
IMMP-27, repl. 2, 24h	23402662	2094144	8,95	21308518	775851341	24,11	18915069	88,77	111,03	59,83	86,95
IMMP-94, repl. 1, 24h	14390413	901227	6,26	13489186	729841441	35,83	12186486	90,34	96,64	54,56	85,41
IMMP-94, repl. 2, 24h	27352727	1655263	6,05	25697464	1235521747	31,84	23084321	89,83	102,74	57,80	84,52
Blank, repl. 1, 48h	22532893	822320	3,65	21710573	1368385387	41,74	19533106	89,97	88,47	51,13	79,64
Blank, repl. 2, 48h	23432237	902055	3,85	22530182	1200860127	35,3	20328428	90,23	97,88	56,61	80,17
IMMP-27, repl. 1, 48h	24004294	4361225	18,17	19643069	1175519132	39,63	16574957	84,38	92,59	42,62	81,24
IMMP-27, repl. 2, 48h	21275156	8177863	38,44	13097293	691799504	34,98	11702497	89,35	98,7	36,19	80,22
IMMP-96, repl. 1, 48h	28735621	11137710	38,76	17597911	1296700269	48,8	14765189	83,9	81,2	27,82	83,92
IMMP-96, repl. 2, 48h	35679453	11638379	32,62	24041074	1511937150	41,65	20304145	84,46	92,59	35,13	85,36
IMMP-96, repl. 1, 6h	18066769	185180	1,03	17881589	390649819	14,47	15736624	88	126,02	73,18	75,76
IMMP-96, repl. 2, 6h	20993700	55456	0,26	20938244	412525567	13,05	18327046	87,53	128,66	74,88	71,55
IMMP-96, repl. 1, 12h	35944607	70088	0,20	35874519	1270799851	23,46	31773426	88,57	113,35	66,80	76,36
IMMP-96, repl. 2, 12h	22773537	64983	0,29	22708554	964291453	28,12	20307424	89,43	106,59	63,37	76,79
IMMP-96, repl. 1, 24h	17241186	1380292	8,01	15860894	385783659	16,11	13984773	88,17	122,24	66,10	87,1
IMMP-96, repl. 2, 24h	19319950	1502708	7,78	17817242	607265419	22,57	15673285	87,97	111,85	60,49	86,3
IMMP-96, repl. 1, 48h	34060532	145402	0,43	33915130	1420500568	27,74	30260173	89,22	107,64	63,75	75,81
IMMP-96, repl. 2, 48h	31992519	119756	0,37	31872763	1450909006	30,15	28116504	88,21	105,1	61,58	72,91
Blank, repl. 1, 0 h	14780189	291372	1,97	14488817	458214833	20,94	12514201	86,37	111,32	62,84	84,1
IMMP-dig27, repl. 1, 6h	21177851	113633	0,54	21064218	1266290906	39,81	17927736	85,11	89,44	50,48	81,34
IMMP-dig27, repl. 2, 6h	14730499	91528	0,62	14638971	464265511	21	12827318	87,62	113,79	66,06	81,04
IMMP-dig27, repl. 1, 12h	19229982	120092	0,62	19109890	353792739	12,26	16872478	88,29	128,38	75,09	81,28
IMMP-dig27, repl. 2, 12h	19183847	184071	0,96	18999776	449569556	15,67	16837679	88,62	123,83	72,46	81,79
IMMP-dig27, repl. 1, 24h	19000262	346531	1,82	18653731	573909821	20,38	16612325	89,06	117,09	68,25	82,71
IMMP-dig27, repl. 2, 24h	18981915	260033	1,37	18721882	214922072	7,6	16320453	87,17	135,84	77,86	83,29
IMMP-dig27, repl. 1, 48h	19947312	1144473	5,74	18802839	508216740	17,9	16742425	89,04	120,37	67,35	88,6
IMMP-dig27, repl. 2, 48h	11159596	475984	4,27	10683612	160119185	9,93	9455765	88,51	132,49	74,84	89,42
Average	**21366411,23**	**1514013,714**	**6,33**	**19852397,51**	**801840435,8**	**25,74**	**17591935,06**	**85,99**	**106,19**	**60,89**	**78,66**
Total	**747824393**	**52990480**		**694833913**	**28064415252**		**615717727**				

Table S2. Assignment of genus-level taxa per cluster, showing the amount of assigned genes and differentially expressed genes

Organism	**Genes assigned**	**Genes differentially expressed**
Cluster 1		
Ruminococcus	9904	6220
Lactococcus	8211	5263
unclassified_Gammaproteobacteria	373	285
Bos	329	263
N/A	208	85
unclassified_Mammalia	198	67
unclassified_Bovidae	107	41
Clostridiales	4	0
Bacteria	3	1
Eukaryota	2	1
Gammaproteobacteria	1	1
Viruses	1	0
Cluster 2		
Lactobacillus	8655	7279
Bifidobacterium	7817	5849
unclassified_Actinobacteria	265	195
unclassified_Bifidobacteriaceae	116	98
Fusobacterium	70	47
unclassified_Lactobacillaceae	52	38
Myoviridae	36	33
Cluster 3		
Enterococcus	2580	1927
unclassified_Lactobacillales	1154	1012
unclassified_Bacilli	537	473
unclassified_Enterococcaceae	139	136
Cluster 4		
unclassified_Enterobacteriaceae	11350	9255
unclassified_Bacteria	6997	4494
Eubacterium	4884	3849
Escherichia	2972	2386
unclassified_Proteobacteria	683	530
Salmonella	59	47
Shigella	51	9
Enterobacter	36	29
Citrobacter	31	14
Vibrio	26	9
Cluster 5		
Bacteroides	53749	49101
unclassified_Bacteroidales	10243	9067
Parabacteroides	1749	919
Prevotella	302	233
Bacteria	193	133
Desulfosporosinus	32	11
Flavobacterium	30	27

Cluster 6		
Clostridium	4244	3228
unclassified_Clostridia	114	69
Cluster 7		
unclassified_Clostridiales	10819	3356
unclassified_Lachnospiraceae	1030	219
Anaerostipes	127	6
Clostridiales	39	5
Cluster 8		
N/A	8225	4083
Sutterella	3598	2645
unclassified_Bacteroidetes	770	681
unclassified_Betaproteobacteria	135	77
Odoribacter	94	76
Ethanoligenens	28	21
Corynebacterium	20	8
Cluster 9		
Bilophila	3853	2614
unclassified_Firmicutes	3152	1898
Phascolarctobacterium	1328	8
unclassified_Selenomonadales	244	2
Acidaminococcus	174	4
unclassified_Acidaminococcaceae	117	0
Selenomonas	109	19
Veillonella	88	8
Megamonas	64	12
unclassified_Veillonellaceae	56	3
Pelosinus	53	4
Megasphaera	53	9
Desulfitobacterium	51	15
Anaeromusa	33	0
Acetonema	32	3
Mitsuokella	20	5

Text mining

Further EC numbers were derived by text mining and matching all InterProScan derived domain names against the BRENDA database (download 13.06.13) [23]. The text mining algorithm included lower casing all characters, removal of non-alphanumerical characters (colons, commas, brackets, apostrophes, dashes, terminal points), removal of partial and generic terms (type, terminal, subunit, domain, enzyme, like, hypothetical, conserved, operon, active site, enzyme, probably, central, 51 kd, respiratory chain, c terminal, n terminal), rejection of overly generic final result terms (kinase, cytochrome, protein, methyltransferase) and reduction of certain terms (deletion of PEP/pyruvate binding; removal of “prokaryotic” in “prokaryotic cytidylate kinase”; “family” in “cytidilate kinase family”; “phosphorylating” in “glyceraldehyde phosphate dehydrogenase phosphorylating”; “iron containing” in “iron containing alcohol dehydrogenase”; “zinc containing” in “zinc containing alcohol dehydrogenase”; “manganese containing” in “manganese containing catalase”; “20 kd” in “nadh ubiquinone oxidoreductase 20 kd”; replacement of “carboxyltransferase” with “carboxylase” in “pyruvate carboxyltransferase”). Furthermore, all terms, which were only of length one, were also removed, in case the remaining name contained more than two words. On some domain names a manual curation was performed, and overly generic identifications (e.g. matching PF12847 “Methyltransferase domain” with e.g. EC 2.1.1.124 with alternative name “Protein Methyltransferase I”) were rejected.

Chapter 8

General Discussion and Future Perspectives

Microbial communities living in the human gastrointestinal (GI) tract are highly diverse, metabolically active, and they have a profound effect on their host's health and well-being. Through the application next generation sequencing based methods targeting microbial DNA and RNA as measures of microbial community composition and activity, we were able to study the development of the GI microbial ecosystem during early life, and to investigate the responses of the faecal microbial community to a set of novel prebiotics during an *in vitro* fermentation experiment.

In the first section of this thesis, I provided a brief review on microbial ecosystems inhabiting different regions of the adult human GI tract, and highlighted the changes in microbiota structure and/or function which are known to be associated with selected diseases. Since the time **Chapter Two** was written and published in March, 2016 [1], nearly five thousand new research and review articles on GI tract microbiota and its relation to health have been published and deposited to PubMed (search performed on February 19th, 2018, using keywords "intestinal microbiota" OR "gut microbiota" AND "health"). Thus, the field is rapidly developing, and new discoveries had been made in the last two years, which were not included in the review, but I believe are worth mentioning here. For example, one of the most important challenges in the field of human microbiota research has been defining a "normal" microbiota, and how it varies in healthy individuals. There is still no clear definition to what is "normal", but the issue has been addressed in a number of large cohort studies revealing the wide scope of variability in faecal microbiota composition in healthy adults [2, 3]. These studies also investigated the effect of a large number of life factors, such as, for example, dietary and lifestyle habits, socioeconomic status, stress, medication use and various health parameters, to evaluate whether any of these factors could explain some of the variability in human GI tract microbiota composition [2-5]. As expected, dietary habits were identified among the most common and important factors influencing GI microbiota in healthy adults. Notable progress has also been made in the area of research on the role of GI microbiota in the aetiology of different diseases, especially in the studies investigating the gut-brain axis and the link between the GI microbiota and multiple sclerosis [6]. Neurological conditions, such as Alzheimer's disease [7] and Parkinson's disease [8] have now also been shown to have a microbial component. Interesting developments have been made in understanding the role of the microbiota in various types of cancers, particularly colorectal cancer, and the effectiveness of anti-cancer immunotherapies [9]. Also, a number of studies looked into mechanisms via which GI microbiota contributes to the development of other, non-GI cancers [10], in particular with regard to liver cancer [11], and breast cancer [12]. Furthermore, the gut-lung axis has been investigated with respect to the translocation of microbial products and their role in lung cancer and other respiratory diseases [13]. Finally, there is accumulating evidence that the genetic make-up of the host determines the presence or absence of certain microbial groups [14], leading to a growing number of studies, which incorporate the aspect of host genetics to better understand the microbiota - host interphase and its effect on host's health [15], for example in relation to Inflammatory Bowel Disease (IBD) [16] (Figure 1). Nevertheless, even with all the progress in the field of microbiota research in the last two years, **Chapter Two** still offers a valid and comprehensive overview.

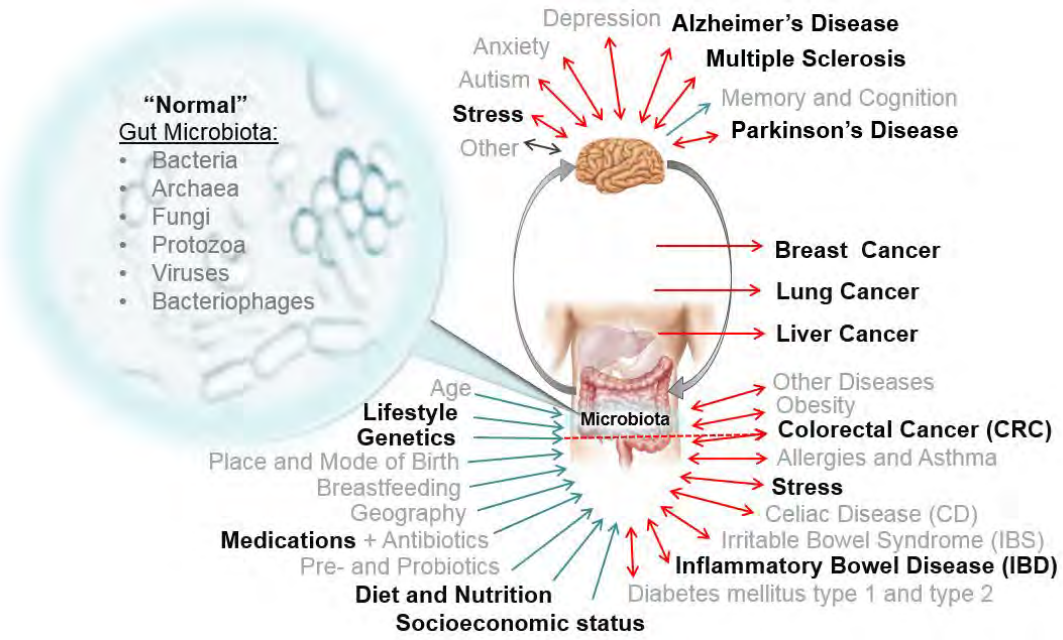


Figure 1. Summary of selected factors and diseases associated with changes in GI tract microbiota structure and function highlighting some of the areas where important progress has been made in the last two years.

The developing microbiota in early life

In the second section of this thesis I presented our research findings that focus on the GI microbiota composition and development in healthy infants mainly in relation to feeding mode, prebiotic use and maternal breastmilk oligosaccharide content. In **Chapter Three** we showed that at one month of age, infants receiving traditional formula without galacto- and/or fructo-oligosaccharides (scGOS, lcFOS; KOALA cohort), had distinctively different faecal microbiota profiles than the breastfed controls. These differences were less pronounced in the BINGO cohort, presumably as a result of prebiotic supplementation that is common in infant formulas nowadays. A number of earlier, mainly culture based studies, showed that formula feeding significantly decreased infant faecal counts of *Bifidobacterium* and increased numbers of enterobacteria. The results from the KOALA cohort were in line with these early findings, as we observed significant reduction in the relative abundance of number of genera associated with breastfeeding (e.g. *Bifidobacterium, Lactobacillus, Staphylococcus, Haemophilus*) and increase in the number and relative abundance of other taxa, including those that are normally found in the more adult – like microbiota (e.g. *Alistipes, Enterococcus, Clostridium*, Lachnospiraceae *Incertae Sedis,* unidentified genera within the families Ruminococcaceae and Enterobacteriaceae) [17-19]. Thus, the overall number of the detected OTU "species" was higher and their distribution was more even, as compared to what was observed in breastfed infants. Even though the prebiotic supplemented formulas used in the BINGO cohort did not reverse all the differences in faecal microbiota richness, we could see an improvement in

species evenness, or distribution, which became more similar to that of breastfed infants. Also, the levels of *Bifidobacterium*, *Lactobacillus, Clostridium* and enterobacteria were more similar to those measured in the breastfed infants. This result is not surprising, as both GOS and FOS are well studied prebiotics with a documented bifidogenic effect *in vitro* and *in vivo*, although it should be noted that these results in infants seem to depend on the exact type of GOS/FOS and doses of the oligosaccharides being tested [20, 21]. In addition, these studies reported a positive effect on other related parameters, such as stool pH, stool frequency and consistency, and production of SCFA [21, 22]. Unfortunately, we do not have data on these parameters, also because of the low quantities of the samples which could be obtained for the analyses.

Here I would like to discuss in more detail the results of the microbial analyses at the OTU level, in particular the composition of the bifidobacterial fraction. In both study cohorts, we found the same three *Bifidobacterium* OTUs which were predominant in the faeces of the one month old infants. The OTU names and numbers, their nucleotide sequences, and the top NCBI blast results are summarised in Table S1. Earlier studies showed that *Bifidobacterium longum, Bifidobacterium longum* subsp. *infantis* (*B. infantis*), *Bifidobacterium bifidum,* and *Bifidobacterium breve* were the four main species detected in the faeces of breastfed infants, while *Bifidobacterium adolescentis* and *Bifidobacterium pseudocatenulatum,* which are found in more adult like microbial communities were associated with formula feeding [23, 24]. In addition, a recent study showed that GOS/FOS fortified formulas resulted in bifidobacterial species distribution similar to that found in breastfed infants and dominated by *B. breve, B. longum* and *Bifidobacterium catenulatum*, while unsupplemented formulas led to increase in relative abundance of *B.catenulatum* and *B. adolescentis* [25]. Thus, based on literature data and NCBI blast analyses of corresponding sequences, it is tempting to speculate that the three main bifidobacterial OTUs in our dataset correspond to *B. bifidum*, one or more members of the unresolved cluster of *B. breve, B. longum (infantis)* and *B.catenulatum,* and possibly *B. adolescentis*. However, it should be kept in mind that these are only predictions and should be confirmed, using for example, species and strain specific qPCRs targeting different genes (e.g. *groEL* in *Bifidobacterium*) allowing unequivocal resolution at the species level [26]. To this end, our results also reinforce the notion that short amplicon sequencing data in many cases does not provide the required resolution to explicitly assign microbial species to observed OTUs [26, 27].

Remarkably, while in the KOALA cohort formula feeding and breastfeeding resulted in significantly different distribution in all three major bifidobacterial OTUs, in the BINGO cohort the relative abundance of the most abundant *Bifidobacterium* OTU L2 was not significantly different between infants fed breastmilk, formula or both (Figure 2). However, just like in the KOALA cohort, *Bifidobacterium* OTU L1 increased significantly in formula fed infants and in spite of the use of prebiotics. When infants received mixed feeding (breastmilk and formula), there were no significant differences in the main bifidobacterial OTUs as compared to breastfed infants in the BINGO cohort, whereas in the KOALA cohort there was a significant decrease in the relative abundance of *Bifidobacterium* OTU L2 (Figure 2).

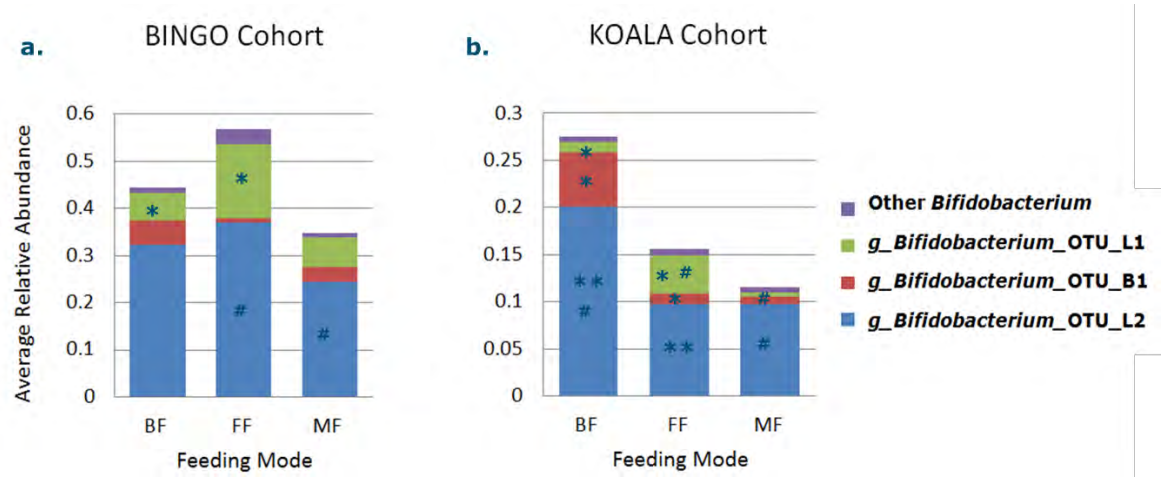


Figure 2. Relative abundance of bifidobacterial Operational Taxonomic Units (OTUs) in different feeding modes. a. Differences of the main *Bifidobacterium* OTUs between formula fed (FF) and breastfed (BF) infants in the BINGO cohort are not significant for OTUs L2 and B1 but significant for OTU L1 ($p<0.05$); b. There are significant differences in the main *Bifidobacterium* OTUs between BF and FF infants in the KOALA cohort (OTU L2 FDR<0.05; OTU L1 and B1 $p<0.05$). In both cohorts mixed feeding (MF) resulted in decrease in relative abundance of bifidobacteria, and there was a significant decrease for OTU L2 ($p<0.05$) between FF and MF in the BINGO cohort and between BF and MF in the KOALA cohort, and a significant decrease in OTU L1 in MF compared to FF in the KOALA cohort. Symbols *, #, ** indicate pairs of taxa that were compared and were significantly different in Kruskal - Wallis test ($p<0.05$).

Few culture based studies up to date looked at the prebiotic effect of supplementation of infant formulas with GOS [22, 28], or mixtures of GOS and FOS [21, 29], and showed an increase in cell counts of *Bifidobacterium* and *Lactobacillus,* and no change in *Bacteroides*, *Clostridium* spp., *E. coli*, *Enterobacter*, *Citrobacter*, *Proteus*, *Klebsiella*, and *Candida* [21]. In addition, several studies investigated the genetic and physiological potential of certain bacteria to utilise complex carbohydrates and reported that different bacterial species or strains exhibit adaptations which lead to high structural specificity in utilising different compounds [30-32]. For example, an *in vitro* fermentation study, which tested growth of 29 *Lactobacillus* spp. strains and 39 bifidobacterial strains from both human and animal origin, showed that various strains of *B. infantis, B. bifidum* and *B. adolescentis* grow very well on GOS and FOS, with *B. adolescentis* strains reaching lower optical densities than those reported for *B. infantis* and *B. bifidum.* On the other hand, *B. adolescentis* grew better on inulin than either *B. infantis* and *B. bifidum* [30]. Also, it has been shown that *B. adolescentis* has a number of genetic adaptations allowing it to thrive in the adult GI tract and utilise a much wider range of carbohydrates, including (resistant) starch and starch-like oligo- and polysaccharides, that are normally present at high concentrations in the diets of adults [31]. Thus, to really understand the effects of different compounds on microbial groups, we should focus more on the responses within individual species or strains. It is also important to remember that the ability of certain strains to grow and degrade carbohydrates may be different when the microbes are grown in pure or defined co-culture or inside their natural ecosystems, within the host [33-35].

Our study methods, although limited in their ability to unambiguously identify individual species and strains, provided us with a comprehensive and sensitive view at the overall microbial diversity, and the relative abundance of individual populations within the faecal community. In order to gain a better understanding of the dynamic patterns within the community structure, and to reveal possible dependencies between community members, we performed Dirichlet multinomial mixture model (DMM) clustering analysis [36]. The resulting DMM model revealed presence of three distinct assemblages that could be found in infants from both study cohorts, and regardless of feeding modes. These assemblages, or clusters, were characterised by either mixed community structure, or by communities with either a high relative abundance of *Bifidobacterium*, or a high relative abundance of both, *Bifidobacterium* and *Bacteroides* (**Chapter Three**, Figure S2). We also noted a significant correlation between the prevalence of the different cluster types and infant age and feeding mode. Older infants, and those receiving the prebiotic fortified formulas were more likely to have faecal communities of the *Bifidobacterium* or *Bacteroides/ Bifidobacterium* dominated clusters, whereas very young infants or those fed traditional formulas, were more likely to be characterized by the mixed cluster type. Similar clusters were detected in another study, which showed that the cluster structure could be linked to adverse health effects, for example higher risk of developing atopy and asthma later in life [37]. In addition, our BINGO data revealed that most infants underwent a gradual and directional progression through the clusters, typically starting from the mixed cluster defined by high levels of the initial colonisers, and towards *Bifidobacterium*, or *Bifidobacterium/Bacteroides* dominated clusters at weeks six and twelve. Contrary to what has been reported previously in a similar study [38], this gradual transition was not observed when infants received the prebiotic fortified formulas, where the transition seemed to occur more abrupt from the mixed community type to the *Bifidobacterium*-dominated cluster. Together, these findings suggest that, as the community structure might be improved with prebiotics, the succession dynamics might be altered and deviate from that observed in breastfed infants. This alteration might be important, as the patterns in the colonisation dynamics might be biologically relevant and could be linked with health outcomes later in life [39]. For example, a recent study which also followed infant faecal microbiota progression through the three microbial clusters, concluded that a delay in reaching the mature cluster defined by high relative abundance of *Bifidobacterium* and low relative abundance of *Streptococcus*, was actually considered suboptimal to infant's health, as it was also associated with a shorter gestation time, caesarean section delivery, and relatively low adiposity at 18 months [40]. The long-term consequences, however, remain unknown. In addition, none of the studies to date looked at the physiological effect of prebiotic supplemented formulas in infants. This aspect should also be taken into consideration, as a recent study showed that in healthy adults high doses (16g/day) of GOS and FOS led to adverse effects on fasting glucose levels and glucose metabolism, a decrease in butyrate-producing bacteria (*Ruminococcus, Phascolarctobacterium, Coprococcus* and *Oscillospira*), and a subsequent decrease in faecal concentrations of short chain fatty acids (SCFA), even though they also led to a significant increase in bifidobacterial populations [41]. Thus, even though the prebiotic concentrations are much lower in the prebiotic fortified infant formulas (e.g. Nutrilon -scGOS:lcFOS (9:1); 0.6 g/100 ml), and these formulas provide a promising alternative to breastfeeding with regard to their bifidogenic effect, a broader effect of these products on the infant's metabolism should also be investigated.

In healthy humans, microbial species within the genera *Bifidobacterium* and *Bacteroides* play important roles in the degradation of complex carbohydrates [42-44]. Based on the results presented in **Chapter Three** we could clearly see that breastfeeding plays a crucial role in driving the GI tract microbiota succession, leading to establishment of the microbial ecosystem strongly dominated with certain *Bifidobacterium* OTUs. Since breastmilk HMOs are indigestible by infants' enzymes, this group of glycans is believed to have evolved specifically to guide the microbial colonisation process in infants [42, 45]. In **Chapter Four** and **Chapter Five** we looked in more detail at the KOALA and BINGO cohort data to investigate the link between selected breastmilk HMOs and the microbial groups involved in their utilisation *in vivo*.

A number of studies up to date reported that infant associated microbial populations, mainly the species and strains of *Bifidobacterium,* have a genetic potential to employ various strategies and mechanisms to efficiently transport and hydrolyse different HMO types, and to colonize the infant GI tract [42, 46]. For example, *B. infantis* has highly conserved gene clusters encoding ABC transporters which show a strong affinity for neutral HMOs containing either lacto-N-biose I (LNB) or Nacetyllactosamine (LacNAc; e.g. in LNT, LNnT), fucosylated HMOs (e.g. 2′FL, 3FL, LNFPI, LNFPII, LNFPIII, LNFPV, LNDFH), acidic HMO (e.g. 3′SL and 6′SL) and Lewis epitopes (please see **Chapter Four**, Table S1 for the full names of the HMO structures). In addition, the high abundance of genes encoding Family 1 solute binding proteins (F1SBPs) found as part of the ABC transporters suggests that these proteins might help *B. infantis* to attach itself to intestinal epithelial cells, and may facilitate the colonisation process [42, 47]. *B. infantis,* and its close relative *B. longum,* the most common bifidobacterial isolates in both, children and adults [48], are described as "selfish" feeders, which tend to import and degrade oligosaccharides inside their cells, without releasing monosaccharides, but instead secreting acetic and lactic acid into their environment. On the other hand, *B. bifidum* relies on a set of diverse membrane-associated extracellular glycosyl hydrolases, α-fucosidases and α-sialidases, which allow for extracellular degradation of a wide range of HMOs, and cross feeding with other resident microbes, such as *B. breve,* which shows a preference for sialylated HMOs and can also grow on free fucose and sialic acid [46]. The different feeding strategies within the main bifidobacterial groups are summarized in Figure 3.

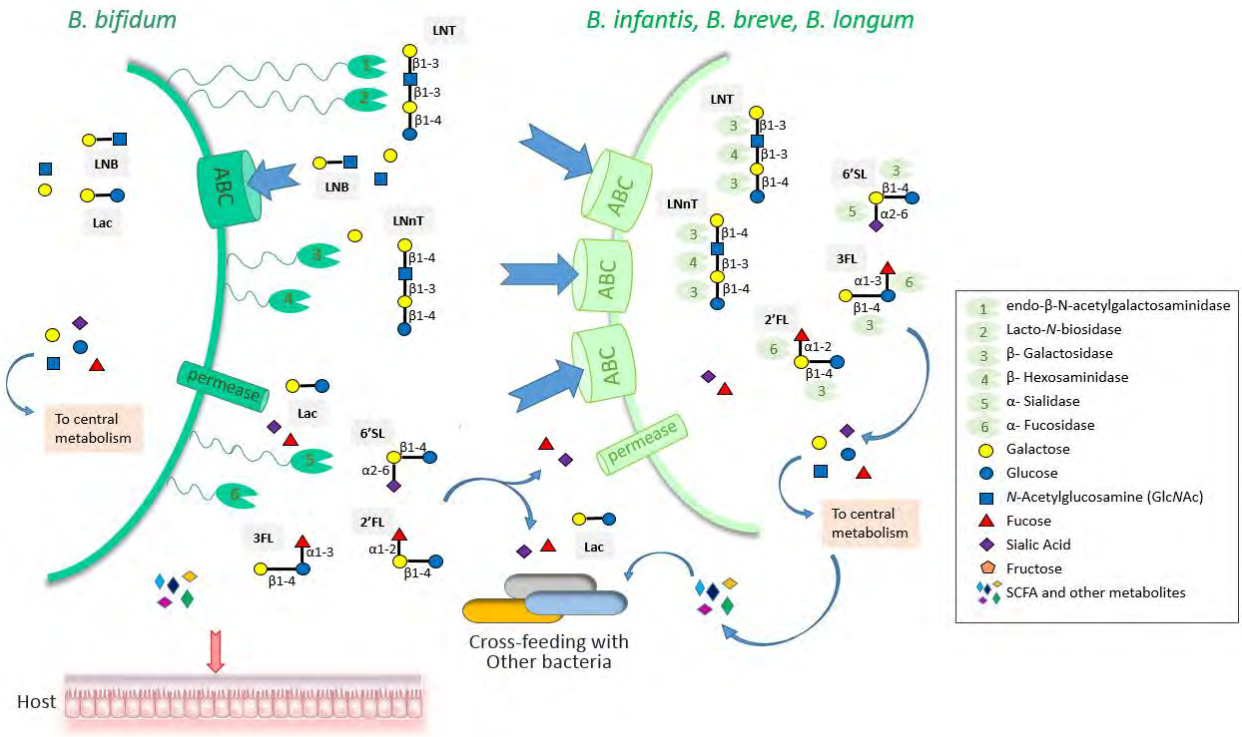


Figure 3. Representation of two major metabolic strategies in infant-associated *Bifidobacterium* spp. for HMO utilization (adapted from [46, 49]).

Based on the available data, as discussed in the previous paragraph, our main hypothesis was that breastmilk HMO content might directly influence the composition of GI microbiota, and that microbial species associated with the infant GI tract specialize in the consumption of specific HMOs [42, 45]. In contrast to this hypothesis, our analyses in both, the KOALA and the BINGO cohorts, showed only a very weak association between specific HMOs in mother's milk and the microbiota composition in infant faeces. More specifically, the lack of a consistent, positive link between any of the HMOs we measured and the main bifidobacterial groups was surprising. Thus, our findings support the idea that, rather than the individual HMOs stimulating individual species, the infant microbiota composition is shaped by a combination of factors related to breastfeeding, including the mixtures of various HMO types, antimicrobial factors present in breastmilk (lysozyme, lactoferrin, secretory IgA), and the microbiota of the breastmilk itself [50-55]. In addition, our hypothesis was largely based on the assumption that the ability of different microbial species to effectively utilise different HMOs would directly translate into a higher relative abundance of these species in the community. This may not be so simple, as studies showed that growth of some microbial groups, such as *Staphylococcus*, can be enhanced just by the presence of HMOs by activating growth promoting signalling, and without the need of that species to actually consume HMOs [56]. Knowing how important the initial colonisation of the GI tract is for human health, it is feasible to think that the mechanisms that evolved to control this process are complex and multifactorial, and that there exists some level of functional redundancy between different factors, for example, between various HMO structures.

In both study cohorts we estimated HMO consumption levels of selected HMOs based on their measured amounts in breastmilk, and in corresponding infant faeces. In both cohorts (**Chapters Four** and **Five**), we showed that higher HMO consumption was significantly associated with higher relative abundance of bifidobacteria, including the main *Bifidobacterium* OTU L2, and longitudinal data from the BINGO indicated that HMO consumption also increased with infant age (**Chapter Five**). Our data showed that *Bifidobacterium* OTU L2 was associated with high consumption of a wide range of HMOs, including neutral HMOs (LNH, pLNH, LNnH, LNT, LNnT), fucosylated HMOs (2′FL, DFL, 3FL, LNFPI, LNFPII, LNFPIII, LNFPV, LNDFHII) and 3′SL. *Bifidobacterium* OTU B1 was linked with high consumption of LNH, pLNH, LNnH, LNT, LNnT, 2′FL, LNDFHI, LNFPI, LNFPII, LNFPIII, and 3′SL. Interestingly, *Bifidobacterium* OTU L1, which we hypothesized to be one of the infant nonspecific strains (i.e. most closely related to *B. adolescentis*), was also not significantly associated with consumption of any of the HMOs, in neither of the two cohorts studied here. Finally, we also saw that the degradation of LSTa, LSTb, LSTc and 6′SL was not linked with any of the major *Bifidobacterium* OTUs, but was weakly correlated with Lachnospiracaeae, low abundance bifidobacterial OTUs, *Actinomyces,* and few other low abundance groups.

The fermentation of HMOs provides a substantial amount of energy, and thus, it likely plays a key role in shaping the network of metabolic dependencies between the members of the microbial community in the infant GI tract. Based on the available data we built a microbial interaction network by identifying statistically significant positive and negative correlations between different OTU level groups across all samples. Using BINGO cohort data, we noticed that as infants were getting older, two main network centres emerged (Figure 4).

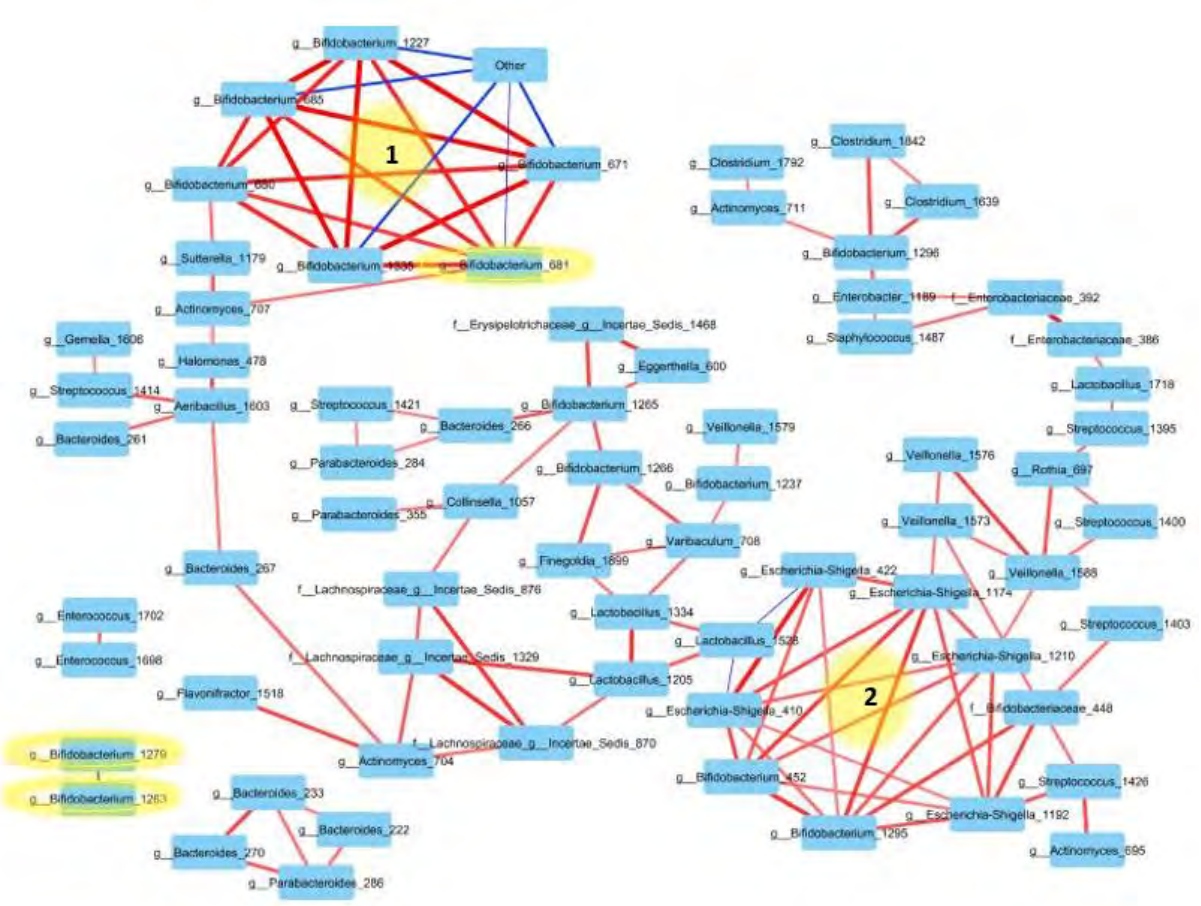


Figure 4. The OTU network showing statistically significant ($p<0.05$) positive and negative associations which pass Spearman correlation threshold of 0.5 in 12 week old infants from the BINGO study cohort. Centres and *Bifidobacterium* OTUs of interest are highlighted in yellow.

The first centre was characterised by the presence of positively correlating *Bifidobacterium* OTUs, including *Bifidobacterium* OTU B1 (681). As already mentioned, OTU B1 likely corresponds to *B. bifidum,* which can degrade HMOs extracellularly, allowing it to cross feed with other microbial groups [46]. Indeed, we could show that *Bifidobacterium* OTU B1 was positively linked with a high number of other bifidobacterial OTUs, and also with *Actinomyces* OTU 707. In contrast, *Bifidobacterium* OTU L2 (1263), which was predicted to be likely related with "selfish" *B.longum* or *B. infantis*, was correlating with only one other group, showing a negative correlation with *Bifidobacterium* OTU L1 (1279). The second centre was dominated by OTUs of *Escherichia – Shigella*, few lower abundance *Bifidobacterium* OTUs, and was negatively associated with *Lactobacillus* OTU 1528. Earlier i*n vitro* fermentation studies showed that besides *Bifidobacterium*, only selected *Bacteroides* spp. (e.g. *B. fragilis* and *B. vulgatus*) could consume HMOs, whereas monocultures of *Escherichia coli* and selected species and strains within the genera *Lactobacillus*, *Clostridium*, *Eubacterium*, *Streptococcus*, and *Veillonella* grew poorly, or not at all, on common HMOs as the only carbon source [57]. This might explain why in the *Escherichia – Shigella* OTU dominated centre all *Escherichia – Shigella* OTUs were positively correlating with at least one *Bifidobacterium* OTU. In fact, in the entire network a large number of OTUs displayed direct positive correlation with either *Bifidobacterium*, and/or *Bacteroides* OTUs, suggesting the central role that these

groups play in the infant microbial community. Only few studies to date described mutualistic interactions between various bifidobacterial strains and other intestinal bacteria and showed the importance of this microbial cross-talk on the overall community composition and metabolism, mainly in relation to the modulation of SCFA production [58]. The findings presented in **Chapters Three** and **Four** might provide clues on microbial co-occurrences and the types and extent of interactions between different members of the microbiota, particularly in relation to HMO utilisation inside the infant GI tract.

Evaluating the potential of novel prebiotics

The first two sections of this thesis highlight the importance of the GI tract microbiota in human health and investigate how prebiotics could modulate the development and the composition of the microbiota in the infant GI tract. The last section, which include **Chapters Six** and **Seven,** investigated the prebiotic potential of different IMMPs. The IMMPs tested in our study were characterized by differences in the distribution of the side chains containing α-(1→6) glyosidic linkages. The Etenia^{TM} starch derived IMMP-96 was characterised by the highest content of α-(1→6) glyosidic linkages, and bimodal distribution of side chains as compared to IMMP-27 and IMMP-94. Despite this difference, both IMMP-94 and IMMP-96 showed a similar degradation pattern, with the utilisation of the polysaccharide fraction at 12-24 h, and subsequent utilisation of the α-(1→6) linked oligosaccharides with DP>7 between 24-48 h. The IMMP-27 and IMMP-dig27 contained a mix of α-(1→6) glyosidic linkages and α-(1→4) glyosidic linkages and showed different degradation patterns. IMMP-27, which contained over 70% of α-(1→4) linkages, was utilised in two stages. First the α-(1→4) linked maltodextrin fraction was used in the first 12 h of incubation, followed by the utilisation of the α-(1→6) glyosidic linkages between 12 and 48 h. IMMP-dig27, which contained mostly short α-(1→6) linked side chains and was depleted of α-(1→4) glyosidic linkages, was completely degraded between 12 and 24 h of incubation. Taken together our results suggest that the linkage type, the length of α-(1→6) linked side chains, and the overall composition all play a role in the degradation dynamics. We hypothesized that the members of our starting microbiota already had the enzymatic capacity to degrade α-(1→4) glyosidic linkages, and that changes in the community composition, or the activation of a different metabolic machinery, was needed to utilise the α-(1→6) glyosidic linkages.

In **Chapter Seven** we presented microbial gene expression data to explain the degradation of IMMPs and the associated metabolic changes. By combining findings presented in **Chapters Six** and **Seven**, together with our current understanding of degradation of similar polymers by human GI tract associated microbes [35, 59-62] we propose a model of degradation of IMMPs (Figure 5).

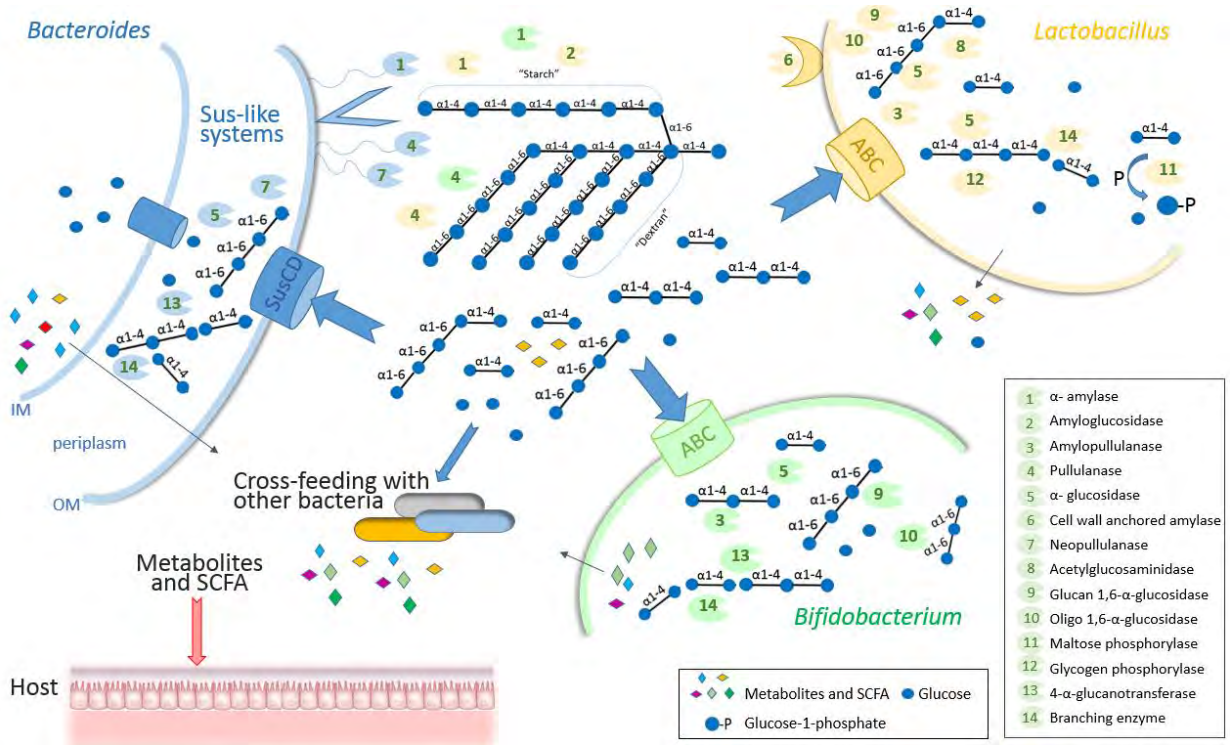


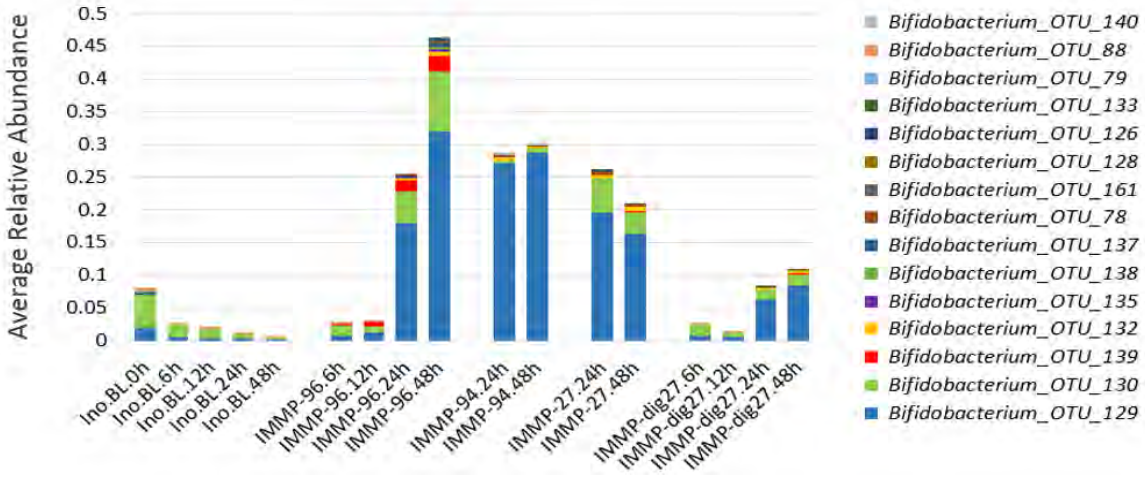
Figure 5. Generalised model of IMMP degradation by the three main groups: *Bacteroides*, *Bifidobacterium* and *Lactobacillus*. The model is based on RNA expression of selected genes discussed in **Chapter Seven**.

First the α-(1→6) and α-(1→4) glycosidic bonds in the IMMP polymers are hydrolysed with extracellular endo acting α-amylases and pullulanases (debranching enzymes) to produce oligosaccharides. In earlier studies, α-amylase activity was described in several lactobacilli, including *L. acidophilus*, *L. fermentum, L. plantarum, L. mannihotivorans, L. amylovorus*, and *L. gasseri* [59]. Our analyses showed that activity of *L. fermentum* and *L. gasseri* was related to the presence of IMMPs, with *L. fermentum* being already active at 6 h, while *L. gasseri* showing the maximum activity at later timepoints (data not shown). *Bacteroides* spp. specialize in degradation of complex polysaccharides, activate their starch utilisation systems (Sus-like systems) and trap the starch chains of the IMMPs to hydrolyse the internal α-(1→4) glycosidic bonds, which releases maltooligosacharides, isomaltose, dextrin and α-limit dextrins. Members of the genus *Bacteroides* have a wide array of enzymes to hydrolyse different glycans, and they can prioritize uptake of specific structures, for example as a strategy to avoid competition [35]. Among the main competitors for *Bacteroides* spp. are *Bifidobacterium* spp. Some bifidobacteria can produce and secrete α-amylases, but the majority of species rely on the presence of primary degraders to break down large molecules. During IMMP degradation, the levels of transcription of α-amylase (3.2.1.1) encoding genes were high at 0 h and 6 h, and observed transcripts were mainly assigned to unclassified Enterobacteriaceae and *Lactobacillus*. In contrast, at 12 and 24 h of incubation, the expressed α-amylase genes were assigned to *Bacteroides, Lactobacillus* and *Bifidobacterium*. Among all IMMPs, the highest expression of the α-amylase genes assigned to members of the genus *Bifidobacterium* (16%) was in the incubations containing

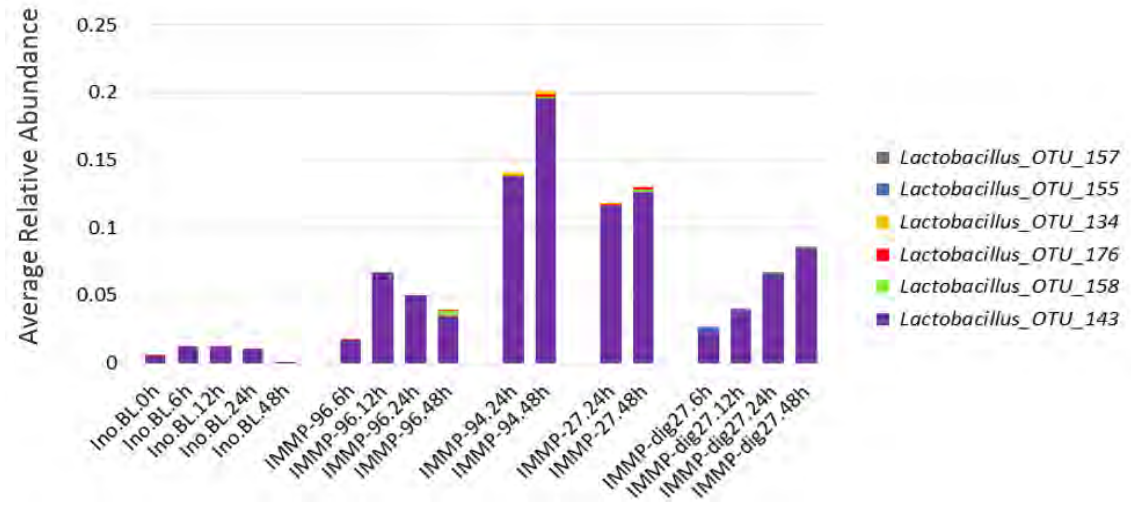
IMMP-94. Once IMMPs were degraded to smaller fragments, they could be imported inside of the bacterial cells and further fermented in the cytosol, or in the periplasm (*Bacteroides*). Smaller oligomers and glucose units could also be taken up and metabolised by other microbial groups. Oligomer fermentation continued via action of exoamylases, such as α-glucosidases (EC 3.2.1.20), capable of cleaving α-(1→6) and α-(1→4) glycosidic bonds, and glucan 1,6-α-glucosidase (EC 3.2.1.70) and oligo 1,6-α-glucosidase (EC 3.2.1.10) which cleave α-(1→6) glyosidic linkages to release terminal glucose units. High expression of genes encoding α-glucosidases was found in *Bacteroides* and *Lactobacillus,* while genes encoding oligo 1,6-α-glucosidase (EC 3.2.1.10) were almost exclusively originating from *Lactobacillus* and *Bifidobacterium.* Among all IMMPs, incubations with IMMP-94 at 24 h and 48 h had the highest fraction of the oligo 1,6-α-glucosidase encoding mRNA coming from *Bifidobacterium* (60%). Besides being cleaved, oligomers that enter cells could also be phosphorylated, or modified via action of transferases, such as transglucosylase (4-α-glucanotransferase, EC 2.4.1.25) and branching enzymes, such as 1-4-α-glucan branching enzyme (EC 2.4.1.18), which move 1-4-α-glucans and form new α-(1→6) and α-(1→4) glycosidic bonds. The expression of genes encoding for both of these enzymes was highest in *Bacteroides* and *Bifidobacterium* at 24 h.

The analyses at OTU level were not presented in **Chapter Six**, but provided interesting insights, in particular with respect to the three main microbial groups, i.e. *Bacteroides*, *Bifidobacterium* and *Lactobacillus,* which based on the data presented in this thesis were the key players in the degradation of the IMMPs. Our analyses indicated presence of 15 different bifidobacterial OTUs, six OTUs within the *Lactobacillus*, and 16 in the genus *Bacteroides* (Figure 6, Table S1).

a.



b.



c.

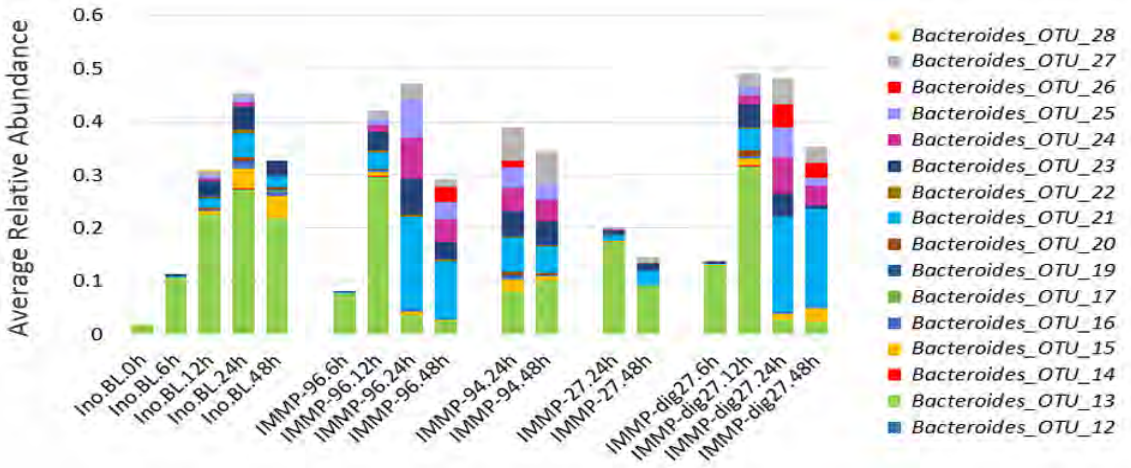


Figure 6. Relative abundance of OTU level taxa detected at different time points during *in vitro* fermentation of IMMPs. a. *Bifidobacterium* OTUs; b. *Lactobacillus* OTUs; c. *Bacteroides* OTUs

The most abundant *Bifidobacterium* OTUs were *Bifidobacterium* OTU 129, *Bifidobacterium* OTU 130, and *Bifidobacterium* OTU 139 (Figure 6a). Surprisingly, *Bifidobacterium* OTU 129 had a 100 % nucleotide sequence match to the earlier described *Bifidobacterium* OTU L2, which was also the most abundant bifidobacterial OTU in breastfed infants in the KOALA and BINGO studies described in the **Chapters Three, Four** and **Five.** Similarly, *Bifidobacterium* OTU 130 had a 100% nucleotide match to the sequence of *Bifidobacterium* OTU L1, which in the KOALA and BINGO studies was significantly enriched in the formula fed infants (Figure 2). Finally, *Bifidobacterium* OTU 139 matched several different bifidobacterial species and strains, with the highest total match score to *Bifidobacterium animalis*, including *B. animalis* subsp. *lactis* (Table S1). *B. animalis* subsp. *lactis* is normally found in large intestine of humans, and is commonly used in probiotic supplements. A recent study also showed that *B. animalis* subsp. *lactis* has a carbohydrate specific ABC transporter system which allows it to simultaneously utilise a wide range of substrates, for example the Isomalto-oligosaccharides (IMO) α-(1,6)-glucosides and soybean

α-(1,6)-galactosides [62]. It is tempting to speculate that the closely related *Bifidobacterium* OTU 139 has the same system, as this might explain its growth on a substrate such as IMMP-96 which contains a mix of long and short side chains. In addition, *B. animalis* subsp. *lactis* is known to produce a wide range of carbohydrate hydrolases, which further expands its range of the possible carbon sources, and might benefit this species in its competition with other microbial groups [63]. Furthermore, our analyses on enzymatic activity of extracellular and cell associated fractions indicated that endo- acting enzymes, which degrade polysaccharides into oligosacharides, were mostly associated with the extracellular fraction. Smaller oligosaccharides could be then degraded using exo- acting enzymes, which were associated mainly with the cellular fraction, suggesting that most of the oligosaccharides were transported and degraded inside the bacterial cell. This is in line with our prediction of possible species to belong to *B. longum (infantis*), *B. adolescentis* and *B. animalis* subsp. *lactis* dominating the ecosystem. Comparative genome studies showed that despite the fact that these species belong to three different *Bifidobacterium* clusters, they all share the same "core" of 67% of genes, including genes encoding the necessary machinery to transport and degrade mono- and oligosaccharides intracellularly, via carbohydrate uptake systems such as ATP-binding cassette (ABC) transporters or phosphotransferase systems (PTS) [63].

Within the genus *Lactobacillus*, *Lactobacillus* OTU 143 was the most predominant, comprising more than 80% of all *Lactobacillus* OTUs and showing a 100% match only to one strain - *Lactobacillus mucosae* strain S32 (DSM 13345) [64] (Figure 6b). *L. mucosae* strain S32 is a pig isolate, but another *L. mucosae* strain ME-340 was found in humans [65]. Metabolic studies on pig isolates showed that *L. mucosae* could ferment a wide range of simple sugars and produce lactic acid [64]. However, its ability to utilize complex sugars has not been yet characterised.

Finally, the 16S rRNA sequencing data indicated that members of the genus *Bacteroides* comprised up to 50% of the microbial communities in the different in vitro incubations (Figure 6c). *Bacteroides* are Gram negative bacteria, which comprise a significant fraction of the microbial community in the large intestine of adult humans, and also are the main microbial group responsible for glycan catabolism, as they are able to ferment a wide range of monosaccharides and polysaccharides [33]. This is probably the reason why, when analysing community composition at the genus level, we could not detect significant differences in the relative abundance of *Bacteroides* between the control and prebiotic treatment groups. At the OTU level, however, we could see clear changes in the distribution of different *Bacteroides* OTUs within the different treatment groups (Figure 6c). In particular, *Bacteroides*_OTU_13 was highly dominant in the control and in the groups supplemented with IMMP-27. In the groups supplemented with IMMP-96, IMMP-94 and IMMP-dig27 the relative abundance of *Bacteroides* OTU 13 gradually decreased during fermentation, while there was an increase in the relative abundance of *Bacteroides* OTU 21, *Bacteroides* OTU 24, *Bacteroides* OTU 25, and *Bacteroides* OTU 27. The NCBI blast revealed close matches of the above nucleotide sequences to the corresponding 16S rRNA gene fragment of several well characterized species of *Bacteroides* (Table S1).

Cultivation studies suggest that *Bacteroides vulgatus*, *Bacteroides distasonis* and *Bacteroides thetaiotaomicron* are the most abundant *Bacteroides* spp. in the human colon (10^{10} per g dry weight of faeces), followed by *Bacteroides fragilis*, *Bacteroides ovatus*, *Bacteroides eggerthii*, *B. uniformis* (10^{9} per g dry weight of faeces) [66]. Members of the genus *Bacteroides* specialize in extracellular polysaccharide degradation by producing a wide range of carbohydrate-active enzymes (CAZymes), which allow them to degrade a wide range of dietary substrates [35]. Genome sequencing showed that the types and numbers of the CAZymes vary greatly between species, and that in general, over 80% of gene clusters which are involved in carbohydrate metabolism (mainly glycoside hydrolases and polysaccharide lyases), are coupled with signal sequences to form specialised gene clusters for export of the enzymes to the surface of the cell. These gene clusters, encoding so-called Sus-Like Systems, which are neither found in bifidobacteria nor in lactobacilli (or Firmicutes in general), allow *Bacteroides* spp. to effectively capture and hydrolyse large polysaccharides outside of their cells, and transport the hydrolysis products inside their cells [35].

The results presented in **Chapter Seven** verified our OTU data analyses and aligned with metabolite analyses that showed that the degradation of the polysaccharide fraction of the IMMPs took place mainly between 12-24 h, except for IMMP-27, for which the degradation of α-(1→6) linked oligosaccharides was delayed. This aligns well with the *Bifidobacterium* gene expression data, showing a peak at 24 h for all IMMPs except IMMP-27, where the *Bifidobacterium* gene expression was highest at 48 h (Figure 7).

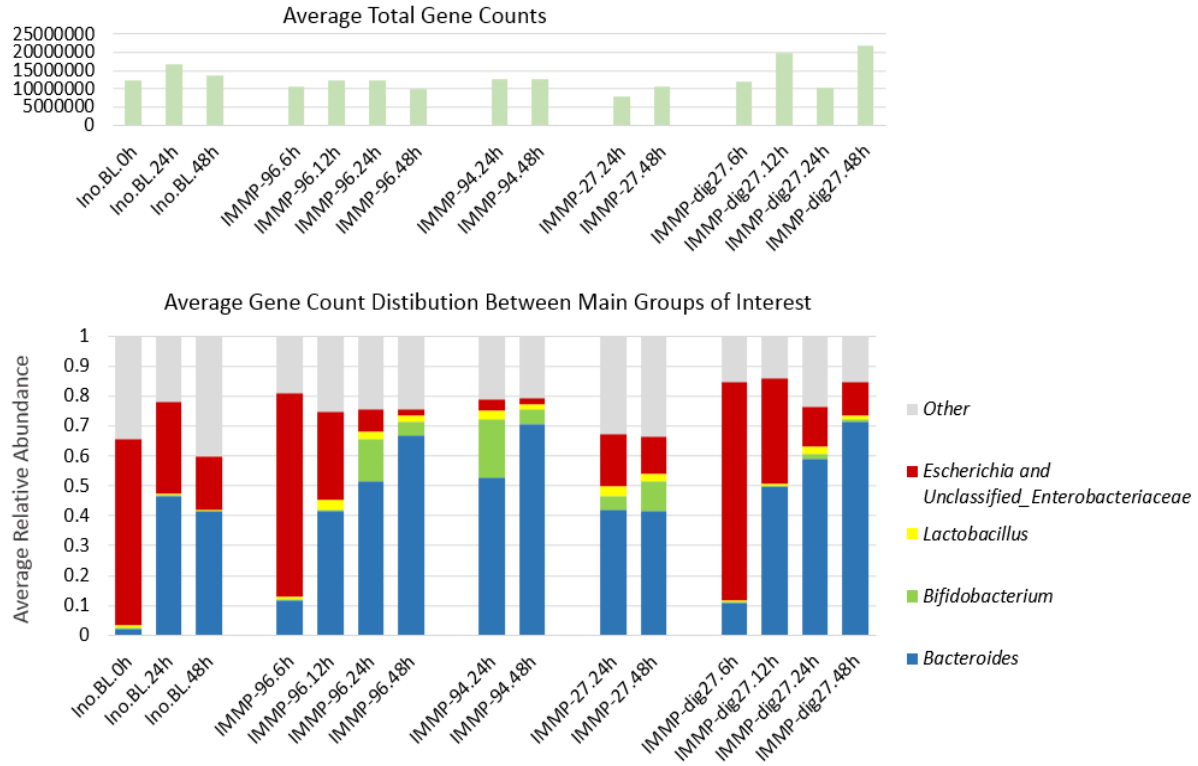
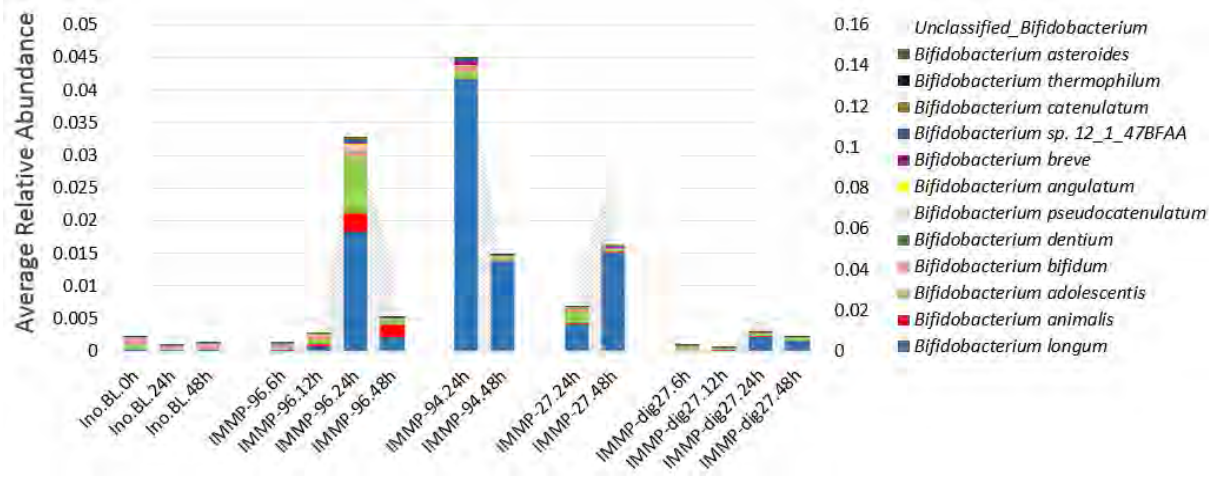


Figure 7. Average total gene counts and gene distribution among main bacterial taxa of interest in all treatment groups at different time points during fermentation. Details on gene expression data processing and analysis are discussed in detail in **Chapter Seven**.

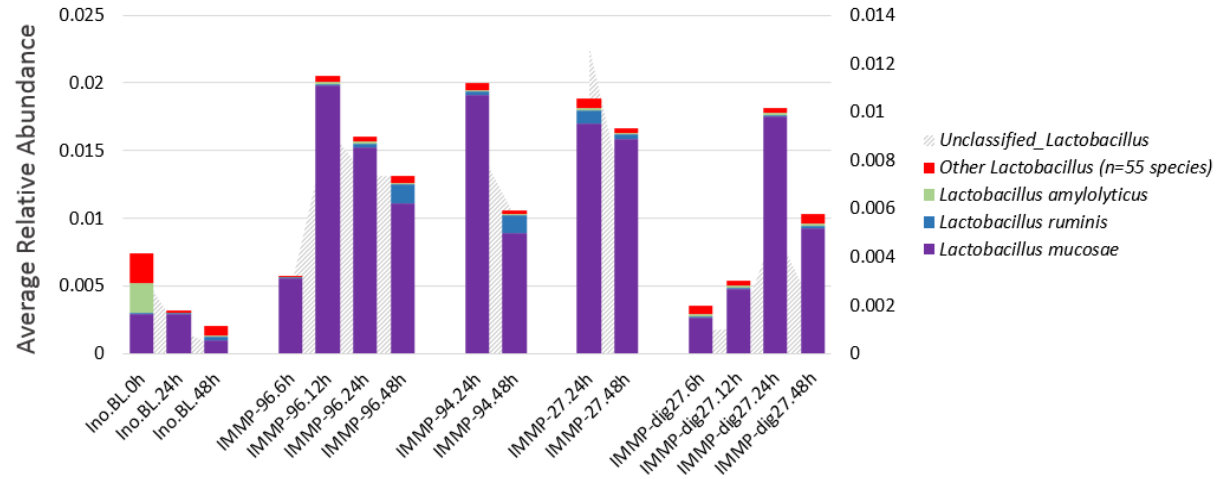
Overall, we identified 190,000 bacterial genes being expressed (**Chapter Seven)**, of which more than 25 % were assigned to the genus *Bacteroides*. Furthermore, the activity of *Lactobacillus* and *Bifidobacterium* appeared highly dependent on the presence of any of the different IMMPs in the fermentation media. The lowest relative abundance, and the lowest level of upregulation of expression of *Lactobacillus* and *Bifidobacterium* genes was observed during incubation with IMMP-dig27. Notably, incubations in the presence of this substrate showed very high activity of *Bacteroides*, and also a rapid accumulation of succinic acid, as discussed in **Chapter Six**.

Finally, the metatranscriptome data could also be used to confirm our species predictions made in **Chapter Six,** particularly within the three main groups of interest: *Bacteroides*, *Lactobacillus* and *Bifidobacterium* (Figure 8).

a.



b.



c.

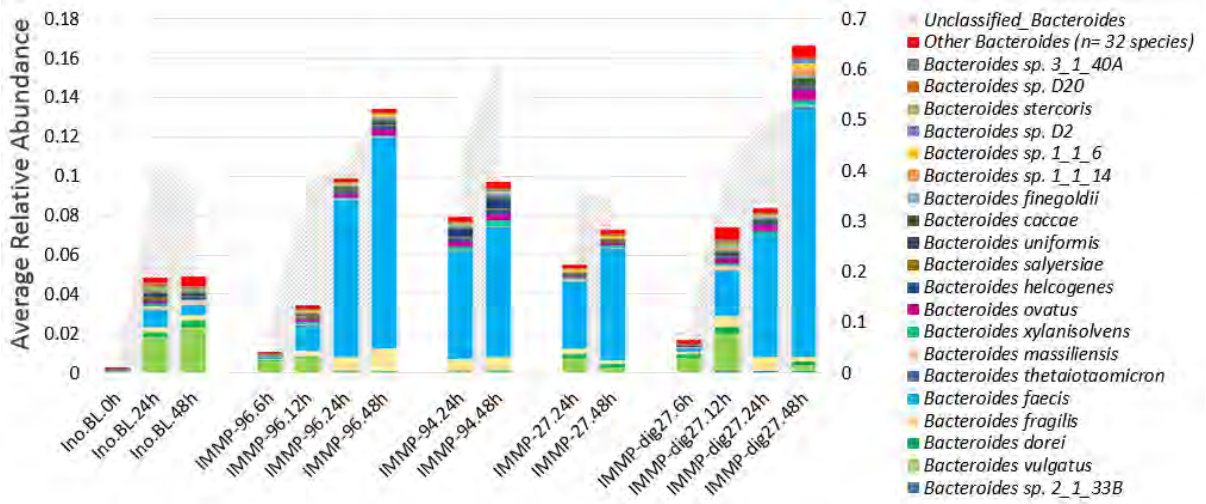


Figure 8. Species composition and the relative abundance of their gene counts detected at different time points during *in vitro* fermentation of IMMPs. Relative abundance of the unclassified taxa are shown as a patterned area with a corresponding scale on the right. a. *Bifidobacterium*; b. *Lactobacillus*; c. *Bacteroides*. Please note the different scales used for the different genera.

When we compared the results obtained with the two approaches, we noted that a large proportion of transcripts detected in the metatranscriptome analyses could not be assigned to a known species. This is likely due to ambiguous assignments, for example when the same hits correspond to a number of different species (Bastian Hornung, personal communication). However, of the twelve species of bifidobacteria which could be identified, a large proportion of reads was assigned to *B. longum, B. animalis* and *B. adolescentis*, which agrees with our OTU blast interpretation (Table S1). There was also overall agreement between the two data sets in the identification of *L. mucosae* as the main *Lactobacillus* species. However, based on our OTU data, this species was highly dominant, and we only identified five other *Lactobacillus* OTUs. In contrast, the metatranscriptome analysis indicated presence of three highly abundant species in addition to a very large number (n=58) of species with low transcript levels, and those which could not be identified beyond genus level. Nevertheless, our data suggests that this genus is very active in the metabolism of the IMMPs (**Chapter Seven**). In addition, we noted a relatively high activity of lactobacilli in the control samples at the start of the experiment, concomitant with very high activity of *Escherichia* and unclassified Enterobacteriaceae. This is an interesting observation, especially because it has been shown that growth of certain strains of *Lactobacillus* might be facilitated in the presence of *E. coli* [33]. Finally, the metatranscriptome data indicated presence of 52 active species of *Bacteroides*, in addition to a very high number of reads that could not be assigned beyond the genus level. The activity, and the relative abundance of *B. feacis* was specifically related to the presence of the IMMPs in the media. On the other hand, the activity of *B. vulgatus* showed an opposite trend and could be linked with the control and early fermentation stages of the IMMP-dig27 incubations.

The studies presented in the last section of this thesis had some advantages as well as limitations, which I will briefly describe here. One of the appealing characteristics of the *in*

vitro batch fermentation model used here is its experimental simplicity, allowing us to investigate the effect of different variants of a novel substrate class on human colonic microbiota without any ethical constraints and difficulties associated with organising a human trial. Another advantage was that we could collect samples at different time points and thus, we were able to investigate the temporal dynamics of IMMP degradation and the associated changes in the microbial community composition and activity. However, there were also several important limitations. For example, we could not account for the influence of the host, and, related to that, the simulation time had to be limited to 48 h because of microbial metabolite accumulation, as well as substrate depletion. Furthermore, the fact that at the beginning of the experiment we observed a large proportion of *Escherichia-Shigella,* which is not normally found in such high amounts in faeces of healthy adults, indicated that most probably the faecal inoculum was exposed to oxygen during the activation procedure. The overgrowth of this group might have been also related with the use of SIEM medium, which is designed to imitate the material that enters the colonic environment [67]. The SIEM medium contains peptone and bile salts which made the fermentation system eutrophic and selective towards bile tolerant microbiota, which also has been shown to stimulate growth of *Escherichia-Shigella* and *Bilophila in vitro* [67]. Both, the initial infiltration of oxygen, such as found in the GI tract of new-born infants, and the culture media which resulted in a nutritionally rich environment (perhaps, in this sense resembling the large intestines of adults on western diets) were not intentional, however, our results showed that oxygen depletion and the addition of IMMPs lead to a number of positive changes and a restoration of more desirable microbial community composition. As the incubation progressed, we could see a reduction in the relative abundance of *Escherichia-Shigella* and the recovery of *Bacteroides* to the levels that are normally found in a health adult human colon [68]. In addition, the supplementation of IMMPs resulted in a significant increase in the relative abundance of *Bifidobacterium* and *Lactobacillus,* higher production of SCFA, and the corresponding drop in the pH, all of which considered as beneficial to host's health. In a sense, these initial fermentation conditions, and the changes we observed resembled processes that we think also occur in the infant GI tract, which were covered in the first half of my thesis.

Future perspectives

There is no doubt the microbiota research is nothing less than "exciting, important, and growing". However, as we continue to explore the microbial world and its role in human health, we only begin to realize how complex and full of nuances the microbial ecosystems are, and how essential they are to our own survival and well-being. We also begin to understand that despite a great progress in methods currently used in microbiota research, there are still important limitations, and overcoming these limitations will be a major challenge for future research and application of the findings.

In **Chapter Two** the role of the human GI tract microbiota in health and disease was addressed. Despite many developments in this area, there is still no consensus on what defines a normal and healthy microbiota, and making any claims on causality is still extremely difficult

[69, 70]. Thus, the overwhelming majority of studies report an association between "dysbiosis" and health outcomes, but they are unable to determine if changes in microbiota were the underlining cause or the result of a disease. Determining such causalities is problematic, especially in humans, where experimentation, including the use of highly invasive procedures, is restricted due to ethical concerns. One of the common ways to address this is to use *in vitro* or animal models, but those methods are not optimal as they cannot fully replicate the complexity and the specificity of a human host, especially in a context of well-being or disease. Thus, there is a need for more mechanistic approaches, well-designed human trials and a development of methods that would allow, for example, collecting GI tract microbiota samples in non-invasive ways. One of the promising solutions might be the use of a capsule that could collect multiple samples while it passes through the GI tract. A prototype of such device was already developed in the sixties [71]. A recently developed electronic capsule is based on this idea and can accurately measure the gastrointestinal gases to measure gut microbiota function for diagnostic, therapeutic or screening purpose [72].

In **Chapters Three** through **Six** we assessed the development and dynamics of the GI tract microbiota using next generation sequencing of PCR-amplified 16S rRNA gene fragments. Despite many advantages of this approach, it also has some limitations. For example, the resulting data is compositional and provides no information regarding the actual numbers of microbes in a sample, making it impossible to evaluate the extent or directionality of changes in abundance or metabolic potential of taxa in a sample. Recently, a new method has been developed which addresses this limitation and allows for quantitative microbiome profiling of faecal material [73]. Another limitation is that the taxonomic identification of microbial groups is based on short read fragments which only span one or two hypervariable regions within the 16S rRNA gene. The resolution within the variable regions often varies between taxa or it can be conserved between closely related groups to the point at which the separation between taxa is no longer possible [27]. Thus, a choice of a single primer set may not be sufficient as it may not allow simultaneous taxonomic identification of all genera present in the sample. In addition, identification of species or strains is not reliable [74]. Species-specific primer pairs targeting a hypervariable sequence of a fast evolving gene can be used instead, or in addition to universal primers to provide information about specific species or strains [24, 75]. Other approaches have been developed recently that have the potential to provide a better resolution of the sequencing data by generating sequencing reads which span nearly the entire length of the 16S rRNA gene [76], and also do not require PCR amplification, thus avoiding a primer bias [77]. In the studies such as those presented in my thesis, these methods could provide more detailed information on community structure at higher taxonomic resolution, which is of particular importance as we begin to understand that even within the same species, microbes can vary genetically and they can perform quite different functions [78]. This in turn may have profound consequences for the host health and immune function, and at the same time it can provide the basis for developing more targeted-approaches for screening and therapeutic applications.

The objectives of our studies, besides characterising GI tract microbiota composition, were also to i) improve our insight with respect to possible associations between different microbial groups, and ii) try to reconstruct microbial interaction networks. It is a common

practice to approach this by using statistical correlations of taxa (relative) abundances across samples, but unfortunately such correlation - based inference networks do not offer explanation for mechanisms that underline the identified correlations [79]. One approach to address this is to apply whole-genome shotgun (WGS) metagenomic sequencing, followed by read annotation and mapping of entire metabolic pathways. In addition, this method also allows for taxonomic identification on bacterial species and strains [78]. Another option could be to incorporate transcriptome data to study gene expression patterns. However, both of these approaches also have some limitations. For example, in our study we obtained a metatranscriptome dataset which contained information on 190,000 genes from over 2,000 different taxonomic groups. Analysing a dataset of this magnitude in great detail is very laborious, if not impossible. In addition, a large fraction of species and gene assignments are still unknown and thus, our understanding of many metabolic and regulatory networks remains fragmentary. Thus, during the analyses of the IMMP fermentation study data described in **Chapters Six** and **Seven** we decided to focus on the three microbial groups, which we found most interesting because of their known role in degradation of similar compounds and their well-recognised probiotic function in the human GI tract. In order to take a full advantage of large datasets like this and analyse them efficiently, there is a need for developing better computational tools and more complete databases, although it should be noted that continuous advances in computational power, as well as efficiency of associated algorithms continue to help in further facilitating such analyses.

As our knowledge of the role of GI microbiota is rapidly increasing, the obvious next stage is to apply this knowledge to improve and maintain gut and systemic health of humans and animals. One of the ways is through the development of novel functional foods. Thus, the research on prebiotics is gaining popularity among public and private sectors worldwide [80]. In particular, the field of glycoscience - the science and technology of carbohydrates - is emerging globally and is of a particular interest, as it brings new technologies for synthesis and purification of prebiotics in quantities that are sufficient to enter the food and/or feed industry. The research presented in this thesis offers an example on how the areas of glycoscience and microbiota research are intimately intertwined, from the initial step of identifying natural substrates that act as prebiotics (HMOs), through testing novel substrates (IMMPs), and up to evaluating the effects of commercially available products in human populations (GOS/FOS fortified infant formulas). The area that still needs to be addressed is the long-term effects of the prebiotic supplementation on human health. There is no doubt we will be hearing more about these in the upcoming years.

References

1. M. Cox, L., *The Human Microbiome Handbook, JasonTetroEmmaAllen-Vercoe DEStech Publications (March 2016), 978-1-60595-159-1*. Anaerobe. 2016.
2. Zhernakova, A., et al., *Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity.* Science, 2016. **352**(6285): p. 565-569.
3. Falony, G., et al., *Population-level analysis of gut microbiome variation.* Science, 2016. **352**(6285): p. 560-564.
4. Smits, S.A., et al., *Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania.* Science, 2017. **357**(6353): p. 802-806.
5. Harrison, C.A. and D. Taren, *How poverty affects diet to shape the microbiota and chronic disease.* Nat Rev Immunol, 2017.
6. Hindson, J., *A possible link between multiple sclerosis and gut microbiota.* 2017. **13**: p. 705.
7. Jiang, C., et al., *The Gut Microbiota and Alzheimer's Disease.* J Alzheimers Dis, 2017. **58**(1): p. 1-15.
8. Parashar, A. and M. Udayabanu, *Gut microbiota: Implications in Parkinson's disease.* Parkinsonism Relat Disord, 2017. **38**: p. 1-7.
9. Kroemer, G. and L. Zitvogel, *Cancer immunotherapy in 2017: The breakthrough of the microbiota.* Nat Rev Immunol, 2018. **18**(2): p. 87-88.
10. Pevsner-Fischer, M., et al., *Role of the microbiome in non-gastrointestinal cancers.* World J Clin Oncol, 2016. **7**(2): p. 200-13.
11. Yu, L.-X. and R.F. Schwabe, *The gut microbiome and liver cancer: mechanisms and clinical translation.* 2017. **14**: p. 527.
12. Goedert, J.J., et al., *Investigation of the association between the fecal microbiota and breast cancer in postmenopausal women: a population-based case-control pilot study.* J Natl Cancer Inst, 2015. **107**(8).
13. Marsland, B.J., A. Trompette, and E.S. Gollwitzer, *The Gut-Lung Axis in Respiratory Disease.* Ann Am Thorac Soc, 2015. **12 Suppl 2**: p. S150-6.
14. Zhao, P., D.M. Irwin, and D. Dong, *Host genetics is associated with the gut microbial community membership rather than the structure.* Mol Biosyst, 2016. **12**(5): p. 1676-86.
15. Dabrowska, K. and W. Witkiewicz, *Correlations of Host Genetics and Gut Microbiome Composition.* Front Microbiol, 2016. **7**: p. 1357.
16. Imhann, F., et al., *Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease.* Gut, 2018. **67**(1): p. 108-119.
17. Ringel-Kulka, T., et al., *Intestinal Microbiota in Healthy U.S. Young Children and Adults—A High Throughput Microarray Analysis.* PLOS ONE, 2013. **8**(5): p. e64315.
18. Tap, J., et al., *Towards the human intestinal microbiota phylogenetic core.* Environmental Microbiology, 2009. **11**(10): p. 2574-2584.
19. Guaraldi, F. and G. Salvatori, *Effect of Breast and Formula Feeding on Gut Microbiota Shaping in Newborns.* Frontiers in Cellular and Infection Microbiology, 2012. **2**: p. 94.
20. Bakker-Zierikzee, A.M., et al., *Effects of infant formula containing a mixture of galacto- and fructo-oligosaccharides or viable Bifidobacterium animalis on the intestinal microflora during the first 4 months of life.* British Journal of Nutrition, 2005. **94**(5): p. 783-790.
21. G Moro, I.M., M Mosca, S Fanaro, J Jelinek, B Stahl, G Boehm, *Dosage-Related Bifidogenic Effects of Galacto- and Fructooligosaccharides in Formula-Fed Term Infants.* Journal of Pediatric Gastroenterology and Nutrition, 2002. **34**(3): p. 291-5.
22. Ben, X.-M., et al., *Low level of galacto-oligosaccharide in infant formula stimulates growth of intestinal Bifidobacteria and Lactobacilli.* World Journal of Gastroenterology, 2008. **14**(42): p. 6564.
23. Garrido, D., D.C. Dallas, and D.A. Mills, *Consumption of human milk glycoconjugates by infant-associated bifidobacteria: mechanisms and implications.* Microbiology, 2013. **159**(Pt 4): p. 649-64.
24. Turroni, F., et al., *Diversity of Bifidobacteria within the Infant Gut Microbiota.* PLOS ONE, 2012. **7**(5): p. e36957.

25. Oozeer, R., et al., *Intestinal microbiology in early life: specific prebiotics can have similar functionalities as human-milk oligosaccharides.* The American Journal of Clinical Nutrition, 2013. **98**(2): p. 561S-571S.
26. Junick, J. and M. Blaut, *Quantification of human fecal bifidobacterium species by use of quantitative real-time PCR analysis targeting the groEL gene.* Appl Environ Microbiol, 2012. **78**(8): p. 2613-22.
27. Ghyselinck, J., et al., *The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies.* PLoS One, 2013. **8**(8): p. e71360.
28. Silvia Fanaro, B.M., Rossana Bagna, Vittorio Vigi, Claudio Fabris, Luis Pena-Quintana, Federico Arguelles, Katharina E. Scholz-Ahrens, Gunther Sawatzki, Richard Zelenka, Jurgen Schrezenmeir, Michael de Vrese, and Enrico Bertino, *Galacto-oligosaccharides Are Bifidogenic and Safe at Weaning: A Double-blind Randomized Multicenter Study.* Journal of Pediatric Gastroenterology and Nutrition, 2008. **48**: p. 82–88
29. Hojsak, I. and A. MocicPavic, *Supplementation of prebiotics in infant formula.* Nutrition and Dietary Supplements, 2014: p. 69.
30. Watson, D., et al., *Selective carbohydrate utilization by lactobacilli and bifidobacteria.* J Appl Microbiol, 2013. **114**(4): p. 1132-46.
31. Duranti, S., et al., *Evaluation of genetic diversity among strains of the human gut commensal Bifidobacterium adolescentis.* Sci Rep, 2016. **6**: p. 23971.
32. Kim, J.H., et al., *Proteomic analysis of Bifidobacterium longum subsp. infantis reveals the metabolic insight on consumption of prebiotics and host glycans.* PLoS One, 2013. **8**(2): p. e57535.
33. Mizuno, K., et al., *Adjacent-possible ecological niche: growth of Lactobacillus species co-cultured with Escherichia coli in a synthetic minimal medium.* Sci Rep, 2017. **7**(1): p. 12880.
34. Houghteling, P.D. and W.A. Walker, *Why is initial bacterial colonization of the intestine important to the infant's and child's health?* Journal of pediatric gastroenterology and nutrition, 2015. **60**(3): p. 294-307.
35. Cockburn, D.W. and N.M. Koropatkin, *Polysaccharide Degradation by the Intestinal Microbiota and Its Influence on Human Health and Disease.* J Mol Biol, 2016. **428**(16): p. 3230-3252.
36. Ian Holmes, K.H., Christopher Quince, *Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics.* PLoS ONE, 2012. **7**(2): p. 1-15.
37. Fujimura, K.E., et al., *Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation.* Nat Med, 2016. **22**(10): p. 1187-1191.
38. Timmerman, H.M., et al., *Intestinal colonisation patterns in breastfed and formula-fed infants during the first 12 weeks of life reveal sequential microbiota signatures.* Sci Rep, 2017. **7**(1): p. 8327.
39. Matamoros, S., et al., *Development of intestinal microbiota in infants and its impact on health.* Trends Microbiol, 2013. **21**(4): p. 167-73.
40. Dogra, S., et al., *Dynamics of infant gut microbiota are influenced by delivery mode and gestational duration and are associated with subsequent adiposity.* MBio, 2015. **6**(1).
41. Liu, F., et al., *Fructooligosaccharide (FOS) and Galactooligosaccharide (GOS) Increase Bifidobacterium but Reduce Butyrate Producing Bacteria with Adverse Glycemic Metabolism in healthy young population.* Sci Rep, 2017. **7**(1): p. 11789.
42. Medicine, I.o., *Microbial Ecology in States of Health and Disease: Workshop Summary*, ed. E.R. Choffnes, L. Olsen, and A. Mack. 2014, Washington, DC: The National Academies Press. 548.
43. Flint, H.J., et al., *The role of the gut microbiota in nutrition and health.* Nature Reviews Gastroenterology &Amp; Hepatology, 2012. **9**: p. 577.
44. Flint, H.J., et al., *Microbial degradation of complex carbohydrates in the gut.* Gut Microbes, 2012. **3**(4): p. 289-306.
45. Yu, Z.T., C. Chen, and D.S. Newburg, *Utilization of major fucosylated and sialylated human milk oligosaccharides by isolated human gut microbes.* Glycobiology, 2013. **23**(11): p. 1281-92.

46. Thomson, P., D.A. Medina, and D. Garrido, *Human milk oligosaccharides and infant gut bifidobacteria: Molecular strategies for their utilization.* Food Microbiology, 2017.
47. Sela, D.A., et al., *The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome.* Proc Natl Acad Sci U S A, 2008. **105**(48): p. 18964-9.
48. Gavini, F., et al., *Differences in the Distribution of Bifidobacterial and Enterobacterial Species in Human Faecal Microflora of Three Different (Children, Adults, Elderly) Age Groups.* Microbial Ecology in Health and Disease, 2009. **13**(1): p. 40-45.
49. Pacheco, A.R., et al., *The impact of the milk glycobiome on the neonate gut microbiota.* Annu Rev Anim Biosci, 2015. **3**: p. 419-45.
50. Li, S.W., et al., *Bacterial Composition and Diversity in Breast Milk Samples from Mothers Living in Taiwan and Mainland China.* Front Microbiol, 2017. **8**: p. 965.
51. Smilowitz, J.T., et al., *Breast milk oligosaccharides: structure-function relationships in the neonate.* Annu Rev Nutr, 2014. **34**: p. 143-69.
52. Solis, G., et al., *Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut.* Anaerobe, 2010. **16**(3): p. 307-10.
53. Jost, T., et al., *Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding.* Environ Microbiol, 2014. **16**(9): p. 2891-904.
54. Maga, E.A., B.C. Weimer, and J.D. Murray, *Dissecting the role of milk components on gut microbiota composition.* Gut Microbes, 2013. **4**(2): p. 136-9.
55. Rockova, S., et al., *Growth of bifidobacteria and clostridia on human and cow milk saccharides.* Anaerobe, 2011. **17**(5): p. 223-5.
56. Hunt, K.M., et al., *Human milk oligosaccharides promote the growth of staphylococci.* Appl Environ Microbiol, 2012. **78**(14): p. 4763-70.
57. Marcobal, A., et al., *Consumption of human milk oligosaccharides by gut-related microbes.* J Agric Food Chem, 2010. **58**(9): p. 5334-40.
58. Hidalgo-Cantabrana, C., et al., *Bifidobacteria and Their Health-Promoting Effects.* Microbiol Spectr, 2017. **5**(3).
59. Ganzle, M.G. and R. Follador, *Metabolism of oligosaccharides and starch in lactobacilli: a review.* Front Microbiol, 2012. **3**: p. 340.
60. Kamasaka, H., et al., *Bacillus stearothermophilus Neopullulanase Selective Hydrolysis of Amylose to Maltose in the Presence of Amylopectin.* Applied and Environmental Microbiology, 2002. **68**(4): p. 1658-1664.
61. Hii, S.L., et al., *Pullulanase: role in starch hydrolysis and potential industrial applications.* Enzyme Res, 2012. **2012**: p. 921362.
62. Abou Hachem, M., et al., *A Snapshot into the Metabolism of Isomalto-oligosaccharides in Probiotic Bacteria.* Journal of Applied Glycoscience, 2013. **60**(2): p. 95-100.
63. Pokusaeva, K., G.F. Fitzgerald, and D. van Sinderen, *Carbohydrate metabolism in Bifidobacteria.* Genes Nutr, 2011. **6**(3): p. 285-306.
64. S Roos, F.K., L Axelsson, H Jonsson, *Lactobacillus mucosae sp. nov., a new species with in vitro mucus-binding activity isolated from pig intestine.* International Journal of Systematic and Evolutionary Microbiology, 2000. **50**: p. 251-258.
65. 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, 2010. **109**(3): p. 927-35.
66. Salyers, A.A., *Bacteroides of the human lower intestinal tract.* Annu. Rev. Microbiol., 1984. **38**: p. 293-313.
67. Long, W., et al., *Differential responses of gut microbiota to the same prebiotic formula in oligotrophic and eutrophic batch fermentation systems.* Sci Rep, 2015. **5**: p. 13469.
68. Arumugam, M., et al., *Enterotypes of the human gut microbiome.* Nature, 2011. **473**(7346): p. 174-180.
69. Fritz, J.V., et al., *From meta-omics to causality: experimental models for human microbiome research.* Microbiome, 2013. **1**(1): p. 14.
70. de Vos, W.M. and E.A.J. de Vos, *Role of the intestinal microbiome in health and disease: from correlation to causation.* Nutrition Reviews, 2012. **70**(suppl_1): p. S45-S56.

71. Simacek, I.S.a.J., *Device for obtaining samples of intestinal contents for microbiological and biochemical examinations.* Gut, 1968. **9**: p. 246-248.
72. Kalantar-Zadeh, K., et al., *A human pilot trial of ingestible electronic capsules capable of sensing different gases in the gut.* Nature Electronics, 2018. **1**(1): p. 79-87.
73. Vandeputte, D., et al., *Quantitative microbiome profiling links gut community variation to microbial load.* Nature, 2017. **551**(7681): p. 507-511.
74. Rintala, A., et al., *Gut Microbiota Analysis Results Are Highly Dependent on the 16S rRNA Gene Target Region, Whereas the Impact of DNA Extraction Is Minor.* J Biomol Tech, 2017. **28**(1): p. 19-30.
75. Hu, L., et al., *Assessment of Bifidobacterium Species Using groEL Gene on the Basis of Illumina MiSeq High-Throughput Sequencing.* Genes (Basel), 2017. **8**(11).
76. Burke, C.M. and A.E. Darling, *A method for high precision sequencing of near full-length 16S rRNA genes on an Illumina MiSeq.* PeerJ, 2016. **4**: p. e2492.
77. Karst, S.M., et al., *Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias.* Nature Biotechnology, 2018. **36**: p. 190.
78. Marx, V., *Microbiology: the road to strain-level identification.* 2016. **13**: p. 401.
79. Sung, J., et al., *Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis.* Nat Commun, 2017. **8**: p. 15393.
80. Claire Doherty, e.a., *A roadmap for Glycoscience in Europe.* Future of European Glycoscience Workshop, 2014.

Supplementary Table

Table S1. Selected V4 16S rRNA gene based OTUs from KOALA, BINGO and IMMP studies (**Chapters Three** to **Six)**, showing a list of matching taxa with the highest total scores (NCBI Blast on 20/2/2018)

***g*_Bifidobacterium_L2: KOALA OTU#:614; BINGO OTU#: 1263; IMMP OTU#129; Total BLAST Score: 262**
TACGTAGGGTGCAAGCGTTATCCGGAATTATTGGGCGTAAAGGGCTCGTAGGCGGTTCGTCGCGTCCGGTGCCTGTTCGCT CCCCACGCTTTCGCTCCTCAGCGTCAGTAACGGCCCAGAGACCTGCCTTCGCCATTGGTG
NR_043437.1 *Bifidobacterium longum* subsp. *infantis* strain ATCC 15697 16S ribosomal RNA gene, partial sequence
NR_037118.1 *Bifidobacterium gallicum* strain P6 16S ribosomal RNA gene, partial sequence
NR_040783.1 *Bifidobacterium breve* strain DSM 20213 16S ribosomal RNA gene, partial sequence
NR_041875.1 *Bifidobacterium catenulatum* strain DSM 16992 16S ribosomal RNA gene, partial sequence
NR_117506.1 *Bifidobacterium longum* strain KCTC 3128 16S ribosomal RNA gene, partial sequence
NR_112779.1 *Bifidobacterium kashiwanohense* strain HM2-2 16S ribosomal RNA gene, partial sequence
NR_037117.1 *Bifidobacterium pseudocatenulatum* strain B1279 16S ribosomal RNA gene, partial sequence
NR_113174.1 *Bifidobacterium stellenboschense* strain AFB23-3 16S ribosomal RNA gene, partial sequence
NR_145535.1 *Bifidobacterium longum* subsp. *suillum* strain Su 851 16S ribosomal RNA, partial sequence
***g*_Bifidobacterium_B1: KOALA OTU#:418; BINGO OTU#:681; Total BLAST Score: 262**
TACGTAGGGCGCAAGCGTTATCCGGATTTATTGGGCGTAAAGGGCTCGTAGGCGGCTCGTCGCGTCCGGTGCCTGTTCGCT CCCCACGCTTTCGCTCCTCAGCGTCAGTGACGGCCCAGAGACCTGCCTTCGCCATCGGTG
NR_044771.1 *Bifidobacterium bifidum* strain KCTC 3202 16S ribosomal RNA, partial sequence
NR_117505.1 *Bifidobacterium bifidum* strain KCTC 3202 16S ribosomal RNA gene, partial sequence
NR_117764.1 *Bifidobacterium bifidum* strain DSM 20456 16S ribosomal RNA gene, partial sequence
NR_113873.1 *Bifidobacterium bifidum* strain NBRC 100015 16S ribosomal RNA gene, partial sequence
***g*_Bifidobacterium_L1: KOALA OTU#:622; BINGO OTU#: 1279; IMMP OTU#130; Total BLAST Score: 262**
TACGTAGGGTGCAAGCGTTATCCGGAATTATTGGGCGTAAAGGGCTCGTAGGCGGTTCGTCGCGTCCGGTGCCTGTTCGCT CCCCACGCTTTCGCTCCTCAGCGTCAGTGACGGCCCAGAGACCTGCCTTCGCCATTGGTG
NR_036857.1 *Bifidobacterium ruminantium* strain Ru 687 16S ribosomal RNA gene, partial sequence
NR_041348.1 *Bifidobacterium tsurumiense* strain OMB115 16S ribosomal RNA gene, partial sequence
NR_116746.1 *Bifidobacterium stercoris* strain Eg1 16S ribosomal RNA gene, partial sequence
NR_118589.1 *Bifidobacterium stercoris* strain Eg1 16S ribosomal RNA gene, partial sequence
NR_114397.1 *Bifidobacterium moukalabense* strain GG01 16S ribosomal RNA gene, partial sequence
NR_133982.1 *Bifidobacterium faecale* strain CU3-7 16S ribosomal RNA, partial sequence
NR_037115.2 *Bifidobacterium dentium* strain B764 16S ribosomal RNA gene, partial sequence
NR_074802.2 *Bifidobacterium adolescentis* strain ATCC 15703 16S ribosomal RNA, complete sequence
***Bifidobacterium*_OTU_139; Total BLAST Score: 262**
TACGTAGGGTGCGAGCGTTATCCGGATTTATTGGGCGTAAAGGGCTCGTAGGCGGTTCGTCGCGTCCGGTGCCTGTTCGCT CCCCACGCTTTCGCTCCTCAGCGTCAGTGACGGCCCAGAGACCTGCCTTCGCCATTGGTG
NR_040867.1 *Bifidobacterium animalis* subsp. *lactis* strain YIT 4121 16S ribosomal RNA gene, partial sequence
***Lactobacillus*_OTU_143; Total BLAST Score: 262**
TACGTAGGTGGCAAGCGTTATCCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTTTGATAAGTCTGATGCCTGTTCGCTA CCCATGCTTTCGAGCCTCAGCGTCAGTTGCAGACCAGACAGCCGCCTTCGCCACTGGTG
NR_024994.1 *Lactobacillus mucosae* strain S32 16S ribosomal RNA, partial sequence
***Bacteroides*_OTU_13; Total BLAST Score: 262**
TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGATGGATGTTTAAGTCAGTTGCCTGTTTGATA CCCACACTTTCGAGCCTCAATGTCAGTTGCAGCTTAGCAGGCTGCCTTCGCAATCGGAG
NR_112143.1 *Bacteroides vulgatus* strain JCM 5826 16S ribosomal RNA gene, partial sequence
NR_041351.1 *Bacteroides dorei* strain 175 16S ribosomal RNA gene, partial sequence
NR_112946.1 *Bacteroides vulgatus* strain JCM 5826 16S ribosomal RNA gene, partial sequence
NR_074515.1 *Bacteroides vulgatus* strain ATCC 8482 16S ribosomal RNA, partial sequence
***Bacteroides*_OTU_21; Total BLAST Score: 262**
TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGTGGACAGTTAAGTCAGTTGCCTGTTTGATA CCCACACTTTCGAGCATCAGTGTCAGTTGCAGTCCAGTGAGCTGCCTTCGCAATCGGAG
NR_117387.1 *Bacteroides faecis* strain MAJ27 16S ribosomal RNA gene, partial sequence
NR_112944.1 *Bacteroides thetaiotaomicron* strain JCM 5827 16S ribosomal RNA gene, partial sequence
NR_113067.1 *Bacteroides faecis* strain MAJ27 16S ribosomal RNA gene, partial sequence
NR_113206.1 *Bacteroides faecichinchillae* strain JCM 17102 16S ribosomal RNA gene, partial sequence
NR_074277.1 *Bacteroides thetaiotaomicron* strain VPI-5482 16S ribosomal RNA, partial sequence
***Bacteroides*_OTU_24; Total BLAST Score: 262**

TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGTGGATTGTTAAGTCAGTTGCCTGTTTGATA CCCACACTTTCGAGCATCAGTGTCAGTAACAGTCTAGTGAGCTGCCTTCGCAATCGGAG

NR_040865.1 *Bacteroides ovatus* strain JCM 5824 16S ribosomal RNA gene, partial sequence

NR_115301.1 *Bacteroides ovatus* strain CIP 103756 16S ribosomal RNA gene, partial sequence

NR_116181.1 *Bacteroides ovatus* strain JCM5824 16S ribosomal RNA gene, partial sequence

NR_112940.1 *Bacteroides ovatus* strain JCM 5824 16S ribosomal RNA gene, partial sequence

***Bacteroides*_OTU_25; Total BLAST Score: 262**

TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGTGGATTGTTAAGTCAGTTGCCTGTTTGATA CCCACACTTTCGAGCATCAGTGTCAGTGACAGTCTAGTGAGCTGCCTTCGCAATCGGAG

NR_040865.1 *Bacteroides ovatus* strain JCM 5824 16S ribosomal RNA gene, partial sequence

NR_115301.1 *Bacteroides ovatus* strain CIP 103756 16S ribosomal RNA gene, partial sequence

NR_116181.1 *Bacteroides ovatus* strain JCM5824 16S ribosomal RNA gene, partial sequence

NR_112940.1 *Bacteroides ovatus* strain JCM 5824 16S ribosomal RNA gene, partial sequence

***Bacteroides*_OTU_27; Total BLAST Score: 262**

TACGGAGGATCCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGTGGATTGTTAAGTCAGTTGCCTGTTTGATA CCCACACTTTCGAGCATCAGTGTCAGTTGCAGTCTAGTGAGCTGCCTTCGCAATCGGAG

NR_042499.1 *Bacteroides xylanisolvens* strain XB1A 16S ribosomal RNA gene, partial sequence

NR_112947.1 *Bacteroides xylanisolvens* strain XB1A 16S ribosomal RNA gene, partial sequence

Summary

The colon is the most densely colonized area within the human gastrointestinal (GI) tract. The colonic microbiota can be considered an "invisible organ" influencing metabolism, normal immune and nervous system functions, and the overall health and well-being of the host. The first section of this thesis offers an overview of the structure and function of the GI microbiota, in particular its bacterial fraction, in health and with relation to selected diseases. Although chapter two focuses mainly on adult microbiota, accumulating evidence suggests that the early life microbiota development and perturbations might also play an important role in health outcomes throughout the entire lifespan.

The GI microbiota development starts before or at birth, and the microbial colonization progresses through a sequence of well-orchestrated events, which eventually lead to the establishment of a stable microbial ecosystem adapted for milk digestion. What guides the microbial colonization in early life is not fully known, but a number of factors, including mode of delivery, infant gender, and medications had been identified to play an important role in this process. However, one of the most fundamental modulators of the GI microbiota in both infants and adults is diet.

In breastfed infants, breastmilk is the sole source of nutrition, which evolved to nourish the growing infant and to facilitate microbiota colonization in the developing GI tract. Thus, in healthy, breastfed infants the GI microbial ecosystem is highly adapted to breastmilk consumption and dominated by *Bifidobacterium* spp., in particular *B. longum, B. longum* subsp. *infantis* (*B. infantis*), *B. bifidum,* and *B. breve.* Contrary, formula fed infants have a different microbiota composition with a reduced relative abundance of *Bifidobacterium* in their faeces.

Breastmilk is a source of microbes and bioactive components, such as free human milk oligosaccharides (HMOs). HMOs are a highly variable group of glycans and the third most abundant component of breastmilk. Breastmilk HMOs are indigestible by the infant's own digestive enzymes, but instead they can be utilized by certain groups of bacteria. Thus, these HMOs are believed to have evolved specifically to facilitate microbial colonization of the GI tract and to act as prebiotics by promoting growth and activity of bacterial species that are beneficial to a developing infant.

Knowing that the HMO composition in breastmilk varies between mothers and across lactation stages, we studied nearly two hundred infant-mother pairs to investigate the link between maternal breastmilk HMO profiles and the microbiota composition in the faeces of corresponding infants. However, we did not find any strong and consistent correlations, suggesting that there is no clear one to one relationship between specific breastmilk HMOs and microbial species that colonize the infant GI tract. Instead, it is likely that the microbial community in early life is shaped through a combined effect of many different milk HMOs, as well as other modulatory factors in breastmilk, including breastmilk's own microbiota.

We detected high variability in the faecal microbiota composition in the infants during their first 12 weeks of life, which allowed us to stratify the observed microbial patterns into

three distinct microbial cluster types. Furthermore, we showed that as infants developed, their microbiota had a tendency to progress towards *Bifidobacterium/Bacteroides* and *Bifidobacterium* rich faecal microbial clusters (cluster B and C), and that the ability of these infants to utilize HMOs also increased. Thus, even though we could not predict infant microbiota profiles based on the HMO composition in maternal milk, we detected strong correlations between infant faecal microbial composition and an infant's ability to utilize specific HMOs. Furthermore, there was a strong link between high relative abundance of specific *Bifidobacterium* phylotypes (OTUs) and utilization of the main HMOs that we measured.

We also noted that infants who received formulas fortified with prebiotics, mainly galactooligosaccharides (scGOS) and/or fructooligosaccharides (lcFOS), showed faecal profiles that closely resembled those of healthy breastfed infants, in particular with regard to the levels of *Bifidobacterium* and *Lactobacillus*. In addition, these infants also showed an accelerated transition towards the bifidobacteria dominated state, as compared to their breastfed counterparts. This effect was not noted in infants which received mixed feeding, for whom the transition was delayed. Thus, the type and extent of prebiotic supplementation (breastmilk HMOs vs. prebiotics in infant formula) had an impact on both, the microbial composition and the dynamics of microbial colonization.

The exact mode of action and effect of most prebiotic supplementations on GI microbiota community structure and function is still largely unknown. There is accumulating evidence suggesting that microbial species and strains show a high degree of specificity in their preference to utilize different prebiotic compounds. This specificity, together with the advances in glycosciences, offer leads for developing prebiotic supplementations for targeted approaches in modulating gut microbiota function for a particular health, preventative, or therapeutic purpose. The last section of this thesis presents our findings on the modulatory effect of a novel group of starch-derived prebiotics, isomalto/malto-polysaccharides (IMMPs), on adult faecal microbiota. Using an *in vitro* fermentation model, we showed that differences in the molecular composition of different IMMPs with regard to the predominant glycoside linkage types and the chain length distribution had an effect on the speed of degradation and the dynamics within the microbial community. Using metatranscriptome data we were able to identify microbial species with high levels of activity in the presence of different IMMP structures, and proposed a simple model describing the ways the three main microbial groups of interest, namely *Bifidobacterium, Bacteroides,* and *Lactobacillus,* contribute to breaking down the IMMP molecules. Furthermore, microbiota activity, as measured at the metatranscriptome level, was also reflected at the metabolic level through the accumulation of SCFAs in the fermentation media, and the corresponding decrease in pH.

The research presented in this thesis shows how natural prebiotics (HMOs) and prebiotic supplementations (IMMPs, scGOS/lcFOS) can influence human GI microbiota structure and function during infancy and adulthood. Developments in the field of glycosciences together with a better understanding of the mode of action of prebiotics with regard to microbial community structure and function could eventually lead to development of substrates offering attractive and safe ways to modulate microbiota to achieve specific health outcomes. Our

research provided insights into both, the infant and adult large intestinal ecosystems, but additional studies are needed and should address the long-term effects of the prebiotic supplementation on human health.

Appendices

Co-author affiliations

Hauke Smidt[1]
Bastian Hornung[1,5]
Edoardo Saccenti[2]
Maria Suarez-Diez[2]
Peter J. Schaap[2]
Fangjie Gu[3]
Henk Schols[3]
Bernadette Richter[3]
Pieter H. van der Zaal[4]
Piet Buwalda[4,6]
Christine Hechler[7]
Roseriet Beijers[7]
Carolina de Weerth[7]
Sander S. van Leeuwen[8]
Ilja Arts[9,10]
John Penders[9,11]
Carel Thijs[9]
Arjen Nauta[12]
Cordula Linder[12]

[1] Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

[2] Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeneng 4, 6708 WE, Wageningen, The Netherlands

[3] Laboratory of Food Chemistry, Wageningen University & Research, Bornse Weilanden 9, 6708 WG Wageningen, The Netherlands

[4] Biobased Chemistry and Technology, Wageningen University & Research, The Netherlands

[5] Leiden Universitair Medisch Centrum LUMC, Albinusdreef 2, 2333 ZA Leiden, The Netherlands

[6] Coöperatie AVEBE u.a., P.O. Box 15, 9640 AA Veendam, The Netherlands
[7] Department of Developmental Psychology, Behavioral Science Institute, Radboud University Nijmegen, Postbus 9104, 6500 HE Nijmegen, The Netherlands

[8] Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

[9] Department of Epidemiology Maastricht University, CAPHRI School for Public Health and Primary Care, Minderbroedersberg 4-6, 6211 LK Maastricht, The Netherlands

[10] CARIM School for Cardiovascular Diseases, Universiteitssingel 50 6229ER Maastricht; and Maastricht Center for Systems Biology (MaCSBio), Universiteitssingel 60, 6229 ER Maastricht, The Netherlands

[11] Department of Medical Microbiology, Maastricht University Medical Center, P. Debyelaan 25, 6229 HX Maastricht; NUTRIM School for Nutrition, Toxicology and Metabolism, Universiteitssingel 40, 6229 ER Maastricht, The Netherlands

[12] FrieslandCampina Domo, Bronland 20, 6708 WH Wageningen, The Netherlands

Acknowledgments

Looking back at my time spent in Wageningen I would like to thank many people who shared this chapter of my life with me. Those are the coworkers, the dancers I have met while training in the Netherlands, Germany and in England, my parents in Poland and my Minnesotan friends who stayed in touch over the last years. All of these people, directly or indirectly became involved in my PhD, either through working together in the lab, taking or teaching courses together, sharing the research projects, brainstorming ideas, or simply talking about the "things at work".

Getting the PhD is a long and difficult process, but I believe that having an adviser who is kind and supportive can make a huge difference and is the key to persisting and completing the program. I was lucky to have Hauke to be my supervisor, and I want to thank him for guiding my research, for his trust in me, and for always having my back when a situation required it. Thank you for being the genuine and caring person you are, and always willing to help, whether it was giving a last minute feedback for a presenation, or lending me tools to fix things in my car or around the apartment. Thank you!

The lab and the office often become a second home during our studies, so I want to thank my officemates for their friendship and company. Teresita, Ying and Susana - you were the first colleagues who I got to meet and I am thankful that you made me feel welcomed and comfortable from the day I joined your office. Thank you girls for your lasting friendship and all the fun memories of our small trips, exploring towns and castles in the Netherlands, running around Paris at midnight, or biking through tulip fields in the spring. Indra (the favorite Indonesian and the brand ambassador) and Hugo ("the man") - thank you for being the goofs that you are, and making the whole office laugh every time you discuss possible business ideas. Hugo, thank you also for all your efforts in trying to make me be more Dutch with the "Drop". Tika, thank you for your company during long evenings of writing, and Carrie - thank you for all your help and for being my paranymph. Alex, thank you for the walks, chats and of course all your help with R - you saved me from a lot of frustration when working with it. Tom, Clara, Floor, Gerben, Romy, LooWee, Arjen, Ellen, Cordula, Henk, Fangije - thank you for all the fruitful discussions about our microbiota research. Detmer, Mark – thank you for introducing me to fish gut microbiota and involving me in the Tilapia project. Benoit huge thanks for saving me when my car broke down in Germany. I also want to give special thanks to Ineke, Hans, Monika, Tom, Wim and Wilma for their support during working in the lab and their guidance when trying out new protocols or equipment, and to Prokopis - thank you for your dedication and all your hard work on the piglet project. Maria – thank you for all your hints regarding not only the data analyses, but also polishing the thesis. Edo, Leo, Cajo - thank you interesting discussions and all your help with statistics. Bastian, Alexia, Fangjie, Rima, Mark and everyone involved in the CCC – thank you for working together on the projects, I truly enjoyed being part of the consortium and getting to know all of you.

Dziekuje wam, mamo i tato za wasz optymizm i wiare w to ze "wszystko sie uda, bo dlaczego mialo by nie". Jestem wdzieczna za to ze przez te ostatnie piec lat moglismy

wspolnie spedzac swieta i wakacje, i zwiedic odrobine swiata. Rowniez podziekowania dla mamy i Eli (firma Relex) za pomoc przy projekcie okladki.

Finally, I want to thank the special group of dance friends – Phil and Nicky, Andreas, Bert, Oliver, Jasper, Clara, Andrej, Paul and Cynthia, Daniel, Remi, Iris, Martin, Johen, Farrida, Laura, WuBDA students, and many others… Thank for all for the time we spent on a dance floor training together or traveling to lessons and competitions. There are many memories we shared that I will always remember when I think about my life in the Netherlands. Andreas and Phil, thank you for being my competition partners, for your motivation and dedication to practice and to compete together, for taking the sacrifices of the long international commutes, for believing that we can be the best, but also for putting up with all my PhD commitments and time limitations. I learned a lot about dancing from both of you. Daniel, thank you for your friendship and introducing me to the Dutch dance scene, and for all the trips we did together. Nicky, thank you for all your support in England, you are "fab" woman and I am very happy that I got to meet you. Oliver, thank you for being our trainer, but also a friend, and for all the interesting chats we had over the years. Jill, Mariusz, Andreas, Ben, Adam – thank you for keeping the friendships going, even across the Atlantic – and thank you for coming to visit me in Wageningen and for travelling around Europe together. I also want to thank Anna, who has been my "Minnesota mom" – I truly value your friendship, your kindness and care over all these years that we have known each other.

I also want to thank Irene and Nam for sending the pictures of their babies to be used in the cover of this thesis, also Alex, Indra, Carrie, Hugo, Ran, and Anja for being the models for the photos. Anja, thank you for your help and advice on how the things work in the Netherlands.

To all others who are reading this thesis and the acknowledgments, but whose names are not listed – I am sorry, you name must have slipped my mind, but thank you for taking your time to read this booklet, and I hope you enjoyed it.

Klaudyna

About the author

Klaudyna Aneta Borewicz was born in Warsaw, Poland. After completing her secondary education, she received her Bachelor's degree in Natural Resources and Wildlife Management /Pre-veterinary Sciences at the University of Minnesota, Twin Cities. In 2005 she completed her Master's degree in Veterinary Entomology and Public Health, after which she joined the laboratory of Prof. R. Isaacson in the Veterinary Biosciences at the University of Minnesota to work as a scientist in the area of microbiota research. In 2013 she moved to the Netherlands and started her PhD at the laboratory of Microbiology under the supervision of Prof. H. Smidt. Currently she continues her work at the laboratory of Microbiology under the supervision of Prof. D. Sipkema.

List of publications

Ecology of the human microbiome.
Klaudyna Borewicz and Hauke Smidt; The Human Microbiome Handbook; Jason Tetro, Emma Allen-Vercoe, DEStech Publications, Inc, 15 Mar 2016, ISBN 1605951595

The effect of prebiotic fortified infant formulas on microbiota composition and dynamics in early life.
Klaudyna Borewicz, Maria Suarez-Diez, Christine Hechler, Roseriet Beijers, Carolina de Weerth, Ilja Arts, John Penders, Carel Thijs, Arjen Nauta, Cordula Lindner, Ellen van Leusen, Hauke Smidt *(manuscript in preparation)*

The association between infant faecal microbiota composition and the degradation of human milk oligosaccharides in one month old, healthy breastfed infants.
Klaudyna Borewicz*, Fangjie Gu*, Edoardo Saccenti, Ilja Arts, John Penders, Carel Thijs, Sander S. van Leeuwen, Arjen Nauta, Cordula Lindner, Ellen van Leusen, Henk A. Schols, Hauke Smidt *(manuscript in preparation)*

The association between infant faecal microbiota composition and the degradation of human milk oligosaccharides in healthy breastfed infants at two, six and twelve weeks of age.
Klaudyna Borewicz*, Fangjie Gu*, Edoardo Saccenti, Christine Hechler, Roseriet Beijers, Carolina de Weerth, Sander S. van Leeuwen, Henk A. Schols, Hauke Smidt *(manuscript in preparation)*

In vitro fermentation behaviour of isomalto/malto-polysaccharides using human faecal inoculum indicates prebiotic potential.
Fangjie Gu*, **Klaudyna Borewicz***, Bernadette Richter, Pieter H. van der Zaal, Hauke Smidt, Pieter L. Buwalda, Henk A. Schols. Molecular Nutrition & Food Research, 2018

Isomalto/malto-polysaccharides maintain normal gut functioning while promoting growth and activity of beneficial bacteria**.**
Klaudyna Borewicz*, Bastian Hornung*, Fangjie Gu, Pieter H. van der Zaal, Henk Schols, Peter J. Schaap, Hauke Smidt *(manuscript submitted)*

The combination long-chain inulin with *Lactobacillus acidophilus* W37 confers microbiota associated protective effects through weaning and against *Salmonella Typhimurium* in neonate piglets.
Alexia F.P. Lépine*, Prokopis Konstanti*, **Klaudyna Borewicz***, Jurriaan Mes, Paul de Vos, Hauke Smidt *(manuscript in preparation)*

Association between psychosocial stress and faecal microbiota in pregnant women.
Christine Hechler, **Klaudyna Borewicz**, Roseriet Beijers, Edoardo Saccenti, Marianne J. Riksen-Walraven, Hauke Smidt, Carolina de Weerth *(manuscript in preparation)*.

Effects of pectin supplementation on the fermentation patterns of different structural carbohydrates in rats.
Lingmin Tian, Jan Scholte, **Klaudyna Borewicz**, Hauke Smidt, Anton J. Scheurink, Harry Gruppen, Henk Schols. Molecular Nutrition & Food Research. 2016 Jun; 60(10) 2256-2266.

Effects of pectin on fermentation characteristics, carbohydrate utilization and microbial community composition in the gastrointestinal tract of weaning pigs.

Lingmin Tian, Geert Bruggeman, Marco van den Berg, **Klaudyna Borewicz**, Anton J. Scheurink, Eric Bruininx, Paul de Vos, Hauke Smidt, Henk Schols, Harry Gruppen. Molecular Nutrition & Food Research. 2017 Jan;61(1).

Whole grains increase betainized compounds and affect energy metabolism.
Olli Kärkkäinen, Jenna Pekkinen, Tomi Tuomainen, **Klaudyna Borewicz**, Ville Koistinen, Jukka Leppänen, Marilena Vitale, Maria Lankinen, Jussi Paananen, Tuomo Laitinen, Jaana Rysä, Marko Lehtonen, Anna-Marja Aura, Natalia Rosa-Sibakov, Rosalba Giacco, Emilia Nordlund, Gabriele Riccardi, Valerie Micard, Marjukka Kolehmainen, Kaisa Poutanen, Ursula Schwab, Angela Rivellese, Seppo Auriola, Antti Poso, Hannu Mykkänen, Hauke Smidt, Matti Uusitupa, Pasi Tavi, Kati Hanhineva *(manuscript in preparation)*

Modification of the equine gastrointestinal microbiota after feeding of Jerusalem artichoke meal.
Maren Glatter, **Klaudyna Borewicz**, Bart van den Bogert, Monika Wensch-Dorendorf, Mandy Bochnia, Martin Bachmann, Hauke Smidt, Gerhard Breves, Annette Zeyner *(manuscript in preparation)*

Changes in the porcine intestinal microbiome in response to infection with *Salmonella enterica* and *Lawsonia intracellularis*.
Klaudyna A. Borewicz, Hyeun Bum Kim, Randall S. Singer, Connie J. Gebhart, Srinand Sreevatsan, Timothy Johnson, Richard E. Isaacson. PLoS One. 2015 Oct 13;10(10): e0139106.

Effects of tylosin administration on C-reactive protein concentration and carriage of *Salmonella enterica* in pigs.
Hyeun Bum Kim, Randall S. Singer, **Klaudyna Borewicz**, Bryan A. White, Srinand Sreevatsan, Timothy J. Johnson, Luis A. Espejo, Richard E. Isaacson. American Journal of Veterinary Research, May 2014, Vol. 75, No. 5, Pages 460-467

A combined enrichment and aptamer pulldown assay for *Francisella tularensis* detection in food and environmental matrices
Elise A. Lamont, Ping Wang, Shinichiro Enomoto, **Klaudyna Borewicz**, Ahmed Abdallah, Richard E. Isaacson, Srinand Sreevatsan. PLoS One. 2014 Dec;9(12)

Longitudinal analysis of the lung microbiome in lung transplantation.
Klaudyna Borewicz, Alexa Pragman, Hyeun Bum Kim, Marshal Hertz, Christine Wendt, Richard E. Isaacson. FEMS Microbiol Lett. 2013: 339(1), 57-65

Microbial shifts in the swine distal gut in response to the treatment with antimicrobial growth promoter, tylosin.
Hyeun Bum Kim, **Klaudyna Borewicz**, Bryan A. White, Randall S. Singer, Srinand Sreevatsan, Zheng Jin Tu, Richard E. Isaacson. PNAS 2012:109 (38): 15485-15490

Characterization and differential gene expression between two phenotypic phase variants in *Salmonella enterica* serovar *Typhimurium*.
Sheila K. Patterson, **Klaudyna Borewicz**, Timothy Johnson, Wayne Xu, and Richard E. Isaacson. PLoS One. 2012; 7(8): e43592.

Longitudinal investigation of the age-related bacterial diversity in the feces of commercial pigs.
Hyeun Bum Kim, **Klaudyna Borewicz**, Bryan A. White, Randall S. Singer, Srinand Sreevatsan, Zheng Jin Tu, Richard E. Isaacson. Veterinary Microbiology 2011:153(1-2):123-133

Overview of completed training activities

Discipline specific courses	Graduate school/institute	Year
Genetics and physiology of food-associated micro-organisms	VLAG	2013
The intestinal Microbiome and Diet in Human and Animal Health	VLAG	2014
Biological Network Analysis	BioSB	2015
Essential Skills in Data Intensive Research	eScience Centre	2016
Discipline specific meetings		
Gut day	Gut Flora Foundation	2014
The Gut Microbiota Throughout Life	ENGIHR Conference	2014
Gut Microbiota for Health Summit	GMFH Conference	2015
General ASM meeting	ASM Conference	2015
Gut Microbiota for Health Summit	GMFH Conference	2017
Microbiology Centennial Symposium	WUR	2017
CCC open day symposium	CCC	2016
CCC closing symposium	CCC	2017
General courses		
VLAG PhD week	WUR	2014
Career assessment	WUR	2017
Information Literacy including EndNote Introduction	WUR	2013
Project and Time Management	WUR	2014
Communication with the Media and the General Public	WUR	2014
Scientific Writing	WUR	2016
Introduction to the Netherlands	WUR	2013
Optionals		
PhD/postdoc meetings	WU-Microbiology	2013-2017
2 week PhD trip to USA	WU-Microbiology	2015
Laboratory of Microbiology group meetings	WU-Microbiology	2013-2017
Strategy Days	WU-Microbiology	2013-2017
Quarterly CCC project meetings	CCC	2013-2017
CCC sub-project meetings	CCC	2013-2017

The research presented here was performed in the public-private partnership CarboHealth coordinated by the Carbohydrate Competence Center (CCC, www.cccresearch.nl) and financed by participating partners and allowances of the TKI Agri&Food program, Ministry of Economic Affairs of the Netherlands.

Colophon
Cover design: Ela Zawadzka, Aneta Borewicz, Klaudyna Borewicz
Layout design: Klaudyna Borewicz
Printed by: GVO drukkers & vormgevers B.V, Ede, The Netherlands