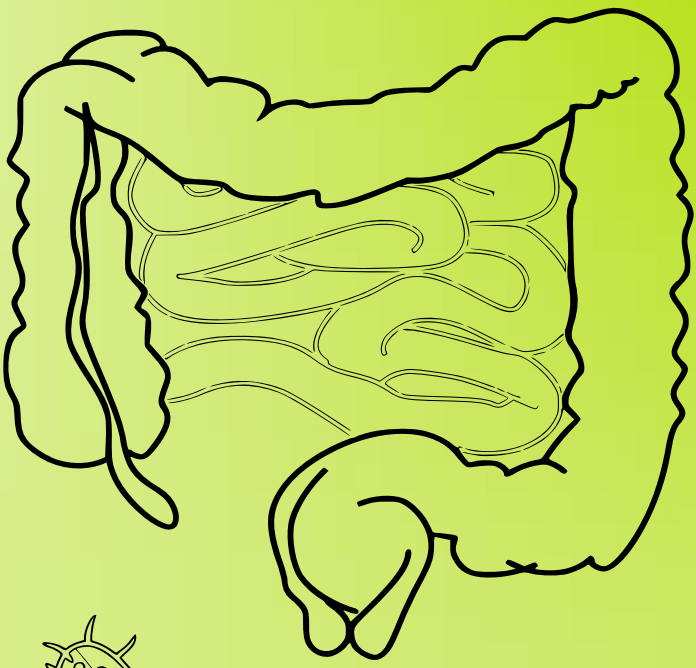
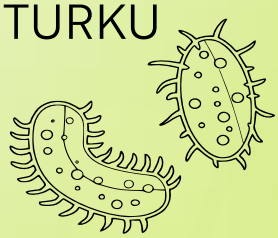
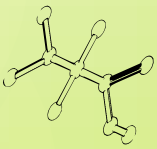
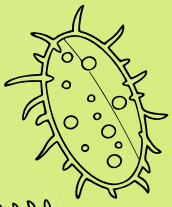
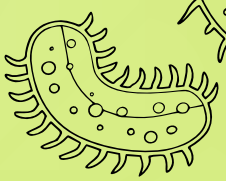
TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

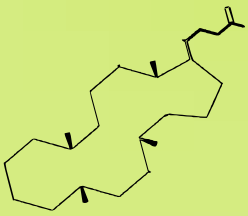


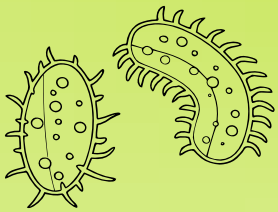


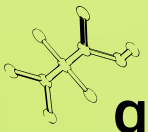












Utilization of next generation sequencing and metabolomics in human microbiome studies

Optimized laboratory methods and exploratory findings on gut microbiome and metabolome development in early life

Heidi Isokääntä

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1900 | MEDICA – ODONTOLOGICA | TURKU 2025

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

UTILIZATION OF NEXT GENERATION SEQUENCING AND METABOLOMICS IN HUMAN MICROBIOME STUDIES

Optimized laboratory methods and exploratory findings on gut microbiome and metabolome development in early life

Heidi Isokääntä

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1900 | MEDICA – ODONTOLOGICA | TURKU 2025

University of Turku

Faculty of Medicine
Institute of Biomedicine
Medical Microbiology
Doctoral Programme in Clinical Research

Supervised by

Docent Alex Dickens
Turku Bioscience
University of Turku
Turku, Finland

PhD Teemu Kallonen
Microbiome Biobank and
Clinical microbiology
TYKS, VARHA
Turku, Finland

MD, PhD Anna-Katariina Aatsinki
Centre for Population Health research
Faculty of medicine
University of Turku and
Turku university hospital
Turku, Finland

Reviewed by

Docent Marko Lehtonen, PhD
School of Pharmacy
Faculty of Health Sciences
University of Eastern Finland
Kuopio, Finland

PhD Mirjam Bloemendaal
Medical Sciences Curie institute and
Goethe University Hospital
Frankfurt, Germany

Opponent

Docent Mikael Niku, PhD
Faculty of Veterinary Medicine,
Biosciences, Anatomy and
Developmental Biology
University of Helsinki
Helsinki, Finland

Cover Image: Heidi Isokääntä

ISBN 978-952-02-0289-7 (PRINT)
ISBN 978-952-02-0290-3 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Painosalama, Turku, Finland 2025

To my family

UNIVERSITY OF TURKU
Faculty of Medicine
Institute of Biomedicine
Medical Microbiology
HEIDI ISOKÄÄNTÄ: Utilization of next generation sequencing and metabolomics in human microbiome studies: Optimized laboratory methods and exploratory findings on gut microbiome and metabolome development in early life
Doctoral Dissertation, 220 pp.
Doctoral Programme in Clinical Research
September 2025

ABSTRACT

The proportion of the gut microbiota dominates the total human microbiota, and its diversity and dynamic properties pose certain methodological challenges. With the development of methods, significant strides have been made in the field, such as population-level sample collections and high-performance analysis methods. On the other hand, although many health responses have been found, there is little information about the development of the microbiota and its metabolites. Timely developmental stages lay the foundation for a balanced gut microbiota that produces vital metabolic products to ensure growth and development. Early childhood microbial exposures prepare the individual's tolerance to various exposures, and the immune defence is programmed to function according to prevailing conditions. In early childhood, breast milk plays a significant role in the development of the microbiota.

In this dissertation, optimized protocols for DNA isolation in a microplate format and sample collection for faecal microbiota and metabolome studies were developed, and developmental patterns in early childhood gut microbiota and metabolome were mapped together with breast milk metabolites. The optimized sample collection and DNA isolation achieved seamless sample pre-processing, sufficient reproducibility, breakdown of hard-to-lyse cell membranes, and minimization of contamination. The connections between breastfeeding, human milk composition and the child's gut microbiota and metabolome confirmed previously reported findings.

KEYWORDS: Gut microbiota, metabolomics, 16s rRNA sequencing, method development, sample collection, sample preservative, high-throughput DNA extraction, early life, maturation of microbiome, fecal metabolites, human milk metabolites

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Biolääketieteen laitos
Lääketieteellinen mikrobiologia
HEIDI ISOKÄÄNTÄ: Uuden sukupolven omiikka-menetelmien hyödyntäminen kliinisessä mikrobistotutkimuksessa: optimoidut laboratoriomenetelmät ja kartoittavat löydökset suolistomikrobiston ja -metabolomin kehityksestä varhaislapsuudessa
Väitöskirja, 220 s.
Turun kliininen tohtoriohjelma
Syyskuu 2025

TIIVISTELMÄ

Suolistomikrobiston osuus dominoi ihmisen kokonaismikrobimäärää ja sen monimuotoisuus ja dynaamiset ominaisuudet asettavat tiettyjä menetelmällisiä haasteita. Menetelmäkehityksen myötä alalla on otettu merkittäviä harppauksia kuten väestötason näytekeräykset ja korkean suorituskyvyn analysointimenetelmät. Toisaalta vaikka löydettyjä terveysvasteita on paljon, mikrobiston ja sen metaboliittien kehityksestä on vain vähän tietoa. Oikea-aikaiset kehitysvaiheet luovat pohjan tasapainoiselle suolistomikrobistolle, joka tuottaa elintärkeitä aineenvaihdunnan tuotteita kasvun ja kehityksen takaamiseksi. Varhaislapsuuden mikrobialtistukset valmistelevat yksilön toleranssia erilaisille altisteille ja immuunipuolustus ohjelmoituu toimimaan vallitsevien olosuhteiden mukaan. Varhaislapsuudessa äidinmaito on merkittävässä roolissa mikrobiston kehittymisessä.

Tässä väitöskirjassa kehitettiin optimoidut protokollat DNA-eristykseen kuoppalevyformaatissa ja näytteenkeräykseen ulosteen mikrobisto- ja metabolomitutkimukseen sekä kartoitettiin varhaislapsuuden kehityksellisiä reittejä suolistomikrobistossa ja -metabolomissa yhdessä äidinmaidon metaboliittien kanssa.

Optimoidulla näytesäilytyksellä ja DNA-eristyksellä saavutettiin sujuva näytteenesikäsittely, riittävä toistettavuus, sitkeidenkin solukalvojen hajotus ja kontaminaatioiden minimointi. Imetyksen yhteydet lapsen suolistomikrobiston kehittymiseen vahvistivat jo aiemmin raportoituja löydöksiä.

AVAINSANAT: Suolistomikrobisto, metabolomiikka, uuden sukupolven sekvensointi, menetelmäkehitys, ulostenäytteen keräys, näytteen säilytys, korkean suorituskyvyn DNA-eristys, suolistomikrobiston kypsyminen, varhaislapsuus, ulostemetaboliitit, äidinmaidon metaboliitit

Table of Contents

Abbreviations **8**

List of Original Publications **10**

1 Introduction **11**

2 Review of the Literature **13**
2.1 Microbiome research in health and disease 15
2.2 Methodological challenges 16
2.3 Methodology for next generation sequencing 22
2.3.1 Sample storage and DNA extraction 22
2.3.2 16S rRNA based NGS libraries and sequencing 23
2.4 Methodology for measuring faecal and milk metabolites 25
2.4.1 GC-MS 25
2.4.2 LC-MS 26
2.4.3 NMR 27
2.5 Host-microbe interaction via microbial metabolites 27
2.5.1 Human milk and gut microbiome in early life 29
2.5.2 Metabolites in early life 31
2.5.2.1 SCFA metabolism in gut 34
2.5.2.2 Bile acid metabolism and gut microbiome 35
2.6 Summary of literature 37

3 Aims of the Study **39**

4 Materials and Methods **40**
4.1 Methodological experiments 40
4.1.1 Study design for optimizing DNA extraction 40
4.1.2 Study design for stability measurements 41
4.2 Study design for applied cohort studies (III and IV) 42
4.3 Microbiome methods 44
4.4 Metabolomics 47
4.5 Data modulation and statistical methods 50
4.6 Ethical considerations 55

5 Results **56**
5.1 Optimized high-throughput DNA extraction method for covering different bacteria types and avoiding contaminants 56

5.1.1 Differential abundances and higher alpha diversity by bead beating ... 56
5.2 Sample preservative for stability of gut metabolites and microbiota ... 58
5.3 Faecal metabolites change during early development ... 61
5.3.1 Infant microbiota shows more diverse microbiota community types compared with toddlers ... 64
5.3.2 Microbiota alpha diversity and genera abundances are associated with faecal metabolites ... 66
5.3.3 Microbial community types associate with different levels of metabolites ... 69
5.3.4 Interactions Between Breastfeeding, Gut Microbiota, and Metabolites ... 70
5.4 Human milk directing the colonization process of gut microbiome ... 72

6 Discussion ... 82
6.1 Methodological choices (studies I and II) associate with bacterial genera and microbial metabolites ... 82
6.2 Successional patterns in developing gut microbiota and metabolome (studies III and IV) ... 84
6.2.1 Human milk metabolites influencing infant gut microbiome and metabolome ... 86
6.3 Strengths and Limitations ... 89
6.4 Future perspectives ... 90

7 Conclusions ... 91

Acknowledgements ... 92

References ... 94

Original Publications ... 115

Abbreviations

ASV	Amplicon sequence variant
BA	Bile acid
BCFA	Branched-chain fatty acid
BF	Breastfeeding
bp	Base pair
CA-d4	Cholic acid -d4
CDCA-d4	Chenodeoxycholic acid -d4, ,
CLR	Centered log ratio
DCA	Deoxycholic acid
DCA-d4	Deoxycholic acid -d4,
DF	Dietary fibre
DNA	Deoxyribonucleic acid
ECC	Endocannabinoid
FDR	False discovery rate
GC	Gas chromatography
GCA-d4	Glycocholic acid-d4
GCDCA-d4	Glycochenodeoxycholic acid -d4
GLCA-d4	Glycolithocholic acid -d4
GM	Gut microbiota
GUDCA-d4	Glycoursodeoxycholic acid -d4
HM	Human milk
HMO	Human milk oligosaccharide
LC	Liquid chromatography
LCA	Lithocholic acid
LCA-d4	Litocholic acid -d4
Microbiome	The combined genetic material of all microorganisms in a given host
Microbiota	The sum total of all organisms presents in the human body
MS	Mass spectrometry
NCD	Non-communicable chronic disease
NGS	Next generation sequencing
NMR	Nuclear magnetic resonance

OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PERMANOVA	Permutational Analysis of Variation
ppb	Parts per billion
ppm	Parts per million
RCLR	Robust centered log ratio, similar to regular clr but allows data with zeroes and avoids the need to add pseudo count
rRNA	Ribosomal ribonucleic acid
RT	Room temperature
SCFA	Short chain fatty acid
SNP	Single nucleotide polymorphism, a single base substitution, limited to germline DNA
TCA-d4	Taurocholic acid -d4
UDCA-d4	Ursodeoxycholic acid -d4

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Isokääntä, H., Tomnikov, N., Vanhatalo, S., Munukka, E., Huovinen, P., Hakanen, A.J. and Kallonen, T. High throughput DNA extraction strategy for faecal microbiome studies. *Microbiology spectrum*, 2024; 12, 6, eLocator: e02932-23.

II Heidi Isokääntä, Lucas Pinto da Silva, Naama Karu, Teemu Kallonen, Anna-Katariina Aatsinki, Thomas Hankemeier, Leyla Schimmel, Edgar Diaz, Tuulia Hyötyläinen, Pieter C. Dorrestein, Rob Knight, Matej Orešič, Rima Kaddurah-Daouk, Alex M. Dickens, Santosh Lamichhane. Comparative metabolomics and microbiome analysis of Ethanol vs. OMNImet/gene®•GUT faecal stabilization. *Analytical Chemistry*, 2024; 96, 22, 8893–8904.

III Anna-Katariina Aatsinki#, Santosh Lamichhane#, Heidi Isokääntä#, Partho Sen, Matilda Kråkström, Marina Amaral Alves, Anniina Keskitalo, Eveliina Munukka, Hasse Karlsson, Laura Perasto, Minna Lukkarinen, Matej Orešič, Henna-Maria Kailanto, Linnea Karlsson, Leo Lahti, Alex M Dickens. Dynamics of Gut Metabolome and Microbiome Maturation during Early Life. Accepted in *iScience*. #Shared first authorship

IV Heidi Isokääntä, Laura Perasto, Santosh Lamichhane, Minka Ovaska, Teemu Kallonen, Eveliina Munukka, Hasse Karlsson, Linnea Karlsson, Henna-Maria Kailanto, Ulrik Sundekilde, Matej Orešič, Alex M. Dickens and Anna-Katariina Aatsinki. Human milk metabolites association with infant gut microbiome and metabolome development. Manuscript.

1 Introduction

The amount of microbiome research has been on the rise for the past decade, and the results have demonstrated that it is a key area of study for understanding the health effects of microbes on humans. As research knowledge increases and healthcare advances, the impacts on human health from infectious diseases have been replaced by various chronic gastrointestinal disorders, autoimmune diseases and atopic conditions. One possible reason behind this shift in disease burden is the reduction in microbial diversity and an imbalance in the microbiome, known as dysbiosis which can originate from disturbances in the primary colonization of the infant.(Bull & Plummer, 2014) Disturbances of microbiome may have more deleterious effects on infants, when the immune system is still immature (Tamburini et al., 2016). It is evident that microbiota in our environment decreases over time due to evolution, and consequent changes in human microbiota might contribute to the evolving burden of disease (Blaser & Falkow, 2009).

Human microbiota consists of bacteria, viruses, fungi, and archaea. Since bacteria are considered the most abundant in terms of biomass in the microbial community, microbiota mainly refers to bacteria, although viruses, including bacteriophages, might outnumber the bacteria in terms of numbers due to their smaller size (Gilbert et al., 2018). It has been found that the genetic material of bacterial cells in the gut surpasses human genetic material a hundred-fold (Gilbert et al., 2018; Sender et al., 2016). The composition of this diverse community varies based on the body site, surrounding environment, and numerous other internal and external factors (Falony et al., 2016; Zhernakova et al., 2016). The gut is the most extensively studied body site, with the predominant members of gut microbiota being classified under the *Bacteroidetes and Firmicutes* phyla. However, the *Actinobacteria, Proteobacteria, Fusobacteria, and Verrucomicrobia* phyla are also found to be prevalent, especially in well-studied Western populations (Huttenhower et al., 2012). Additionally, a significant portion of unknown species is present in the gut, particularly in non-Western populations (Pasolli et al., 2019). It has been determined that most of the variation in gut microbiota originates from environmental sources, with genetic background playing a limited role in shaping its composition (Rothschild et al., 2018).

The human microbiome can vary greatly between individuals, which is why it is said to be like a fingerprint. This phenomenon presents challenges for microbiome research because it is difficult to define what constitutes a healthy microbiome (Joos et al., 2025). Moreover, changes in the microbiome can be small, but their effects on human well-being can be significant. Due to this and the large individual variation, there is a need for extensive data collections (Siezen & Kleerebezem, 2011; Tomasello et al., 2017). Classifying bacteria as either good or bad is also problematic because the usefulness of a bacterium can be linked to the conditions in the gut, such as the amount of fibre (Desai et al., 2016). The functional aspect of microbes is partially reflected in the metabolites secreted into the gut. While studying the microbial ecosystem, it is also important to investigate gut metabolome to understand the metabolic significance of bacteria and host-microbe interaction in the human body (Lamichhane et al., 2018).

The early stages of life provide the foundation for the development of the microbiome. Harmful exposures in early childhood may be associated with later morbidity through the development of an unfavourable microbiome. The first 1000 days of life are considered as a critical period for the development and establishment of the gut microbiome, which may have long-term implications for host metabolism and physiology. Early life nutrition has been highlighted according to factors shaping infant microbiome. Human milk is known to feed the early life gut bacteria, e.g. *Bifidobacterium* by human milk oligosaccharides (HMOs) and it is tailor-made for the infant, meaning that milk composition can vary by the developmental stage and infant needs.(Ames et al., 2023) Despite that early-life factors have been shown to affect infant gut microbiota and faecal metabolites, our current understanding of human gut colonization in early life remains limited (Beller et al., 2021).

This thesis assesses the role of preanalytical and biological factors in shaping human gut microbiome and faecal metabolome assembly. Specifically, the effect of sample preservative, storage time, storage temperature and optimized DNA-extraction, coupled with bead-beating and a plate format for high throughput analysis, were examined. In the early life studies, associations between infant gut microbiome, faecal metabolome and human milk metabolites were investigated, along with demographic factors, such as delivery mode, use of antibiotics, siblings and duration of breastfeeding.

2 Review of the Literature

The field of microbiome research has its roots in microbiology, dating back to the seventeenth century when first observed microbes were seen through microscope (Kutschera, 2023). Advances in this field have often been propelled by the development of innovative techniques and equipment. Notably, many technological breakthroughs have significantly advanced microbiological research, leading to paradigm shifts in our understanding of health and disease. Given the historical impact of infectious diseases on human populations, medical microbiology emerged as the initial focus of both research and public attention. (Berg et al., 2020)

In Western literature, the initial discovery of gut microbiota happened in the 19th century, while the technological advancements that promoted the field occurred in the 20th century. The discovery of DNA, the development of sequencing technologies, PCR, and cloning methods equipped the investigation of microbial communities utilizing DNA- and RNA-based approaches which replaced bacterial cultivation to some extent (Brul et al., 2008; Mathews, 2024). At the turn of the 19th and 20th century, the study, characterization, and even therapeutic use of "protective" microbes reached its scientific peak in central Europe (Anukam & Reid, 2007; Farré-Maduell & Casals-Pascual, 2019). In scientific literature, the terms microbiota and microbiome can be traced back at least to 1927 when "The unseen life of the soil" was studied (Baudoin et al., 2019; Prescott, 2017). And already in 1958, the first faecal transplants (enemas) were successfully used to treat four patients with Clostridium difficile infection (Eiseman et al., 1958).

Despite the early effort of improve human health by modulating the microbiota, this area of research was deferred for many decades. However, in 1988, Whipps and colleagues defined the term "microbiome," as a "characteristic microbial community" inhabiting a "well-defined habitat" with specific physio-chemical properties, forming distinct ecological niches. The definition was a stepping stone towards emphasizing unique properties, functions, and environmental interactions of microbiome. Over the years, various definitions of the microbiome have emerged. The most widely cited, proposed by Lederberg (2001), describes the microbiome as a community of commensal, symbiotic, and pathogenic microorganisms within a body space or environment, highlighting its ecological context (Berg et al., 2020). Recently, the

definition has been refined; microbiome refers to the combined genetic material of all microorganisms in a given host, while microbiota refers to the sum total of all organisms present in the human body (Khan, 2021). Often, microbiome is referred to as both microbes and their metabolites. (Levy et al., 2017; Oliphant & Allen-Vercoe, 2019).

For a long time in 20th century, the bacteria were only seen as pathogens. However, the view was challenged during late 20th century and early 21st century, when scientists noted the increasing rates of noncommunicable diseases (NCD), such as obesity (Ng et al., 2014; Ogden et al., 2014; Vuorela et al., 2011), allergic asthma (Backman et al., 2017) and inflammatory bowel disease (Alatab et al., 2020). The increased burden of NCD has been coupled with urbanization and industrialized populations (Ezzati et al., 2005; Goryakin et al., 2017), contributing to decreased human microbiome diversity (Mancabelli et al., 2017; Sun et al., 2022) and also general decline in biodiversity (Haahtela et al., 2024). The less diverse microbiome has been associated with increased use of antibiotics (McDonnell et al., 2021), Western diets poor in fibre (G. D. Wu et al., 2011), improved sanitation and limited exposure to natural biodiversity (Bloomfield et al., 2016; Flies et al., 2020; Karkman et al., 2017) in the human body or their immediate surroundings. Human microbiota is mainly consisted of bacteria. It has estimated that human body has more bacterial cells (90%) than human cells (10%) and the load of genes is even larger. Microbes in human body are mainly located in gut; 10^{10}-10^{12} microbial cells per gram of feces, however the estimates of total microbial mass varies from 200 grams to two kilograms (Walker & Hoyles, 2023). These numbers are not confirmed and they can vary from person to person depending on factors such as body size and age. (Walker & Hoyles, 2023) The gut is defined as a part of gastrointestinal (GI) tract, which in humans is approximately 9 meters long including the mouth, oesophagus, stomach, small intestine, large intestine, and anus. Most of the absorption of nutrients from food takes place in small intestine formed by the duodenum, jejunum, and ileum. The large intestine refers to the cecum, appendix, colon and rectum. Most bacteria are located in colon where the pH value is more preferable for most microbes (O'May et al., 2005).

Bacteria are studied through taxonomical levels: Phylum, Class, Order, Family, Genus and Species. Recent taxonomic updates concern the phylum naming. The phylum *Bacillota*, formerly known as *Firmicutes*, is a group of mostly Gram-positive bacteria (Chukwudulue et al., 2023). This phylum has been the subject of recent taxonomic debates, with proposals to retain the name *Firmicutes* (Pallen, 2023). Frequently, genus-level classification is preferred over phylum-level. While phylum-level can reveal broad differences in microbiota, genus-level or finer resolution often provides more informative results. However, the appropriate taxonomic resolution depends on the specific study question. (Z. Xu et al., 2014).

Enterotypes are classifications of gut microbiota based on dominant bacterial genera. Research has identified two to three main enterotypes in Western adult

samples, typically dominated by *Bacteroides, Prevotella, or Ruminococcaceae* (Costea et al., 2018). These enterotypes show stability over time and are influenced by factors such as genetics, diet, and geographical location (Yatsunenko et al., 2012). Although, three robust enterotypes of the human gut microbiome have been observed that are not nation or continent specific (Arumugam et al., 2011). The *Prevotella*-dominated enterotype has been associated with high-fiber diets and certain vitamins and minerals (David et al., 2014). Enterotypes may also be linked to health outcomes, with different types showing varying associations with cardiovascular disease risk factors (Costea et al., 2018; de Moraes et al., 2017). Genetic differences between enterotypes have been observed, particularly in genes related to amino acid and carbohydrate metabolism. Additionally, host genetic variations, such as those linked to body fat distribution, may influence the abundance of specific bacterial genera like *Prevotella*. (J. Li et al., 2018) Recently, a novel approach of enterosignatures, has been created to characterize gut microbiome composition, representing common bacterial guilds that co-occur in the human gut (Frioux et al., 2023). This concept builds upon the earlier enterotype model (Arumugam et al., 2011) but offers a more nuanced view of microbiome variability. Frioux et al. (2023) identified five generalizable enterosignatures dominated by *Bacteroides, Firmicutes, Prevotella, Bifidobacterium, or Escherichia*. The *Bacteroides*-associated enterosignature appears to be core in westernized gut microbiomes, while combinations with other enterosignatures complement the functional spectrum. This approach enables detection of gradual shifts in community structures and atypical gut microbiomes associated with adverse health conditions. (Arumugam et al., 2011; Frioux et al., 2023).

2.1 Microbiome research in health and disease

In recent years, microbiome research has expanded to population-based studies (Abdill et al., 2025; Falony et al., 2016). Speculation on causality has moved to more mechanistic approach. There are already studied links between microbiome and human disease. Emerging evidence highlights the critical role of the gut microbiota (GM) in regulating metabolic, immune, and endocrine functions, as well as in priming the immune system's response to pathogens. This interplay is bidirectional, with the immune system also influencing microbiota composition (Zheng et al., 2020). Alterations in the GM, such as changes in overall abundance or shifts in species and family ratios, have been linked to numerous health conditions (de Vos et al., 2022; Sommer & Bäckhed, 2013), particularly non-communicable chronic diseases (NCDs). The most evidence is found for conditions such as diabetes mellitus, chronic kidney disease, hypertension, inflammatory bowel disease (IBD), multiple sclerosis, allergic asthma, and rheumatoid arthritis (Kim et al., 2017). Gut dysbiosis, an imbalance in the microbial communities within the gastrointestinal tract, contributes to the pathogenesis

of these chronic inflammatory diseases by disrupting the interplay between gut microbiota and host immune cells (Kim et al., 2017). Additionally, gut microbiota imbalance has been associated with obesity and metabolic dysfunction-associated steatotic liver disease (MASLD) (Abdelhameed et al., 2025; Gangarapu et al., 2014), cardiovascular diseases (Tang et al., 2019; Tang & Hazen, 2014), neurodegenerative disorders (Ghezzi et al., 2022; Roy Sarkar & Banerjee, 2019) and several cancers, such as gastrointestinal cancers (Brennan & Garrett, 2016; Tözün & Vardareli, 2016). Chronic inflammation and toxin production from dysbiotic microbiota can promote carcinogenesis (García-Castillo et al., 2016). The microbiome may influence cancer progression and treatment efficacy. Moreover, gut microbiota can exert both stimulation or reduction of tumorgenicity and growth over time . (Jaye et al., 2021; Tözün & Vardareli, 2016; Vivarelli et al., 2019)

Additionally, most of these conditions are accompanied with low-grade chronic inflammation which can elevate oxidative stress (Biswas, 2016; Menzel et al., 2021). Therapeutic approaches targeting the gut microbiota, such as probiotics, prebiotics, synbiotics, and faecal microbiota transplantation (FMT), have shown promise in managing NCDs, although more extensive clinical trials are needed to fully evaluate their efficacy (López-Tenorio et al., 2024; Noce et al., 2019).

Recently, more focus has gone into how the microbes can be both beneficial and protective in promoting health in humans. In addition to commonly used probiotics, novel protective functions of certain bacteria have been found, i.e. *Faecalibacterium prauznitzii* has been demonstrated to have therapeutic value through metabolic and anti-inflammatory properties (Martín et al., 2023). The metabolic aspect of microbiota has been increasingly studied. Since gut microbiome has a dynamic role to host metabolism, it has also been termed as a metabolic organ and therefore there is increased interest in studying the microbiome specific metabolism. However, the therapeutic value of probiotics may be missed if the colonization resistance is strong. This has raised an approach “feed your bacteria” and certain dietary recommendations for balanced gut microbiome. (Zmora et al., 2019)

2.2 Methodological challenges

The use of methods always involves choices and compromises. The methods in microbiome research are highly variable and often are based on in-house or experimental protocols. Workflows and experimental protocols can be complicated and prone to bias and errors at all steps (Fig. 1), from sample collection to storage (Choo et al., 2015; Watson et al., 2019), DNA extraction (Lim et al., 2018; Yang et al., 2020), primer selection (Z. Chen et al., 2019; Rintala et al., 2017; Walker et al., 2015), sequencing, and bioinformatics analyses (Clooney et al., 2016; Y. Wang & LêCao, 2020; Ye et al., 2019). In one study, DNA extraction method explained 5.7% of

microbiome variability, which was comparable to the observed interindividual differences (7.4%) in the microbiota (Bartolomaeus et al., 2021). For quality improvement, the relative effect sizes of biological and technical covariates, technical bias, and present quantitative effect sizes alongside P-values should be considered (Debelius et al., 2016). Including technical replicates can help minimize unwanted variations and ensure accurate biological interpretations (Fachrul et al., 2022). Establishing standard methods in this field is challenging due to the need for optimization, implementation, constant updates, and equal access to resources among scientists. Consequently, the objective is to employ optimized, well-controlled methods that can be cross-referenced to other studies (Bharucha et al., 2020).

Another challenge for methods is the longitudinality (Martínez Arbas et al., 2021). Microbiome research and laboratory methods have developed rapidly in recent years. New instruments and reagents for sample collection, DNA-extraction, primers, sequencing and metabolite assays have been added to selection, without forgetting new bioinformatic tools. It is necessary to determine whether to apply consistent methods during follow-up studies, or to update the most recently improved methods at the cost of reduced comparability. (Martínez Arbas et al., 2021)

Due the expanded population-based microbiome studies, the recent challenge is also to create high-throughput methods. Large sample collection requires automatization and high-volume capacity instruments. Although, faecal sample is a feasible sample material regarding non-invasive aspect, it can be challenging to handle by pipetting robots. Faecal material is a semi-solid mixture composed of fibre, fat, undigested food residues, mucus and other viscous components that may lead to technical issues with pipette aspiration.(Karu et al., 2018; Shakeri et al., 2022)

Additionally, the practical challenge of collecting faecal samples from human volunteers is a significant limiting factor in microbiome and metabolomics studies. Microbes and metabolites are susceptible to bias during sample storage. The time and temperature of storage are critical for preserving samples for microbiota or metabolome analysis. The ratio of microbes may change because some bacteria can grow at varying temperatures while producing metabolites. Certain metabolites degrade rapidly if the sample is not frozen shortly after collection. Furthermore, fluctuating temperatures can degrade specific metabolites due to their molecular instability or enzymatic activity. (Stevens et al., 2019) Microbial metabolites may increase at ambient temperatures if microbial growth is not inhibited. Bacteria adapted to room temperature are more likely to be overrepresented during sample storage. Environmental factors, such as the cold winter in Finland, can affect samples through freeze-thaw cycles during logistics. Generally, immediately freezing samples at −80°C is considered the gold standard for preserving metabolites by halting enzymatic activity, hydrolysis, oxidation, and other degradative processes (Stevens et al., 2019). Recently, for human faecal samples especially in large human

cohorts, considerations of cost, practicality, donor privacy, and convenience have driven a strong demand for at-home collection. However, at-home sample collection involves multiple steps with temperature fluctuations. Moreover, storing and shipping frozen field samples can be inconvenient for participants and prohibitively expensive for researchers, creating a need for ambient-temperature storage options (Zierer et al., 2018). Multiple preservatives have been evaluated (Table 1) for their suitability in sample collection for metabolite assays and DNA extraction (Natarajan et al., 2021; Vogtmann et al., 2017).

Bias in DNA extraction can arise from variations in bacterial characteristics, such as resilient cell walls or envelopes. Gram-positive bacteria are particularly difficult to lyse due to their thicker peptidoglycan cell walls. Mechanical lysis has been shown to enhance DNA yield, although it may also lead to DNA shearing, which can cause reduced sequencing efficiency, loss of long-read information and bias in the quantification of microbial abundances. With an ideal DNA extraction method, all bacteria should be recovered equally well. (Costea et al., 2017; X. Li et al., 2020; Robe et al., 2003)

Methodological bias can lead to significant discrepancies in observed microbiome composition, resulting in variability across studies and laboratories that use different protocols (Han et al., 2020; Roesch et al., 2009; Salter et al., 2014; Sinha et al., 2017). To enhance comparability and ensure the accuracy of measurements, the standardization of microbiome analysis methods has been identified as a critical need by academic, diagnostic, industrial, and regulatory sectors. (Greathouse et al., 2019; Stulberg et al., 2016; Tourlousse et al., 2021; W.-K. Wu et al., 2019)

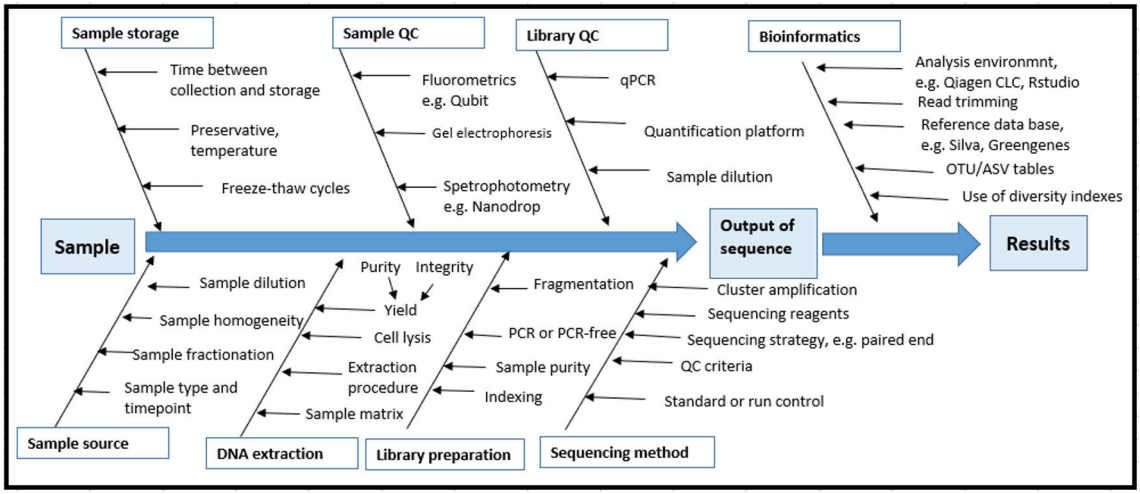


Figure 1. Possible sources of methodological bias in microbiome studies. Modified from Bharucha et al. 2020.

To combine the sample collection tube for microbiome and metabolites, research groups have studied several accessible sampling devices (Table 1), for instance DNA-stabilizing tubes, faecal immunochemical test (FIT) tubes and FOBT cards (Lim et al., 2020; Ramamoorthy et al., 2021). However, the abundant additives, buffers, salts, and other detergents in the collection containers may render them incompatible with liquid chromatography (LC) and mass spectrometry (MS) if the ions form precipitates. Although, it is seen that high inorganic salt concentrations are acceptable in reversed-phase LC-MS as long as ions do not form precipitates. Moreover, it is known that some of these storage conditions can distort the metabolomic profile of faecal samples compared to flash-freezing (Lim et al., 2020; Ramamoorthy et al., 2021). Ethanol 95% as sample preservative has been found feasible in metabolomics. (Lim et al., 2020; Loftfield et al., 2016) The ethanol prevents microbial growth and stabilizes the microbiome until downstream analysis is performed. Furthermore, it is expected to stabilize the metabolome, to some degree, because of enzymatic metabolism of gut microbiota is prevented by protein precipitation. Recently, OMNImetGUT tube including 95% EtOH (DNA Genotek, Canada) was launched as a stool storage and stabilization kit developed specifically for metabolomics. The corresponding kit for microbiome OMNIgeneGUT has been used in previous studies for gut microbiome (de Goffau et al., 2022; Williams et al., 2019).

A shared collection tube with preservative has been seen as the optimal solution which has not yet been found. The problem regarding the ethanol storage of bacterial cells may originate from osmotic pressure, where the inner cell liquid, which is ionic, is forced out of the cell to the less ionic liquid outside the cell. This causes drying, shrinkage and death of the cells. This does not necessary disturb the DNA extraction since live bacteria are not needed. However, cell shrinkage and drying can affect microbiome composition. To maintain fitness of cells, glycerol has been used for decades. Glycerol is widely used as a cryoprotectant for long-term bacterial storage at low temperatures. It prevents cell membrane damage and rupture by inhibiting ice crystal formation (Fowler & Toner, 2006), however, glycerol is not suitable for metabolomics since it may lead to misinterpretation of metabolic pathways (Nastasi et al., 2023; F. Xu et al., 2009). Altogether, methodological studies in preanalytical factors in the field are predominantly low in sample size, not done systematically and lacking cross-sectional or longitudinal check points (Table 1).

Table 1. Preanalytical factors affecting microbiota and metabolite results.

Approach	Studied method	Effect on results	Samples	Reference
Microbiota				
Sample preservative	no preservative	unstable at RT at 4°C for up to 72 hours minor alterations >Actinobacteria phyla < Firmicutes	n=4, n=1	(Choo et al., 2015; Roesch et al., 2009)
	OMNIgeneGUT	>Bacteroides (vulgatus, uniformis), Prevotella copri, Ruminiclostridium siraeum, <Agathobacter rectalis, Blautia_A wexlerae	n=5, n=5	(Z. Chen et al., 2019; Pribyl et al., 2021)
	Ethanol	<Alpha diversity <Bacteroides, Firmicutes	n=5, n=50	(Z. Chen et al., 2019; Vogtmann et al., 2017)
	DNA/RNA shield	>Faecalibacterium	n=5	(Z. Chen et al. 2019)
	Faecal cards (FOBT, FTA, FIT)	FOBT: stable after 3 days, <read count, <Firmicutes, FTA: comparable to OMNIgene, stable up to 8 weeks, <0.5% changes in rel. abund, FIT: low stability during 4 days	n=3, n=15, n=50	(Dominianni et al., 2014; S. J. Song et al., 2016; Vogtmann et al., 2017)
	RNA later	< Alpha diversity, < Bifidobacterium, no inhibition of bacterial activity, <DNA purity and yield	n=1, n=3	(Choo et al., 2015; Dominianni et al., 2014)
Storage temperature	RT, up to 72h	<>Bacteroides, <Clostridium, >Enterobact, Bifidobact and Citrobacter, diversity changed at 12 h (3.06%) at 24h (8.61%), 48h (9.72%) and 72 h (10.14%)	n=4, n=1	(Choo et al., 2015; Roesch et al., 2009)
	4°C	<Bacteroidetes and Actinobacteria, >Firmicutes (Faecalibacterium, Coprococcus, Eubacterium, Ruminococcus, and Blautia), stable until 24h	n=11, n=2, n=28	(Nagata et al., 2019; C. S. Poulsen et al., 2021; Tedjo et al., 2015)
	-20°C	stable after 14 days, >Bacteroidetes,Proteobacteria, >lysis of cell wall by freezing	n=2, n=2 n=1	(Lauber et al., 2010; C. S. Poulsen et al., 2021; Ray et al., 1976)

Approach	Studied method	Effect on results	Samples	Reference
	-80°C	stable after 14 days, minimal changes after 12mo, >Bacteroidetes and Proteobacteria, >ratio of Firmicutes to Bacteroidetes	n=20, n=1, n=2, n=2	(Barko et al., 2024; Choo et al., 2015; Lauber et al., 2010; C. S. Poulsen et al., 2021)
Pre-treatment of DNA extraction	pre-treatment/lysis	> diversity with beads, >GRAM+ bacteria	n=2, n=745, n=1, n=1, n=2 n=26	(Costea et al., 2017; Fernández-Pato et al., 2024; X. Li et al., 2020; Lim et al., 2018; Walker et al., 2015; Watson et al., 2019)
Metabolomics				
Sample preservative	Ethanol, OMNImetGUT	stable profile, high overlap with flash freezing	n=18, n=16	(Loftfield et al., 2016; Ramamoorthy et al., 2021)
	OMNIgeneGUT, RNA later	>distort the metabolic profile, lower detection rates vs EtOH	n=8, n=8	(Guan et al., 2021; Z. Wang et al., 2018)
Storage time and temperature	no preservative	detectable changes in 6h RT, unstable profiles in RT, >SCFA, <amino acids and nicotinate	n=11, n=5	(Chiu et al., 2023; Gratton et al., 2016)

2.3 Methodology for next generation sequencing

Next-generation sequencing (NGS) technologies have revolutionized gut microbiome studies, enabling deeper insights into microbial communities and their functions. Two primary approaches are commonly used: 16S rRNA sequencing for taxonomic identification and shotgun metagenomics for species/strain-level characterization (Jovel et al., 2016; Wensel et al., 2022). These techniques have overcome limitations of traditional culture-based methods, allowing for comprehensive analysis of complex microbial ecosystems (Gao et al., 2021). Bioinformatics tools play a crucial role in processing and interpreting the massive amounts of data generated by NGS (Pereira et al., 2020). Microbiome composition is often investigated by α-diversity (within samples) and β-diversity (between samples) to understand microbial community structures and their correlations with health and disease (Jovel et al., 2016). Moreover, metagenomics has an essential role in many phases of microbiome-based product development and identification of microbial targets for clinical trials. While significant progress has been made, challenges remain in fully understanding the ecology of diseases and developing targeted treatments based on microbiome data (Maheshwari et al., 2024).

Next generation sequencing (NGS) workflow is preceded by sample collection, DNA isolation and library preparation. Steps are explained in more detail below.

2.3.1 Sample storage and DNA extraction

Faecal samples for microbiome studies are commonly collected in sterile, RNA/DNase-free containers and stored at -80°C to preserve microbial DNA. Sample is usually collected at home and stored at +4°C until the study visit and handling in laboratory. More advanced way to collect samples, is to use collection tubes with preservative and send sample by post to an operating laboratory. (Costea et al., 2017; Thomas et al., 2015; W.-K. Wu et al., 2019)

The extraction process involves lysing microbial cells to release their DNA, followed by purification to remove contaminants like proteins, lipids, and inhibitors (e.g., bile salts or polysaccharides) commonly found in faecal matter. Cell lysis can be performed by i) Mechanical disruption: bead beating (with silica or zirconium beads) to physically break open tough microbial cell walls, including Gram-positive bacteria, ii) Chemical lysis: lysis buffers containing detergents (e.g., SDS) to solubilize membranes and disrupt cells. iii) Enzymatic lysis (optional): use of lysozyme or proteinase K to digest specific cell wall components and proteins. Next, the lysate is centrifuged to remove insoluble debris like cell wall fragments and undigested materials. Typically, DNA purification involves 1) binding of DNA: Apply the lysate to a silica-based column or magnetic beads, which bind DNA while other components are washed away. 2) Washing with ethanol-based or specific wash

buffers to remove salts and impurities. 3) Elution of pure DNA from the column or beads using a low-salt buffer or water. The quality of eluate can be evaluated by DNA concentration and purity using a spectrophotometer (e.g., Nanodrop) or fluorometric assay (e.g., Qubit). Integrity can be measured by gel electrophoresis or e.g. a bioanalyzer (Agilent, US). DNA isolates are more stable regarding storage condition compared to original sample material (e.g. feces), however the recommended temperature for long-term storage has been -20 or -80 degrees (Gavriliuc et al., 2021; Widjaja & Rietjens, 2023).

The key challenges in faecal DNA extraction are (1) High microbial diversity: Different microbes have varying cell wall structures, requiring methods that can lyse all cell types. (2) Presence of inhibitors: Compounds in feces can inhibit downstream enzymatic reactions, so thorough purification is essential. The principal goal is to isolate intact, pure DNA that accurately represents the microbial community for downstream applications like 16S rRNA sequencing. (Fernández-Pato et al., 2024)

2.3.2 16S rRNA based NGS libraries and sequencing

NGS technique involves PCR amplification of the 16S rRNA gene, sample barcoding, and sequencing, followed by bioinformatic analysis (Couper & Swei, 2018). 16S rRNA is a small subunit of ribosomal RNA, which originates from the subunit of the prokaryotic (30S) ribosome (Stern et al., 1989). 16S rRNA encoding genes can be found in all bacteria and archaea, however, one limitation is the variability in copy numbers within bacterial genomes, as well as sequence variations among closely related taxa and within individual genomes. (Větrovský & Baldrian, 2013).

The microbial 16S rRNA gene is close to 1500 base pairs (bp) in length. Some regions of this gene are highly conserved among microbial species and exhibit a very slow rate of evolution (Van de Peer et al., 1996). Besides these conserved regions, the 16S rRNA gene contains nine hypervariable regions, known as target regions (V1 to V9), which show notable variability in sequence between different microbial species (Van de Peer et al., 1996, Chakravorty et al., 2007). Due the mentioned properties, the 16S rRNA gene serves as a feasible tool for taxonomic analyses and species identification, of the micro-organisms (Chakravorty et al., 2007; Van de Peer et al., 1996). In next-generation sequencing (NGS), one or multiple hypervariable target areas of the bacterial 16S rRNA gene (i.e. V3V4 or V4) are amplified, and the sequences from every sample are identified using unique index sequences specific to each sample (Klindworth et al., 2013). The length of the amplicon generated with i.e. the V3V4 region typically ranges from 460 to 480 bp depending on the specific primers used (Regueira-Iglesias et al., 2023). Library preparation involves diluting of DNA isolates (template DNA), one- or two-phased PCR and purification of PCR

products. PCR program and number of cycles are usually modified based on the sample material. For example, low biomass samples (e.g. skin) require more PCR cycles than high biomass (e.g. feces). Moreover, amount of template DNA can be modified to reach higher yield of the PCR product. However, the cycles and concentration must be balanced against the risk of generating artefacts, such as chimeric sequences (Haile et al., 2019). The PCR products also called library preps are then frozen until the preparation of sequencing. Libraries are diluted (normalized), pooled and denatured before the sequencing (www.illumina.com).

The 16S rRNA gene libraries are commonly sequenced with a high-throughput instrument, such as Illumina MiSeq (Illumina, USA) or Ion Torrent (Thermo Fischer Scientific, USA). Illumina typically produce reads of up to 300 bp (Johnson et al., 2019). This makes the V3V4 region suitable for high-throughput sequencing applications, allowing for the generation of paired-end reads that can effectively cover the entire region (Z. Chen et al., 2019). Illumina sequencing is based on flow cell technology, which has millions of small nanowells. Nanowell is etched into the class surface for optimal cluster spacing. All nanowells contain DNA probes used to capture sample DNA strands for amplification during cluster generation. The procedure warrants that DNA clusters are only formed within the nanowells, providing even, consistent spacing between adjacent clusters and assuring accurate resolution of clusters in imaging phase. This can be an issue in Miseq therefore the loading concentration and the passfilter (%) should be always verified. Maximal use of the flow cell surface leads to overall higher clustering. (Aird et al., 2011; Sato et al., 2019)

The sequence data analysis is performed with large variety of bioinformatic tools. Briefly, the sequences can be exported in different kinds of file formats (e.g. fastq), next they are trimmed and clustered, and then compared to known sequences in reference databases (Molano et al., 2024), e.g. Greengenes2 (DeSantis et al., 2006; D. McDonald et al., 2024) or SILVA (Quast et al., 2013). For the pre-processing of NGS data, there are few pipelines available, e.g. DADA2. (Regueira-Iglesias et al., 2023)

The 16S rRNA NGS has certain limitations; genus level is commonly the taxonomic level that can be reached. Therefore, shotgun metagenomics has advantage on deeper insight into taxonomy by species or even strain level on top of information on functional genes. Metagenomics is not limited to the 16S gene, which creates new possibilities for studying other microbes as well. However, metagenomics has difficulties in eliminating human host genetic background. Additionally, hypersensitivity may lead to false positive results.(Y. Chen et al., 2022)

2.4 Methodology for measuring faecal and milk metabolites

Metabolomics is a comprehensive approach used to study the small-molecule metabolites present within biological samples. The key techniques employed in metabolomics are gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR) spectroscopy. These methods take advantage of the fact that the separation and identification of each individual compound is based on their chemical differences and properties. (Koek et al., 2011; Munjal et al., 2022). Prior these techniques, metabolite extraction is typically performed to remove proteins and other macromolecular debris, which is particularly important for complex matrices such as fecal samples (Zierer et al., 2018).

Faecal metabolomics has emerged as a complementary tool to microbiome analysis, offering insights into microbial functions and their metabolic outputs. Metabolomics is a promising approach to unveil host–microbiota interactions in the context of disease risk. The crosstalk and dynamics between the host and gut microbiome are mediated by metabolites, which serve as vital signalling molecules and essential nutritional value. (Lamichhane et al., 2018) The most studied microbial metabolites are short-chain fatty acids (SCFA) and bile acids (BA), since there is substantial evidence supporting the prominence of these metabolites as key microbial metabolites. Previous studies have highlighted their significant roles in various physiological and pathological processes (Ganesan & Suk, 2022; Visekruna & Luu, 2021) .

2.4.1 GC-MS

GC separates volatile compounds in a sample based on their boiling points and affinities for the stationary phase of the column. The sample is vaporized and carried through a column by an inert gas, usually helium. As the components interact with the stationary phase, they elute at different times, known as retention times. This means that only small volatile metabolites can be measured, and often chemical derivatisation is needed to reduce the boiling point of the compound during the gas phase. (Fiehn, 2016; Garcia & Barbas, 2011)

Mass Spectrometry (MS) identifies and quantifies the separated metabolites by measuring their mass-to-charge ratios. After elution from the GC, the compounds enter the mass spectrometer, where they are ionized, fragmented, and detected. The resulting mass spectrum provides a unique fingerprint for each metabolite. GC-MS enables the identification and quantification of small-molecule metabolites (<650 Da) (Fiehn, 2016)

GC-MS is used to generate comprehensive profiles of metabolites in biological samples, such as urine, blood, stool or tissues. The application can identify potential biomarkers for diseases by comparing metabolite profiles between healthy and diseased states. The technique aids in understanding metabolic pathways by tracking the presence and concentration of specific metabolites. (Papadimitropoulos et al., 2018)

Recent advances in GC-MS have provided medium to high sensitivity and specificity, allowing for the detection of low-abundance metabolites. It can provide reliable quantitative data, essential for comparing metabolite concentrations across samples. There are some limitations regarding sample preparation: Samples must be volatile or derivatized to be compatible with GC, which can introduce variability. Moreover, there can be issues with data complexity; the interpretation of mass spectra can be complex due to overlapping peaks and the need for extensive databases for identification. (Garcia & Barbas, 2011; Kanani et al., 2008)

GC-MS is a powerful tool in metabolomics, providing detailed insights into metabolic changes and the biochemical state of organisms. Its applications in clinical research and environmental studies continue to expand, making it a cornerstone of modern metabolomics investigations (Fiehn, 2016; Koek et al., 2011)

2.4.2 LC-MS

Liquid chromatography (LC), widely used in metabolomics, works by separating complex mixtures of metabolites based on their chemical properties. LC enables the identification, quantification, and characterization of these compounds, offering critical insights into cellular processes, disease states, and physiological responses. (Furlani et al., 2021; Gika et al., 2014) Liquid chromatography operates by passing a liquid sample through a column packed with a solid stationary phase, with a liquid mobile phase flowing through it. The interaction between the metabolites and these two phases causes different compounds to elute (exit the column) at different times, known as retention times. These retention times are used to separate and identify the metabolites. (Gika et al., 2014; Theodoridis et al., 2012)

Several types of LC are commonly used in metabolomics, each suited for specific types of metabolites or analytical goals. For example, ultra-high-performance liquid chromatography (UHPLC) operates at higher pressures and uses smaller particle sizes in the stationary phase than HPLC, leading to improved separation. It has faster run times, better resolution, higher sensitivity than HPLC. It is applicable to high-throughput metabolomics studies, where rapid and efficient analysis is critical. The separation power of liquid chromatography is often combined with highly sensitive detection methods like mass spectrometry (LC-MS) (Petras et al., 2017). This operates similarly to the MS on the GC but uses softer ionisation techniques which

results in less fragmentation of the molecular ion removing some of the data complexity (Furlani et al., 2021). LC-MS yields information of the compound structure such as accurate masses, isotope ratios, and product ion spectrums (Pitt, 2009).

2.4.3 NMR

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful, non-destructive analytical technique utilized in metabolomics to identify, quantify, and characterize metabolites in biological samples. NMR provides detailed structural information about metabolites and requires minimal sample preparation, making it a valuable tool in both untargeted and targeted metabolomics studies. (Nagana Gowda & Raftery, 2021)

NMR spectroscopy has the ability to detect metabolites through one or more types of atomic nuclei such as ^{1}H, ^{13}C, ^{31}P, or ^{15}N (Nagana Gowda & Raftery, 2021), and possess a property called nuclear spin. When inserted in a strong magnetic field, these nuclei align with or against the field, creating discrete energy states. Upon applying a radiofrequency (RF) pulse, the nuclei are excited to a higher energy state. As they return to their lower energy state, they emit RF signals, which are detected and transformed into a spectrum. The resulting NMR spectrum provides information about chemical shifts, coupling constants and signal intensity correlating with the concentration of metabolites.(Kruk et al., 2017)

^{1}H-NMR (Proton NMR) is the most common type of NMR technique in metabolomics. It detects protons (hydrogen nuclei) in metabolites and provides information about the hydrogen-containing functional groups. NMR is a robust tool for metabolomics, offering advantages in terms of reproducibility, quantification, and structural elucidation. A key draw back, however, is that NMR is far less sensitive than mass spectrometry. However, its non-destructive nature, minimal sample preparation, and ability to provide detailed molecular information make it a popular technique for studying metabolism in health and disease. (Marshall & Powers, 2017)

2.5 Host-microbe interaction via microbial metabolites

Recently, the microbiome field has been moving towards multiomic approaches. So far microbiome research has answered the “Who is there?” question by NGS. The next step has been to understand “What do they do?” by metabolomics. Further studies will show “Why and how they do that?” by a range of analytical techniques including shotgun metagenomics.

Functional side of gut microbiome is seen in metabolites. Microbes produce large variety of microbial metabolites, which are mostly beneficial to the host, for example Bacteroides can digest dietary fibre, producing short-chain fatty acids that nourish colon cells (T. Chen et al., 2017).

Host-microbe interactions refer to the complex relationships between a host organism (e.g., humans) and the microorganisms that live on or within it. These interactions play key roles in immune system development, metabolism and nutrition, protection against pathogens, pathogenesis, and microbiome-host communication. The gut microbiome influences immune system programming by affecting tolerance, stability, and reactivity. Commensal microbes stimulate immune system maturation and help maintain immune homeostasis. Gut microbes aid in the digestion of complex carbohydrates, synthesis of vitamins (e.g., vitamin K and folate), and regulation of lipid metabolism. Commensal microbes also compete with pathogenic bacteria for nutrients and space, providing a protective effect known as colonization resistance. Pathogenic microbes can cause infections, inflammation, and tissue damage, leading to diseases, such as sepsis, pneumonia, or gastrointestinal infections. Additionally, microbes can cause low-grade inflammation, which is not as easily detectable as an acute bacterial infection. Inflammation is often linked with weight gain. (Tremaroli & Bäckhed, 2012) Moreover, obesity and inflammation are recognized as early symptoms of metabolic dysregulation and may contribute to the development of cardiometabolic diseases. The mechanisms are not entirely understood but may relate, in part, to the effects of microbiota on energy harvesting capacity (Sanmiguel et al., 2015). Gut microbes also influence blood glucose levels, fat storage, and the host's response to satiety hormones. An incorrect assembly of gut microbes may direct to obesity, chronic inflammation, and metabolic disorders (Allin et al., 2018; Delzenne et al., 2015; Sommer & Bäckhed, 2013).

Some metabolites act as signalling molecules that mediate host-microbiome interaction and dynamics (Visconti et al., 2019). Gut microbiota produces bioactive metabolites, such as short-chain fatty acids, aromatic amino acid metabolites, and bile acids, which have been associated widely in human health and disease. A balanced healthy state, homeostasis of gut (Fig. 2) is maintained by commensal bacteria, microbial metabolites, mucus layers, unbroken epithelium with tight junctions and other immunological factors. In an unbalanced state of gut (dysbiosis), pathogenic bacteria take over the gut. They reach the inner mucus layer reducing the tolerance of gut lumen, and epithelial barrier is damaged leading to leaky gut (Fig. 2). This can result in systemic low-grade inflammation. Furthermore, a leaky gut has been associated with disease states including obesity, insulin resistance, cardiovascular disease, autism-spectrum disorder and Parkinson’s disease. GM has also role in metabolism of certain medicine (Zimmermann et al., 2019), which can be absorbed only by bacterial metabolism. Therefore, microbiome-metabolome

crosstalk can also affect the treatment response of the later disease state. (Weersma et al., 2020; Wilson & Nicholson, 2017)

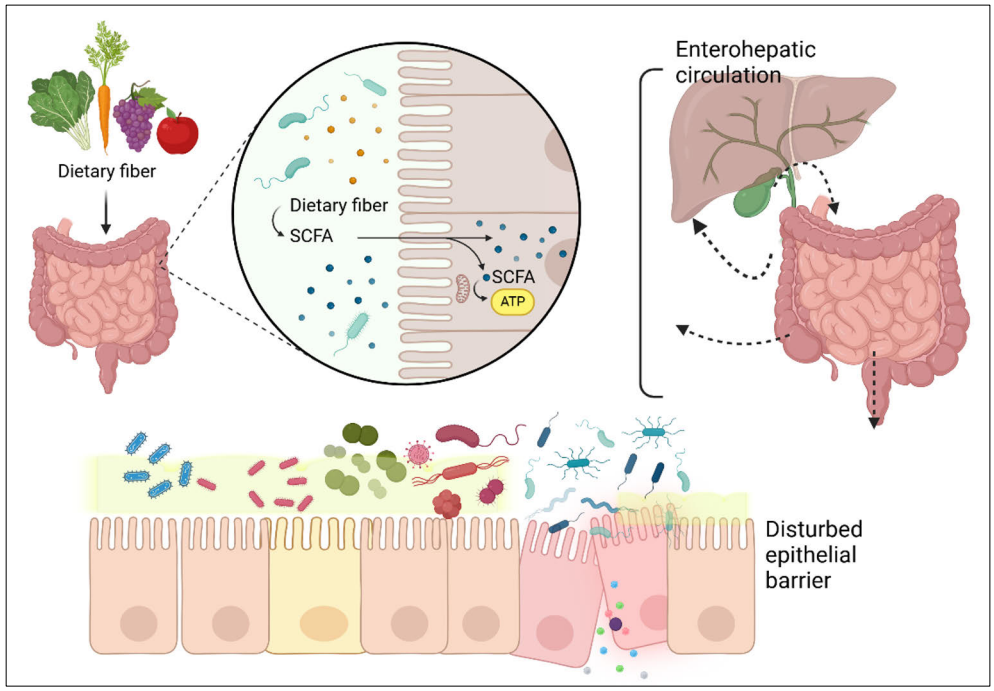


Figure 2. Simplified overview of metabolism in gut. Source and absorption of short-chain fatty acids in gut. Enterohepatic circulation of bile acid metabolism. Healthy and disturbed epithelial barrier of gut. Created in Biorender (www.biorender.com).

2.5.1 Human milk and gut microbiome in early life

An increasing number of studies on gut microbiota (GM) have focused on the early life and the vulnerable time of development, already noted in 1884, when paediatrician Escherich expressed concern over the high mortality rates associated with gastrointestinal infections in the first months of life (Shulman, 2004). It is stated that development of gut microbiome starts during birth and early life factors, such as breastfeeding, delivery mode, and gestational age, significantly shape the gut microbiota (Ruiz et al., 2016). It is known that composition evolves from low alpha diversity in infancy, stabilizes in adulthood, and declines with age. Human milk oligosaccharides (HMOs) promote beneficial microbial colonization, while disruptions in early microbiome assembly can have long-term impacts on immunity, metabolism, and behaviour. Accurate microbial profiling remains essential for understanding host-microbiota interactions and finding early-life trajectories for microbiome maturation through the microbial metabolites (Roager et al., 2023)

Human milk (HM) is regarded as the most optimal way to nourish a new-born during the first months of life, as it provides a balanced mix of nutrients and bioactive compounds essential for the optimal growth and development of the infant's organs, immune system, and gut microbiota. Additionally, breastfeeding is linked to numerous short- and long-term health benefits. In the short term, it helps protect infants against gastrointestinal and respiratory infections, as well as atopic diseases. (Peila et al., 2022; K. O. Poulsen & Sundekilde, 2021) In the long term, it is associated with a reduced risk of obesity and type 2 diabetes (Victora et al., 2016).

Breastfeeding is recognized for its role in nurturing a healthy microbiota in infants. Human milk contains macro- and micronutrients, as well as a range of bioactive components, i.e. carbohydrates, amino acids, lipids, enzymes, antibodies, immunoglobulins, hormones, growth factors, cell debris, prebiotics and HMOs. HMOs, produced by mother, are key components of breastmilk known to influence the composition of the infant's microbiome. (Ames et al., 2023; Azad, Robertson, et al., 2018). Secretion of HMO is regulated by mother's secretor status. FUT2 (secretor) and FUT3 (Lewis type) genes determine if mother is a secretor or non-secretor of specific HMOs (Samuel et al., 2019).

Certain gut microbes, at least *Bifidobacterium* and *Bacteroides* species, are capable to metabolize dietary oligosaccharides including HMOs that human cannot digest. This metabolic route produces microbial metabolites (i.e. short-chain fatty acids, acetate, propionate, butyrate, folate, amino acids, vitamins, hydrogen, and carbon dioxide) that provide nutritional value and serve as substrates for enzymes, i.e. acetylases, methylases, and glucuronidases. (Esch et al., 2020; Rowland et al., 2018; Zhang et al., 2022) These metabolites are connected to the immune system through their roles in regulating antigen tolerance, promoting mucus layer maturation, and inducing epigenetic modifications in intestinal cells and peripheral tissues. Disruptions in these processes have been associated with conditions such as inflammatory bowel disease (IBD), obesity, and type 2 diabetes. In early childhood, improper microbiome colonization or antibiotic-induced imbalances can interfere with immune system development, potentially increasing the risk of atopic or allergic diseases. (Stiemsma & Turvey, 2017) Moreover, insufficient microbial exposure during early life may heighten the risk of immune-mediated diseases, including autoimmune disorders and type 1 diabetes (Sarkar et al., 2021). Hence, understanding how early life nutrition among other factors influences the gut microbiome is crucial to developing methods to avoid disrupted colonization and preventing immune-mediated diseases. (Ames et al., 2023)

In addition to human milk, other influencing factors of gut microbiome colonization have been studied (Suárez-Martínez et al., 2023), such as delivery mode (Mueller et al., 2015), gestational age (Dougherty et al., 2020; Ruiz et al., 2016), maternal BMI, antibiotics, genetic factors (sex, FUT/ABO blood group), pets,

environmental factors (greenness, urban/rural), lifestyle (physical activity, diet) and generational transmission (Suárez-Martínez et al., 2023). Most studied factors and their influence on gut microbiota and metabolites are presented in Table 2. Although, there is a broad spectrum of influenced microbes, the early colonizers are well-presented and replicable seen in those studies. There is limited research on microbial metabolites, although there are already knowledge of metabolites linking to gestational age and feeding mode.

Infant microbiome develops through the two main routes that involve the vertical transmission or mother-new-born transmission (Suárez-Martínez et al., 2023; Vatanen et al., 2022). First route is linked to the type of delivery. This route starts with vaginal and faecal microbe colonisation such as *Lactobacillus, Bacteroides, Bifidobacterium, Streptococcus spp and Prevotella*. Next, infant is exposed to more facultative anaerobes from environment, like *Enterobactericeae.* Later, anaerobic bacteria are more prevalent, such as *Bifidobacterium, Clostridium and Bacteroides*. Close to one year of age, microbiome has already relatively high diversity and start to resemble of adult composition. The second route is linked to breast- and formula-feeding. In infancy, inter-individual variability increases and bacterial diversity may decrease by the human milk embracing certain bacteria (Brink et al., 2020; Mercer et al., 2024). Human milk has immunoglobulins (IgA), HMOs and antimicrobial factors that affect the gut colonization process and bacterial composition. Human milk itself has microbiome, which commonly consists of *Bifidobacterium, Lactobacillus, Streptococcus and Staphylococcus*. In addition to these routes, controversial results have been found related to microbial exposures in uterus (Mishra et al., 2021; Stinson et al., 2019) but still birth should be considered as the key moment for colonization of the new-born by viable microorganisms (Banchi et al., 2024; de Goffau et al., 2019). However, microbial metabolites can pass the placenta and a mouse model has shown association between circulating microbial metabolites and mammalian prenatal gene expression, which further highlights the significant influence of maternal microbial status on genes essential for fetal immune function, translation, metabolism, and neurophysiology (Husso et al., 2023). Although gut microbiome is highly diverse between individuals, microbial maturation in the gut during early life follows consistent patterns in humans worldwide (Fahur Bottino et al., 2025). Finnish cohort study showed that infants' GM development is highly predictable; it follows one of five trajectories, dependent on infant exposures, and predictive of later health outcomes (Hickman et al., 2024).

2.5.2 Metabolites in early life

Metabolites in early life have been studied but results are still relatively sparse (Table 2). Meta-analysis of SCFA showed association between breastfeeding and changes

in propionate and butyrate concentrations during the first 9 months, which was not observed in formula-fed children (Łoniewski et al., 2022). They observed a higher concentration of faecal SCFAs in formula fed compared to breastfed infants at different age. Only in the first month of life, the concentration of acetic acid was lower in formula fed group. Similarly in Bäckhed's study, the feeding mode had a major impact on the composition and functions of the gut microbiota, but it was the cessation of breastfeeding rather than the introduction of solid foods that determined the GM maturation (Bäckhed et al., 2015). Another study found positive association between faecal butyrate and *Clostridiales* and breastfeeding cessation, and that diverse and personalised assemblage of *Clostridiales* species possessing the acetyl-CoA pathway played a primary role in gut butyrate production (Tsukuda et al., 2021). Milk metabolites are important in early life development and HMO metabolism has been shown to associate also with other than *Bifidobacteria*. Borewicz et al. (2020) showed that duration of lactation had a major impact on breastmilk HMO content, which decreased by time, with exception of 3-fucosyllactose (3-FL) and lacto-N-fucopentaose III (LNFP III). GM composition was associated with infant age and influenced by delivery mode and human milk LNFP III concentration at two weeks, and next at six weeks with infant sex, mode of delivery, and concentrations of 3-SL (HMO), and then lastly at 12 weeks of age with infant sex and lacto-N-hexaose (HMO). Correlations were weak between the levels of individual HMOs and the relative abundance of OTUs in infant feces, including the most predominant *Bifidobacterium*, which varied with age. The results suggested that the HMO composition represents only one of several factors regulating the colonization of the infant GM (Borewicz et al., 2020).

Table 2. Main factors shaping gut microbiome and metabolome in early life.

Factor	Phase	Influenced microbes	Influenced metabolites	Reference
Gestational age/preterm vs term	prenatal	Preterm: *>Atopobium, Gardnerella, Candida alb. and Ureaplasma* *<Lactobacillus spp., Bacilli* delayed colonization	Preterm: < SCFA >vitamins, amino acid derivatives and unconjugated bile acids	Ruiz et al., 2016, Rougé et al., 2010, Moles et al., 2013; Henderickx et al., 2019, Dougherty et al., 2020, Ardissone et al., 2014
Delivery mode/ vaginal delivery	neonatal	*>Lactobacillus, Escherichia, Bacteroides, Bifidobacterium londum and catenulatum species, Streptococcus spp., Prevotella and Sneathia*	> carbohydrate metabolism including glycolysis/gluconeogenesis and glyoxylate and dicarboxylate metabolism	Penders et al., 2006; Dominguez-Bello et al., 2010; Fallani et al., 2010; Azad et al., 2013; Laursen et al., 2015, Biasucci et al., 2008; Dogra et al., 2015, Hoen et al. 2021
Delivery mode/ C-section delivery	neonatal	*>Corynebacterium, Propionibacterium Clostridium (cluster I), Clostridium difficile, Staphylococcus species and Klebsiella* *< Bacteroides and Bifidobact, depleted Bacteroidetes*	>galactose metabolism >ABC transporters	Biasucci 2008, Dominguez-Bello 2010, Munyaka 2014, Penders 2006, Ege 2011, Tun 2018, Shi 2018, Bokulich 2016, Hoen 2021
Feeding mode/Breastfeeding	postnatal	*>Staphylococcus, Streptococcus, Serratia,Pseudomonas, Corynebacterium, Ralstonia, Propionibacterium, Sphingomonas, Bradyrhizobiaceae, Bifidobacterium, Lactobacillus spp,* <diversity at 1 month more stable and uniform population	Exclusive BF: < total SCFA (Butyrate, Propionate, Valerate, Isobutyrate, and Isovalerate) > Lactate, Acetate, tryptophan, N-acetyltryptophan and riboflavin Differences in butyric acid, d-sphingosine, betaine& kynurenic acid	Oddy, 2017, Moossavi et al., 2018, Kumbhare et al., 2019, Porro et al. 2023, Brink et al. 2020, Lordan et al., 2024, Ding et al., 2022, Yelverton et al., 2023
Feeding mode/Infant formula	postnatal	*> E. coli, Clost. difficile, Bacteroides, Lactobacilli, Staphylococci, Atopobium,* age-dependent diversity deviations	> SCFAs and secondary bile acids > Amino acid and proteolytic metabolism < Acetate	Bezirtzoglou et al., 2011, Guaraldi and Salvatori, 2012; Gritz and Bhandari, 2015; Azad et al., 2016, Martin et al. 2016, Azad 2018, O'Sullivan 2015; van den Elsen 2019, Hoen 2021, Sillner 2021
Siblings	postnatal	*>Bifidobact* >ratio of anaerobe/facultative anaerobe		Penders 2006, Alderberth 2007, Laursen 2015
Dietary pattern/Solid foods	postnatal	**6-12 months** *<Veillonella,Lactobacilli, Bifidobacteria, and Enterobacteriaceae,* *>Ruminococcaceae, Lachnospiraceae,* **12-24 months** *>Faecalibacterium prausnitzii and Akkermansia muciniphila*		Homann et al. 2021, Laursen et al. 2017, Bergström et al. 2014, Fallani et al. 2011, Thompson et al. 2015, van den Elsen et al., 2019, Milani et al., 2017
Use of antibiotics (mother/ infant)	pre-/postnatal	**Mother:** disturpt overall ecology, *>Staphylococcus aureus* **Infant:** delayed colonization of *Bifidobacterium and Lactobacillus species, > Clostridium, < Staphylococcus, >Enterococcus, Clostridiales/ Lachnospiraceae,* delayed microbiome development	Infant: >Amino acid metabolism	Arrieta 2014, Kumbhare 2019, Dierikx 2020, Mueller 2015, Johnson 2005; Penders 2006; Fallani 2010; Chu 2015; Rasmussen 2018, Kozyrskyj 2007; Metsälä 2015, Lu 2020

2.5.2.1 SCFA metabolism in gut

Short-chain fatty acids (SCFAs) are microbiota-gained modulator molecules, for instance a nutrient that can act as communicating molecules between the gut microbiome (GM) and the host (Sakata, 2019). The colon, where most microbes reside, functions as a fermenter system for dietary components that escape host digestion. Non-digestible dietary carbohydrates (NDCs) form the primary energy sources for bacterial growth in the colon. It has been calculated that 20 to 60 grams of NDCs, including oligosaccharides and sugar alcohols, bypass digestive enzymatic degradation and reach the human colon. Microbial fermentation of NDCs primarily produces SCFAs, such as acetate, propionate, and butyrate, along with lactate, succinate, ethanol, methane, carbon dioxide, and hydrogen (Cummings & Macfarlane, 1991). The GM enterotype, which refers to a specific microbiome composition enriched in certain genera, affects the amount of SCFA produced, while human digestive enzymatic activity may alter microbial communities (Rios-Covian et al., 2020; Sakata, 2019).

SCFAs are volatile fatty acids with a structure of six or fewer carbon atoms. They can be in a straight-chain form (C1, formate; C2, acetate; C3, propionate; C4, butyrate; C5, valerate; C6, caproate) or a branched-chain form (like isobutyrate, isovalerate, and 2-methyl-butanoate). Acetate, propionate, and butyrate constitute 90–95% of the total SCFA output from the gut microbiome. (Macfarlane & Macfarlane, 2012). Previously considered only as food components, caproate and valerate (J. A. K. McDonald et al., 2018) are possible also be GM products, with caproate being notable increased in stool samples of severe obesity humans (BMI ≥ 40) (Rios-Covian et al., 2020).

Branched-chain fatty acids (BCFA), mainly isobutyrate, isovalerate and 2-methylbutanoate, contribution of total SCFA production to is approximately 5%, and arise from the metabolism of the amino acids valine, leucine, and isoleucine, respectively (Macfarlane & Macfarlane, 2012; Rios-Covian et al., 2020). Increased amount of BCFA reflects high protein intake, such as from meat-based diet and a decreased dietary fiber intake, which are further linked with negative health outcomes and ageing related health issues (Rios-Covian et al., 2020).

Oligosaccharides and pectin are the DF compounds thought to contribute the most to SCFA production in the colon. Butyrate, acetate, and propionate are absorbed after entering the enterocyte layer, whereas lactate and succinate seem to be intermediate products of dietary fibre fermentation. Butyrate serves as the primary energy source for colonocytes and regulates the maturation of mucosa-associated lymphoid tissue (MALT) (Furusawa et al., 2013; P. A. Gill et al., 2018; Louis et al., 2014), described with an elevated presence of immune cells such as macrophages, B and T cells and that it has an important role in antigen sensing. Only a minor part of the produced SCFA end up in the host's systemic circulation. (Boets et al., 2017)

However, SCFA levels exhibit significant inter-individual variation, as well as intra-individual variability, including factors such as dose-response, time-course, and circadian fluctuations (Sakata, 2019). SCFAs have seen as strong influencers of immune regulation, regarding findings of asthma and atopy in infants, as well as in animal models, or gastrointestinal disease in adults (Blacher et al., 2017; S. K. Gill et al., 2021; Sasaki et al., 2024)

In the portal circulation, SCFAs undergo first-pass effects, with most propionate being metabolized, influencing gluconeogenesis and lipogenesis. Acetate inhibits lipolysis in adipose tissue and can cross the blood-brain barrier. SCFAs are effective against microglial oxidative stress and have cellular signalling properties, impacting insulin release and the hypothalamic-pituitary-adrenal axis. Additionally, these signals affect ghrelin, leptin, and peptide YY release, promoting balanced appetite and satiety, improved insulin sensitivity and glucose metabolism, and lower serum lipid levels. (Ramos Meyers et al., 2022)

The infant gut microbiota undergoes distinct phases of SCFA production, progressing from low acetate and high succinate to high propionate and butyrate (Tsukuda et al., 2021). SCFAs can induce the differentiation and expansion of regulatory T cells, which are critical for neonatal health (Chun & Toldi, 2022). The establishment of SCFA production in infants is influenced by factors such as diet and antibiotic exposure. Total faecal SCFA concentrations increase until 12 months and stabilize after that in healthy children. (Łoniewska et al., 2023)

Disruptions in SCFA production are associated with various paediatric health issues, including obesity and allergies (Hsu et al., 2024). The ability of gut microbiota to produce SCFAs and essential vitamins is crucial for supporting metabolic progress for optimal growth and development (Kane et al., 2015). Olga et al. (2023) found that higher butyrate levels from human milk are inversely related to infant weight and adiposity, suggesting it may reduce the risk of childhood obesity. Animal studies have linked butyrate and its producing bacteria to lower risks of obesity and metabolic issues, such as liver fibrosis and insulin resistance (Arnoldussen et al., 2017; Lin et al., 2012). Additionally, studies in mice showed that acute oral administration of butyrate can quickly increase satiety and reduce food intake (Jin et al., 2015; Yadav et al., 2013). These studies underscore the potential efficacy of butyrate in human milk on infant metabolic health, weight regulation, and growth patterns. (Hsu et al., 2024; Olga et al., 2023)

2.5.2.2 Bile acid metabolism and gut microbiome

In the human body, bile acids (BAs) are part of metabolism and cholesterol catabolism, which functions to remove cholesterol from the body via urine and feces. Bile acids are typically present as taurine and glycine conjugates. BAs play a crucial

role in lipid metabolism and have hormonal functions, as they act as signalling molecules in the liver and activate the nuclear hormone receptor FXR. Additionally, they promote the absorption of fat-soluble vitamins in the small intestine. The presence of bile acids in the body is tightly regulated and mediated by various enzymes involved in BA metabolism. Excessive or insufficient secretion of BAs can lead to metabolic issues, such as malabsorption of nutrients. (Hofmann & Hagey, 2008; Hylemon et al., 2009)

When examining the metabolic pathways in cells, BAs are part of cholesterol catabolism as they contribute to the elimination of cholesterol from the body. BAs found in the human body can be categorized based on their structure and function into primary and secondary bile acids. Primary BAs include cholic acid and chenodeoxycholic acid. These BAs are produced as part of the enterohepatic circulation in the liver, where cholesterol serves as their precursor. With the help of glycine and taurine conjugates, primary BAs are transported to the gallbladder, where they are concentrated and stored. (Hofmann & Hagey, 2008)

When primary BAs are exposed to bacteria in the intestine, they are converted into secondary BAs, such as deoxycholic acid (DCA) and lithocholic acid (LCA). Intestinal anaerobic bacteria produce secondary BAs through bile salt hydrolysis and dehydroxylation reactions (Ridlon et al., 2006). Both deoxycholic and lithocholic acids are amphipathic (possess both hydrophobic and hydrophilic regions) and reduce the surface tension between fat droplets, allowing them to interact with bile salts for efficient emulsification in the small intestine. This process is crucial for breaking down large fats into smaller micelles that can be easily absorbed through intestinal walls. In addition, the emulsifying action creates a better environment for the absorption of fat-soluble vitamins (A, D, E, and K) and for removal of waste products; the emulsified fats, aided by secondary BAs, become soluble in water due to their interaction with DCA/LCA and form larger micellar structures. This increased water solubility facilitates the removal of waste products along with the digested fats through feces. Moreover, secondary BAs exhibit antimicrobial activity against certain bacteria (e.g., Escherichia coli) in the gut microbiome. However, LCA can potentially be toxic under specific circumstances, such as high doses, pre-existing health conditions, medical conditions (impaired liver or gallbladder function) or gut microbiome imbalances.(Sheng et al., 2022)

In the enterohepatic circulation (Fig. 2), BAs follow the pathway intestine–portal vein–liver–gallbladder–intestine. When fatty foods leave the stomach and enter the small intestine, bile is secreted from the gallbladder. Once the fats are absorbed in the intestine, most of the released BAs are transported back to the gallbladder. However, a fraction of BAs from the intestine is excreted with urine and feces. This cycle typically occurs 6–15 times per day. BAs are primarily reabsorbed in the ileum, with approximately 5% excreted in feces or entering peripheral blood (Browning &

Campos, 2017; Lu et al., 2010). This 5% loss amounts to about 500 mg/day in adult humans and is replaced by newly synthesized BAs (Lu et al., 2010).

Recent studies have linked BA deficiency to constipation, with patients showing reduced faecal BA levels, decreased BA synthesis markers, and increased fibroblast growth factor. This deficiency may result from increased colonic transit time leading to enhanced passive absorption of faecal BAs or a decreased proportion of secretory faecal BAs (Vijayvargiya et al., 2018). Understanding BA physiology and transport mechanisms is crucial for understanding the development of gut homeostasis and metabolic disorders (Halilbasic et al., 2013; Hofmann & Hagey, 2008; Hylemon et al., 2009).

During early life faecal BAs undergo significant changes. In neonates, primary BAs dominate, accounting for 98% of total BAs by day 7 after birth. However, secondary BAs gradually increase from 2% to 90% between 6 months and 6 years of age (J.-J. Xiong et al., 2021). Diet influences BA patterns, with breast-fed infants showing lower concentrations of cholic acid and secondary BAs compared to formula-fed infants. Secondary BAs can appear shortly after birth but their concentrations are low. (Hammons et al., 1988)

2.6 Summary of literature

To date, it is well-established that methodological choices can significantly impact microbiome research outcomes. The microbial activity of bacteria renders faecal samples susceptible to biases in bacterial composition and the levels of microbial metabolites produced and degraded. Practical challenges associated with faecal sample collection and high-throughput laboratory handling further complicate the process. Standardized, cost-efficient methods are still under investigation. While it is not possible to eliminate all methodological biases, it is essential to recognize their effects to address them in interpretation and potentially mitigate them in the future through computational tools

Numerous studies have explored the associations between disease states and microbiome- metabolome profiles. It is evident that disturbances in the microbiome are linked to inflammation in various body parts and organs. Disease onset is typically initiated by an unbalanced microbiome in conjunction with genetic predispositions. Additionally, critical life stages, such as extreme prematurity, have been extensively studied. However, there is a paucity of knowledge regarding the normal developmental trajectories during infancy. Previous research suggests that both prenatal and postnatal environments influence a child's development and future health, as infancy represents a highly sensitive period for growth and development. Various early life factors are known to shape the microbiome, with substantial evidence pointing to the influence of delivery mode, gestational age, early nutrition,

and antibiotic exposure. Few studies have already highlighted that under optimal conditions, the neonate and the microbiota develop in a coordinated manner, influenced by the nutritional, immunological, hormonal, and prebiotic effects of maternal milk.

Based on previous studies, it could be hypothesised that human milk associates with faecal metabolites through the infant gut microbiota. Despite this, faecal metabolites are rarely studied in conjunction with milk metabolites and gut microbiota, meaning the functional potential of microbes is not systematically utilized. The precise mechanisms and trajectories of early life gut maturation remain relatively elusive, although the potential role of human milk oligosaccharides (HMOs) and gut microbial metabolites has been recognized.

This thesis aims to address the knowledge gaps regarding the effects of sample preservatives on downstream analyses of the microbiome and metabolome, as well as to identify an optimal DNA extraction protocol suitable for high-throughput applications. The infant gut microbiome, in conjunction with human milk and faecal metabolites, is rarely studied in human samples, and longitudinal data from large birth cohorts are lacking. Therefore, these exploratory descriptive studies are essential for revealing dynamics of the early life gut microbiome and metabolome maturation.

3 Aims of the Study

The thesis project had two main focuses: to develop reliable preanalytical methods in metagenomics and metabolomics for clinical microbiome samples and to investigate the association and relation between gut microbiome development, faecal metabolites and human breast milk metabolome in early childhood.

The specific aims were:

1. To investigate what is the best practice for automated DNA extraction for downstream sequencing applications and handling of large sample collections in microbiome research and clinical microbiology
2. To establish the validity of an off-site faecal collection method that preserves the integrity of both the metabolome and the microbiome during the freight and until processing
3. To explore the longitudinal patterns of gut metabolites during early life, and how they are related to gut microbiota composition in FinnBrain Birth Cohort Study
4. To investigate associations between the human milk metabolome, infant gut microbiome and faecal metabolome in early childhood in FinnBrain Birth Cohort Study.

4 Materials and Methods

The study materials were sourced from human faecal samples (I-IV), human milk samples (IV) and questionnaire and registry data (III-IV). Thesis consists of three original publications (I-III) and one manuscript (IV) and in which the materials and methods are described in more detail. AI tools were used for R code correction, information retrieval, and text processing for possible grammatical errors according to the guideline on the responsible use of generative AI in research (European Commission, 2024).

4.1 Methodological experiments

For methodological studies (I and II), faecal samples were collected from healthy volunteers. They signed consent form and were informed that samples will be used only for method development studies anonymously without personal information. In study I, samples (n=3) were collected in plastic tubes at home and brought to laboratory within 24h. The original bulk samples were divided into the studied preservative containing tubes in the laboratory. In the study II, samples (n=4) were given next to operating laboratory, since an aliquot for refence method "golden standard" had to be frozen immediately (here in 15-20 min). The rest of the sample was then divided in aliquots by storage temperature, preservative and storage time. In the study II, aliquot tubes were spiked with stability standards.

4.1.1 Study design for optimizing DNA extraction

This work was set out to investigate the impact of human faecal sample collection (preservative), pre-treatment of DNA extraction, and further, choice of 16S rRNA gene hypervariable region to microbiome results. The main focus was to validate the optimal pre-treatment method in a high-throughput manner for DNA extraction. The workflow of optimization is represented in Fig. 3. Two commercial sample preservatives, titled OMNIgeneGUT (DNA Genotek, Canada) and DNA/RNA shield fluid (Zymo Research, USA), were tested for stool sample collection. In addition, four different pre-treatment protocols were tested with all samples including chemical lysis (lysis buffer), bead beating (mechanical lysis) and

incubation with proteinase K. In addition, two 16S rRNA target regions (V3V4 and V4) were compared. All faecal replicates were included in V3V4, while in V4 two replicates of the faecal samples were sequenced.

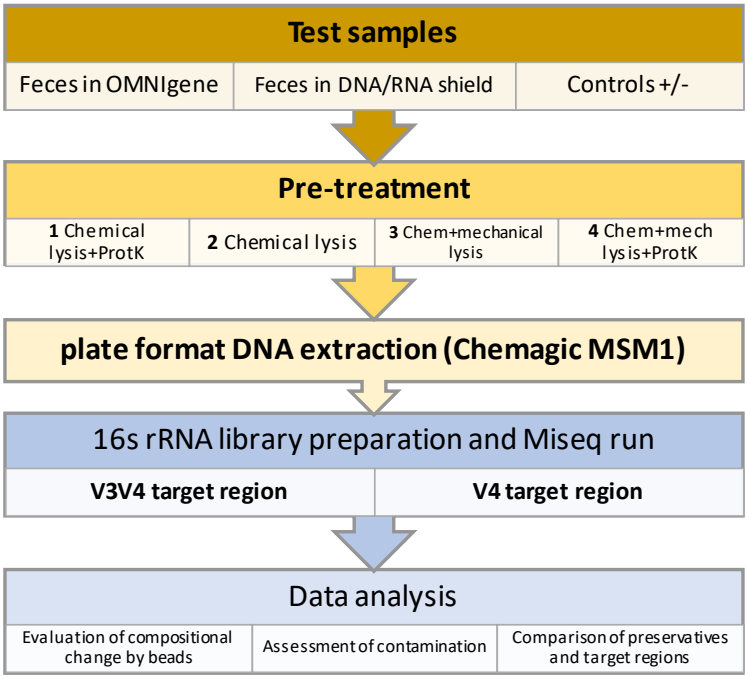


Figure 3. Study design for optimized DNA extraction protocol. Modified from original publication I.

4.1.2 Study design for stability measurements

Faecal samples (n=4) were collected in the morning next to the operating laboratory (Fig. 4). Faecal metabolites and lipids were analysed from a total of 168 faecal samples (aliquoted from four individuals) simulating different conditions of the sample storage matrix: (a) crude feces without any solvent, (b) feces in 95% EtOH, and (c) feces in OMNImetGUT solvent. For each sampling condition, samples were aliquoted from 3 parts of the bulk faecal specimen to represent biological replicates. To investigate the effects of storage time and temperature, initially, one aliquot of the homogenized faecal sample was obtained for immediate freezing at −80°C, which was used as a reference sample (golden standard). Other corresponding aliquots were tested for different durations at following temperatures: 24h at +4 °C (except OMNImetGUT), 24h at room temperature (RT), 36h at RT, 48h at RT, 48h at +4 °C (except OMNImetGUT), and 7 days at RT (Fig. 4).

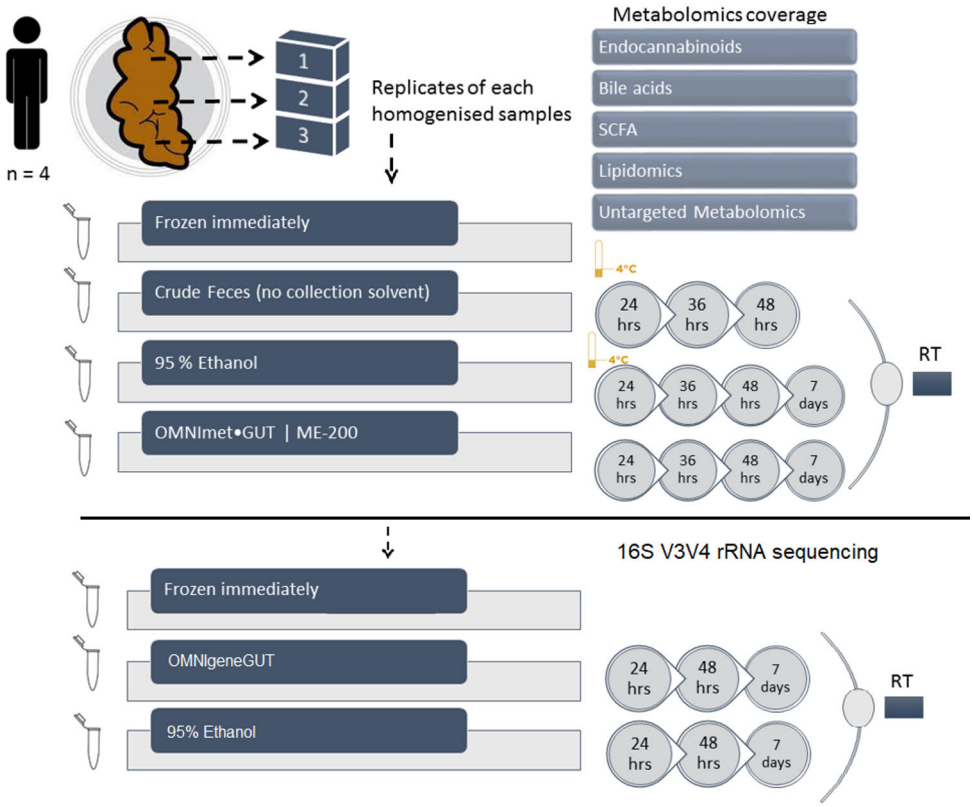


Figure 4. Study design of stability experiment for metabolites and microbiome. Published with the permission of Lamichhane, who drew the figure with Isokääntä.

4.2 Study design for applied cohort studies (III and IV)

The subjects in studies III and IV are children from the FinnBrain Cohort Study, a general population birth cohort based in southwestern Finland (Fig. 5). This cohort recruited families with sufficient fluency in Finnish or Swedish and a normal first trimester ultrasound. A subset of participants took part in study visits, with no exclusion criteria for faecal sample collection. Recruitment occurred between December 2011 and April 2015, while faecal samples were collected from May 2013 to May 2018. Parents collected infant and toddler faecal samples at home, following both written and oral instructions, at 2.5, 6, 14, and 30 months postpartum. The samples were collected in plastic containers without preservatives, and parents were instructed to store them in a refrigerator and bring them to the research facilities within 24 hours. The time of sample collection was also recorded. The longitudinal faecal metabolome was analysed, using both targeted and untargeted methodologies,

and microbiota in infant stool samples collected at 2.5 (n=272), 6 (n=232), 14 (n=289), and 30 (n=157) months (mo) of age.

A subgroup of mothers gave their milk samples and following timepoints of milk samples were included in this study: 2 (n=406), 6 (n=176) and 14 (n=80) months of infant age. The human milk samples were collected during the visits in the presence of a study nurse. Mothers were instructed to breastfeed their infant from the right breast 1.5–2 hours before the study visit, though breastfeeding from the left breast was permitted based on the infant's needs on the same day. Wearing latex gloves, mothers manually expressed 10 mL of foremilk from their right breast into a sterile cup (Pundir et al., 2017). Lower volumes were also accepted. The milk was promptly transferred into plastic tubes, transported to the laboratory, and stored at −70 °C. Prior to processing, thawed milk was gently shaken for 1 minute. (Nolvi et al., 2018)

Clinical data for the study were gathered through parental reports during and after pregnancy at 14, 24, and 34 weeks of gestation, as well as at 3, 6, 12, and 24 months postpartum, and during study visits at 2.5, 6, and 14 months. Additionally, data on maternal pre-pregnancy body mass index (BMI; kg/m²), gestational duration, and mode of delivery (caesarean section vs. vaginal) were obtained from the National Birth Registry, provided by the National Institute for Health and Welfare of Finland (www.thl.fi). Information on maternal perinatal and infant neonatal intravenous antibiotic use was collected from hospital records. Other demographic variables were siblings, primipara, sex, pets and duration of exclusive and partial breastfeeding (BF). Breastfeeding was categorized in two ways: 1) any current BF (yes vs. no); 2) overall adequate breastfeeding of exclusive BF for at least 4 months and partial BF for at least 6 months (breastfeeding criteria, yes vs. no).

Infant secretor status (FUT2) was obtained from genetic SNPs from blood samples as in previous study (Korhonen et al., 2021). Briefly, DNA from cord blood was extracted as instructed by standard procedures at the Finnish Institute for Health and Welfare (THL). Extracted DNAs were genotyped with Illumina Infinium PsychArray BeadChip (Illumina, San Diego, CA) comprising 603132 SNPs at Estonian Genome Centre (Tartu, Estonia), and quality control was carried out with PLINK (www.cog-genomics.org/plink/1.9/). Markers were deleted for missingness (> 5%) and Hardy–Weinberg equilibrium ($P < 1 \times 10–6$). Subjects were checked for missing genotypes (> 5%), relatedness (identical by descent calculation, PI_HAT > 0.2) and population stratification (scaled multidimensionally). Two SNPs of FUT2 gene rs3894326_T (non-functional FUT2 gene) and rs812936_G (non-functional FUT2 gene) were investigated to determine the secretor status of infant. The SNPs were coded as minor/minor=2, major/minor=1 and major/major=0 , meaning that subjects with zero values have normally functioning FUT2 gene (secretors) and other combinations were considered as non-secretors.

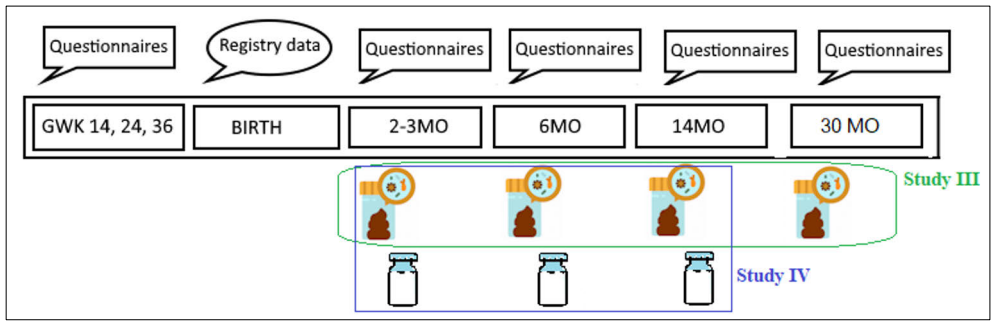


Figure 5. FinnBrain sample collection. Studies III-IV.

4.3 Microbiome methods

In study I, microbial DNA was extracted from human faecal samples (n=3) of different ages (infant, adult and senior) as well as the ZymoBIOMICS Gut Microbiome Standard and negative controls. Extraction method for study I and II was a Magnetic Separation Module I (MSM I) extraction robot with DNA Stool 200 H96 kit (PerkinElmer, Finland).

For Study I, the samples in DNA/RNA shield fluid were stored at a 1:10 ratio of sample to preservative, while the OMNIgeneGUT samples were stored at a 1:4 ratio. Four different sample preparation techniques, including bead beating and chemical lysis, were applied to evaluate the effects of sample lysis and homogenization. Three technical replicates were extracted from adult and senior samples, and two replicates were extracted from the infant sample. In addition, the extraction process included negative controls and the ZymoBIOMICS Gut Standard. Negative controls were placed between faecal samples to check for potential cross-contamination, while the Gut Standard was used to assess the extraction method's sensitivity. Pre-treatment steps were adapted from the manufacturer's protocol, "Purification Protocol for Human Feces Material Using the Chemagic Magnetic Separation Module I." To begin, 800 µL of Lysis Buffer 1 was added to 200 µL of faecal sample, followed by 925 µL of Lysis Buffer 1 added to 75 µL of the ZymoBIOMICS Gut Standard. For negative controls (OMNIgeneGUT and DNA/RNA shield fluid), 800 µL of Lysis Buffer 1 was added to 200 µL of each collection fluid. For the Lysis Buffer control, only 1 mL of buffer was used.

There were four groups of tested pre-treatments: Group (1) included the original protocol of MSM I manufacturer with proteinase K incubations. After adding the lysis buffer, the tubes were vortexed, and 15 µL of proteinase K was introduced. The samples were then incubated in a thermo shaker at 70°C for 10 minutes, followed by a 5-minute incubation at 95°C. The samples were centrifuged at maximum speed for 5 minutes, and the lysate (800 µL) was transferred to a sample plate for further

extraction according to the manufacturer's protocol. Group (2) followed the MSM I manufacturer's original protocol without proteinase K incubation. Group (3) involved bead beating using PowerBead Pro Plates (Glass beads 0.1 mm) and a TissueLyser II (Qiagen, USA). The faecal samples, Gut Standard, negative controls, and lysis buffer were added to the bead plates, which were sealed with a film. The plates were shaken at 15 Hz for 5 minutes twice in the TissueLyser II. Afterward, the plate was centrifuged at 4,500 × g for 6 minutes, and 800 µL of the lysate was transferred to a clean plate for continued extraction as per the manufacturer's protocol. Group (4) combined bead beating with proteinase K incubation. After bead beating, the plate was centrifuged, and 800 µL of the lysate was transferred to 2 mL screw cap tubes containing 15 µL of proteinase K. The tubes were mixed and incubated as in Group 1. Following a brief spin, the lysate was transferred to a clean sample plate, and isolation was completed according to the manufacturer's protocol.

The concentration of isolated DNA was measured using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA) with a Qubit dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific, USA). DNA integrity was assessed via 1% TBE agarose gel electrophoresis. The DNA was then divided into two 100-µL aliquots and stored at −80°C. DNA extraction and subsequent analyses were conducted using DNase-/RNase-free plasticware.

In the study II, the above-described extraction protocol was used with pre-treatment group 4. The samples were distributed into aliquots of three sample types: i) feces with no preservative (crude feces), ii) feces in EtOH 95%, and iii) feces with OMNIgeneGUT in the same ratio as earlier mentioned.

Infant faecal samples (study III and IV) were extracted earlier with former extraction protocol. The samples were transferred into cryotubes and stored at -80°C within 2 days of arrival at the laboratory, with samples being kept at +4°C prior to freezing. Only samples that were frozen within 48 hours of collection were included in the studies. Approximately 100 mg of each sample was used for DNA extraction, and the masses were recorded. To each sample, 1 mL of lysis buffer was added, and the samples were homogenized with glass beads at 1000 rpm for 3 minutes using a PowerLyzer (Qiagen). The samples were then centrifuged at high speed (> 13,000 rpm) for 5 minutes. The lysate (800 µL) was transferred to clean tubes, and the extraction proceeded according to the manufacturer's protocol using a semi-automatic extraction instrument, GenoXtract, with the DNA stool kit (HAIN Life Science, Germany) and later with corresponding spare instrument Arrow (NorDiag, Norway).

In study I, NGS libraries were amplified with V3V4 and V4 target regions to see the influence of the region and PCR protocol. In study II, V3V4 target region was used with corresponding protocol. NGS libraries used in infant studies (III and IV), were amplified with V4 target region, since V4 was found to detect Bifidobacterium

better than V3V4 in previous in-house optimization. The sequence libraries were constructed according to the Illumina library preparation protocol (*Illumina*, 2014) and V4 library with an in-house protocol (Rintala et al., 2017).

In the V4 library preparation, amplicon PCR and index PCR were united (Rintala et al., 2017). The DNA was diluted with PCR-grade water to a concentration of 10 ng/μL before PCR. PCR amplification was carried out using the KAPA HiFi High Fidelity PCR kit with dNTPs (Roche, USA). The final concentrations of each component were as follows: 1x for the 5x KAPA HiFi Fidelity Buffer, 0.3 mM for the dNTP mix, 0.5 U for KAPA HiFi DNA polymerase, and PCR-grade water. The sequences of the forward and reverse primers (at 0.3 μM each) were: 5′- AATGATACGGCGACCACCGAGATCTACAC -i5- TATGGTAATT-GTGT GCCAGCMGCCGCGGTAA-3′ (forward) and 5′- CAAGCAGAA-GACGGCATACGAGAT -i7- AGTC AGTCAG-GCGGACTAC-HVGGGTWTCTAAT-3′ (reverse), with i5 and i7 representing sample-specific indices. The template DNA used was 50 ng, and the final reaction volume was 25 μL. The combined PCR conditions were as follows: an initial denaturation at 98°C for 4 minutes, followed by 30 cycles of denaturation at 98°C for 20 seconds, annealing at 65°C for 20 seconds, and extension at 72°C for 35 seconds, ending with a final extension at 72°C for 10 minutes.

The V3V4 protocol differed from Illumina's suggestion of the PCR final volumes and the DNA visualization. Prior to PCR, the DNA templates were diluted to 2.5 ng/μL in PCR-grade water. In short, amplicon PCR included 2x KAPA HiFi HotStart ReadyMix (Roche, USA), Illumina amplicon forward and reverse primers (6.6 μM), PCR-grade water, and microbial DNA (16.5 ng). The final volume of the amplicon PCR was 33 μL and index PCR was carried out following the instructions provided by Illumina.

After PCR, 8 μL of the quality of product was analysed with 1.5% TAE agarose gel (120V, 1h). The concentration of the library samples was determined using a Qubit Fluorometer with a Qubit dsDNA High Sensitivity Assay kit. The 4 nM library pool was denatured, diluted to a concentration of 4 pM, and 8% denatured PhiX control (Illumina, USA) was added. The V3V4 library samples were sequenced using a MiSeq Reagent Kit v3, 600 cycles (Illumina, USA), on a MiSeq system, generating 2 × 300 base pair (bp) paired-end reads, in accordance with the manufacturer's instructions. The V4 library samples were sequenced in similar manner with 2 × 250 bp paired ends. A positive control (plasmid DNA 7-mock) and a negative control (PCR-grade pure water) were along in library preparation to control the PCR. Studies I and II had also positive and negative controls of DNA-extraction.

4.4 Metabolomics

Metabolomics methods were used in studies II-IV. **In the study II**, the aliquot tubes were spiked with stability standards (sodium butyrate-13C4 5 parts per million (ppm), cholic acid-24−13C 2,5 parts per billion (ppb), palmitic acid (1−13C, 99%) 5 ppm, hippuric acid-d5 5 ppm, indole-2,4,5,6,7-d5−3-acetic acid 5 ppm, nicotinamide-d4 5 ppm, sucrose-1−13Cfru 5 ppm, L-tyrosine-d4 5 ppm, and cortisol-1,2-d2 5 ppm) and dried using a SpeedVac (Thermo Fisher Scientific) prior the sample collection. For the aliquots, 150 ± 10mg of stool was weighed to each spiked tube. Next, storage fluid was added to the prepared tubes at a ratio of 1:4 (600 µL of OMNImetGUT or 95% ethanol). To facilitate homogenization, 100 µL of ultrapure water was added to samples without preservative. All aliquots were homogenized using a bullet homogenizer (Next Advance) at medium speed for 2 minutes. During sample processing, priority was given to freezing one aliquot per biological replicate immediately. These aliquots were frozen within 15–20 minutes of defecation and are referred to as the "golden standard." Following initial processing, aliquots without preservative (the crude sample), with OMNImetGUT solvent, and with 95% ethanol were stored at room temperature (RT) for 24, 36, 48 hours, and 7 days. Additionally, aliquots containing 95% ethanol were stored at +4°C for 24, 36, 48 hours, and 7 days. The crude samples were kept at +4°C for 24h. Additional set of crude sample aliquots were freeze-dried to measure the water content of samples prior to extraction to determine the dry weight for later normalization.

In the study III, aliquots of faecal samples were freeze-dried in the same manner mentioned above. Another set of aliquots was homogenized by adding ultrapure water (10 µL per mg of dry weight) and ceramic beads to wet feces. The same homogenizer was used as in study II. Bile acids were extracted by adding 40 µL the faecal homogenate to crash solvent 400 µL (methanol containing 62,5 ppb) including each of the internal standards LCA-d4, TCA-d4, GUDCA-d4, GCA-d4, CA-d4, UDCA-d4, GCDCA-d4, CDCA-d4, DCA-d4 and GLCA-d4) and filtering them using the filter plates mentioned above. The filtered samples were dried under a gentle flow of nitrogen and resuspended using 20 µL resuspension solution (methanol : water 40:60 with 5 ppb Perfluoro-n-[13C9] nonanoic acid as in injection standard). Quality control (QC) sample was created by pooling an aliquot of each sample into a tube. Next, QC was vortexed it and prepared the same way as the other samples. Blank samples only included 400 µL crash solvent and they were filtered, dried and resuspended as the other samples. Calibration curves were generated by pipetting 40 µL of standard dilution into vials, adding 400 µL of crash solution, and then drying and resuspending the samples in the same manner as the other samples. The standard dilution concentrations ranged from 0.0025 to 600 ppb. LC separation was carried out on a Sciex Exion AD 30 (AB Sciex, Framingham, MA) LC system,

which included a binary pump, an autosampler set at 15 °C, and a column oven set at 35 °C. A Waters Aquity UPLC HSS T3 column, paired with a pre-column of the same material, was used for separation. The flow rate was set to 0.5 mL/min, and the injection volume was 5 µL. The used mass spectrometer instrument was a Sciex 5500 QTrap, operating in scheduled multiple reaction monitoring mode in negative mode. Data was pre-processed using Sciex MultiQuant.

Quantification of SCFA

Analysis for the targeted SCFA was adapted and modified from previous work (Trimigno et al., 2017). Analysis of SCFAs was conducted on faecal homogenate (50 µL) that was mixed with 500 µL of methanol containing internal standards (propionic acid-d6 and hexanoic acid-d3 at 10 ppm). The samples were vortexed for 1 minute and then filtered using a 96-well filter plate as previously described. A retention index (RI) composed of 8 ppm C10-C30 alkanes and 4 ppm 4,4-Dibromooctafluorobiphenyl in hexane was added to the samples. Gas chromatography (GC) separation was conducted on an Agilent 5890B GC system with a Phenomenex Zebron ZB-WAXplus column, along with a short blank pre-column of the same dimensions. Samples (1 µL of each) were injected into a split/splitless inlet at 285°C using a 2:1 split ratio via a PAL LSI 85 sampler. Mass spectrometry was carried out on an Agilent 5977A MSD, with mass spectra acquired in Selected Ion Monitoring (SIM) mode.

Analysis of polar metabolites

Polar metabolites were extracted using methanol, following a method adapted from a previous study (Lamichhane et al., 2018). Faecal homogenate (60 µL) was mixed with 600 µL of methanol crash solvent containing internal standards (heptadecanoic acid at 5 ppm, valine-d8 at 1 ppm, and glutamic acid-d5 at 1 ppm). After precipitation, the samples were filtered using the same filter plates as previously described. One aliquot (50 µL) was transferred to a shallow 96-well plate to create a QC sample, while the remaining filtered sample was dried under a gentle stream of nitrogen and stored at -80°C until further analysis. After thawing, the samples were re-dried to remove any residual water. Derivatization was performed using a Gerstel MPS MultiPurpose Sampler before injection, with automatic derivatization controlled by Gerstel Maestro 1 software. GC separation was performed on an Agilent 7890B GC system equipped with an Agilent DB-5MS column. A 1 µL sample was injected into a split/splitless inlet at 250°C in splitless mode. The mass spectrometry (MS) was carried out on a LECO Pegasus BT system (LECO). Data acquisition was done with ChromaTOF software. The samples were analysed in 9 batches, each consisting of 100 samples and

a calibration curve. To monitor the run, a blank, a QC sample, and a standard sample with a known concentration were included after every 10 samples. Between batches, the septum and liner of the GC were replaced, the precolumn was trimmed whenever necessary, and the instrument was primed.

The retention index was determined using ChromaTOF with the reference method function. A reference file was formed with the spectra and approximate retention times of alkanes from C10 to C30, using manual determination. A reference method was applied to each sample to accurately determine the retention time of the alkanes. Untargeted data processing was performed with MSDIAL. Identification was carried out based on retention index, utilizing the GCMS DB-Public-kovatsRI-VS3 library available on the MSDIAL website. The results were exported as peak areas and further processed in Excel, where they were normalized using heptadecanoic acid as the internal standard. Features with a coefficient of variation (CV) of less than 30% in the QC samples were selected. Additional filtering removed alkanes and duplicate features. The identified features that passed the CV check were then cross-verified using the Golm Metabolome Database.

In the study IV, in addition to previously measured faecal metabolites, human milk samples were analysed with ^{1}H nuclear magnetic resonance (NMR). The samples order was randomized prior the NMR. NMR based metabolomics were processed using standard protocol for milk-based metabolomics (Sundekilde et al., 2016). The samples were gently thawed in a water bath and kept on ice while the Amplicon Ultra 0.5-ml 10-kDa spin filters (Millipore, Billerica, MA, USA) were washed three times. Skimmed milk was obtained by centrifuging the samples at 4,000 g and 4°C for 10 minutes, followed by removal of the fat layer. Subsequently, 500 µL of the skimmed milk was transferred into individual Amplicon Ultra 0.5-ml 10 kDa spin filters. The samples were then centrifuged at 10,000 g and 4°C for 30 minutes. After centrifugation, 400 µL of the supernatant milk from each sample was transferred into separate 5-mm NMR tubes. To each tube, 200 µL of D2O containing 0.05% 3-(trimethylsilyl)propanoic acid (TSP, Sigma-Aldrich, Saint-Louis, MO, USA) was added. Spectra were acquired following the protocol described in a previous study (Sundekilde et al., 2016).

1H-NMR spectra were obtained using a Bruker Avance III 600 MHz spectrometer, fitted with a 5-mm 1H TXI probe (Bruker BioSpin, Rheinstetten, Germany), at a temperature of 298 K and a 1H frequency of 600.13 MHz. A single 90° pulse experiment (Bruker pulse sequence: noesypr1d) was performed to reach one-dimensional spectra, with a relaxation delay of 5 seconds. During the relaxation delay, water suppression was applied. A total of 64 scans were collected, with 32,768 data points and a spectral width of 12.15 ppm. The resulting 1H-NMR spectra were referenced to the TSP signal at 0 ppm. A line-broadening function of 0.3 Hz was applied to each spectrum before Fourier transformation. Pre-processing, including

phase and baseline corrections, was carried out both automatically and manually using Topspin 3.2 (Bruker Biospin, Germany).

Milk metabolite data consists of 44 compounds. Data was only partly normal distributed and therefore data was log-transformed. Milk metabolites were grouped into five groups based on their chemical classification: HMOs, Energy metabolism, Amino acids & Protect nutrients, Bacterial fermentation and Lipids & fatty acids. Milk samples represent a snapshot of compound concentrations and therefore normalization was needed. Normalization for milk metabolites was done by lactose content since the lactose is relatively stable in mature milk and is correlated with milk volume (Dror & Allen, 2018) and has previously been used for normalization (Sundekilde et al., 2016). Mother´s secretor status of 2'-fucosyllactose (FUT2) was determined by the concentration of 2'-FL in milk. Infant secretor status was determined by the FUT2 gene snips from blood sample as mentioned earlier (4.2)

4.5 Data modulation and statistical methods

In all four studies (Fig. 6), sequence data was exported as fastq-files from the sequencing instrument. P-values (two-tailed) smaller than 0.05 were interpreted as statistically significant. In studies I and II, the raw sequence data was processed and analysed with a CLC Microbial Genomics Module (Qiagen, USA) which complies with QIIME2 (Bolyen et al., 2019). In studies III and IV, all analyses were performed in R, and dada2 pipeline was used for pre-processing and creating ASV-table.

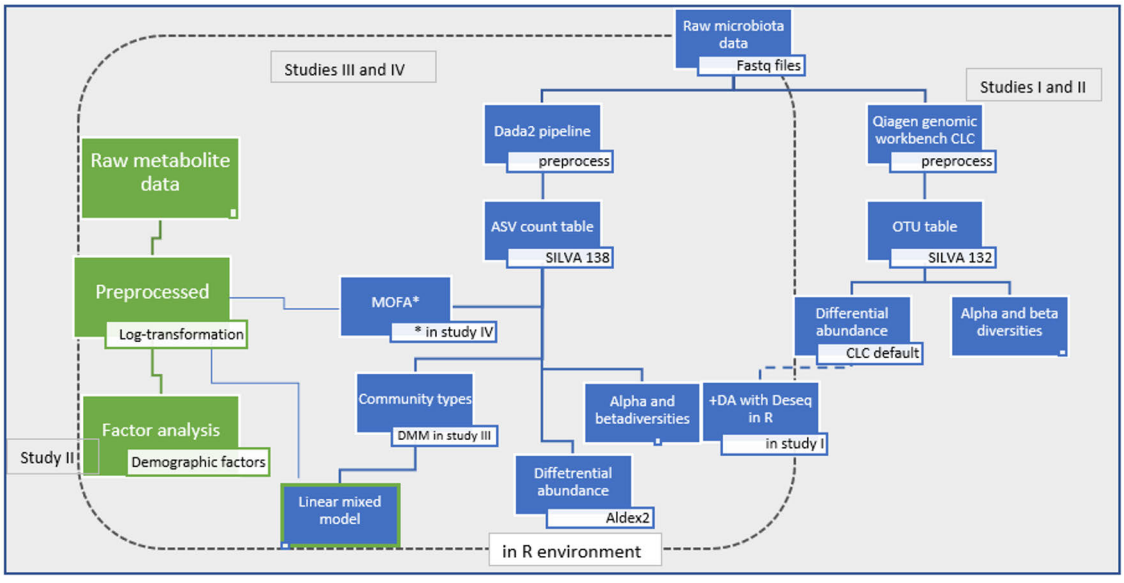


Figure 6. Simplified flow chart of data analysis and statistical methods for studies I-IV.

In the study I and II, the data was analysed in CLC with default settings by the workflows namely "Data quality control (QC) and operational taxonomic unit (OTU) clustering" including read trim and "Alpha and beta diversities". Merged read pairs were used. The cut-off number of reads for the faecal samples was 200,000 sequences, which ensures sufficient sequencing depth for reliable taxonomic profiling. This was cut-off was not used for negative controls. One replicate of infant sample (study I) from V3V4 sequencing was excluded based on the low number of reads. Both libraries had an index and adapter trims from the 5′ end. Database of SILVA 16S version 132 with a 97% similarity for OTU clustering was used for sequence mapping. In the diversity analysis, low abundance OTUs (<100 reads) were excluded to remove noise and sequencing errors and improve statistical power. OTUs were adjusted using MUSCLE (MUltiple Sequence Comparison by Log-Expectation). Alpha diversity was calculated by a neighbour-joining tree. In the study I, Chao1 and observed OTUs (the total OTU number) were chosen to display alpha diversity. Alpha diversity of the total OTU number represents richness meaning the number of species observed in each sample while Chao1 estimates the total richness that accounts for unobserved species (Deng et al., 2024). Two-sample t-test was used to test if there is a difference in alpha diversity between pre-treatment groups. Analysis of beta diversity was carried out using the principal coordinate analysis (PCoA) with Jaccard and Bray-Curtis. OTU tables and diversity indices were exported to a RStudio and GraphPad Prism for graphics. Differential abundances analysis (DAA) was done with both CLC Microbial Genomics Module and Deseq2 (Love et al., 2014) in R environment to amplify the findings. The reported DAA results are consensus of the two performed methods.

In the study II, OTU tables were exported from CLC with same settings as in study I and downstream analysis were done in R Bioconductor ecosystem (R version 4.2.3). Dissimilarities in microbiota (β-diversity) by storage condition were estimated using Bray−Curtis, distance-based redundancy analysis and PERMANOVA. Alpha diversity was calculated by the Shannon index, number of observed OTUs, Chao1 and the Simpson Index. (Borman et al., 2024) All those indices were included in the intraclass correlation (ICC) testing. Difference in the Shannon index by storage condition was estimated using two-sample t-tests, with equal variance as determined by Levene's test. For metabolite data, all multivariate statistical analyses were performed with log2-transformed intensity data, which were auto-scaled before analysis to enhance global interpretability. As the water content of faecal samples vary, the measured metabolites were normalized to the dry weight of the feces. The PLS Toolbox (Eigenvector Research Inc., Manson, WA) in MATLAB 2017b (MathWorks, Natick, MA) was used for multivariate analysis. ASCA, a multivariate extension of ANOVA, was performed to interpret the variation caused by different factors, including time, individual differences, and collection matrix. For univariate analysis, the level of each metabolite in each sample storage method (crude feces, EtOH 95%

and OMNImetGUT) was compared to the level of the same metabolite in the paired immediately frozen sample (gold standard). To describe group differences, the fold changes were represented. And those were calculated by dividing the mean concentration of an observed metabolite in one group by another. A multivariate linear model with the MaAsLin2 package in R was utilized to test group differences accounting for random effects within individual samples or subjects. Due to multiple comparisons, the Benjamini and Hochberg method were used to correct nominal p-values. Adjusted p-values <0.25 (q-values) were considered significant.

In the study III, the data analyses were performed with R with packages including phyloseq, mia, vegan, DirichletMultinomial, lme and pheatmap. ASVs was used instead of OTUs. Alpha diversity was estimated by the Shannon index and Inverse Simpson using the mia package from the untransformed ASV-table. Metabolite data was log2-transformed with a pseudocount (minimum value /2). The Dirichlet Multinomial Mixtures (DMM) model, using rarefied counts in genus-level, was employed for microbiome clustering by community types, which were determined by the Laplace criteria. Factor analysis was conducted to estimate the relative contribution of clinical and demographic factors to the total variance of metabolite classes using a fitted linear regression model. The total metabolite concentrations of a specific metabolite class were regressed against a clinical/demographic factor of interest, and the median marginal coefficient of determination (R^2) and percentage of explained variance were calculated. The 'Scater' package in R was used for this factor analysis. Statistical analyses included the Wilcoxon test, Chi-square test, and Kruskal-Wallis test with Dunn's posthoc test. Spearman correlation was used for network analysis. Linear mixed models, with child ID as a random effect and sampling age as a fixed effect, were used to study: a) metabolite age-trends, b) associations between metabolite concentrations and demographic factors, c) associations between microbiome community type membership and demographic factors, d) associations between metabolite concentrations and community types, and e) associations between metabolite concentrations and the interplay with breastfeeding and rclr-transformed prevalent genera abundances, as BF was found to promote the development of microbiome in preliminary analysis. The mixed model was performed using the nlme package. Model singularity was checked with the lme4 package. The ALDEx2 package had the clr-module, which was utilized for differential abundance analysis (Fernandes et al., 2013). Explained variances in the metabolome assays by demographic factors were computed using the Scater package (McCarthy et al., 2017). As in study II, p-values were adjusted using the Benjamini-Hochberg procedure.

In Study IV, milk metabolites were log2 transformed with a pseudocount of half the minimum values. The dissimilarity of the milk metabolome was analysed using Euclidean distance. Principal Components Analysis (PCA) was used to reduce the dimensionality of the milk metabolites, and the first three principal components

(PC1-PC3) were used in the PERMANOVA analysis with microbiome abundance assay as the dependent variable. PCA was conducted using the prcomp function, and PERMANOVA was performed using the adonis2 function, calculating both marginal and by-terms effects.

Differential abundance analysis (DAA) was performed using the Wilcoxon signed-rank test between genus-level presence-absence data and milk metabolites. Genera with at least ten occurrences in both groups were included. P-values were adjusted for multiple testing using the FDR method. The analysis was stratified by timepoints (2.5, 6, and 14 months). Additionally, DAA was conducted with all genera using Aldex2, which uses probabilistic modelling and considers the count-compositional nature of the data.

Associations between each milk metabolite and each faecal metabolite were examined using Spearman's correlation. The 95% bias-corrected and accelerated (BCa) bootstrap confidence intervals were calculated for the correlation coefficient (based on 1000 bootstrap samples). The analysis was stratified by timepoints (2.5, 6, and 14 months).

Linear mixed effects models were used to examine the longitudinal association between faecal metabolites and milk metabolites:

M:faecal metabolites~intercept+maternal BMI+birth month+parity+gestational age +maternal distress+delivery mode+maternal education+age terms×milk metabolites

Faecal metabolites (untargeted polar metabolites, SCFA, BA) and milk metabolites were used in the model one at a time. Interaction between each milk metabolite and the child's age at the sampling time was included. Covariates included breastfeeding (partial, full), parity (primipara, multipara), delivery mode (vaginal, section), and maternal education (low, mid, high). Continuous variables were maternal BMI, child's birth month, gestational age, and maternal distress. Maternal distress was calculated as the sum of scaled mean EPDS and SCL scores at various timepoints. Missing values were handled by calculating the mean sum score based on available data. Age terms refer to segments of the piecewise linear function modelling the child's age at sampling. Covariates were selected based on expected causal relationships. The breakpoint was at 6 months. Milk and faecal metabolites were scaled (mean=0, sd=1).

Multiomic factor analysis was conducted using the MOFA2 function to discover principal sources of variation in multi-omics datasets. Milk and faecal metabolites and genus-level microbiome data were used in MOFA2 simultaneously. Only prevalent genera were included (detection=0.01, prevalence=0.1). Genera were examined using the centered log ratio-transformation (CLR) with a pseudocount of one. Timepoints were treated as groups to compare sources of variability driving each timepoint. Table 3 summarizes the data modulation and analyses of the thesis.

Table 3. Summary of data analysis.

	Study I	Study II		Study III		Study IV		
Raw data	Fastq files	Fastq files	Mass spectrum for metabolites	Fastq files	Mass spectrum for metabolites	Fastq files	Mass spectrum for fecal metabolites	NMR spectrum for milk metabolites
Preprocess	Qiagen CLC for OTUs	Qiagen CLC for OTUs	Normalization by dry weight, log2 transf, auto-scaled	Dada2 in R for ASV	Normalization by dry weight, log2 transf	Dada2 in R for ASV	Normalization by dry weight, log2 transf	Normalization by lactose conc., log2 (+minimum/2)
Database	Silva132	Silva132		Silva138		Silva138		
Alpha diversity	Observed OTUs and Chao1 +two sample t-test by pre-treatment	Shannon, observed OTUs, Chao1 and Simpson and two sample t-test by preservation		Shannon		Shannon		
Beta diversity	Bray Curtis and Jaccard	Bray Curtis, RDA, PERMA-NOVA, Jaccard, UniFrac, and Unweighted UniFrac	ASCA	DMM and Bray Curtis		Bray Curtis		Euclidean, PERMANOVA between milk PCA and microbiota
Differential abundance	CLC and Deseq2			ALDEx2		Wilcoxon signed-rank test between genus-level presence-absence data and milk metabolites with FDR, and ALDEx2		
Other analyses		Intraclass correlation (ICC) testing including all the diversity indices and three abundant genera	A multivariate linear model using the MaAsLin2 taking into account random effects within an individual sample or subject	Linear mixed model (Child ID= random effect, sampling age =fixed effect), to study: a) metabolite age-trends, b) associations between metabolite concentrations and demographic factors, c) associations between community type and demographic factors, d) associations between metabolite concentrations and community types, and e) associations between metabolite concentrations and the interplay with breastfeeding and rclr-transformed prevalent genera abundances		Linear mixed model: faecal metabolites~intercept+maternal BMI+birth month+parity+gestational age +maternal distress+delivery mode+maternal education+age terms×milk metabolites. Multiomic factor analysis(MOFA): genera using CLR with a pseudocount of one, timepoints as groups. Spearman correlation for milk and faecal metabolites	
FDR		Benjamini-Hochberg		Benjamini-Hochberg		Benjamini-Hochberg		

4.6 Ethical considerations

Faecal samples for method development (studies I and II) were received from anonymous faecal donations of healthy individuals, and they were asked to sign a research consent form. Research using anonymized biological material is not restricted by the Finnish legislation and ethical committee in University of Turku on research involving human beings. Consequently, an ethical review statement was not required.

For FinnBrain studies (III and IV) ethical issues have been considered and there is a research permit for the project from VARHA ethical committee (ETMK: 57/180/2011), which has approved Cohort profile and research protocol (Karlsson et al., 2018). The cohort parents have signed a consent form regarding their children's participation in research and given permission to use their samples for scientific purposes. Samples were handled with pseudonymized ID code during the laboratory processes and all methods were performed in accordance with Helsinki declaration (Resneck, 2025).

5 Results

5.1 Optimized high-throughput DNA extraction method for covering different bacteria types and avoiding contaminants

All four tested pre-treatment methods produced sufficient concentration of DNA (>7 ng/μL) in the faecal samples. For adult samples, the highest concentration was achieved in the non-bead-beating group 2 with both stabilizers. In senior samples collected in OMNIgeneGUT, the highest concentration was observed in group 4 (bead beating and proteinase K). For infant samples, the highest concentration was obtained without bead beating (groups 1 and 2). Adult faecal samples showed the lowest standard deviation (SD) in concentrations across pre-treatment groups, while infant samples had the highest SD. OMNIgeneGUT preservation resulted in slightly higher concentrations for senior and infant samples compared to DNA/RNA shield fluid, but there was no notable difference for adult samples. The integrity of the DNA isolates in gel electrophoresis appeared high, since all faecal samples had visible amounts of intact DNA. However, senior samples in OMNIgeneGUT showed some fragmentation and also had the highest DNA concentrations among all samples. Extraction controls (EC) were pure in the gel, however, ECs among other negative controls showed traces of cross-contamination from faecal samples in sequence reads.

5.1.1 Differential abundances and higher alpha diversity by bead beating

The relative abundances of adult samples were relatively similar for both sequencing targets. Bead beating (groups 3 and 4) introduced signatures from hard-to-lyse bacteria, primarily gram-positive genera such as *Blautia, Bifidobacterium*, and *the Ruminococcus* torques group. Samples preserved with DNA/RNA shield fluid had a higher abundance of *Faecalibacterium*, while those preserved with OMNIgeneGUT showed a higher abundance of Bacteroides. The impact of bead beating was less noticeable in samples with DNA/RNA shield fluid. The relative abundances in senior samples were fairly consistent within the same sequencing target; however, there were differences between the V3V4 and V4 regions. *Klebsiella* was absent in V4 sequenced samples, and

Enterobacter was missing in V3V4 sequenced samples. In the infant samples, V3V4 sequencing favoured genera *Bacteroides and Klebsiella*, whereas *Veillonella and Enterobacter* had higher abundance with V4. Minor increase was detected in the abundances of *Enterobacter and Pantoea* genera in bead groups (3 and 4) in the V4 sequencing. Within the same sequencing target, the profiles were repeatable. The target region had a high impact on the infant sample. While the profiles were dominated with the same bacteria, the ratios differed greatly between the V4 and V3V4.

Alpha diversity levels (Fig. 7) varied by pre-treatment method; in adult and senior samples, the bead-beating groups (3 and 4) had higher Chao indices compared to the non-bead-beating groups (1 and 2), with significant differences (adult $p = 0.0003$, senior $p = 0.0017$). The highest alpha diversity in adult samples was observed in group 3, while in senior samples, it was highest in group 4. Both bead-beating groups showed greater variation compared to the non-bead-beating groups in adult samples. For infant samples, on average, the bead-beating group 3 produced the highest diversity, although the deviation between replicates was relatively high. Overall, the effect of pre-treatment was smaller in the infant samples and the Gut Standard compared to the adult and senior samples in the compositional diversity.

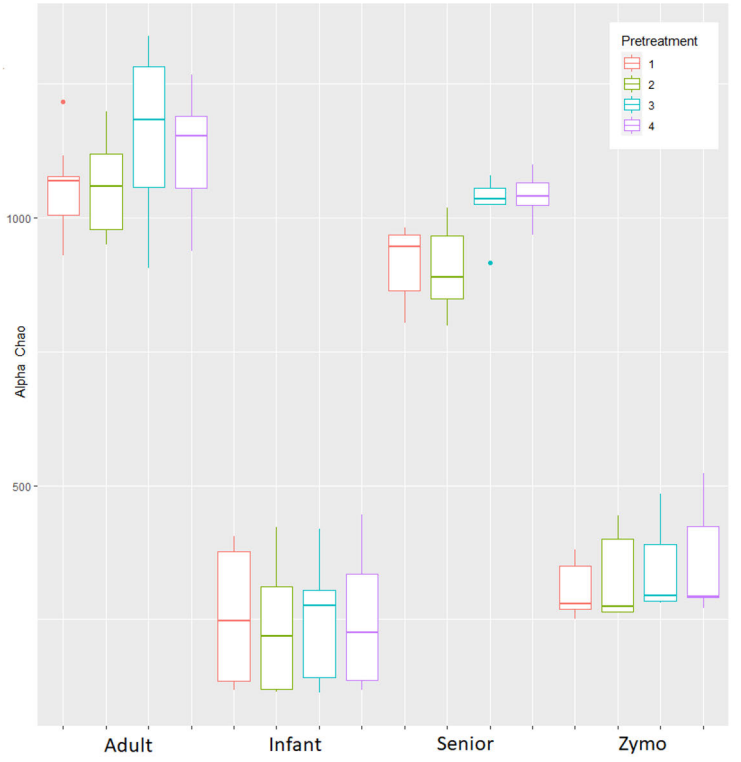


Figure 7. Alpha diversity (Chao1 index) by pretreatment method in faecal samples and gut standard (Zymo). Modified from original publication I.

Altogether, bead beating increased the abundances of several gram-positive bacteria using differential abundance analysis (DAA) with all faecal samples together (Fig. 8). The differences were significant in genera present in fig 8 (FDR-P ≤ 0.05). However, DAA has limitations, especially given the high inter-individual variation in the microbiome, which may explain the high variability in abundances and prevalence. Figure 8 displays the genera identified by both the CLC and Deseq2 methods.

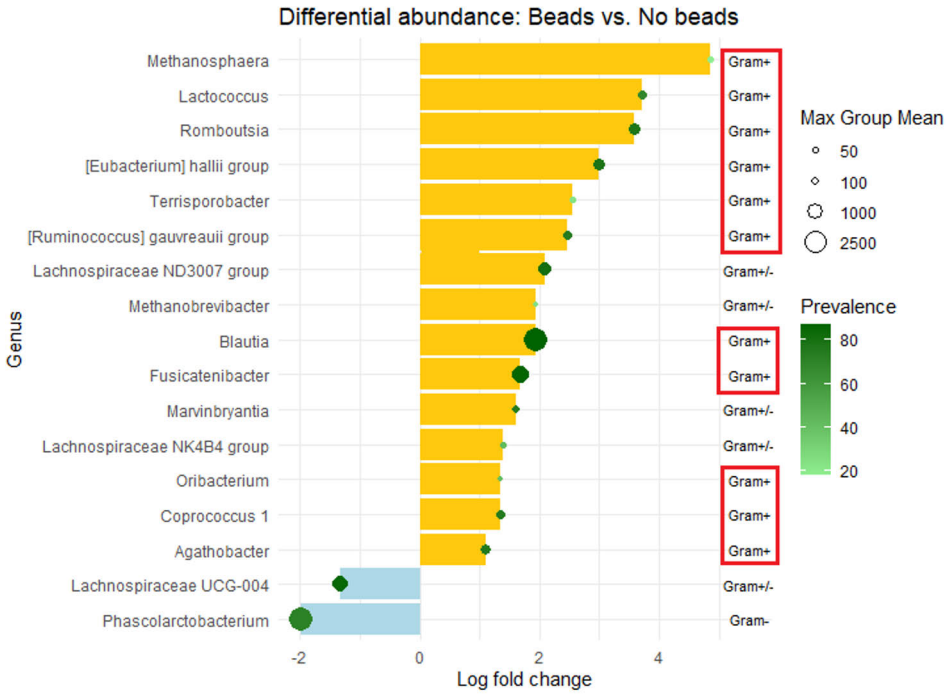


Figure 8. Differential abundances of bacterial genera by bead beating. Modified from original publication I.

5.2 Sample preservative for stability of gut metabolites and microbiota

The impact of sampling factors to the faecal metabolome was investigated by analysis of variance (ANOVA)-simultaneous component analysis (ASCA) and it was performed with the following factors: subjects; sample storage matrix (crude, 95% EtOH, OMNImetGUT, immediately frozen); time (24, 36, 48h, 7 days); and temperature (RT, +4 °C). Interindividual differences (Model Effect (%) 71.15, 65.10, 77.70, p = 0.0010) and sample storage matrix (Model Effect (%) 8.2, 9.6, 2.5, p = 0.0010) were found to have the highest impact on BAs, SCFAs, and lipids in faecal

profiles. In comparison, significant influence of the duration of storage (Model Effect (%) 2.1, 2.0, 1.7, $p > 0.05$) and storage temperature (Model Effect (%) 1.60, 1.09, 1.34, $p > 0.05$) were not found. As expected, in crude samples RT storing was found to increase the concentration of SCFAs (Fig. 9A). In particular, butyric acid, isobutyric acid, and valeric acid increased over 1.5-fold in raw feces at RT from 24h until 48h ($q < 0.25$). Ethanol storing was comparable to OMNImetGUT fluid and both preservatives had minor change compared to immediately frozen samples with exception of tauro- and/or glycoconjugated bile acids (GLCA, THDCA, GCDCA), lipids (mainly TGs), and unknown polar features increased over time ($p < 0.05$) and those differed in one to three of sample storage matrices. However, none of these metabolites reached the significance level at the chosen FDR threshold of 0.25 (Fig. 9B).

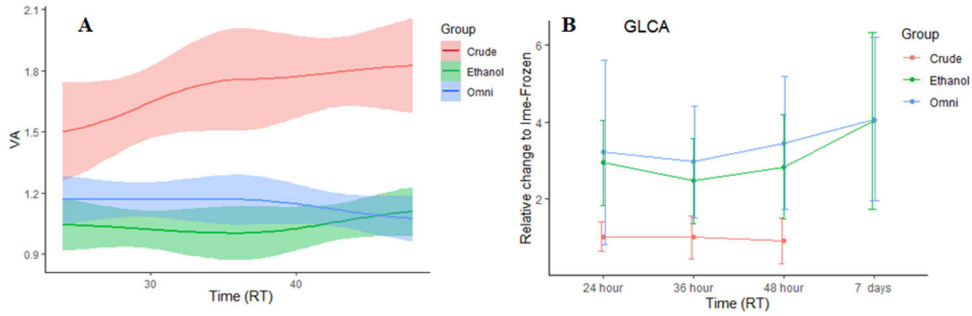


Figure 9. Alterations in metabolites during storage at room temperature. The change in the level of (A) Valeric acid over time (24, 36, and 48h) and (B) Glycolithocholic acid over time (24, 36,48h and 7days) in faecal samples collected as crude, in 95% EtOH, and with OMNImetGUT solvent. The X-axis shows the sample storage duration The Y-axis shows the relative change of each metabolite compared to the gold standard (immediately frozen). Modified from original publication II.

Likewise, interindividual differences were found to have a dominant effect on microbiome alpha (α) diversity (Fig. 10). Both storage solvents led to lower α diversity in comparison to immediately frozen samples. A significant decrease in α diversity was also observed in the longitudinal sample series collected in 95% EtOH. Figure 9 shows that α diversity decreased over time in 95% EtOH compared to OMNIgeneGUT ($p = 0.007$). Similarities between storage types were compared using intraclass correlation coefficients (ICC) with immediately frozen samples as the reference. Similarity in alpha diversity metrics (Shannon, Simpson, Chao1, number of observed OTUs) and the three prevalent genera (*Bacteroides, Bifidobacterium, and Faecalibacterium*) were compared between storage types. OMNIgeneGUT showed higher similarity in alpha diversity with immediately

frozen samples than 95% EtOH samples. Next, differences in overall gut microbiota composition (beta diversity) between samples collected in 95% EtOH and OMNIgeneGUT were analysed. PCoA with Bray-Curtis dissimilarity showed minor differences visually between storage types and time points (Fig. 11). PCoA with Jaccard, UniFrac, and Unweighted UniFrac gave similar results. However, no significant difference in β diversity was found between storage types. Same finding was with distance-based redundancy analysis with Bray-Curtis dissimilarity (PERMANOVA, $p = 0.3$). As expected, interindividual differences were significant ($p = 0.001$), while the variation among biological replicates of a single individual was low.

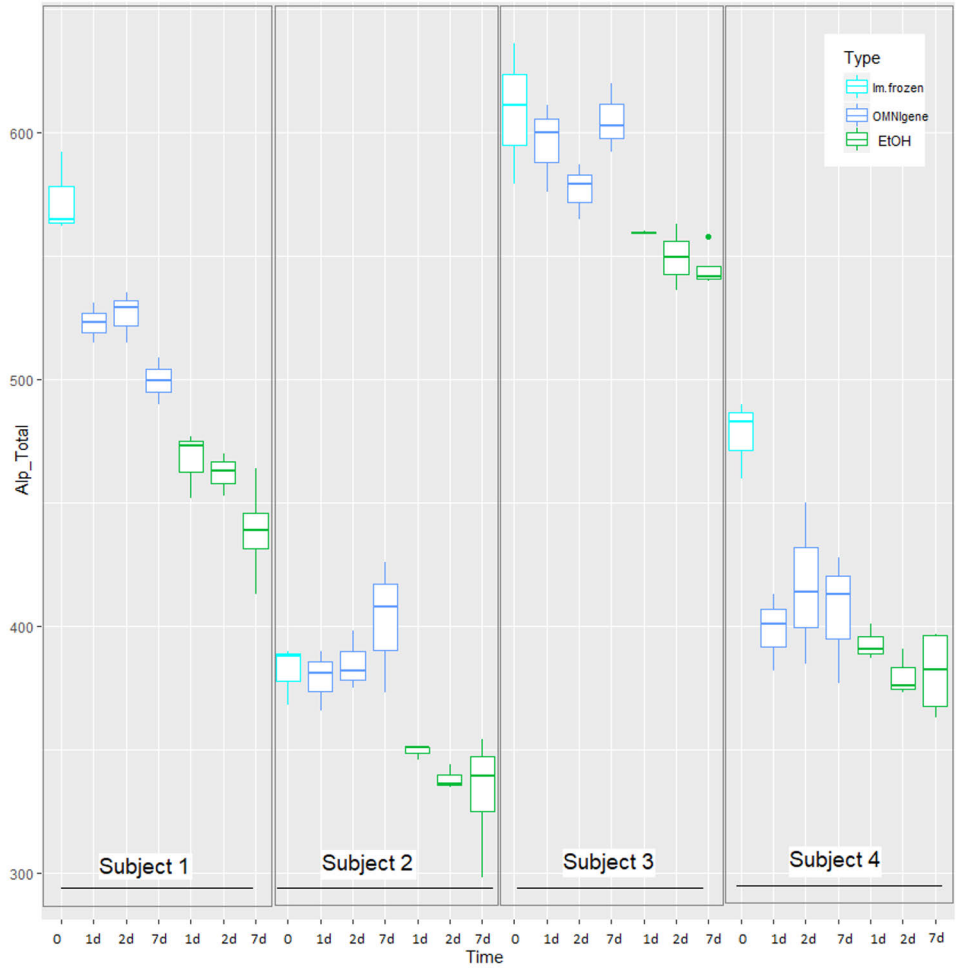


Figure 10. Alpha diversity by time, subjects and sample storage types. Modified from original publication II.

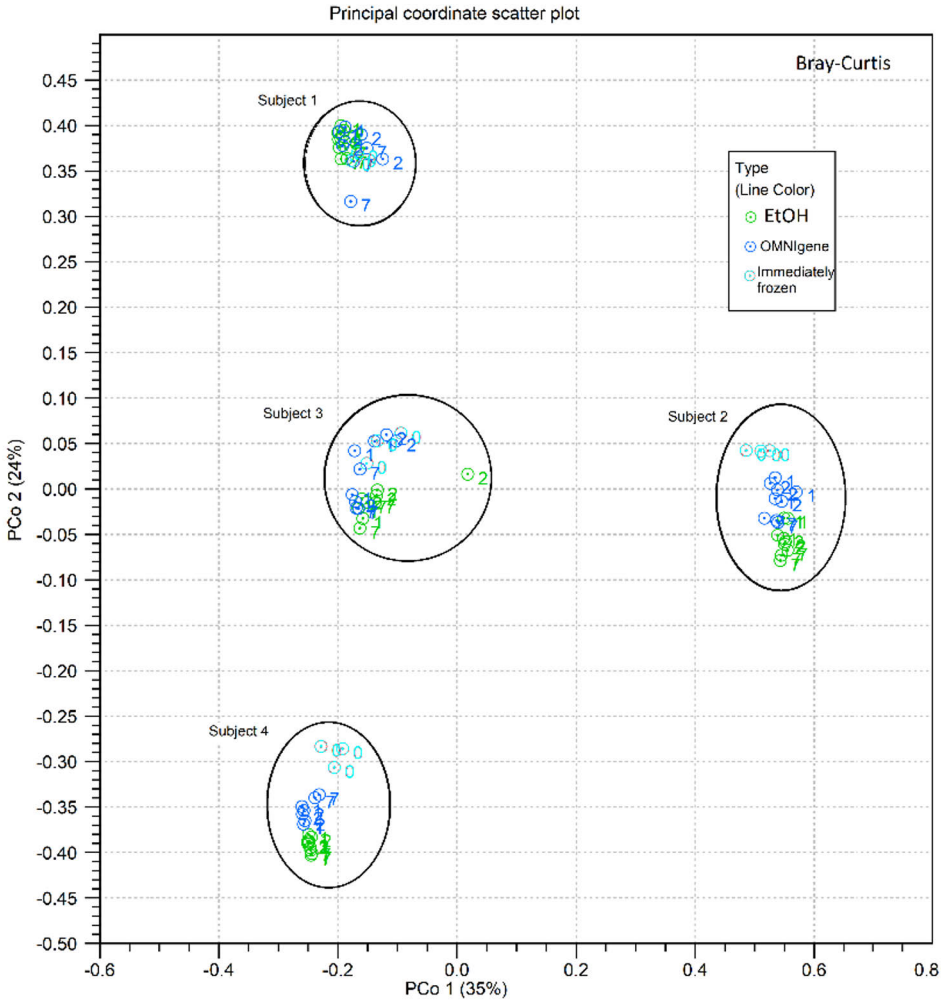


Figure 11. Principal coordinate analysis with Bray-Curtis dissimilarity with two dimensions. Colors indicate the storage types and numbers indicate storage days. Subjects are rounded by ellipses. Modified from original publication II.

5.3 Faecal metabolites change during early development

Age-related variation had a significant impact on the gut metabolome. Most short-chain fatty acids (SCFAs), except for acetic acid, increased with age (Fig. 12A). Individual bile acids (BAs) and polar metabolites did not show clear age-related patterns. However, secondary BAs were positively associated with age, while primary and tauroconjugated BAs were negatively associated with age (Fig. 12B-C). Glycoconjugated BAs were positively associated with age, but this association

weakened when adjusted for breastfeeding (BF). Multiple polar metabolites, such as 5-Hydroxyindoleacetate and 4-Hydroxyphenylacetic acid that had a significant age trend (Fig.12C), but were no longer associated with age when adjusted for BF. In addition, the SCFA and BA trends were explored in the subsample that had all the timepoints available (n=37). The trends were similar to those in the whole sample set.

To study in more detail how gut metabolites relate to demographic exposures, a linear mixed-effect model was implemented with metabolite concentration as the response variable, age and demographic variables as fixed effects, and child as a random effect. In general, demographic exposures explained on average <1 % of variance in polar metabolites, SCFAs, and BA concentrations (Fig.12D-F). It was found that breastfed infants had lower concentrations of secondary and individual tauro- and glycoconjugated bile acids, particularly at an early age(Fig. 13A-B).

Further, it was investigated if duration of exclusive BF associated with metabolite concentrations in 6-, 14- and 30-month-olds (median 4.5 months, mean = 3.96, SD 1.95). The metabolite concentration was modelled by duration of exclusive BF, age, and any current BF (fixed effects), and child identity (random effect). Specifically, the duration of exclusive BF negatively associated with pinitol, lauric acid, ribonic acid, 1,2,3,4,5,6-hexatrimethylsilylinositol, 7-oxo-HDCA, propionic acid and iso-butyric acid, whereas succinic acid was positively associated ($q < 0.05$).

Vaginal delivery was associated with lower concentrations of hydroxyindoleacetate (Fig. 13C), while exposure to intravenous antibiotics during the neonatal period was linked to higher concentrations of butyric acid (Fig. 13D). In the cross-sectional group comparison, infants born vaginally had lower concentrations of 7-oxo-converted BAs at 14 months. Additionally, primary BAs at 14 months and tauroconjugated BAs at 2.5 months were also lower. Similarly, breastfed infants had lower concentrations of secondary and primary bile acids at 2.5 months. Having pets was positively associated with tauroconjugated bile acid concentrations at 14 months, whereas having siblings was positively associated with secondary BA concentrations at 6 and 14 months. These findings suggest that factors related to optimal microbiota development, particularly breastfeeding, are associated with faecal metabolite concentrations.

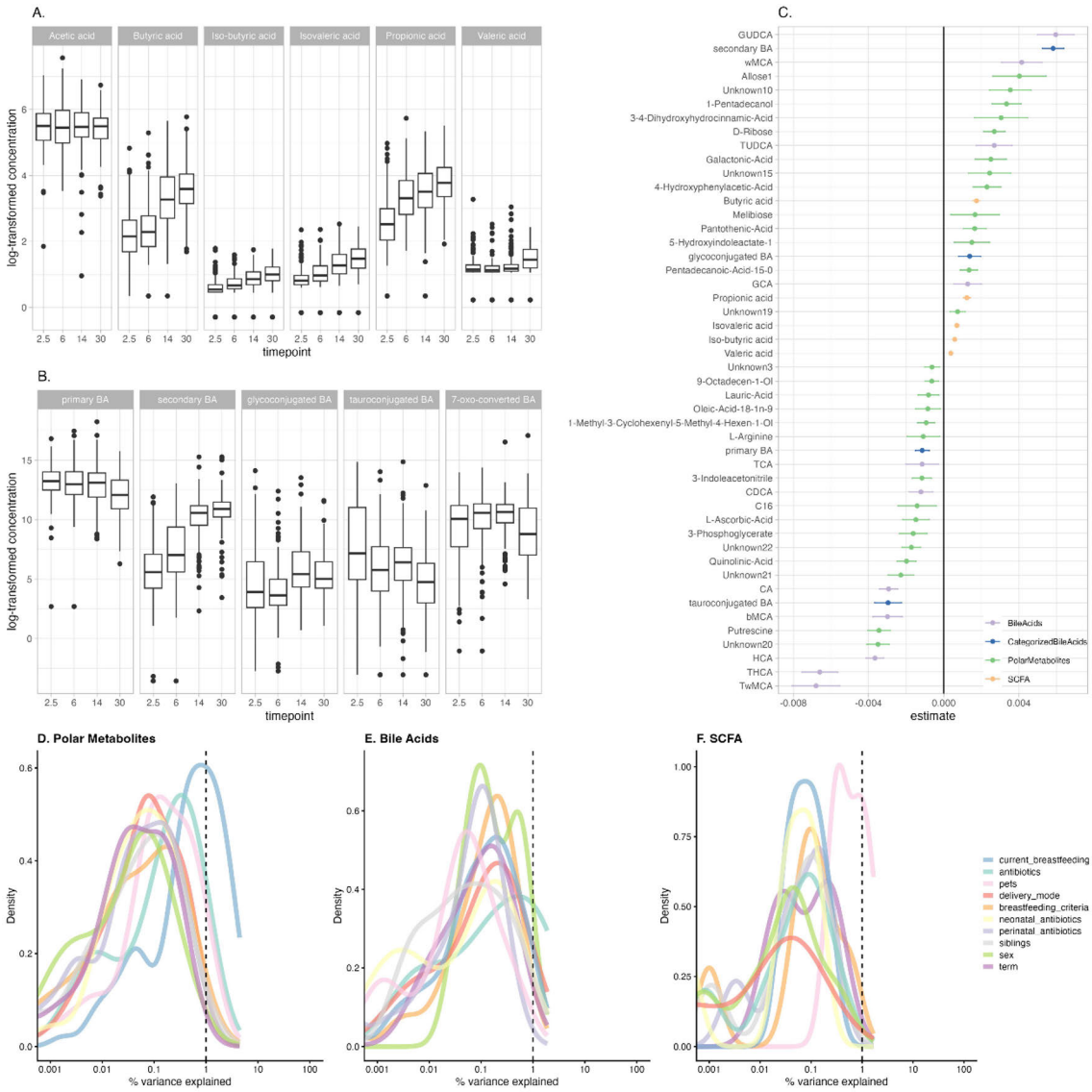


Figure 12. Metabolites varied in their age trends. SCFAs increased, while conjugated BA tended to decrease. A, B. The average changes in SCFAs and BAs levels across different age groups. C. Fixed effect effect-size (age coefficients) for individual metabolites as estimated from the linear mixed models, with metabolite concentration as the response variable, age as the fixed effect and child as the random effect. Density plots showing explained variances (%) by total D. polar metabolites, E. BAs and F. SCFAs associated with the clinical and demographic factors. From original publication III.

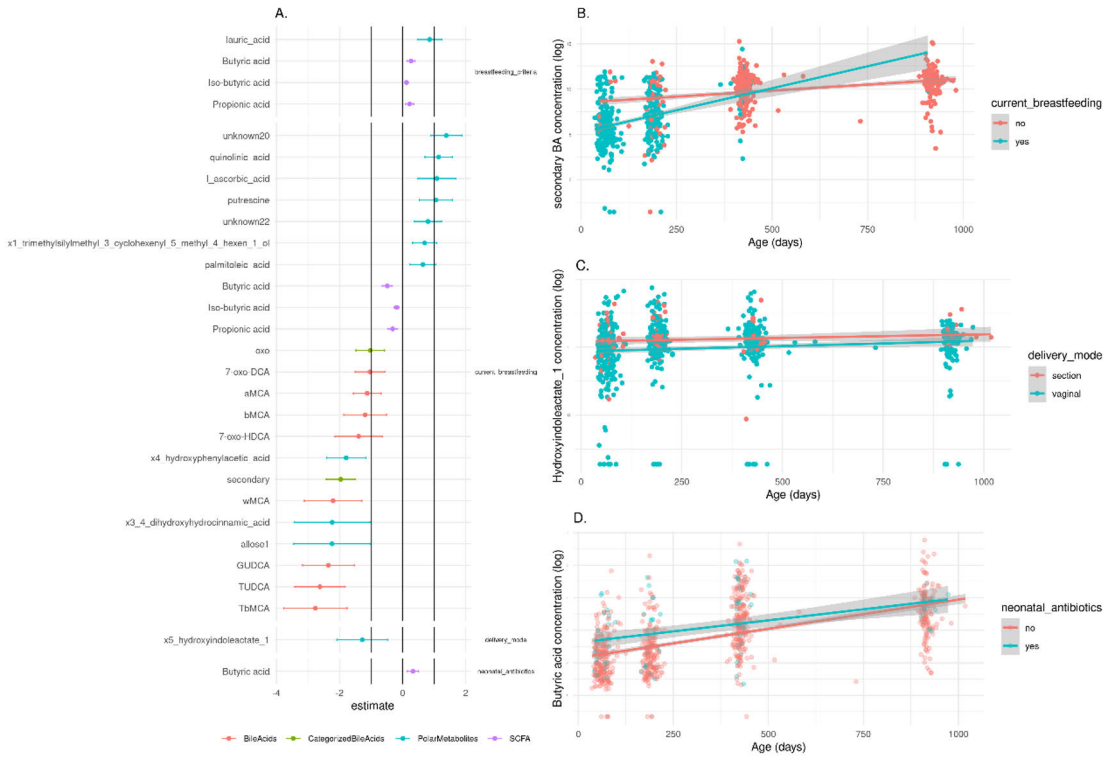


Figure 13. Out of comprehensive list of background factors, breastfeeding associated with concentrations of multiple metabolites from different assays. A. Estimates for each demographic variable from a mixed model with the metabolite concentration as the dependent variable, demographic variable and age as fixed effects, and child ID as random effect. Error bars represent 95% confidence intervals. B. Secondary BA concentrations were lower among breastfed infants in the 2.5-6-month timepoints and higher in the 30-month timepoint. C. Vaginally born infants had consistently lower concentration of hydroxyindoleacetate-1 across all timepoints. D. Concentration of butyric acid was higher in infants who received antibiotics in the neonatal period. B-D. Grey area depicts 95% confidence intervals. From original publication III.

5.3.1 Infant microbiota shows more diverse microbiota community types compared with toddlers

Next, the aim was to confirm the previously suggested successional development of infant taxonomic composition and patterns (Beller et al., 2021; Stewart et al., 2018) in our data. To examine the patterns of gut microbiota succession during early-life, Dirichlet Multinomial Mixture (DMM) model was performed to identify gut microbiota community types and stratify the individuals accordingly. Seven community types were identified based on Laplace criteria when analysing samples from all time points together (Fig. 14A-B). At the first time point, three community

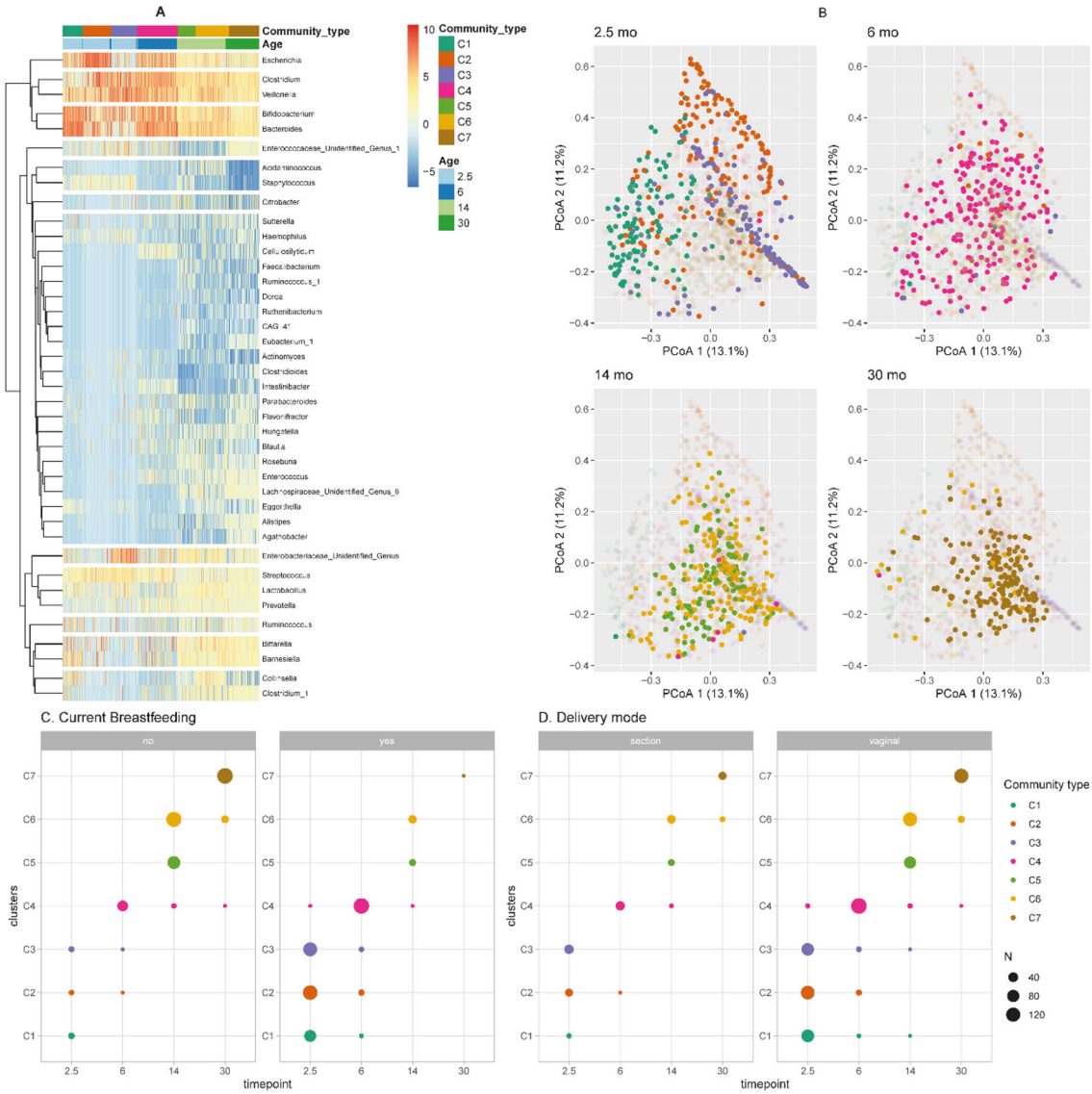


Figure 14. Gut microbiota community composition was stratified into distinct community types with DMM model and linked the community types with breastfeeding and delivery mode. **A**. Seven community types were identified. The breaks between rows are derived from hierarchical clustering of clr-transformed abundances of genera with prevalence over 10 % and abundance over 0.1 %. Here, 50 % height of the maximum dendrogram branch height is used to visualize clusters of taxa in the heatmap. **B**. The community membership is indicated on the PCoA ordination. It seems that C7 was the most homogenous as indicated by DMM theta. The data points with non-transparent color belong to the timepoint indicated above the figures, and the partially transparent points belong to other time points. The color represents the community type. **C-D**. In a mixed model with the community type membership as the dependent variable, demographic variable and age as fixed effects, and child identity as random effect, C. current breastfeeding and D. delivery mode explained transition between community types. From original publication III.

types were predominant, driven by the abundances of *Bacteroides and Bifidobacterium (C1), Escherichia (C2), and Veillonella* along with an unidentified genus in *Enterobacteriaceae* (C3) (Fig. 14A). The later time points were dominated by one community type mostly, influenced by varying proportions of *Bacteroides, Clostridium, or Veillonella* (C4-7, Fig. 14A). Consistent with earlier studies, the gut microbiome community varied according to background factors such as delivery mode and breastfeeding (Fig. 14C-D). Some additional trends aligned with earlier reports but did not reach statistical significance. Conversely, factors like infant sex, having pets, overall duration of exclusive breastfeeding, and intravenous neonatal or recent antibiotic intake were not associated with gut microbiome community membership. When stratified by time point, delivery mode at 2.5 months (C1 3.3%, C2 15.1%, C3 29.4% of C-section born infants, X^2 $q < 0.005$) and preterm birth at 6 months (C1 50%, C2 89%, C3 67%, C4 98%) were enhanced in specific community types. However, perinatal (2.5 mo, 30 mo) and recent antibiotic treatments (6 mo, 30 mo), as well as having siblings (2.5 mo, 14 mo), were not significant. In a mixed model with the community type membership as the dependent variable, demographic variable and age as fixed effects, and child identity as random effect, the vaginal delivery and current breastfeeding were negatively associated with community type progression (Fig 14C, D).

5.3.2 Microbiota alpha diversity and genera abundances are associated with faecal metabolites

Next, the association between microbiota composition and metabolome profiles was investigated. It was found that microbiota alpha diversity correlated with multiple metabolite classes (Fig. 15A), with SCFA concentrations showing constantly positive associations with alpha diversity. A linear mixed model indicated that THCA, TMCA, and several polar metabolites (arachidonic acid, 2-methylpentadecanoic acid, putrescine) were negatively associated with the Shannon Index, adjusted for age (fixed effect) and child identity (random effect) ($q < 0.015$, Fig. 13). Conversely, butyric, propionic, isovaleric, and iso-butyric acids, MCA, MCA, UDCA, and certain polar metabolite concentrations had positive association with the Shannon Index when adjusted for age ($p = 0.039$, Fig. 15B). Additionally, the Shannon Index was positively correlated with 7-oxo-converted BA concentrations (estimate = 0.3, 95%-CI 0.3-0.56, $q = 0.047$). In differential abundance testing, *Clostridium* and *Bifidobacterium* had associations with butyric acid, P-hydrophenyllactic acid, and conjugated BAs in reverse directions. In 30-month-olds, unidentified genera in the *Oscillospirales* order associated negatively with BAs, such as 7-oxo-converted BA.

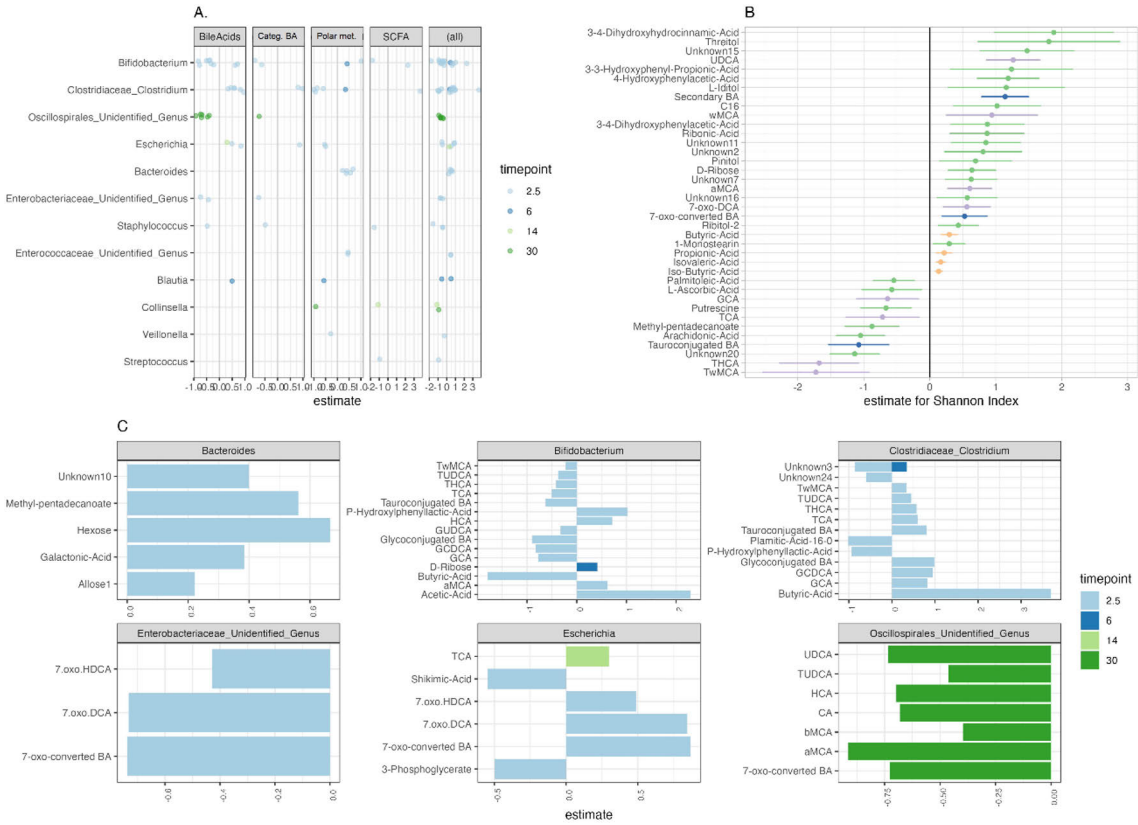


Figure 15. Gut microbiota composition associated with faecal metabolite concentrations, and the associations with genera differed based on age. **A**). Differential abundance analysis (ALDEx2) showing associations between genus abundances and metabolites. Only significant associations ($q < 0.05$) are shown here. Bifidobacterium (number of significant associations n=22), Clostridium (n=18), unidentified genus in Oscillospirales (n=11), Bacteroides (n=9), Escherichia (n=9) had most significant associations. **B**) SCFA (orange) tended to positively correlate with alpha diversity, whereas individual polar metabolites (green) and BAs (blue and gray) correlated both negatively and positively. **C**) Most significant associations between genera and metabolites were at 2.5 mo timepoint. Bifidobacterium at 2.5 mo associated negatively and Clostridium at 2.5 mo associated positively with conjugated BAs and butyric acid. Streptococcus was associated negatively with propionic acid. Unidentified genus in Oscillospirales at 30 mo associated negatively with several BAs, especially 7-oxo-converted and tauroconjugated BAs. From original publication III.

In the network analysis for each timepoint (Fig. 16), the node and edge numbers were higher at the 30-month timepoint suggesting a richer or more complex metabolic profile. However, the density was highest in the first timepoint, possibly indicating that a greater proportion of possible connections are actually present. Additionally, the degree distribution was more left-skewed (negatively skewed) in

Figure 16. Networks of microbiome and metabolite inter-correlations dependent on the age (Spearman correlation). Visually, the 30 months was dominated by microbial inter-correlations whereas correlations with metabolites were limited. At 2.5 months, Bifidobacterium and Clostridium both associated with BAs and SCFAs. As expected, both SCFA and BA had strong correlations within the group. Black outer circle indicates ”high impact”, i.e. nodes that have high degree and betweenness, and are in the top 25 % in both. A. 2.5 months, B. 6 months, C. 14 months, D. 30 months. From original publication III.

the 30 months compared with 6 and 14 months (Kolmogorov-Smirnov test, p= 0.046 for 6 months, p= 0.024 for 14 months), i.e. at 30 months, the network has more nodes with higher number of connections compared to the earlier timepoints. There were most sub-communities in the 14 months, and it had the highest modularity score (sub-communities 2.5 months n= 5, 6 months n=7, 14 months n=13, 30 months n=10, 2.5 months modularity = 0.55, 6 months modularity = 0.53, 14 months modularity =0.57, 30 months modularity = 0.37). Metabolites and genera with highest degree and betweenness were different depending on the timepoint, e.g. *Bifidobacterium and Clostridium* had high degree and betweenness in the 2.5-month-olds, whereas *Oscillospirales and Ruminococcaceae* had high degree and betweenness in the 30-month-olds (Fig. 16). Additionally, BAs had high score, and CA was among the most connected metabolites at the 6, 14 and 30-month timepoints.

5.3.3 Microbial community types associate with different levels of metabolites

The microbial communities associated with differing levels of faecal metabolites per timepoint, and the largest effect sizes were for TwMCA, TCA, THCA, GCA as well as succinic acid and an unknown polar metabolite at 2.5 months. For all those BAs, C1 had a lower concentration compared with C2 and/or C3. Moreover, both glycoconjugated and tauroconjugated BA concentrations were lower in C1 at 2.5 months. At 14 months, concentration of butyric acid was higher in C6 compared with C5. In addition, at 30 months, C7 had elevated concentrations of valeric acid, MCA, succinic acid and MCA with moderate effect size.

Specific BAs (TMCA, THCA, TCA, GCA) and arachidonic acid had positive association with community type membership. Whereas multiple polar metabolites, UDCA, propionic acid, and branched SCFA had negative associations with community type membership (Fig. 17). Similarly, glycoconjugated and tauroconjugated were both positively associated with community types C2-C6 and C2, C3 and C6, respectively (FDR < 0.05, C1 as reference,). In addition, findings on these community types and metabolite concentrations were similar in the subsample of subjects with the whole timeseries (n=37).

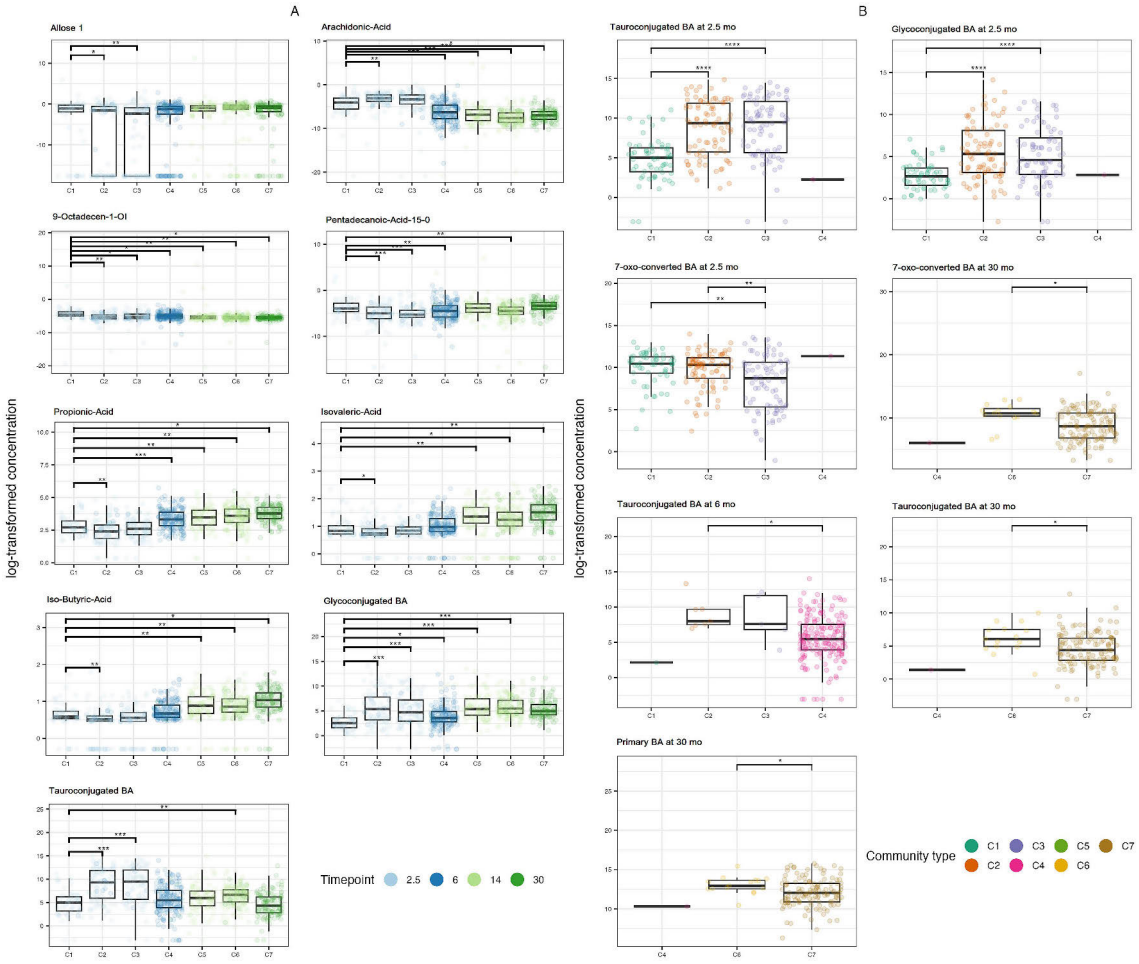


Figure 17. Community type showed altered levels of metabolites. A. Several associations remained after adjusting for BF in the mixed model with the metabolite concentration as the dependent variable, community type, child age and current breastfeeding as the fixed effects, and child identity as random effect, color indicating the timepoint. B. Community types at 2.5, 6 and 30 months had different levels of BA based on cross-sectional group comparison and post hoc testing. * q < 0.05 & q > 0.01, ** q <= 0.01 & q > 0.001, ***q <= 0.001. From original publication III.

5.3.4 Interactions Between Breastfeeding, Gut Microbiota, and Metabolites

Breastfeeding (BF) showed the strongest correlations with metabolite levels. Therefore, further exploration was done for the interactions between gut microbe abundances, metabolite levels and BF. As the prevalent genera were driving the community types and they showed the most associations metabolite levels, next it was studied how the interaction between prevalent genera and BF status associated

with metabolite levels. The analysis had metabolite concentration as the dependent variable, age and the interaction between any current BF and rclr-transformed genus abundance as the fixed effects and child identity as the random effect.

Bifidobacterium abundances were correlated negatively with tauroconjugated BAs only in breastfed infants (Fig. 18). In turn, *Bacteroides* was positively associated with secondary BA in the breastfed infants. Additionally, the less there was Escherichia in breastfed infants gut microbiota, the less there was 7-oxo-HDCA. From polar metabolites, *Bacteroides* abundance was positively associated with pinitol concentrations in the breastfed infants (Fig. 18). In addition to above described analysis, it was tested further if cumulative exclusive BF duration interacts similarly with prevalent genera abundances, with a model that had metabolite concentration as the dependent variable, age, any current BF and the interaction between duration of BF and rclr-transformed genus abundance as the fixed effects and child identity as the random effect. The interaction between exclusive BF duration and *Bacteroides* associated with 7-oxo-converted BA and particularly 7-

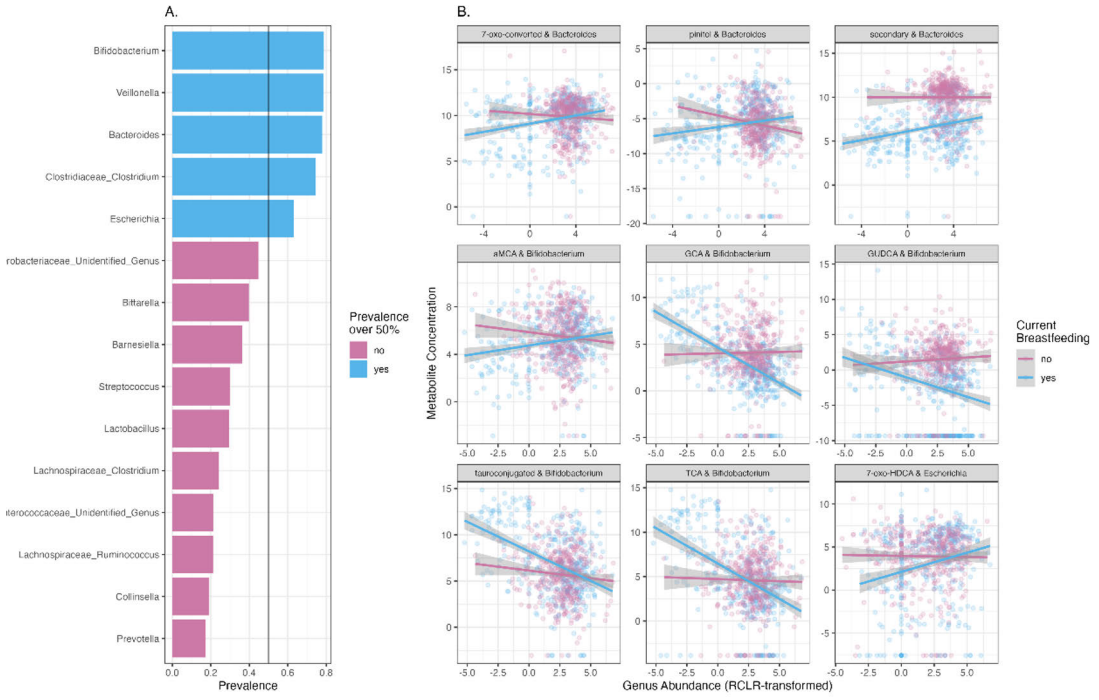


Figure 18. Prevalent genera abundances interaction with breastfeeding status associated with microbial metabolites. A. Only the 5 most prevalent taxa were observed in >50 % of the study subjects, and those were included for the interaction analyses. B. Escherichia, Bifidobacterium and Bacteroides had significant interaction with BF. Scatterplots presented for significant interactions. From original publication III.

oxo-DCA. The interaction between exclusive BF duration and *Escherichia* abundances associated with 7-oxo-HDCA.

In addition, it was studied if the alpha diversity associates with breastfeeding (adequate/non-adequate) in the subset of infant with all four timepoints (n=37) to see the longitudinal development in an individual level. It seemed that non-adequate BF group had higher alpha diversity in the first timepoint whereas adequate BF group had higher diversity in the latest timepoint (Fig. 19). Additionally, alpha diversity seemed to increase more linearly in adequate BF group. However, this was small sample subset with low statistical power.

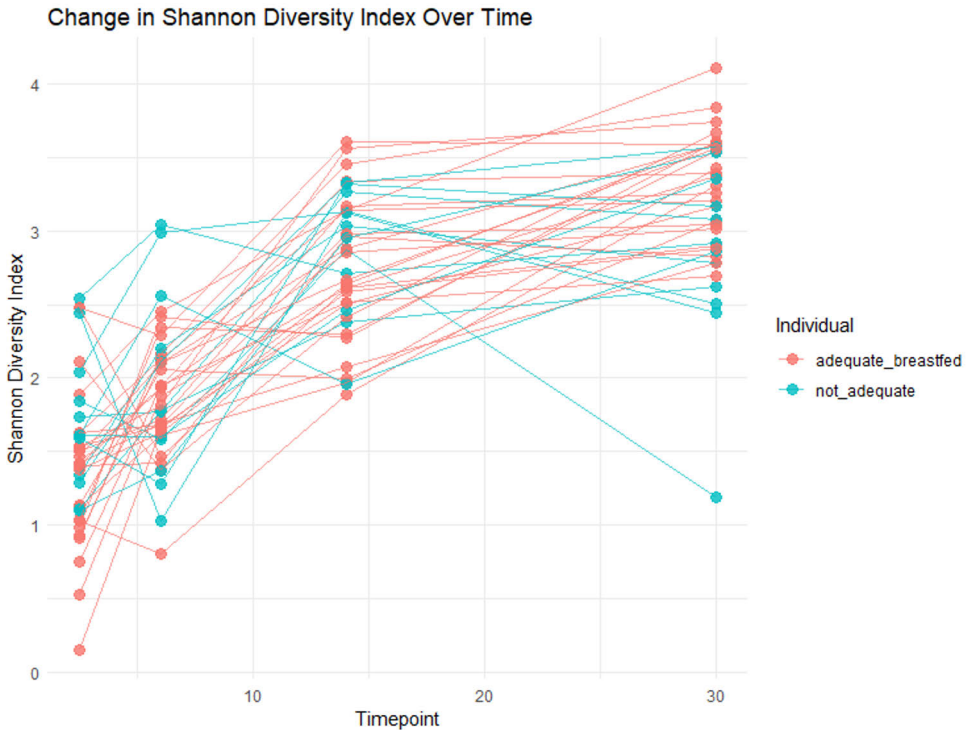


Figure 19. Alpha diversity (Shannon index) by time and breastfeeding criteria in 37 full-series samples. Adequate breastfeeding: exclusive BF for at least 4 months and partial BF for at least 6 months (breastfeeding criteria, yes vs. no). From original publication III.

5.4 Human milk directing the colonization process of gut microbiome

Most infants (78% of) in the cohort were fully breastfed in the 2-month timepoint. In the 6-month timepoint, 71% were partially breastfed. At 14 months only 14% were receiving breast milk. The overlap of milk and stool samples got lower by age due to decreasing breastfeeding; 2mo (n=283), 6mo (n=129) and 14mo (n=65).

Nearly 84% of mothers were secretors regarding the presence of 2'-FL in milk which is slightly higher than previously observed internationally (75-80%) (Azad, Wade, et al., 2018; Soyyılmaz et al., 2021).

The background characteristics of study population have been listed in each timepoint (Table 4).

Table 4. The background characteristics of study population in each timepoint. *Post in current feeding denotes recently ceased breastfeeding or the late stages of weaning.

	2.5mo	6mo	14mo	Overall
	(N=444)	(N=256)	(N=302)	(N=1002)
Sex				
boy	229 (51.6%)	143 (55.9%)	163 (54.0%)	535 (53.4%)
girl	215 (48.4%)	113 (44.1%)	139 (46.0%)	467 (46.6%)
Delivery mode				
section	74 (16.7%)	43 (16.8%)	49 (16.2%)	166 (16.6%)
vaginal	363 (81.8%)	212 (82.8%)	250 (82.8%)	825 (82.3%)
Missing	7 (1.6%)	1 (0.4%)	3 (1.0%)	11 (1.1%)
Gestational age				
Mean (SD)	40.1 (1.39)	40.0 (1.37)	39.9 (1.50)	40.0 (1.42)
Median [Min, Max]	40.1 [34.6, 42.3]	40.0 [34.1, 42.3]	40.1 [32.9, 42.3]	40.1 [32.9, 42.3]
Antibiotics prior 6 months				
no	302 (68.0%)	128 (50.0%)	111 (36.8%)	541 (54.0%)
yes	2 (0.5%)	4 (1.6%)	22 (7.3%)	28 (2.8%)
Missing	140 (31.5%)	124 (48.4%)	169 (56.0%)	433 (43.2%)
Current feeding				
full	267 (60.1%)	10 (3.9%)	0 (0%)	277 (27.6%)
none	7 (1.6%)	1 (0.4%)	4 (1.3%)	12 (1.2%)
partial	57 (12.8%)	159 (62.1%)	37 (12.3%)	253 (25.2%)
post*	11 (2.5%)	55 (21.5%)	233 (77.2%)	299 (29.8%)
Missing	102 (23.0%)	31 (12.1%)	28 (9.3%)	161 (16.1%)
Mother age				
Mean (SD)	30.8 (4.38)	30.7 (4.50)	31.2 (4.31)	30.9 (4.39)
Median [Min, Max]	31.0 [19.0, 45.0]	30.0 [19.0, 42.0]	31.0 [19.0, 45.0]	31.0 [19.0, 45.0]
Mother education				
Low	109 (24.5%)	64 (25.0%)	71 (23.5%)	244 (24.4%)
Medium	135 (30.4%)	72 (28.1%)	90 (29.8%)	297 (29.6%)
High	173 (39.0%)	105 (41.0%)	125 (41.4%)	403 (40.2%)
Missing	27 (6.1%)	15 (5.9%)	16 (5.3%)	58 (5.8%)
Primipara				
no	202 (45.5%)	117 (45.7%)	150 (49.7%)	469 (46.8%)
yes	242 (54.5%)	139 (54.3%)	152 (50.3%)	533 (53.2%)
Siblings				
no	150 (33.8%)	95 (37.1%)	116 (38.4%)	361 (36.0%)
yes	187 (42.1%)	110 (43.0%)	147 (48.7%)	444 (44.3%)
Missing	107 (24.1%)	51 (19.9%)	39 (12.9%)	197 (19.7%)
SCFA data				
yes	239 (53.8%)	195 (76.2%)	243 (80.5%)	677 (67.6%)
Missing	205 (46.2%)	61 (23.8%)	59 (19.5%)	325 (32.4%)

	2.5mo	6mo	14mo	Overall
	(N=444)	(N=256)	(N=302)	(N=1002)
BA data				
yes	237 (53.4%)	194 (75.8%)	243 (80.5%)	674 (67.3%)
Missing	207 (46.6%)	62 (24.2%)	59 (19.5%)	328 (32.7%)
Polar metabolite data				
yes	238 (53.6%)	193 (75.4%)	242 (80.1%)	673 (67.2%)
Missing	206 (46.4%)	63 (24.6%)	60 (19.9%)	329 (32.8%)
Microbiome data				
yes	444 (100%)	256 (100%)	302 (100%)	1002 (100%)
Milk metabolite data				
yes	406 (91.4%)	177 (69.1%)	80 (26.5%)	663 (66.2%)

Milk metabolome clustered according to mothers' secretor status (Fig. 20B). At corresponding timepoints, gut microbiome got more uniform by time (Fig. 20A) as already seen in study III. Although, the milk composition was clustered separately between secretors and non-secretors, the mean concentration of total HMOs was similar between the groups (Fig.21).

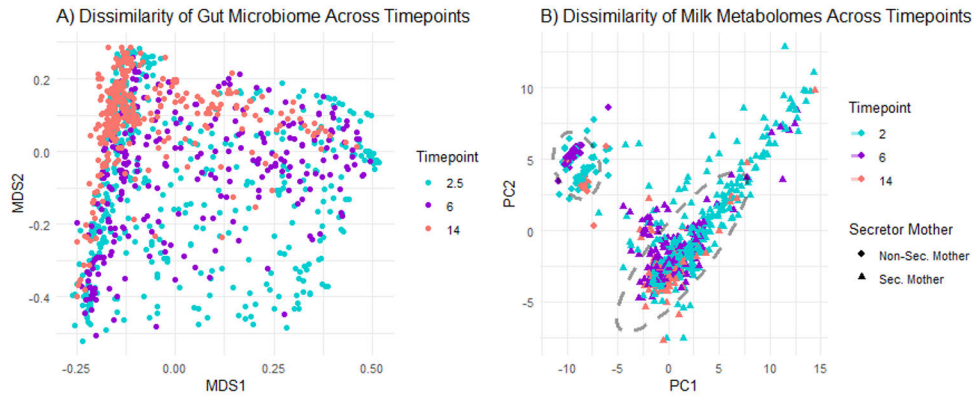


Figure 20. Dissimilarities. A) Infant gut microbiome beta diversity by bray Curtis across three timepoints with multidimensional scaling. Microbiome got more uniform by time. B) Milk metabolomes by Euclidean and principal components by secretor status across timepoints. Non-secretors shaped as square and secretors as triangle. Grey ellipses represent the two clusters determined by secretor status. From the study IV manuscript.

Total HMO concentration differed only at the 2-month timepoint (mean, secr. 16.9, non-secr. 14.9, two sample t-test p=0.0002) by secretor status. HMO-profiles (Fig. 21) showed that non-secretors had more 3-FL while they were lacking 2'-FL, LNFP I, LDFT, and LNDFH I.

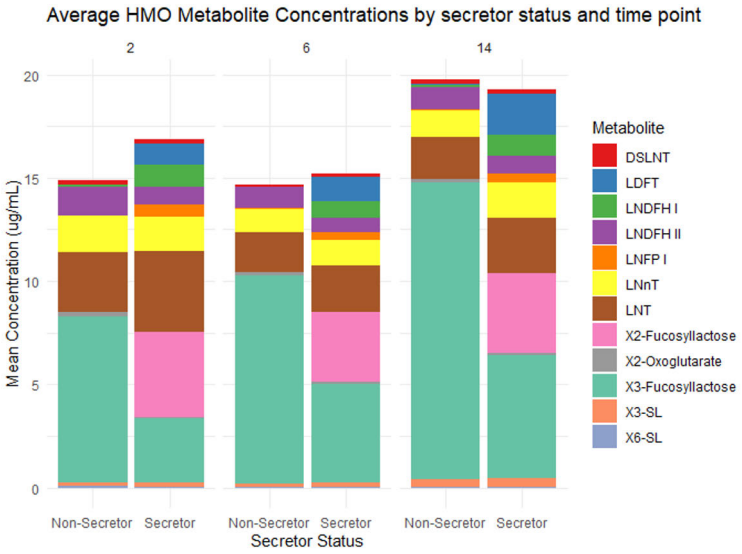


Figure 21. HMO profiles by secretor status and timepoint. From the study IV manuscript.

Overall, the milk composition (principal components (PC) 1-3) did not drive the differences observed in the GM beta diversity, except PC1 of milk metabolome in 6 months explained a minor of part (2%) of GM beta diversity (PERMANOVA, p=0.02). The taxonomic diversity (Shannon) was not associated with either the secretor status or the milk metabolites (p>0.4). This indicates that the milk metabolome was not related to GM in the overall community composition level.

However, there were moderate nominal associations between specific bacterial genera and milk metabolites by using differential abundance analysis (DAA). At the 2-month timepoint, 3-SL had positive correlation to unidentified genus of *Enterobacteriaceae* (q-value 0.16). LNFP I was correlated positively with Bifidobacterium in 6-month timepoint (Fig. 22). Glutamine had positive correlation with butyrate-producers (Sabater et al., 2023) *Faecalibacterium and Roseburia* in 14-month timepoint, while 2'-FL correlated negatively with *Ruminococcus, Clostridioides and Citrobacter* (Fig. 22). Altogether, there were more associations in later timepoints, which also showed more diverse microbiome composition as expected. Sn-glycero-3-phosphocholine, which is a precursor in the synthesis of acetylcholine (a neurotransmitter), was associated in all three timepoints but with different genera. In the last timepoint, it was positively associated with *Bifidobacterium and Lactobacillus* (Fig. 22). However, these associations did not remain after adjusted p-value correction, with exception of the negative correlation between *Bifidobacterium* and ethanolamine/methionine (q-value 0.02/0.03) at the 6-month timepoint. *Bifidobacterium* had also negative association to citrate (q-value 0.19) in 14-month-olds. Similar negative correlation was between phenylalanine and *Collinsella* (q-value 0.14) at the last timepoint.

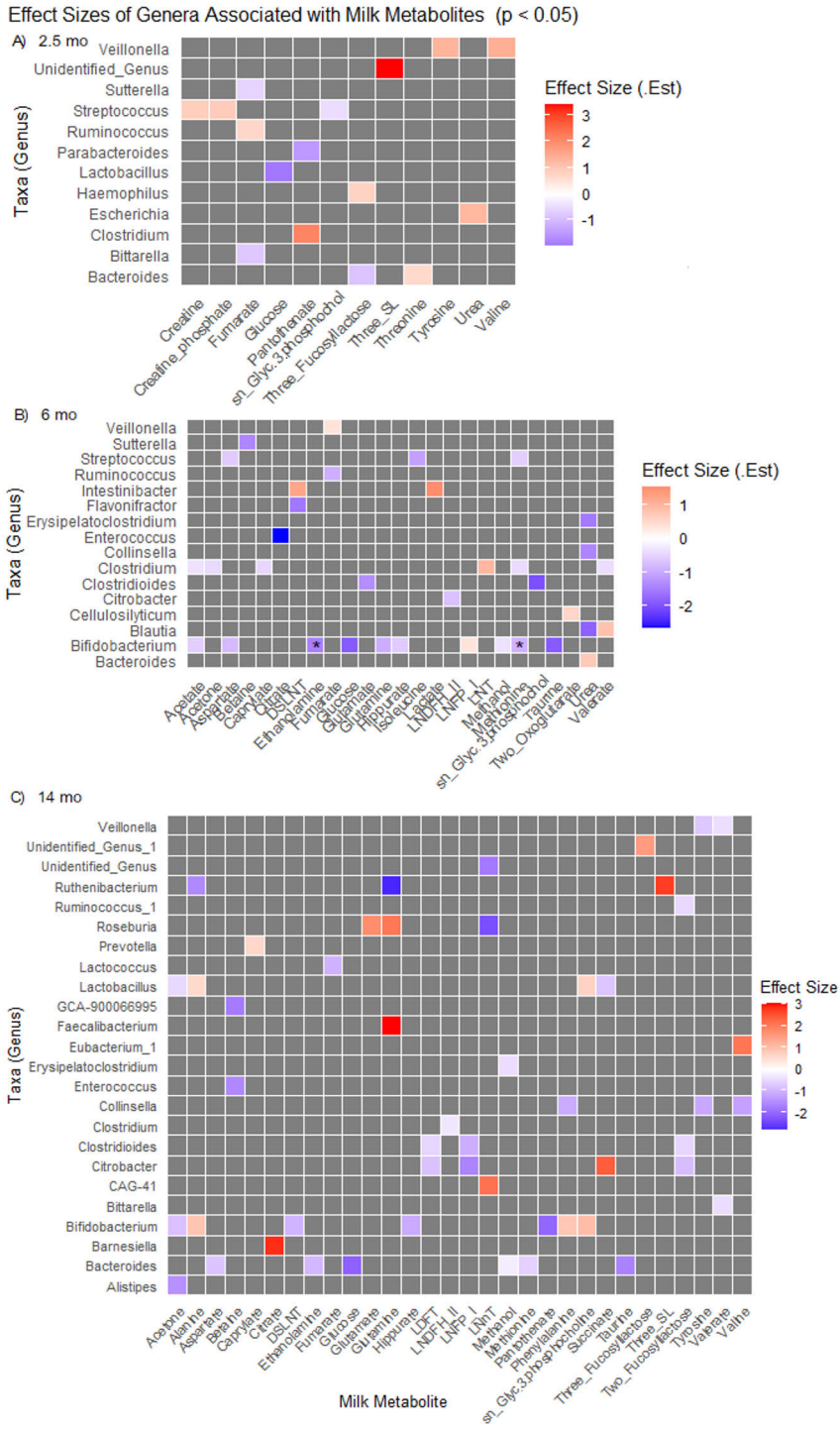


Figure 22. Differential abundances in three timepoints by Aldex2. Effect size by colors. Adjusted p-values (FDR) <0.05 marked with asterisk*. From the study IV manuscript.

Several associations were found between milk metabolites and faecal SCFA, but the results did not withstand p-value correction (Fig. 23). LDFT, which was missing with non-secretors, correlated positively with several faecal SCFAs, especially at the 2.5- and 14-month timepoint. Milk amino acids had negative correlations to faecal acetic acid in 6-month timepoint. Milk amino acids associated with faecal branched-chain fatty acids (isovaleric and isobutyric acid) at the 2- and 6-month timepoints.

Similarly, BAs were associated with several milk metabolites, but the results did not withstand p-value correction (Fig. 24). At the 2.5-month timepoint, milk metabolites had positive associations with primary BAs, while oxo- and secondary BAs were negatively associated. Interestingly, 2'-FL and 3-FL had opposite direction in BA correlation.

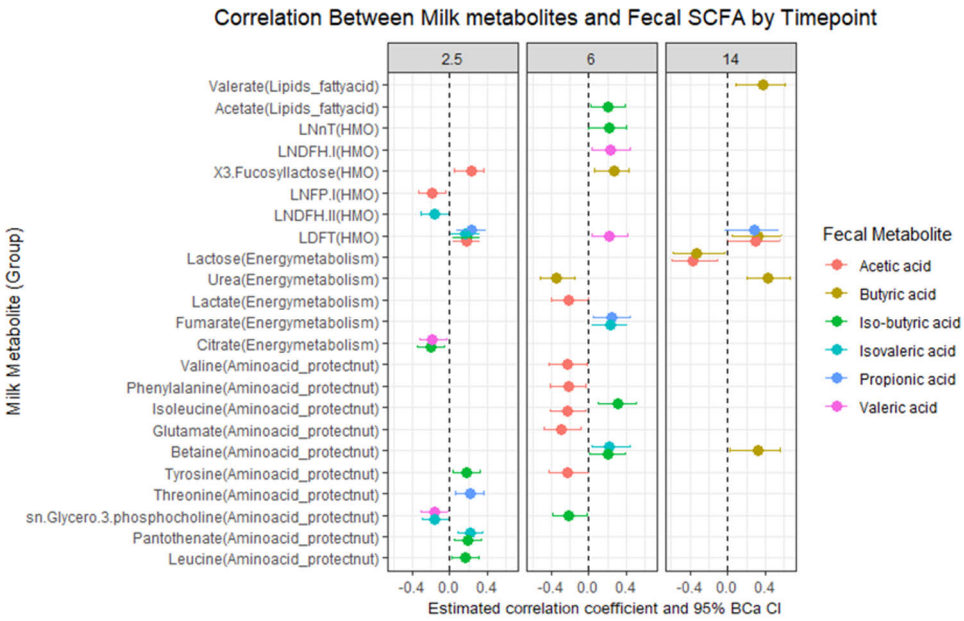


Figure 23. Milk metabolites associated (Spearman's correlation) with faecal short chain fatty acids (SCFAs, see colors). Dots indicate estimated correlation coefficient with lines of 95% confidence interval. These correlations have unadjusted p-values <0.05. From the study IV manuscript.

At the 6-month timepoint, correlations were mainly in glycoconjugated BAs which were not seen in other timepoints. In 14-month timepoint, secondary BAs had several positive correlations with HMOs and milk energy metabolites. HMOs shifted the directions of correlation from 2.5month to 6month, i.e. 2'-FL and LNFP associated positively with tauroconjugated BA at 2.5 months and negatively at 6 months, whereas an opposite pattern was observed for 3-FL.

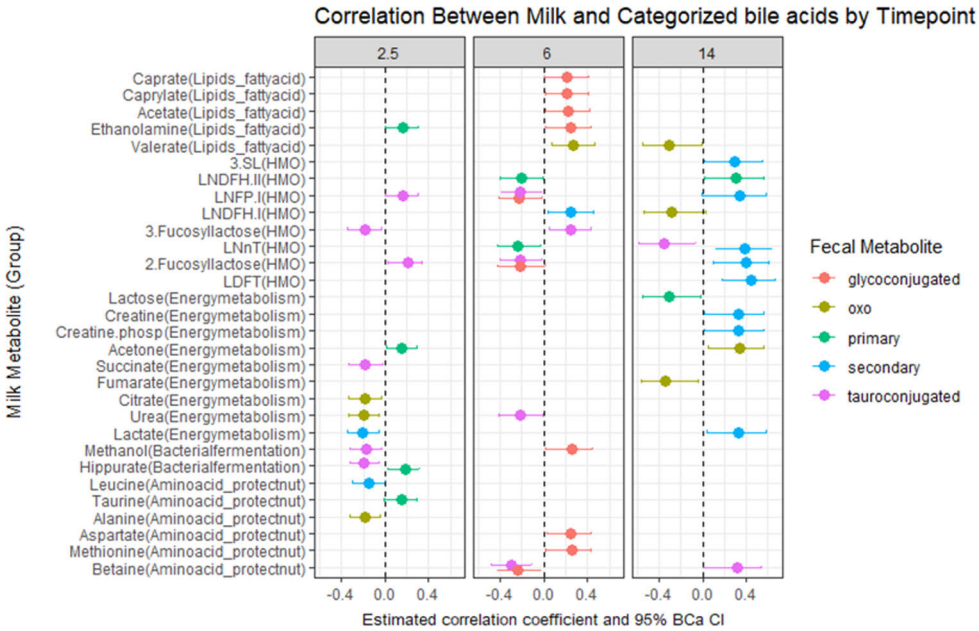


Figure 24. Milk metabolites associated (Spearman's correlation) with faecal categorized bile acids, see colors). Dots indicate estimated correlation coefficient with lines of 95% confidence interval. These correlations have unadjusted p-values <0.05. Correlations by timepoints 2.5, 6 and 14 months. From the study IV manuscript.

Next, multi-omic factor analysis (MOFA, Fig. 25A.) with all omic-data sets was set out to reduce features and see how much variance milk, faecal metabolites and gut bacterial genera explain in the data. Eight factors explained the variance and factors were loaded by different features (Factors 1-4 presented in Fig. 25B-E). Overall, milk metabolites and stool-based omics explained variance in different latent factors. Bile acids and milk metabolites explained the largest part of overall variance in the multiomic datasets. With milk metabolites, variance was mainly explained by factors 1 (R^2 21.4), 3 (R^2 23.3) and 7 (R^2 10.6) at the 2-month timepoint and those remained largest throughout the timepoints. Those factors were loaded by milk energy metabolites, HMOs and faecal propionic acid.

In bile acid data set, variance was explained by latent factor 2 (R^2 20.9) and to some extent by 4 (R^2 5.2) and 5 (R^2 7.0) at the 2-month timepoint and those factor values fluctuated by time. Other assays explained variance to lesser degree. Factor 1 was the only shared factor between milk and faecal metabolites and it explained minor part of SCFAs (2mo: R^2 1.5; 6mo: R^2 1.7).

Additionally, individual factor points were compared with breastfeeding (BF) variables. Overall, categories of current BF did result differences in factor 1 and factors 5 to 8, while secretor status led to differences in factor 1, 3 and 5. Factor 1

and 3 were loaded mainly by lactose and HMOs. Specifically, factors 5 and 6 were higher with full-breastfed in 2.5mo timepoint, which were led by conjugated bile acids, SCFA and Bifidobacterium.

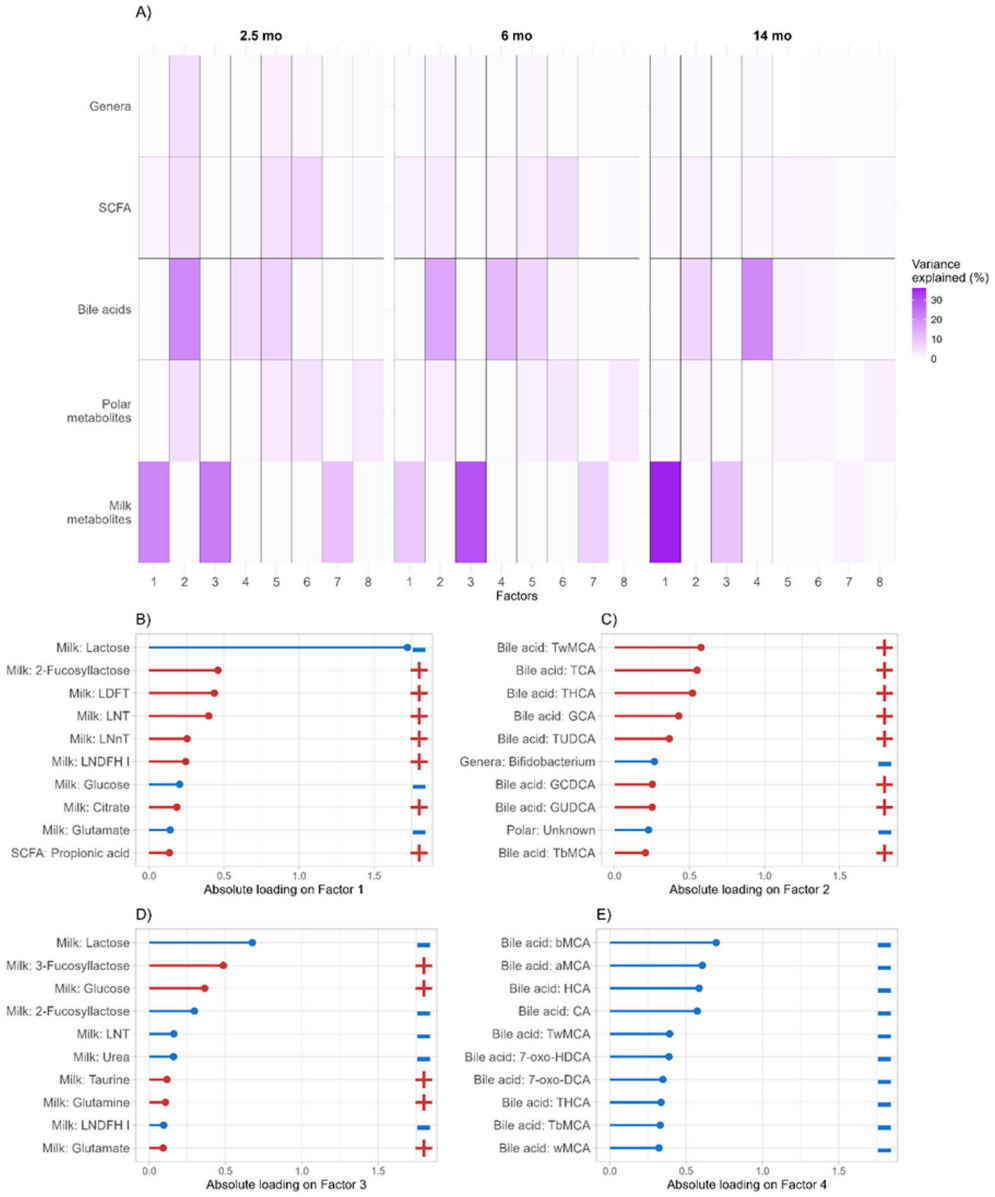


Figure 25. Multi-omic factor analysis showing explained variances (%) by 8 factors. A) Multi-omic factor analysis (MOFA) with five data sets and 8 factors in three timepoints. From study IV. B)-E) Loadings of factors 1, 2, 3 and 4. Top 4 factors and their top 10 features are presented here. Colors (red/blue) and +/- indicate the direction of the loadings. From the study IV manuscript.

Multiple correlations were seen between milk and faecal microbial metabolites in mixed model adjusted for 2 time-intervals by sampling age (Fig 26). From faecal polar metabolites, long-chain carbohydrates fucose and ribonic acid had associations with milk 3-SL and 6-SL (Fig. 26C-D). Betaine and urea, which were already noted in DAA and Spearman, were associated with butyric acid in the later time interval (Fig. 26G-H). LDFT, which associated with multiple SCFAs with Spearman, correlated with 3,4-dihydroxyphenylacetic acid (DOPAC, a metabolite of the neurotransmitter dopamine) in both time intervals (Fig. 26E). LNFP I and caprylate correlated positively with secondary bile acids in the later time-interval (Fig. 26A-B).

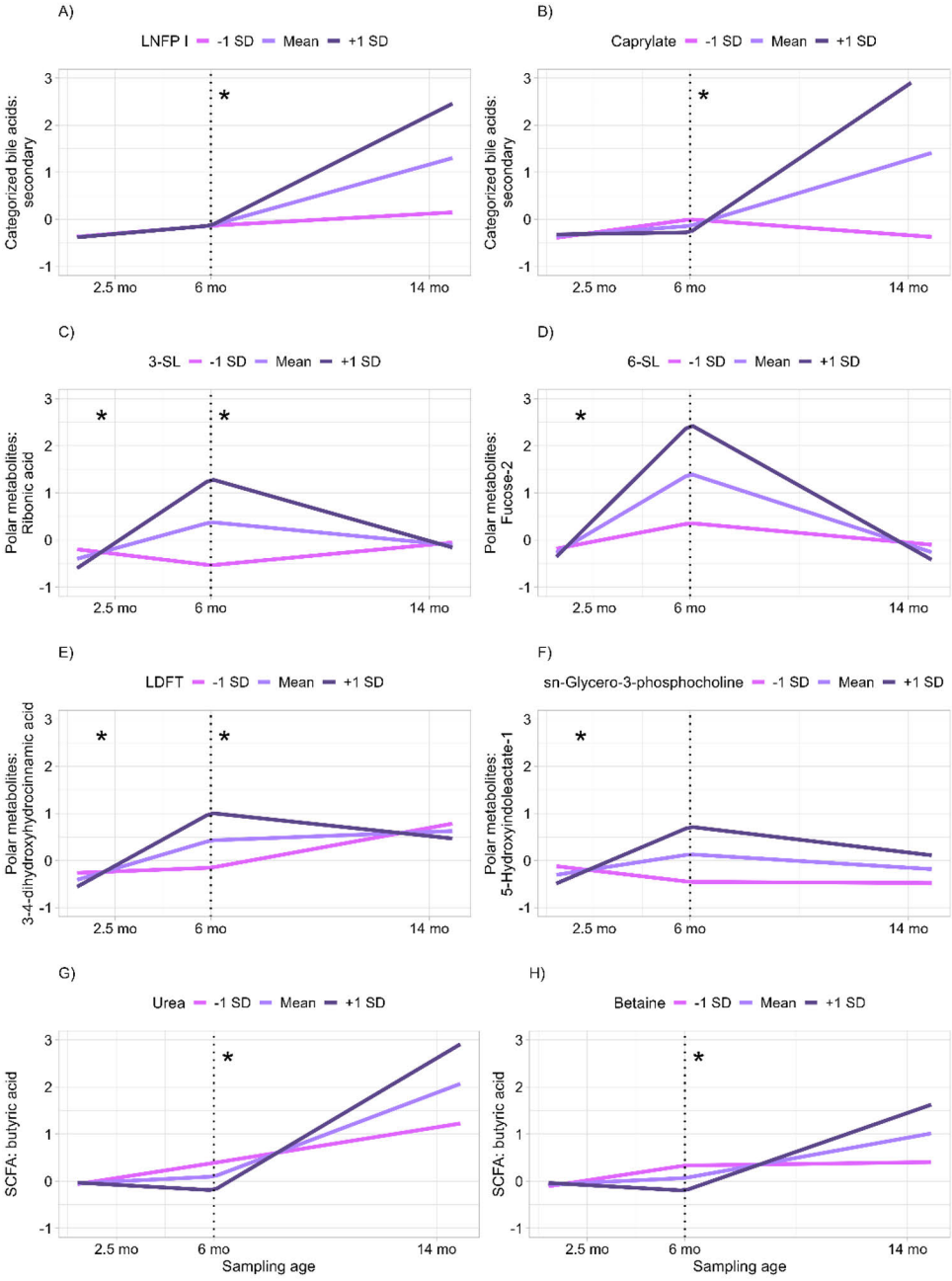


Figure 26. Correlations between faecal and milk metabolites by two time-intervals by sampling age. Mixed model was adjusted for maternal BMI, birth month, parity, gestational age, maternal distress, delivery mode, maternal education and age. Milk metabolite (on top of the plots) and faecal metabolites (y-axis) presented with mean values and standard deviations. Asterisks (*) indicate adjusted p-values <0.05. From the study IV manuscript.

6 Discussion

This thesis utilized next generation sequencing and metabolomics in human microbiome studies. The thesis has two main research orientations, which focuses on methodological improvements and description of early life developmental patterns in gut microbiome and metabolome together milk metabolites. Method optimization aimed to develop reliable preanalytical methods in metagenomics and metabolomics for clinical microbiome samples. Early life studies aimed to investigate the association and relation between gut microbiome development, faecal metabolites and human breast milk metabolome in early childhood.

6.1 Methodological choices (studies I and II) associate with bacterial genera and microbial metabolites

As it is known that choices of laboratory methods can influence results of microbiome composition and faecal metabolites, it is vital to have a critical consideration when setting up new sample collections and laboratory analysis. Follow up-studies are challenging methodologically, since methods are rapidly changing, and reproducibility is required for sample comparability. Therefore, it is important to test the effect of sample collection matrix and DNA extraction protocol with necessary pre-treatment procedure.

Study I showed that bead beating is required for the coverage of hard-to-lyse bacteria. Both preservatives OMNIgeneGUT and DNA/RNA shield yielded sufficient amount of intact DNA although slight differences were found in bacterial composition. The relatively elevated read count in the negative controls can be attributed to the 96-well plate format, where samples are in close proximity and aerosols cannot be completely avoided. Cross-contamination seems evident in the negative control read counts, the relative abundances of the controls, and the beta diversity patterns, which resembled those of the faecal samples. The contamination from faecal sample to another faecal sample was not seen, however, this would be difficult to detect in this study design and therefore was not properly tested. Five from 66 negative controls exceeded the limit value for the faecal samples. In such cases, those negative samples should be closely examined and consider re-analysis

in laboratory if possible. Moreover, it is notable that these laboratory methods require trained and well-experienced staff to minimize human errors and contamination.

Our experiments on sample preservatives showed that preservation is needed for inhibiting the microbial activity when the immediate freezing is not option. Moreover, using the sample collection kit is feasible when sampling is done at home. Although, individual variation dominated the effect of methodological choices, there were some alterations in microbiome by preservative, such as higher abundances of *Bacteroides* and *Faecalibacterium prausnitzii*. It should be noted that preservative may have a favourable effect to some bacteria abundances. In the metabolomics, alterations were also seen, but partly for different reasons. Interestingly, higher levels of BAs were detected during the ethanol-based storage. This may be explained by the fact that ethanol is not only preservative but can also be used as crushing solvent for metabolites and BAs were isolated already during the storage time, which should be noted further in analysis.

Previous studies have shown controversies on feasibility of ethanol preservation for microbiome. This might be originated from bias of testing only one individual sample that may not represent generalized result. OMNIgeneGUT was slightly better than 95% EtOH at preserving microbiota based on alpha diversity, however, on beta diversity EtOH was comparable to OMNIgeneGUT. Based on high individual differences of studied microbiome and metabolic profiles, it can be proposed that methodological testing requires more than one or two subjects and ideally from different sex and age-groups. Previous controversies in feasible methods can originate from limited sample sizes and chosen targeted metabolites. Lim et al. (2020) suggested that it is possible to extract both microbiome and metabolite data from a single sample using OMNIgeneGUT, although they noted this approach has several limitations for metabolite analysis. In contrast, Wang et al. (2018) found poor detection level of metabolites (34%) with OMNIgeneGUT. Therefore, choice of method also depends on the research interests. However, both studies conclude that all studied preservatives were comparable to flash freezing. This raises a question regarding the interpretation of the results and how do we define comparability.

In study II, it was determined that storing faecal samples at RT and stabilizing them in OMNImetGUT or 95% ethanol produced metabolomic profiles that were in generally similar to those stored by flash freezing. Specifically, comparable identities and abundances of observed biochemicals, as well as comparable metabolic profiles of the study subjects, were observed. Additionally, metabolic changes in crude feces over time were characterized, which could originate from microbial activity and nonenzymatic reactions such as oxidation-reduction. Thus, samples can be effectively stored in the tested preservatives at RT for up to 7 days. Using 95% ethanol for faecal collection can provide a more cost-effective method

for collecting samples at home and storing them during the sample logistics. The overall composition of the microbiome was primarily influenced by individual differences rather than the storage type. However, OMNIgeneGUT was slightly more effective than 95% ethanol in preserving microbiota based on alpha diversity. Further investigation into an existing commercial kit is ongoing, which may enhance the assessment of the microbiome and metabolome.

6.2 Successional patterns in developing gut microbiota and metabolome (studies III and IV)

The development of gut microbiota follows a successive pattern during early life (Beller et al., 2021), influenced by factors like breastfeeding and delivery mode (Stewart et al., 2018). In our research, this dynamic progression was observed through compositional shifts in which community types clustered by age. However, there is a knowledge gap in the development of faecal metabolites, which are important mediators of the physiological effects of the gut microbiota. Here, in a population-based cohort, it was observed that the faecal metabolome develops alongside the gut microbiota, and individual variation in GM is associated with faecal metabolome composition. Moreover, observations suggest that breastfeeding, a key factor in modulating microbiota, is linked to metabolite concentration, which varies depending on the composition of the gut microbiota. This indicates that the metabolome is connected to the development of the microbiota and that common exposures may have individualized effects based on the specific microbiota composition. Studied trends of faecal metabolites included SCFA, BA and untargeted polar metabolites in addition to showed links between the metabolites and multiple early life exposures thus extending the current understanding on early life gut metabolome and associating factors.

In the studied cohort, breastfeeding drove the microbiota maturation. Breastfed babies were found to have lower alpha diversity in earlier study (Yelverton et al., 2023), therefore, the aim in this work was to investigate further the mechanism behind this. It was seen that gut microbiome and metabolome co-mature during early life, with breastfeeding and early colonizers like *Bacteroides, Escherichia and Bifidobacterium* shaping this process. Although, the associations did not remain in the gut microbiome in later timepoints, it is possible that the starting point for later health issues is seeded by the early life disturbances in the gut and may show up later in life.

The systematic increase in SCFAs might be sourced from the more complex microbiota and increased intake of indigestible fibre with age. This agrees with an earlier study suggesting an increasing stool SCFA trend after birth (Xiong et al., 2022). The only exception was the relatively stable levels of acetic acid, which might

be related to the lack of very early sampling. Secondary BAs increased with age, while primary and tauroconjugated BAs decreased with age, partially in line with previous findings (Lamichhane et al., 2022). The decrease in bile acids might be related to increased bile salt hydrolase (BSH) activity, possibly driven by *Clostridium and Bacteroides* (Wahlström et al., 2016). Interestingly, glycoconjugated BAs did not increase with age when adjusted for breastfeeding. This could be explained by the high breastfeeding rate at the 2.5-month-olds, who harboured more *Bifidobacteria*, often possessing BSH enzymes with a preference for glycine as a substrate over taurine (Brink et al., 2020).

A negative association between *Bifidobacterium* and butyric acid was noted, which is in contrast to Brink et al. (2020) report. However, specific *Bifidobacterium* strains can compete with butyrate producers for the same substrates (Moens et al., 2016; Nguyen et al., 2021), which may explain the discrepancies in strain-level detection between studies. It is possible that *Bifidobacterium*-dominated microbiota can deconjugate glycine earlier, supported by our observation that *Bifidobacterium* was negatively associated with glycoconjugated BA concentration.

This work suggests that breastfed infants with lower *Bifidobacterium* or higher *Bacteroides* abundance have higher concentrations of microbially modified BAs. Both *Bacteroides* and *Bifidobacterium* are key genera in the gut ecosystem of breastfed infants, with different capacities for BA metabolism. Although our study lacks strain-level information, this thesis confirms previous findings that secondary BA concentration is lower in breastfed infants (Khine et al., 2020), possibly due to slower colonization by BA-metabolizing microbiota.

Consistent with other studies, gut microbiota community types generally align with age, reflecting typical early-life colonization patterns. An exception was observed at the 2.5-month timepoint, when most infants were breastfed, showing three dominant community types: *Bifidobacterium and Bacteroides, Veillonella and Enterobacteriaceae, or Escherichia* dominance. The *Bifidobacterium* and *Bacteroides*-dominated community type was associated with lower C-section rates, consistent with existing literature (Reyman et al., 2019; Stewart et al., 2018). Additionally, breastfeeding was linked to slower community type progression, indicating slower gut microbiota maturation (Stewart et al., 2018). Thus, our data support the observation that the cessation of breastfeeding accelerates the maturation of the gut microbiota.

The nuanced metabolite concentrations among community types in 2.5-month-olds highlight the interaction between microbiota, colonization factors, and metabolites. The community type dominated by *Bifidobacterium and Bacteroides*, which included more vaginally delivered infants, had lower concentrations of conjugated BAs compared to the other major types at the first timepoint, possibly due to differences in BSH enzymatic activity. Conversely, the *Bifidobacterium*-

Bacteroides dominated community type had higher concentrations of propionic acid, isobutyric acid, and isovaleric acid than the *Escherichia*-dominated community type, suggesting increased protein availability for microbial fermentation. This difference may stem from variations in human milk composition (Borewicz et al., 2020), as no differences in breastfeeding were observed between community types at the first timepoint.

Even though BF is crucial in shaping the microbiota (Stewart et al., 2018), individual colonization patterns varied even among breastfed infants. Further analysis explored how the interplay between BF and dominant taxa affected metabolite concentrations. As expected, breastfed infants with higher *Bifidobacteria* had lower conjugated BA concentrations, while those with higher Bacteroides abundances had higher secondary BA concentrations. This shows a dynamic interaction between early nutrition, microbiota, and microbial metabolites. Therefore, further investigation into human milk components related to microbiota colonization patterns as substrates for microbial fermentation is warranted for study IV.

BAs play a role in regulating inflammatory and metabolic processes via FXR and other bile acid-responsive receptors. For example, secondary BAs, more abundant in breastfed infants with high *Bacteroides* abundance, may inhibit pro-inflammatory processes in microglia (Joo et al., 2003) and are necessary to activate the vitamin D receptor, which is important for optimal growth and adaptive immunity development (Ahmad et al., 2019; X. Song et al., 2020). Thus, early-life microbiota-bile acid interactions may influence growth and brain health. However, the exact impact of these complex feedback systems on physiological outcomes remains uncertain, as gut metabolites shape postnatal gut microbiota composition (van Best et al., 2020), and tauroconjugated BAs metabolized by gut bacteria could inhibit BA synthesis via FXR antagonism (Sayin et al., 2013).

6.2.1 Human milk metabolites influencing infant gut microbiome and metabolome

This study was set out to investigate the association between human milk metabolites and infant gut metabolites and microbiota. Human milk is known to foster a healthy infant gut microbiome, which in turn produces beneficial microbial metabolites. However, there is limited longitudinal research simultaneously considering human milk composition, gut microbiota, and fecal metabolite profiles.

Overall, it was found that while milk composition strongly reflected maternal secretor status, it was not directly associated with infant gut microbiome alpha or beta diversity. This aligns with previous findings (Laursen, 2021; Thorman et al., 2023). These findings suggest that milk metabolites act in a structure-specific

manner, shaping infant gut metabolism and certain taxa, but not the overall community composition. Moreover, associations differed between early and late infancy, indicating that the impact of milk metabolites on gut microbial metabolism may depend on the maturity of the microbiome and increasing dietary diversity.

As expected, milk HMO composition was influenced by maternal secretor status, primarily due to the absence of specific metabolites such as 2'-fucosyllactose (2'-FL), lacto-N-difucohexaose I (LNDFH I), and lacto-difucotetraose (LDFT). Interestingly, non-secretor mothers had higher concentrations of 3-FL. Although 3-FL has been reported to exert prebiotic, immunomodulatory, antiadhesive, and antiviral properties (Z. Li et al., 2023), our results indicated that 2'-FL and 3-FL had opposing associations with fecal metabolites, such as secondary BAs. Thus, compensation by higher 3-FL may not fully address the functional roles of 2'-FL. Prior work has suggested that infants of non-secretor mothers may benefit from extended breastfeeding due to lower levels of protective HMOs (Salamone & Nardo, 2020). This idea is supported by studies showing that high 2'-FL levels are associated with lower risk of *Campylobacter* diarrhea (Morrow et al., 2004) and that specific HMOs such as DSLNT protect against necrotizing enterocolitis in animal models (Jantscher-Krenn et al., 2012). Our findings further highlight the structure-specific nature of HMO effects.

Several specific associations between HMOs and microbial taxa or metabolites were identified. LNFP I, absent in non-secretors, correlated positively with *Bifidobacterium* at 6 months and negatively with *Clostridioides* and *Citrobacter* at 14 months, while also associating with secondary BAs in later infancy, possibly reflecting gut maturation. DSLNT at 2 months correlated positively with *Veillonella* and negatively with *Escherichia*, with later associations to *Hungatella* and *Citrobacter*. Interestingly, 2'-FL and other HMOs did not correlate positively with *Bifidobacterium* but instead negatively with opportunistic taxa such as *Clostridioides* and *Ruminococcus*, suggesting HMOs may also restrict potential pathogens. This is consistent with evidence that diverse bacteria beyond *Bifidobacterium* can metabolize HMOs and their by-products (Chapman et al., 2025).

Beyond HMOs, additional milk metabolites associated with infant gut profiles. Higher milk amino acid levels correlated with branched-chain fatty acids (BCFA), e.g., isobutyric acid at 2 months, reflecting bacterial protein fermentation. While BCFAs have been linked to both protective (anti-inflammatory, anti-carcinogenic) and adverse (excess protein intake, overweight risk) outcomes (Macfarlane & Macfarlane, 2012; Rios-Covian et al., 2020), our findings suggest that bacterial metabolism may partially compensate for infants' limited protein digestion capacity. Similarly, milk lipids such as caprylate was positively associated with glycoconjugated and secondary BAs, especially after weaning, consistent with

increased bile production and microbial metabolism during dietary transitions (Schoeler & Caesar, 2019; Yokota et al., 2012). Urea and betaine were also linked to higher butyrate levels, supporting earlier findings of betaine's beneficial metabolic effects and normal growth patterns in healthy infants as opposed to accelerated growth (Ribo et al., 2021). Additionally, urea, a major source of nitrogen in HM, may affect the gut homeostasis and bacterial metabolism since the commensal bacteria with urease activity can use urea to produce SCFA (Firth et al., 2025).

Of particular interest, sn-glycero-3-phosphocholine (GPC) was negatively associated with *Streptococcus* (2 months) and *Clostridioides* (6 months), but positively correlated with *Bifidobacterium* and *Lactobacillus* (14 months). GPC is both a choline source and a microbial growth substrate since some bacteria possess the necessary enzymes to break down GPC into its constituent parts, which can then be further metabolized for growth (Lewis et al., 2017). However, as some bacteria can convert choline into TMA (a precursor of TMAO), health consequences remain complex (Romano et al., 2015).

It was also observed that full breastfeeding at 2.5 months, compared to partial feeding, was associated with higher levels of fecal 7-oxo-HDCA, wMCA, butyric acid, and *Bifidobacterium*—possible timely hallmarks of healthy gut maturation. In contrast, tauro- and glycine-conjugated BAs were reduced, consistent with earlier findings in study III linking breastfeeding, BA metabolism, and *Bifidobacterium* abundance. These results support the idea that breastfeeding fosters distinct gut metabolic trajectories during critical developmental windows (Stewart et al., 2018).

Taken together, our results highlight that while overall milk composition does not determine infant gut microbiota diversity, specific milk metabolites—particularly individual HMOs, amino acids, lipids, and choline derivatives—are associated with biologically relevant microbial metabolites and taxa. These associations evolve with age and dietary diversification, underscoring the dynamic interplay between milk, microbes, and host development. Importantly, the results also suggest that the absence of certain HMOs in non-secretor mothers does not fundamentally impair infant gut microbiome diversity, which may be reassuring from a clinical and public health perspective.

Finally, these findings should be communicated sensitively. While breastfeeding clearly supports optimal microbiome development, it is not always possible, and infant formula remains the best available alternative. Ongoing improvements to formula composition—such as inclusion of selected HMOs and optimized protein balance—may help bridge some of the functional gaps. From a life-course perspective, these results also reinforce the concept that early-life nutrition interacts with microbiome development in ways that may shape later health outcomes, consistent with the Developmental Origins of Health and Disease (DOHaD) framework (Lacagnina, 2019).

6.3 Strengths and Limitations

Methodological studies supported the international goal of standardized methods and optimized high-through put DNA extraction for upcoming population-based studies and follow-ups. Validated collection methods give an opportunity to stabilize field studies and cost-efficient practice. It was necessary to see why ethanol is not an optimal way to store microbiome and how to take it into account if it is used. Moreover, the methodology studies (I and II) had strength in several studied timepoints, different temperatures and diverse age-range. Although the tested preservatives worked well in downstream analysis, it must be noted that bacterial cultivation is no longer possible using those preservatives. This thesis gave us novel insights how the faecal metabolites together with microbiome maturate in early life. Study III and IV had strength in using the unique longitudinal multi-omics data.

However, one major limitation in the thesis was that the method development was done retrospectively, after the cohort samples (used in studies III and IV) were collected and extracted so the optimized methods could not be utilized with those samples. The 16S rRNA sequencing used in the studies has certain limitation on taxonomic capacity and resolution since it cannot reach a species level. Methodological experiments were done with low number of samples which limits the power of the results.

Nevertheless, this thesis benefits from a large and diverse sample population of children, with variations in breastfeeding and delivery modes that represent the Finnish population. However, the sample collection timepoints do not extend to the neonatal period, nor was the sampling frequency high. These limitations may have restricted our ability to detect more nuanced patterns in colonization and metabolome development. Additionally, our sample consisted mostly of infants and children who received some breastmilk, and there is not corresponding sample size in exclusively formula-fed group nor metabolites from formula. The 16s rRNA sequencing used in this thesis provided valuable insights into the overall microbiota profiles. However, the findings highlight the need for future research to focus on gene-level differences in gut microbiota. Leveraging metagenomic sequencing in future research will enhance our understanding of the role of bile acid metabolizing capacity in the developing gut microbiome. Additionally, more comprehensive data on early diet will be beneficial, such as food diary of infant solid food, could also help to describe the differences in microbiota composition and the functional output. Integrating the reported exploratory findings into mechanistic models in future studies will help clarify the clinical potential related to inflammation (Devkota et al., 2012; X. Song et al., 2020) and metabolic programming (Wahlström et al., 2016). In addition, the FinnBrain cohort samples have been collected from geographically restricted area. The cohort is healthy and skewed to higher socio-economic status leading to a

selection bias which can affect the validity and generalizability of the study's findings (Karlsson et al., 2018).

The studies relied on stool samples. It is previously noted that the microbiota in the outer mucus layer of the colonic mucosa differs from the luminal microbiota in the same division, regardless of health or disease status (Zoetendal et al., 2002). The inner mucus layer and crypts, which contain intestinal stem cells, are traditionally thought to be free of bacteria. However, this has been challenged by the discovery of crypt-associated microbiota in mice (Pédron et al., 2012). Despite this limitation, most gut microbiota studies use stool samples, which are easy to collect at home in a non-invasive manner and are considered to reflect overall variations in colonic microbiota. Therefore, it is understandable that there has been a debate about whether we should talk about faecal microbiota instead of gut microbiota in this context.

6.4 Future perspectives

Given the exponential growth in the number of publications in the microbiome field, it can be argued that a strong foundation has been established for future research. Although the field is progressing, standardized methods are still lacking, and methodological development continues to play a critical role in microbiome studies. Cost-effectiveness in sample collection could be further enhanced. Faecal cards may represent a viable solution in future studies, provided that optimal preservation conditions are achieved.

Gut maturation has been studied to such an extent that we have an outline of developmental trajectories. However, we will have more knowledge of functionality and species level differences by shotgun metagenomics. As the cohort gets older, more health outcomes can be investigated, and later life health can be evaluated in contrast to early life factors. Early life exposures may have lifelong consequences which can be partly mediated through the gut microbiota and its metabolites. In future research, establishing "reference values" for the microbiome would be advantageous for assessing the risk levels associated with dysbiotic gut conditions and the potential onset of related diseases. For studies involving infants, understanding the characteristics of enterotypes or dominant microbial community types could elucidate the buffering effects of the microbiome against harmful disturbances. The dynamics of the gut microbiome in early life may be analogous to the plasticity observed in brain and immune system development; adaptation can be robust and pronounced during rapid developmental phases. However, excessively rapid maturation may result in health complications later in life. Consequently, it would be valuable to investigate the buffering mechanisms that facilitate a return to a healthy state.

7 Conclusions

I The optimized protocol fulfilled the need of feasible handling of faecal samples, comparability, coverage of hard-lyse bacteria and minimized contamination level for automated DNA extraction for downstream sequencing applications and handling of large sample collections.

II Ethanol stabilized the fecal metabolites as well as OMNImetGUT. The microbiome was preserved with ethanol as well, however, OMNIgeneGUT was more optimal based on alpha diversity on longer RT storage. Therefore, separate collection tubes are recommended for longer storage times.

III The longitudinal patterns of gut metabolites during early life were closely linked to gut microbiota composition, particularly in breastfed infants. SCFA concentrations, except acetic acid, increased within the first 30 months. Secondary BA concentrations were lower in breastfed infants. The gut microbiota showed progressive maturation during the first 30 months of life. The abundance of prevalent gut microbes was associated with metabolite levels, especially in 2.5-month-olds. Associations between early colonizers (*Bacteroides, Escherichia, and Bifidobacterium)* and microbial BAs were observed, particularly in breastfed infants. Future studies may find that alterations in early-life BA-microbiota interactions are important mechanisms in the developmental programming of health.

IV Milk composition clustered by secretor status, however the overall milk metabolome did not drive the differences in gut microbiome. These results highlight that human milk composition associate not only with abundances of individual bacterial genera, but also with gut microbial metabolites. Importantly, milk composition association with microbiome was different in early versus late infancy, indicating that milk metabolome may have differential effect on microbial metabolism depending on the diversity of diet and/or maturity of microbiome. Milk metabolite-faecal metabolite associations indicated healthy gut maturation.

Acknowledgements

This thesis work was carried out at the Research Center for Infections and Immunity, Institute of Biomedicine, Centre for Population Health Research and Turku Bioscience Centre at University of Turku, Finland. First, I want to thank the professors of microbiology, Jaana Vuopio, Pentti Huovinen and Jukka Hytönen, for allowing me to carry out this work in the subject. Special thanks to Pentti for being my mentor and a pioneer of bacteriology for decades. I would like to extend my greatest thanks to mother and father of FinnBrain Birth Cohort study, Linnea and Hasse Karlsson for the precious work they have done for the study group and beyond.

This thesis project has been an amazing learning opportunity and I was extremely lucky when I got Anna Aatsinki, Alex Dickens and Teemu Kallonen as my supervisors. You are very talented, supportive and encouraging in every way and I would not have been able to complete this thesis without your endless support and faith in me when I needed it most. Eveliina Munukka and Arto Pulliainen are acknowledged for the precious help, expertise and shared knowledge in the follow-up committee.

I am grateful of all the support I got from principal investigators of research groups, my colleges and collaborators, Matej Orešič and other members of systems medicine group. My journey started from the "Mikrobistoryhmä" and I am thankful for creating the warm and joyful working environment Antti Hakanen, Marianne Gunell, Sofia Kalinen, Anniina Keskitalo, Janina Heiskanen, Tanja Orpana, Erkki Eerola and other old and new members of the group. I will always remember our cheerful moments and encouraging atmosphere in the "perälabra" so my warmest thanks go to Minna Lamppu, Katri Kylä-Mattila, Päivi Haaranen and Anna Musku. Coffee breaks with medisiinaD7 personnel have been my source of strength during heavy days, so thank you all for maintaining the meaningful and encouraging atmosphere. In addition, FinnBrain and POPC staff have supported my journey with their enthusiasm and wisdom, so thank you Susanna Kortesluoma, Susanne Sinisalo, Eija Jossandt, Saara Nolvi, Venla Huovinen and Katja Pahkala. I want to thank my dear friends who kept cheering me up all these years and helped me maintain a work-life balance, thank you Salla, Jutta, Petu and my wolfpack Melina, Saara and Elina.

Many thanks to Tarja and Iivari for taking care of our children and Hertta-doggie; congress trips would not have been possible without your help.

I would like to express the humblest thanks to co-authors Santosh Lamichhane, Sanja Vanhatalo, Natalie Tomnikov, Laura Perasto, Leo Lahti, Henna-Maria Kailanto, Matilda Kråkström, Lucas Pinto Da Silva, Marina Alvares, Rob Knight, Naama Karu, Rima Kaddurah-Daouk, Thomas Hankemeier, Leyla Schimmel, Edgar Diaz, Tuulia Hyötyläinen, Pieter Dorrestein and Minka Ovaska. I appreciate your huge input in the sub studies. It has been an honour to have Marko Lehtonen and Mirjam Bloemendaal as preliminary examiners for my thesis. With their expertise the book got improved tremendously. I was delighted when I heard that Mikael Niku will be my opponent. He has impressive competence in the field and I am confident that we will have a fruitful discussion during the dissertation.

I want to thank FinnBrain families and my funders making this thesis work possible: the Finnish cultural foundation, Juho Vainio foundation, Foundation for the Advancement of Laboratory Medicine, Doctoral program of clinical research (DPCR), the state research funding of Turku University Hospital and other funders around the project: The Finnish Foundation for Cardiovascular Research, The Diabetes Research Foundation and Signe and Ane Gyllenberg foundation. In addition, I am thankful for the travel grants I received from TUMBTS, DPCR and Turku university foundation.

My heartfelt thanks go to my husband Santeri. You have been my rock of faith, love and endless support on this demanding journey. I want to thank my mom, dad and my little sister Heli for always being there for me. And my lovely children, Leo and Sonja, you are my precious, the greatest joy in my life and you keep me going even on rainy days.

18.8.2025

Heidi Isokääntä

References

Abdelhameed, F., Mustafa, A., Kite, C., Lagojda, L., Dallaway, A., Than, N. N., Kassi, E., Kyrou, I., & Randeva, H. S. (2025). Gut Microbiota and Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD): Emerging Pathogenic Mechanisms and Therapeutic Implications. *Livers*, *5*(1), Article 1. https://doi.org/10.3390/livers5010011

Abdill, R. J., Graham, S. P., Rubinetti, V., Ahmadian, M., Hicks, P., Chetty, A., McDonald, D., Ferretti, P., Gibbons, E., Rossi, M., Krishnan, A., Albert, F. W., Greene, C. S., Davis, S., & Blekhman, R. (2025). Integration of 168,000 samples reveals global patterns of the human gut microbiome. *Cell*, *0*(0). https://doi.org/10.1016/j.cell.2024.12.017

Ahmad, O., Nogueira, J., Heubi, J. E., Setchell, K. D. R., & Ashraf, A. P. (2019). Bile Acid Synthesis Disorder Masquerading as Intractable Vitamin D-Deficiency Rickets. *Journal of the Endocrine Society*, *3*(2), 397–402. https://doi.org/10.1210/js.2018-00314

Aird, D., Ross, M. G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D. B., Nusbaum, C., & Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, *12*(2), R18. https://doi.org/10.1186/gb-2011-12-2-r18

Alatab, S., Sepanlou, S. G., Ikuta, K., Vahedi, H., Bisignano, C., Safiri, S., Sadeghi, A., Nixon, M. R., Abdoli, A., Abolhassani, H., Alipour, V., Almadi, M. A. H., Almasi-Hashiani, A., Anushiravani, A., Arabloo, J., Atique, S., Awasthi, A., Badawi, A., Baig, A. A. A., … Naghavi, M. (2020). The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Gastroenterology & Hepatology*, *5*(1), 17–30. https://doi.org/10.1016/S2468-1253(19)30333-4

Allin, K. H., Tremaroli, V., Caesar, R., Jensen, B. A. H., Damgaard, M. T. F., Bahl, M. I., Licht, T. R., Hansen, T. H., Nielsen, T., Dantoft, T. M., Linneberg, A., Jørgensen, T., Vestergaard, H., Kristiansen, K., Franks, P. W., Hansen, T., Bäckhed, F., Pedersen, O., & the IMI-DIRECT consortium. (2018). Aberrant intestinal microbiota in individuals with prediabetes. *Diabetologia*, *61*(4), 810–820. https://doi.org/10.1007/s00125-018-4550-1

Ames, S. R., Lotoski, L. C., & Azad, M. B. (2023). Comparing early life nutritional sources and human milk feeding practices: Personalized and dynamic nutrition supports infant gut microbiome development and immune system maturation. *Gut Microbes*, *15*(1), 2190305. https://doi.org/10.1080/19490976.2023.2190305

Anukam, K. C., & Reid, G. (2007). Probiotics: 100 years (1907-2007) after Elie Metchnikoff's Observation. . . *M*.

Arnoldussen, I. a. C., Wiesmann, M., Pelgrim, C. E., Wielemaker, E. M., van Duyvenvoorde, W., Amaral-Santos, P. L., Verschuren, L., Keijser, B. J. F., Heerschap, A., Kleemann, R., Wielinga, P. Y., & Kiliaan, A. J. (2017). Butyrate restores HFD-induced adaptations in brain function and metabolism in mid-adult obese mice. *International Journal of Obesity*, *41*(6), 935–944. https://doi.org/10.1038/ijo.2017.52

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L.,

Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., … Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature*, *473*(7346), 174–180. https://doi.org/10.1038/nature09944

Azad, M. B., Robertson, B., Atakora, F., Becker, A. B., Subbarao, P., Moraes, T. J., Mandhane, P. J., Turvey, S. E., Lefebvre, D. L., Sears, M. R., & Bode, L. (2018). Human Milk Oligosaccharide Concentrations Are Associated with Multiple Fixed and Modifiable Maternal Characteristics, Environmental Factors, and Feeding Practices. *The Journal of Nutrition*, *148*(11), 1733–1742. https://doi.org/10.1093/jn/nxy175

Azad, M. B., Wade, K. H., & Timpson, N. J. (2018). FUT2 secretor genotype and susceptibility to infections and chronic conditions in the ALSPAC cohort. *Wellcome Open Research*, *3*, 65. https://doi.org/10.12688/wellcomeopenres.14636.2

Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., Khan, M. T., Zhang, J., Li, J., Xiao, L., Al-Aama, J., Zhang, D., Lee, Y. S., Kotowska, D., Colding, C., … Wang, J. (2015). Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host & Microbe*, *17*(5), 690–703. https://doi.org/10.1016/j.chom.2015.04.004

Backman, H., Räisänen, P., Hedman, L., Stridsman, C., Andersson, M., Lindberg, A., Lundbäck, B., & Rönmark, E. (2017). Increased prevalence of allergic asthma from 1996 to 2006 and further to 2016—Results from three population surveys. *Clinical & Experimental Allergy*, *47*(11), 1426–1435. https://doi.org/10.1111/cea.12963

Banchi, P., Colitti, B., Opsomer, G., Rota, A., & Soom, A. V. (2024). *The dogma of the sterile uterus revisited: Does microbial seeding occur during fetal life in humans and animals?* https://doi.org/10.1530/REP-23-0078

Barko, P., Nguyen-Edquilang, J., Williams, D. A., & Gal, A. (2024). Fecal microbiome composition and diversity of cryopreserved canine stool at different duration and storage conditions. *PLOS ONE*, *19*(2), e0294730. https://doi.org/10.1371/journal.pone.0294730

Bartolomaeus, T. U. P., Birkner, T., Bartolomaeus, H., Löber, U., Avery, E. G., Mähler, A., Weber, D., Kochlik, B., Balogh, A., Wilck, N., Boschmann, M., Müller, D. N., Markó, L., & Forslund, S. K. (2021). Quantifying technical confounders in microbiome studies. *Cardiovascular Research*, *117*(3), 863–875. https://doi.org/10.1093/cvr/cvaa128

Baudoin, L., Sapinho, D., Maddi, A., & Miotti, L. (2019). Scientometric analysis of the term "microbiota" in research publications (1999–2017): A second youth of a century-old concept. *FEMS Microbiology Letters*, *366*(12), fnz138. https://doi.org/10.1093/femsle/fnz138

Beller, L., Deboutte, W., Falony, G., Vieira-Silva, S., Tito, R. Y., Valles-Colomer, M., Rymenans, L., Jansen, D., Van Espen, L., Papadaki, M. I., Shi, C., Yinda, C. K., Zeller, M., Faust, K., Van Ranst, M., Raes, J., & Matthijnssens, J. (2021). Successional Stages in Infant Gut Microbiota Maturation. *mBio*, *12*(6), e01857-21. https://doi.org/10.1128/mbio.01857-21

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C. C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G. H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J. A., Maguin, E., Mauchline, T., McClure, R., … Schloter, M. (2020). Microbiome definition re-visited: Old concepts and new challenges. *Microbiome*, *8*(1), 103. https://doi.org/10.1186/s40168-020-00875-0

Bharucha, T., Oeser, C., Balloux, F., Brown, J. R., Carbo, E. C., Charlett, A., Chiu, C. Y., Claas, E. C. J., de Goffau, M. C., de Vries, J. J. C., Eloit, M., Hopkins, S., Huggett, J. F., MacCannell, D., Morfopoulou, S., Nath, A., O'Sullivan, D. M., Reoma, L. B., Shaw, L. P., … Field, N. (2020). STROBE-metagenomics: A STROBE extension statement to guide the reporting of metagenomics studies. *The Lancet Infectious Diseases*, *20*(10), e251–e260. https://doi.org/10.1016/S1473-3099(20)30199-7

Biswas, S. K. (2016). Does the Interdependence between Oxidative Stress and Inflammation Explain the Antioxidant Paradox? *Oxidative Medicine and Cellular Longevity*, *2016*(1), 5698931. https://doi.org/10.1155/2016/5698931

Blacher, E., Levy, M., Tatirovsky, E., & Elinav, E. (2017). Microbiome-Modulated Metabolites at the Interface of Host Immunity. *The Journal of Immunology*, *198*(2), 572–580. https://doi.org/10.4049/jimmunol.1601247

Blaser, M. J., & Falkow, S. (2009). What are the consequences of the disappearing human microbiota? *Nature Reviews Microbiology*, *7*(12), 887–894. https://doi.org/10.1038/nrmicro2245

Bloomfield, S. F., Rook, G. A., Scott, E. A., Shanahan, F., Stanwell-Smith, R., & Turner, P. (2016). Time to abandon the hygiene hypothesis: New perspectives on allergic disease, the human microbiome, infectious disease prevention and the role of targeted hygiene. *Perspectives in Public Health*, *136*(4), 213–224. https://doi.org/10.1177/1757913916650225

Boets, E., Gomand, S. V., Deroover, L., Preston, T., Vermeulen, K., De Preter, V., Hamer, H. M., Van den Mooter, G., De Vuyst, L., Courtin, C. M., Annaert, P., Delcour, J. A., & Verbeke, K. A. (2017). Systemic availability and metabolism of colonic-derived short-chain fatty acids in healthy subjects: A stable isotope study. *The Journal of Physiology*, *595*(2), 541–555. https://doi.org/10.1113/JP272613

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Borewicz, K., Gu, F., Saccenti, E., Hechler, C., Beijers, R., de Weerth, C., van Leeuwen, S. S., Schols, H. A., & Smidt, H. (2020). The association between breastmilk oligosaccharides and faecal microbiota in healthy breastfed infants at two, six, and twelve weeks of age. *Scientific Reports*, *10*(1), 4270. https://doi.org/10.1038/s41598-020-61024-z

Borman, Lahti, Shetty, & Erns. (2024). *Orchestrating Microbiome Analysis with Bioconductor*. https://microbiome.github.io/OMA/

Brennan, C. A., & Garrett, W. S. (2016). Gut Microbiota, Inflammation, and Colorectal Cancer. *Annual Review of Microbiology*, *70*, 395–411. https://doi.org/10.1146/annurev-micro-102215-095513

Brink, L. R., Mercer, K. E., Piccolo, B. D., Chintapalli, S. V., Elolimy, A., Bowlin, A. K., Matazel, K. S., Pack, L., Adams, S. H., Shankar, K., Badger, T. M., Andres, A., & Yeruva, L. (2020). Neonatal diet alters fecal microbiota and metabolome profiles at different ages in infants fed breast milk or formula. *The American Journal of Clinical Nutrition*, *111*(6), 1190–1202. https://doi.org/10.1093/ajcn/nqaa076

Browning, M. G., & Campos, G. M. (2017). Bile acid physiology as the potential driver for the sustained metabolic improvements with bariatric surgery. *Surgery for Obesity and Related Diseases*, *13*(9), 1553–1554. https://doi.org/10.1016/j.soard.2017.06.005

Brul, S., Mensonides, F. I. C., Hellingwerf, K. J., & Teixeira de Mattos, M. J. (2008). Microbial systems biology: New frontiers open to predictive microbiology. *International Journal of Food Microbiology*, *128*(1), 16–21. https://doi.org/10.1016/j.ijfoodmicro.2008.04.029

Bull, M. J., & Plummer, N. T. (2014). Part 1: The Human Gut Microbiome in Health and Disease. *Integrative Medicine: A Clinician's Journal*, *13*(6), 17–22.

Chakravorty, S., Helb, D., Burday, M., Connell, N., & Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods*, *69*(2), 330–339. https://doi.org/10.1016/j.mimet.2007.02.005

Chapman, J. A., Masi, A. C., Beck, L. C., Watson, H., Young, G. R., Browne, H. P., Shao, Y., Kiu, R., Nelson, A., Doyle, J. A., Palmowski, P., Lengyel, M., Connolly, J. P. R., Lamb, C. A., Porter, A., Lawley, T. D., Hall, L. J., Embleton, N. D., Perry, J. D., … Stewart, C. J. (2025). *Human milk oligosaccharide metabolism by Clostridium species suppresses inflammation and pathogen growth* (p. 2025.01.21.633585). bioRxiv. https://doi.org/10.1101/2025.01.21.633585

Chen, T., Long, W., Zhang, C., Liu, S., Zhao, L., & Hamaker, B. R. (2017). Fiber-utilizing capacity varies in Prevotella- versus Bacteroides-dominated gut microbiota. *Scientific Reports*, *7*(1), 2594. https://doi.org/10.1038/s41598-017-02995-4

Chen, Y., Fan, L., Chai, Y., & Xu, J. (2022). Advantages and challenges of metagenomic sequencing for the diagnosis of pulmonary infectious diseases. *The Clinical Respiratory Journal*, *16*(10), 646–656. https://doi.org/10.1111/crj.13538

Chen, Z., Hui, P. C., Hui, M., Yeoh, Y. K., Wong, P. Y., Chan, M. C. W., Wong, M. C. S., Ng, S. C., Chan, F. K. L., & Chan, P. K. S. (2019). Impact of Preservation Method and 16S rRNA Hypervariable Region on Gut Microbiota Profiling. *mSystems*, *4*(1), e00271-18. https://doi.org/10.1128/mSystems.00271-18

Chiu, O., Tal, M., Sanmugam, A., Hesta, M., Gomez, D. E., Weese, J. S., & Verbrugghe, A. (2023). The effects of ambient temperature exposure on feline fecal metabolome. *Frontiers in Veterinary Science*, *10*. https://doi.org/10.3389/fvets.2023.1141881

Choo, J. M., Leong, L. E., & Rogers, G. B. (2015). Sample storage conditions significantly influence faecal microbiome profiles. *Scientific Reports*, *5*(1), Article 1. https://doi.org/10.1038/srep16350

Chukwudulue, U. M., Barger, N., Dubovis, M., & Luzzatto Knaan, T. (2023). Natural Products and Pharmacological Properties of Symbiotic Bacillota (Firmicutes) of Marine Macroalgae. *Marine Drugs*, *21*(11), Article 11. https://doi.org/10.3390/md21110569

Chun, J., & Toldi, G. (2022). The Impact of Short-Chain Fatty Acids on Neonatal Regulatory T Cells. *Nutrients*, *14*(18), Article 18. https://doi.org/10.3390/nu14183670

Clooney, A. G., Fouhy, F., Sleator, R. D., O' Driscoll, A., Stanton, C., Cotter, P. D., & Claesson, M. J. (2016). Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis. *PloS One*, *11*(2), e0148028. https://doi.org/10.1371/journal.pone.0148028

Costea, P. I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M. J., Bushman, F. D., de Vos, W. M., Ehrlich, S. D., Fraser, C. M., Hattori, M., Huttenhower, C., Jeffery, I. B., Knights, D., Lewis, J. D., Ley, R. E., Ochman, H., O'Toole, P. W., Quince, C., Relman, D. A., … Bork, P. (2018). Enterotypes in the landscape of gut microbial community composition. *Nature Microbiology*, *3*(1), 8–16. https://doi.org/10.1038/s41564-017-0072-8

Costea, P. I., Zeller, G., Sunagawa, S., Pelletier, E., Alberti, A., Levenez, F., Tramontano, M., Driessen, M., Hercog, R., Jung, F.-E., Kultima, J. R., Hayward, M. R., Coelho, L. P., Allen-Vercoe, E., Bertrand, L., Blaut, M., Brown, J. R. M., Carton, T., Cools-Portier, S., … Bork, P. (2017). Towards standards for human fecal sample processing in metagenomic studies. *Nature Biotechnology*, *35*(11), 1069–1076. https://doi.org/10.1038/nbt.3960

Couper, L., & Swei, A. (2018). Tick Microbiome Characterization by Next-Generation 16S rRNA Amplicon Sequencing. *Journal of Visualized Experiments (JoVE)*, *138*, e58239. https://doi.org/10.3791/58239

Cummings, J. h., & Macfarlane, G. t. (1991). The control and consequences of bacterial fermentation in the human colon. *Journal of Applied Bacteriology*, *70*(6), 443–459. https://doi.org/10.1111/j.1365-2672.1991.tb02739.x

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., & Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), 559–563. https://doi.org/10.1038/nature12820

de Goffau, M. C., Jallow, A. T., Sanyang, C., Prentice, A. M., Meagher, N., Price, D. J., Revill, P. A., Parkhill, J., Pereira, D. I. A., & Wagner, J. (2022). Gut microbiomes from Gambian infants reveal the development of a non-industrialized Prevotella-based trophic network. *Nature Microbiology*, *7*(1), Article 1. https://doi.org/10.1038/s41564-021-01023-6

de Goffau, M. C., Lager, S., Sovio, U., Gaccioli, F., Cook, E., Peacock, S. J., Parkhill, J., Charnock-Jones, D. S., & Smith, G. C. S. (2019). Human placenta has no microbiome but can contain potential pathogens. *Nature*, *572*(7769), 329–334. https://doi.org/10.1038/s41586-019-1451-5

de Moraes, A. C. F., Fernandes, G. R., da Silva, I. T., Almeida-Pititto, B., Gomes, E. P., Pereira, A. da C., & Ferreira, S. R. G. (2017). Enterotype May Drive the Dietary-Associated Cardiometabolic Risk Factors. *Frontiers in Cellular and Infection Microbiology*, *7*, 47. https://doi.org/10.3389/fcimb.2017.00047

de Vos, W. M., Tilg, H., Van Hul, M., & Cani, P. D. (2022). Gut microbiome and health: Mechanistic insights. *Gut*, *71*(5), 1020–1032. https://doi.org/10.1136/gutjnl-2021-326789

Debelius, J., Song, S. J., Vazquez-Baeza, Y., Xu, Z. Z., Gonzalez, A., & Knight, R. (2016). Tiny microbes, enormous impacts: What matters in gut microbiome studies? *Genome Biology*, *17*(1), 217. https://doi.org/10.1186/s13059-016-1086-x

Delzenne, N. M., Cani, P. D., Everard, A., Neyrinck, A. M., & Bindels, L. B. (2015). Gut microorganisms as promising targets for the management of type 2 diabetes. *Diabetologia*, *58*(10), 2206–2217. https://doi.org/10.1007/s00125-015-3712-7

Deng, Y., Umbach, A. K., & Neufeld, J. D. (2024). Nonparametric richness estimators Chao1 and ACE must not be used with amplicon sequence variant data. *The ISME Journal*, *18*(1), wrae106. https://doi.org/10.1093/ismejo/wrae106

Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., Pudlo, N. A., Kitamoto, S., Terrapon, N., Muller, A., Young, V. B., Henrissat, B., Wilmes, P., Stappenbeck, T. S., Núñez, G., & Martens, E. C. (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell*, *167*(5), 1339-1353.e21. https://doi.org/10.1016/j.cell.2016.10.043

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., & Andersen, G. L. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, *72*(7), 5069–5072. https://doi.org/10.1128/AEM.03006-05

Devkota, S., Wang, Y., Musch, M. W., Leone, V., Fehlner-Peach, H., Nadimpalli, A., Antonopoulos, D. A., Jabri, B., & Chang, E. B. (2012). Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10−/− mice. *Nature*, *487*(7405), 104–108. https://doi.org/10.1038/nature11225

Dominianni, C., Wu, J., Hayes, R. B., & Ahn, J. (2014). Comparison of methods for fecal microbiome biospecimen collection. *BMC Microbiology*, *14*(1), 103. https://doi.org/10.1186/1471-2180-14-103

Dougherty, M. W., Kudin, O., Mühlbauer, M., Neu, J., Gharaibeh, R. Z., & Jobin, C. (2020). Gut microbiota maturation during early human life induces enterocyte proliferation via microbial metabolites. *BMC Microbiology*, *20*(1), 205. https://doi.org/10.1186/s12866-020-01892-7

Dror, D. K., & Allen, L. H. (2018). Overview of Nutrients in Human Milk. *Advances in Nutrition*, *9*, 278S-294S. https://doi.org/10.1093/advances/nmy022

Eiseman, B., Silen, W., Bascom, G. S., & Kauvar, A. J. (1958). Fecal enema as an adjunct in the treatment of pseudomembranous enterocolitis. *Surgery*, *44*(5), 854–859.

Esch, B. C. A. M. van, Porbahaie, M., Abbring, S., Garssen, J., Potaczek, D. P., Savelkoul, H. F. J., & Neerven, R. J. J. van. (2020). The Impact of Milk and Its Components on Epigenetic Programming of Immune Function in Early Life and Beyond: Implications for Allergy and Asthma. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/fimmu.2020.02141

European Commission. (2024). *Living guidelines on the responsible use of generative AI in research | Research and innovation*. https://research-and-innovation.ec.europa.eu/document/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en

Ezzati, M., Hoorn, S. V., Lawes, C. M. M., Leach, R., James, W. P. T., Lopez, A. D., Rodgers, A., & Murray, C. J. L. (2005). Rethinking the "Diseases of Affluence" Paradigm: Global Patterns of Nutritional Risks in Relation to Economic Development. *PLOS Medicine*, *2*(5), e133. https://doi.org/10.1371/journal.pmed.0020133

Fachrul, M., Méric, G., Inouye, M., Pamp, S. J., & Salim, A. (2022). *Assessing and removing the effect of unwanted technical variations in microbiome data.* https://doi.org/10.1038/s41598-022-26141-x

Fahur Bottino, G., Bonham, K. S., Patel, F., McCann, S., Zieff, M., Naspolini, N., Ho, D., Portlock, T., Joos, R., Midani, F. S., Schüroff, P., Das, A., Shennon, I., Wilson, B. C., O'Sullivan, J. M., Britton, R. A., Murray, D. M., Kiely, M. E., Taddei, C. R., … Klepac-Ceraj, V. (2025). Early life microbial

succession in the gut follows common patterns in humans across the globe. *Nature Communications*, *16*(1), 660. https://doi.org/10.1038/s41467-025-56072-w

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M. J., Valles-Colomer, M., Vandeputte, D., Tito, R. Y., Chaffron, S., Rymenans, L., Verspecht, C., De Sutter, L., Lima-Mendez, G., D'hoe, K., Jonckheere, K., Homola, D., … Raes, J. (2016). Population-level analysis of gut microbiome variation. *Science*, *352*(6285), 560–564. https://doi.org/10.1126/science.aad3503

Farré-Maduell, E., & Casals-Pascual, C. (2019). The origins of gut microbiome research in Europe: From Escherich to Nissle. *Human Microbiome Journal*, *14*, 100065. https://doi.org/10.1016/j.humic.2019.100065

Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G., & Gloor, G. B. (2013). ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLOS ONE*, *8*(7), e67019. https://doi.org/10.1371/journal.pone.0067019

Fernández-Pato, A., Sinha, T., Gacesa, R., Andreu-Sánchez, S., Gois, M. F. B., Gelderloos-Arends, J., Jansen, D. B. H., Kruk, M., Jaeger, M., Joosten, L. A. B., Netea, M. G., Weersma, R. K., Wijmenga, C., Harmsen, H. J. M., Fu, J., Zhernakova, A., & Kurilshikov, A. (2024). Choice of DNA extraction method affects stool microbiome recovery and subsequent phenotypic association analyses. *Scientific Reports*, *14*(1), 3911. https://doi.org/10.1038/s41598-024-54353-w

Fiehn, O. (2016). Metabolomics by Gas Chromatography–Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Current Protocols in Molecular Biology*, *114*(1), 30.4.1-30.4.32. https://doi.org/10.1002/0471142727.mb3004s114

Firth, I. J., Sim, M. A. R., Fitzgerald, B. G., Moore, A. E., Pittao, C. R., Gianetto-Hill, C., Hess, S., Sweeney, A. R., Allen-Vercoe, E., & Sorbara, M. T. (2025). Urease in acetogenic Lachnospiraceae drives urea carbon salvage in SCFA pools. *Gut Microbes*, *17*(1), 2492376. https://doi.org/10.1080/19490976.2025.2492376

Flies, E. J., Clarke, L. J., Brook, B. W., & Jones, P. (2020). Urbanisation reduces the abundance and diversity of airborne microbes - but what does that mean for our health? A systematic review. *Science of The Total Environment*, *738*, 140337. https://doi.org/10.1016/j.scitotenv.2020.140337

Fowler, A., & Toner, M. (2006). Cryo-Injury and Biopreservation. *Annals of the New York Academy of Sciences*, *1066*(1), 119–135. https://doi.org/10.1196/annals.1363.010

Frioux, C., Ansorge, R., Özkurt, E., Ghassemi Nedjad, C., Fritscher, J., Quince, C., Waszak, S. M., & Hildebrand, F. (2023). Enterosignatures define common bacterial guilds in the human gut microbiome. *Cell Host & Microbe*, *31*(7), 1111-1125.e6. https://doi.org/10.1016/j.chom.2023.05.024

Furlani, I. L., da Cruz Nunes, E., Canuto, G. A. B., Macedo, A. N., & Oliveira, R. V. (2021). Liquid Chromatography-Mass Spectrometry for Clinical Metabolomics: An Overview. In A. V. Colnaghi Simionato (Ed.), *Separation Techniques Applied to Omics Sciences: From Principles to Relevant Applications* (pp. 179–213). Springer International Publishing. https://doi.org/10.1007/978-3-030-77252-9_10

Furusawa, Y., Obata, Y., Fukuda, S., Endo, T. A., Nakato, G., Takahashi, D., Nakanishi, Y., Uetake, C., Kato, K., Kato, T., Takahashi, M., Fukuda, N. N., Murakami, S., Miyauchi, E., Hino, S., Atarashi, K., Onawa, S., Fujimura, Y., Lockett, T., … Ohno, H. (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature*, *504*(7480), 446–450. https://doi.org/10.1038/nature12721

Ganesan, R., & Suk, K. T. (2022). Therapeutic Potential of Human Microbiome-Based Short-Chain Fatty Acids and Bile Acids in Liver Disease. *Livers*, *2*(3), Article 3. https://doi.org/10.3390/livers2030012

Gangarapu, V., Yıldız, K., Ince, A. T., & Baysal, B. (2014). Role of gut microbiota: Obesity and NAFLD. *The Turkish Journal of Gastroenterology: The Official Journal of Turkish Society of Gastroenterology*, *25*(2), 133–140. https://doi.org/10.5152/tjg.2014.7886

Gao, B., Chi, L., Zhu, Y., Shi, X., Tu, P., Li, B., Yin, J., Gao, N., Shen, W., & Schnabl, B. (2021). An Introduction to Next Generation Sequencing Bioinformatic Analysis in Gut Microbiome Studies. *Biomolecules*, *11*(4), Article 4. https://doi.org/10.3390/biom11040530

Garcia, A., & Barbas, C. (2011). Gas chromatography-mass spectrometry (GC-MS)-based metabolomics. *Methods in Molecular Biology (Clifton, N.J.)*, *708*, 191–204. https://doi.org/10.1007/978-1-61737-985-7_11

García-Castillo, V., Sanhueza, E., McNerney, E., Onate, S. A., & García, A. (2016). Microbiota dysbiosis: A new piece in the understanding of the carcinogenesis puzzle. *Journal of Medical Microbiology*, *65*(12), 1347–1362. https://doi.org/10.1099/jmm.0.000371

Gavriliuc, S., Stothart, M. R., Henry, A., & Poissant, J. (2021). Long-term storage of feces at −80 °C versus −20 °C is negligible for 16S rRNA amplicon profiling of the equine bacterial microbiome. *PeerJ*, *9*, e10837. https://doi.org/10.7717/peerj.10837

Ghezzi, L., Cantoni, C., Rotondo, E., & Galimberti, D. (2022). The Gut Microbiome–Brain Crosstalk in Neurodegenerative Diseases. *Biomedicines*, *10*(7), 1486. https://doi.org/10.3390/biomedicines10071486

Gika, H. G., Theodoridis, G. A., Plumb, R. S., & Wilson, I. D. (2014). Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics. *Journal of Pharmaceutical and Biomedical Analysis*, *87*, 12–25. https://doi.org/10.1016/j.jpba.2013.06.032

Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., & Knight, R. (2018). Current understanding of the human microbiome. *Nature Medicine*, *24*(4), 392–400. https://doi.org/10.1038/nm.4517

Gill, P. A., van Zelm, M. C., Muir, J. G., & Gibson, P. R. (2018). Review article: Short chain fatty acids as potential therapeutic agents in human gastrointestinal and inflammatory disorders. *Alimentary Pharmacology & Therapeutics*, *48*(1), 15–34. https://doi.org/10.1111/apt.14689

Gill, S. K., Rossi, M., Bajka, B., & Whelan, K. (2021). Dietary fibre in gastrointestinal health and disease. *Nature Reviews Gastroenterology & Hepatology*, *18*(2), 101–116. https://doi.org/10.1038/s41575-020-00375-4

Goryakin, Y., Rocco, L., & Suhrcke, M. (2017). The contribution of urbanization to non-communicable diseases: Evidence from 173 countries from 1980 to 2008. *Economics & Human Biology*, *26*, 151–163. https://doi.org/10.1016/j.ehb.2017.03.004

Gratton, J., Phetcharaburanin, J., Mullish, B. H., Williams, H. R. T., Thursz, M., Nicholson, J. K., Holmes, E., Marchesi, J. R., & Li, J. V. (2016). Optimized Sample Handling Strategy for Metabolic Profiling of Human Feces. *Analytical Chemistry*, *88*(9), 4661–4668. https://doi.org/10.1021/acs.analchem.5b04159

Greathouse, K. L., Sinha, R., & Vogtmann, E. (2019). DNA extraction for human microbiome studies: The issue of standardization. *Genome Biology*, *20*(1), 212. https://doi.org/10.1186/s13059-019-1843-8

Guan, H., Pu, Y., Liu, C., Lou, T., Tan, S., Kong, M., Sun, Z., Mei, Z., Qi, Q., Quan, Z., Zhao, G., & Zheng, Y. (2021). Comparison of Fecal Collection Methods on Variation in Gut Metagenomics and Untargeted Metabolomics. *mSphere*, *6*(5), e0063621. https://doi.org/10.1128/mSphere.00636-21

Haahtela, T., Bousquet, J., & Antó, J. M. (2024). From biodiversity to nature deficiency in human health and disease. *Porto Biomedical Journal*, *9*(1), e245. https://doi.org/10.1097/j.pbj.0000000000000245

Haile, S., Corbett, R. D., Bilobram, S., Bye, M. H., Kirk, H., Pandoh, P., Trinh, E., MacLeod, T., McDonald, H., Bala, M., Miller, D., Novik, K., Coope, R. J., Moore, R. A., Zhao, Y., Mungall, A. J., Ma, Y., Holt, R. A., Jones, S. J., & Marra, M. A. (2019). Sources of erroneous sequences and artifact chimeric reads in next generation sequencing of genomic DNA from formalin-fixed paraffin-embedded samples. *Nucleic Acids Research*, *47*(2), e12. https://doi.org/10.1093/nar/gky1142

Halilbasic, E., Claudel, T., & Trauner, M. (2013). Bile acid transporters and regulatory nuclear receptors in the liver and beyond. *Journal of Hepatology*, *58*(1), 155–168. https://doi.org/10.1016/j.jhep.2012.08.002

Hammons, J. L., Jordan, W. E., Stewart, R. L., Taulbee, J. D., & Berg, R. W. (1988). Age and Diet Effects on Fecal Bile Acids in Infants. *Journal of Pediatric Gastroenterology and Nutrition*, *7*(1), 30–38. https://doi.org/10.1002/j.1536-4801.1988.tb09465.x

Han, D., Gao, P., Li, R., Tan, P., Xie, J., Zhang, R., & Li, J. (2020). Multicenter assessment of microbial community profiling using 16S rRNA gene sequencing and shotgun metagenomic sequencing. *Journal of Advanced Research*, *26*, 111–121. https://doi.org/10.1016/j.jare.2020.07.010

Hickman, B., Salonen, A., Ponsero, A. J., Jokela, R., Kolho, K.-L., de Vos, W. M., & Korpela, K. (2024). Gut microbiota wellbeing index predicts overall health in a cohort of 1000 infants. *Nature Communications*, *15*(1), 8323. https://doi.org/10.1038/s41467-024-52561-6

Hofmann, A. F., & Hagey, L. R. (2008). Bile acids: Chemistry, pathochemistry, biology, pathobiology, and therapeutics. *Cellular and Molecular Life Sciences: CMLS*, *65*(16), 2461–2483. https://doi.org/10.1007/s00018-008-7568-6

Hsu, C.-Y., Khachatryan, L. G., Younis, N. K., Mustafa, M. A., Ahmad, N., Athab, Z. H., Polyanskaya, A. V., Kasanave, E. V., Mirzaei, R., & Karampoor, S. (2024). Microbiota-derived short chain fatty acids in pediatric health and diseases: From gut development to neuroprotection. *Frontiers in Microbiology*, *15*. https://doi.org/10.3389/fmicb.2024.1456793

Husso, A., Pessa-Morikawa, T., Koistinen, V. M., Kärkkäinen, O., Kwon, H. N., Lahti, L., Iivanainen, A., Hanhineva, K., & Niku, M. (2023). Impacts of maternal microbiota and microbial metabolites on fetal intestine, brain, and placenta. *BMC Biology*, *21*(1), 207. https://doi.org/10.1186/s12915-023-01709-9

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., FitzGerald, M. G., Fulton, R. S., Giglio, M. G., Hallsworth-Pepin, K., Lobos, E. A., Madupu, R., Magrini, V., Martin, J. C., Mitreva, M., Muzny, D. M., Sodergren, E. J., … The Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207–214. https://doi.org/10.1038/nature11234

Hylemon, P. B., Zhou, H., Pandak, W. M., Ren, S., Gil, G., & Dent, P. (2009). Bile acids as regulatory molecules. *Journal of Lipid Research*, *50*(8), 1509–1520. https://doi.org/10.1194/jlr.R900007-JLR200

Illumina. (2014). https://emea.illumina.com/search.html?q=protocol&filter=all&p=1

Jantscher-Krenn, E., Zherebtsov, M., Nissan, C., Goth, K., Guner, Y. S., Naidu, N., Choudhury, B., Grishin, A. V., Ford, H. R., & Bode, L. (2012). The human milk oligosaccharide disialyllacto-N-tetraose prevents necrotising enterocolitis in neonatal rats. *Gut*, *61*(10), 1417–1425. https://doi.org/10.1136/gutjnl-2011-301404

Jaye, K., Li, C. G., & Bhuyan, D. J. (2021). The complex interplay of gut microbiota with the five most common cancer types: From carcinogenesis to therapeutics to prognoses. *Critical Reviews in Oncology/Hematology*, *165*, 103429. https://doi.org/10.1016/j.critrevonc.2021.103429

Jin, C. J., Sellmann, C., Engstler, A. J., Ziegenhardt, D., & Bergheim, I. (2015). Supplementation of sodium butyrate protects mice from the development of non-alcoholic steatohepatitis (NASH). *British Journal of Nutrition*, *114*(11), 1745–1755. https://doi.org/10.1017/S0007114515003621

Johnson, J. S., Spakowicz, D. J., Hong, B.-Y., Petersen, L. M., Demkowicz, P., Chen, L., Leopold, S. R., Hanson, B. M., Agresta, H. O., Gerstein, M., Sodergren, E., & Weinstock, G. M. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-13036-1

Joo, S.-S., Kang, H.-C., Won, T.-J., & Lee, D.-I. (2003). Ursodeoxycholic acid inhibits pro-inflammatory repertoires, IL-1β and nitric oxide in rat microglia. *Archives of Pharmacal Research*, *26*(12), 1067–1073. https://doi.org/10.1007/BF02994760

Joos, R., Boucher, K., Lavelle, A., Arumugam, M., Blaser, M. J., Claesson, M. J., Clarke, G., Cotter, P. D., De Sordi, L., Dominguez-Bello, M. G., Dutilh, B. E., Ehrlich, S. D., Ghosh, T. S., Hill, C.,

Junot, C., Lahti, L., Lawley, T. D., Licht, T. R., Maguin, E., … Ross, R. P. (2025). Examining the healthy human microbiome concept. *Nature Reviews Microbiology*, *23*(3), 192–205. https://doi.org/10.1038/s41579-024-01107-0

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A. L., Madsen, K. L., & Wong, G. K.-S. (2016). Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.00459

Kanani, H., Chrysanthopoulos, P. K., & Klapa, M. I. (2008). Standardizing GC–MS metabolomics. *Journal of Chromatography B*, *871*(2), 191–201. https://doi.org/10.1016/j.jchromb.2008.04.049

Kane, A. V., Dinh, D. M., & Ward, H. D. (2015). Childhood malnutrition and the intestinal microbiome. *Pediatric Research*, *77*(1), 256–262. https://doi.org/10.1038/pr.2014.179

Karkman, A., Lehtimäki, J., & Ruokolainen, L. (2017). The ecology of human microbiota: Dynamics and diversity in health and disease. *Annals of the New York Academy of Sciences*, *1399*(1), 78–92. https://doi.org/10.1111/nyas.13326

Karlsson, L., Tolvanen, M., Scheinin, N. M., Uusitupa, H.-M., Korja, R., Ekholm, E., Tuulari, J. J., Pajulo, M., Huotilainen, M., Paunio, T., Karlsson, H., & FinnBrain Birth Cohort Study Group. (2018). Cohort Profile: The FinnBrain Birth Cohort Study (FinnBrain). *International Journal of Epidemiology*, *47*(1), 15–16j. https://doi.org/10.1093/ije/dyx173

Karu, N., Deng, L., Slae, M., Guo, A. C., Sajed, T., Huynh, H., Wine, E., & Wishart, D. S. (2018). A review on human fecal metabolomics: Methods, applications and the human fecal metabolome database. *Analytica Chimica Acta*, *1030*, 1–24. https://doi.org/10.1016/j.aca.2018.05.031

Khan, I. (2021). Microbiome. *Indian Journal of Medical and Paediatric Oncology*, *42*, 461–465. https://doi.org/10.1055/s-0041-1735599

Khine, W. W. T., Rahayu, E. S., See, T. Y., Kuah, S., Salminen, S., Nakayama, J., & Lee, Y.-K. (2020). Indonesian children fecal microbiome from birth until weaning was different from microbiomes of their mothers. *Gut Microbes*, *12*(1), 1761240. https://doi.org/10.1080/19490976.2020.1761240

Kim, D., Zeng, M. Y., & Núñez, G. (2017). The interplay between host immune cells and gut microbiota in chronic inflammatory diseases. *Experimental & Molecular Medicine*, *49*(5), e339–e339. https://doi.org/10.1038/emm.2017.24

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, *41*(1), e1. https://doi.org/10.1093/nar/gks808

Koek, M. M., Jellema, R. H., van der Greef, J., Tas, A. C., & Hankemeier, T. (2011). Quantitative metabolomics based on gas chromatography mass spectrometry: Status and perspectives. *Metabolomics*, *7*(3), 307–328. https://doi.org/10.1007/s11306-010-0254-3

Korhonen, L. S., Lukkarinen, M., Kantojärvi, K., Räty, P., Karlsson, H., Paunio, T., Peltola, V., & Karlsson, L. (2021). Interactions of genetic variants and prenatal stress in relation to the risk for recurrent respiratory infections in children. *Scientific Reports*, *11*(1), 7589. https://doi.org/10.1038/s41598-021-87211-0

Kruk, J., Doskocz, M., Jodłowska, E., Zacharzewska, A., Łakomiec, J., Czaja, K., & Kujawski, J. (2017). NMR Techniques in Metabolomic Studies: A Quick Overview on Examples of Utilization. *Applied Magnetic Resonance*, *48*(1), 1–21. https://doi.org/10.1007/s00723-016-0846-9

Kutschera, U. (2023). Antonie van Leeuwenhoek (1632–1723): Master of Fleas and Father of Microbiology. *Microorganisms*, *11*(8), 1994. https://doi.org/10.3390/microorganisms11081994

Lacagnina, S. (2019). The Developmental Origins of Health and Disease (DOHaD). *American Journal of Lifestyle Medicine*, *14*(1), 47–50. https://doi.org/10.1177/1559827619879694

Lamichhane, S., Sen, P., Dickens, A. M., Alves, M. A., Härkönen, T., Honkanen, J., Vatanen, T., Xavier, R. J., Hyötyläinen, T., Knip, M., & Orešič, M. (2022). Dysregulation of secondary bile acid metabolism precedes islet autoimmunity and type 1 diabetes. *Cell Reports. Medicine*, *3*(10), 100762. https://doi.org/10.1016/j.xcrm.2022.100762

Lamichhane, S., Sen, P., Dickens, A. M., Orešič, M., & Bertram, H. C. (2018). Gut metabolome meets microbiome: A methodological perspective to understand the relationship between host and microbe. *Methods (San Diego, Calif.)*, *149*, 3–12. https://doi.org/10.1016/j.ymeth.2018.04.029

Lauber, C. L., Zhou, N., Gordon, J. I., Knight, R., & Fierer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiology Letters*, *307*(1), 80–86. https://doi.org/10.1111/j.1574-6968.2010.01965.x

Laursen, M. F. (2021). Gut Microbiota Development: Influence of Diet from Infancy to Toddlerhood. *Annals of Nutrition & Metabolism*, 1–14. https://doi.org/10.1159/000517912

Levy, M., Blacher, E., & Elinav, E. (2017). Microbiome, metabolites and host immunity. *Current Opinion in Microbiology*, *35*, 8–15. https://doi.org/10.1016/j.mib.2016.10.003

Lewis, E. D., Richard, C., Goruk, S., Wadge, E., Curtis, J. M., Jacobs, R. L., & Field, C. J. (2017). Feeding a Mixture of Choline Forms during Lactation Improves Offspring Growth and Maternal Lymphocyte Response to Ex Vivo Immune Challenges. *Nutrients*, *9*(7), 713. https://doi.org/10.3390/nu9070713

Li, J., Fu, R., Yang, Y., Horz, H.-P., Guan, Y., Lu, Y., Lou, H., Tian, L., Zheng, S., Liu, H., Shi, M., Tang, K., Wang, S., & Xu, S. (2018). A metagenomic approach to dissect the genetic composition of enterotypes in Han Chinese and two Muslim groups. *Systematic and Applied Microbiology*, *41*(1), 1–12. https://doi.org/10.1016/j.syapm.2017.09.006

Li, X., Bosch-Tijhof, C. J., Wei, X., de Soet, J. J., Crielaard, W., Loveren, C. van, & Deng, D. M. (2020). Efficiency of chemical versus mechanical disruption methods of DNA extraction for the identification of oral Gram-positive and Gram-negative bacteria. *The Journal of International Medical Research*, *48*(5), 300060520925594. https://doi.org/10.1177/0300060520925594

Li, Z., Zhu, Y., Ni, D., Zhang, W., & Mu, W. (2023). Occurrence, functional properties, and preparation of 3-fucosyllactose, one of the smallest human milk oligosaccharides. *Critical Reviews in Food Science and Nutrition*, *63*(28), 9364–9378. https://doi.org/10.1080/10408398.2022.2064813

Lim, M. Y., Hong, S., Kim, B.-M., Ahn, Y., Kim, H.-J., & Nam, Y.-D. (2020). Changes in microbiome and metabolomic profiles of fecal samples stored with stabilizing solution at room temperature: A pilot study. *Scientific Reports*, *10*(1), 1789. https://doi.org/10.1038/s41598-020-58719-8

Lim, M. Y., Song, E.-J., Kim, S. H., Lee, J., & Nam, Y.-D. (2018). Comparison of DNA extraction methods for human gut microbial community profiling. *Systematic and Applied Microbiology*, *41*(2), 151–157. https://doi.org/10.1016/j.syapm.2017.11.008

Lin, H. V., Frassetto, A., Jr, E. J. K., Nawrocki, A. R., Lu, M. M., Kosinski, J. R., Hubert, J. A., Szeto, D., Yao, X., Forrest, G., & Marsh, D. J. (2012). Butyrate and Propionate Protect against Diet-Induced Obesity and Regulate Gut Hormones via Free Fatty Acid Receptor 3-Independent Mechanisms. *PLOS ONE*, *7*(4), e35240. https://doi.org/10.1371/journal.pone.0035240

Loftfield, E., Vogtmann, E., Sampson, J. N., Moore, S. C., Nelson, H., Knight, R., Chia, N., & Sinha, R. (2016). Comparison of Collection Methods for Fecal Samples for Discovery Metabolomics in Epidemiologic Studies. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *25*(11), 1483–1490. https://doi.org/10.1158/1055-9965.EPI-16-0409

Łoniewska, B., Fraszczyk-Tousty, M., Tousty, P., Skonieczna-Żydecka, K., Maciejewska-Markiewicz, D., & Łoniewski, I. (2023). Analysis of Fecal Short-Chain Fatty Acids (SCFAs) in Healthy Children during the First Two Years of Life: An Observational Prospective Cohort Study. *Nutrients*, *15*(2), Article 2. https://doi.org/10.3390/nu15020367

Łoniewski, I., Skonieczna-Żydecka, K., Stachowska, L., Fraszczyk-Tousty, M., Tousty, P., & Łoniewska, B. (2022). Breastfeeding Affects Concentration of Faecal Short Chain Fatty Acids During the First Year of Life: Results of the Systematic Review and Meta-Analysis. *Frontiers in Nutrition*, *9*. https://doi.org/10.3389/fnut.2022.939194

López-Tenorio, I. I., Aguilar-Villegas, Ó. R., Espinoza-Palacios, Y., Segura-Real, L., Peña-Aparicio, B., Amedei, A., & Aguirre-García, M. M. (2024). Primary Prevention Strategy for Non-

Communicable Diseases (NCDs) and Their Risk Factors: The Role of Intestinal Microbiota. *Biomedicines*, *12*(11), Article 11. https://doi.org/10.3390/biomedicines12112529

Louis, P., Hold, G. L., & Flint, H. J. (2014). The gut microbiota, bacterial metabolites and colorectal cancer. *Nature Reviews Microbiology*, *12*(10), 661–672. https://doi.org/10.1038/nrmicro3344

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, Y., Feskens, E. J. M., Boer, J. M. A., & Müller, M. (2010). The potential influence of genetic variants in genes along bile acid and bile metabolic pathway on blood cholesterol levels in the population. *Atherosclerosis*, *210*(1), 14–27. https://doi.org/10.1016/j.atherosclerosis.2009.10.035

Macfarlane, G. T., & Macfarlane, S. (2012). Bacteria, Colonic Fermentation, and Gastrointestinal Health. *Journal of AOAC INTERNATIONAL*, *95*(1), 50–60. https://doi.org/10.5740/jaoacint.SGE_Macfarlane

Maheshwari, K., Musyuni, P., Moulick, A., Mishra, H., Ekielski, A., Mishra, P. K., & Aggarwal, G. (2024). Unveiling the microbial symphony: Next-Gen sequencing and bioinformatics insights into the human gut microbiome. *Health Sciences Review*, *11*, 100173. https://doi.org/10.1016/j.hsr.2024.100173

Mancabelli, L., Milani, C., Lugli, G. A., Turroni, F., Ferrario, C., van Sinderen, D., & Ventura, M. (2017). Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations. *Environmental Microbiology*, *19*(4), 1379–1390. https://doi.org/10.1111/1462-2920.13692

Marshall, D. D., & Powers, R. (2017). Beyond the paradigm: Combining mass spectrometry and nuclear magnetic resonance for metabolomics. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *100*, 1–16. https://doi.org/10.1016/j.pnmrs.2017.01.001

Martín, R., Rios-Covian, D., Huillet, E., Auger, S., Khazaal, S., Bermúdez-Humarán, L. G., Sokol, H., Chatel, J.-M., & Langella, P. (2023). Faecalibacterium: A bacterial genus with promising human health applications. *FEMS Microbiology Reviews*, *47*(4), fuad039. https://doi.org/10.1093/femsre/fuad039

Martínez Arbas, S., Busi, S. B., Queirós, P., de Nies, L., Herold, M., May, P., Wilmes, P., Muller, E. E. L., & Narayanasamy, S. (2021). Challenges, Strategies, and Perspectives for Reference-Independent Longitudinal Multi-Omic Microbiome Studies. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.666244

Mathews, A. (2024). DNA Sequencing: A Brief History. In *DNA Sequencing—History, Present and Future*. IntechOpen. https://doi.org/10.5772/intechopen.1007844

McCarthy, D. J., Campbell, K. R., Lun, A. T. L., & Wills, Q. F. (2017). Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*, *33*(8), 1179–1186. https://doi.org/10.1093/bioinformatics/btw777

McDonald, D., Jiang, Y., Balaban, M., Cantrell, K., Zhu, Q., Gonzalez, A., Morton, J. T., Nicolaou, G., Parks, D. H., Karst, S. M., Albertsen, M., Hugenholtz, P., DeSantis, T., Song, S. J., Bartko, A., Havulinna, A. S., Jousilahti, P., Cheng, S., Inouye, M., … Knight, R. (2024). Greengenes2 unifies microbial data in a single reference tree. *Nature Biotechnology*, *42*(5), 715–718. https://doi.org/10.1038/s41587-023-01845-1

McDonald, J. A. K., Mullish, B. H., Pechlivanis, A., Liu, Z., Brignardello, J., Kao, D., Holmes, E., Li, J. V., Clarke, T. B., Thursz, M. R., & Marchesi, J. R. (2018). Inhibiting Growth of *Clostridioides difficile* by Restoring Valerate, Produced by the Intestinal Microbiota. *Gastroenterology*, *155*(5), 1495-1507.e15. https://doi.org/10.1053/j.gastro.2018.07.014

McDonnell, L., Gilkes ,Alexander, Ashworth ,Mark, Rowland ,Victoria, Harries ,Timothy Hugh, Armstrong ,David, & and White, P. (2021). Association between antibiotics and gut microbiome dysbiosis in children: Systematic review and meta-analysis. *Gut Microbes*, *13*(1), 1870402. https://doi.org/10.1080/19490976.2020.1870402

Menzel, A., Samouda, H., Dohet, F., Loap, S., Ellulu, M. S., & Bohn, T. (2021). Common and Novel Markers for Measuring Inflammation and Oxidative Stress Ex Vivo in Research and Clinical Practice—Which to Use Regarding Disease Outcomes? *Antioxidants*, *10*(3), Article 3. https://doi.org/10.3390/antiox10030414

Mercer, E. M., Ramay, H. R., Moossavi, S., Laforest-Lapointe, I., Reyna, M. E., Becker, A. B., Simons, E., Mandhane, P. J., Turvey, S. E., Moraes, T. J., Sears, M. R., Subbarao, P., Azad, M. B., & Arrieta, M.-C. (2024). Divergent maturational patterns of the infant bacterial and fungal gut microbiome in the first year of life are associated with inter-kingdom community dynamics and infant nutrition. *Microbiome*, *12*(1), 22. https://doi.org/10.1186/s40168-023-01735-3

Mishra, A., Lai, G. C., Yao, L. J., Aung, T. T., Shental, N., Rotter-Maskowitz, A., Shepherdson, E., Singh, G. S. N., Pai, R., Shanti, A., Wong, R. M. M., Lee, A., Khyriem, C., Dutertre, C. A., Chakarov, S., Srinivasan, K. G., Shadan, N. B., Zhang, X.-M., Khalilnezhad, S., … Ginhoux, F. (2021). Microbial exposure during early human development primes fetal immune cells. *Cell*, *184*(13), 3394-3409.e20. https://doi.org/10.1016/j.cell.2021.04.039

Moens, F., Weckx, S., & De Vuyst, L. (2016). Bifidobacterial inulin-type fructan degradation capacity determines cross-feeding interactions between bifidobacteria and *Faecalibacterium prausnitzii*. *International Journal of Food Microbiology*, *231*, 76–85. https://doi.org/10.1016/j.ijfoodmicro.2016.05.015

Molano, L.-A. G., Vega-Abellaneda, S., & Manichanh, C. (2024). GSR-DB: A manually curated and optimized taxonomical database for 16S rRNA amplicon analysis. *mSystems*, *9*(2), e00950-23. https://doi.org/10.1128/msystems.00950-23

Morrow, A. L., Ruiz-Palacios, G. M., Altaye, M., Jiang, X., Lourdes Guerrero, M., Meinzen-Derr, J. K., Farkas, T., Chaturvedi, P., Pickering, L. K., & Newburg, D. S. (2004). Human milk oligosaccharides are associated with protection against diarrhea in breast-fed infants. *The Journal of Pediatrics*, *145*(3), 297–303. https://doi.org/10.1016/j.jpeds.2004.04.054

Mueller, N. T., Bakacs, E., Combellick, J., Grigoryan, Z., & Dominguez-Bello, M. G. (2015). The infant microbiome development: Mom matters. *Trends in Molecular Medicine*, *21*(2), 109–117. https://doi.org/10.1016/j.molmed.2014.12.002

Munjal, Y., Tonk, R. K., & Sharma, R. (2022). *Analytical Techniques Used in Metabolomics: A Review*. *13*(08).

Nagana Gowda, G. A., & Raftery, D. (2021). NMR-Based Metabolomics. In S. Hu (Ed.), *Cancer Metabolomics: Methods and Applications* (pp. 19–37). Springer International Publishing. https://doi.org/10.1007/978-3-030-51652-9_2

Nagata, N., Tohya, M., Takeuchi, F., Suda, W., Nishijima, S., Ohsugi, M., Ueki, K., Tsujimoto, T., Nakamura, T., Kawai, T., Miyoshi-Akiyama, T., Uemura, N., & Hattori, M. (2019). Effects of storage temperature, storage time, and Cary-Blair transport medium on the stability of the gut microbiota. *Drug Discoveries & Therapeutics*, *13*(5), 256–260. https://doi.org/10.5582/ddt.2019.01071

Nastasi, J. R., Daygon, V. D., Kontogiorgos, V., & Fitzgerald, M. A. (2023). Qualitative Analysis of Polyphenols in Glycerol Plant Extracts Using Untargeted Metabolomics. *Metabolites*, *13*(4), Article 4. https://doi.org/10.3390/metabo13040566

Natarajan, A., Han, A., Zlitni, S., Brooks, E. F., Vance, S. E., Wolfe, M., Singh, U., Jagannathan, P., Pinsky, B. A., Boehm, A., & Bhatt, A. S. (2021). Publisher Correction: Standardized preservation, extraction and quantification techniques for detection of fecal SARS-CoV-2 RNA. *Nature Communications*, *12*, 7100. https://doi.org/10.1038/s41467-021-27392-4

Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., Margono, C., Mullany, E. C., Biryukov, S., Abbafati, C., Abera, S. F., Abraham, J. P., Abu-Rmeileh, N. M. E., Achoki, T., AlBuhairan, F. S., Alemu, Z. A., Alfonso, R., Ali, M. K., Ali, R., Guzman, N. A., … Gakidou, E. (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, *384*(9945), 766–781. https://doi.org/10.1016/S0140-6736(14)60460-8

Nguyen, Q. P., Karagas, M. R., Madan, J. C., Dade, E., Palys, T. J., Morrison, H. G., Pathmasiri, W. W., McRitche, S., Sumner, S. J., Frost, H. R., & Hoen, A. G. (2021). Associations between the gut microbiome and metabolome in early life. *BMC Microbiology*, *21*(1), 238. https://doi.org/10.1186/s12866-021-02282-3

Noce, A., Marrone, G., Di Daniele, F., Ottaviani, E., Wilson Jones, G., Bernini, R., Romani, A., & Rovella, V. (2019). Impact of Gut Microbiota Composition on Onset and Progression of Chronic Non-Communicable Diseases. *Nutrients*, *11*(5), Article 5. https://doi.org/10.3390/nu11051073

Nolvi, S., Uusitupa, H.-M., Bridgett, D. J., Pesonen, H., Aatsinki, A.-K., Kataja, E.-L., Korja, R., Karlsson, H., & Karlsson, L. (2018). Human milk cortisol concentration predicts experimentally induced infant fear reactivity: Moderation by infant sex. *Developmental Science*, *21*(4), e12625. https://doi.org/10.1111/desc.12625

Ogden, C. L., Carroll, M. D., Kit, B. K., & Flegal, K. M. (2014). Prevalence of Childhood and Adult Obesity in the United States, 2011-2012. *JAMA*, *311*(8), 806–814. https://doi.org/10.1001/jama.2014.732

Olga, L., van Diepen, J. A., Chichlowski, M., Petry, C. J., Vervoort, J., Dunger, D. B., Kortman, G. A. M., Gross, G., & Ong, K. K. (2023). Butyrate in Human Milk: Associations with Milk Microbiota, Milk Intake Volume, and Infant Growth. *Nutrients*, *15*(4), Article 4. https://doi.org/10.3390/nu15040916

Oliphant, K., & Allen-Vercoe, E. (2019). Macronutrient metabolism by the human gut microbiome: Major fermentation by-products and their impact on host health. *Microbiome*, *7*(1), 91. https://doi.org/10.1186/s40168-019-0704-8

O'May, G. A., Reynolds, N., & Macfarlane, G. T. (2005). Effect of pH on an In Vitro Model of Gastric Microbiota in Enteral Nutrition Patients. *Applied and Environmental Microbiology*, *71*(8), 4777–4783. https://doi.org/10.1128/AEM.71.8.4777-4783.2005

Pallen, M. J. (2023). Request for an Opinion on the standing and retention of Firmicutes as a phylum name. *International Journal of Systematic and Evolutionary Microbiology*, *73*(7), 005933. https://doi.org/10.1099/ijsem.0.005933

Papadimitropoulos, M.-E. P., Vasilopoulou, C. G., Maga-Nteve, C., & Klapa, M. I. (2018). Untargeted GC-MS Metabolomics. In G. A. Theodoridis, H. G. Gika, & I. D. Wilson (Eds.), *Metabolic Profiling: Methods and Protocols* (pp. 133–147). Springer. https://doi.org/10.1007/978-1-4939-7643-0_9

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M. C., Rice, B. L., DuLong, C., Morgan, X. C., Golden, C. D., Quince, C., Huttenhower, C., & Segata, N. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*, *176*(3), 649-662.e20. https://doi.org/10.1016/j.cell.2019.01.001

Pédron, T., Mulet, C., Dauga, C., Frangeul, L., Chervaux, C., Grompone, G., & Sansonetti, P. J. (2012). A Crypt-Specific Core Microbiota Resides in the Mouse Colon. *mBio*, *3*(3), 10.1128/mbio.00116-12. https://doi.org/10.1128/mbio.00116-12

Peila, C., Sottemano, S., Cesare Marincola, F., Stocchero, M., Pusceddu, N. G., Dessì, A., Baraldi, E., Fanos, V., & Bertino, E. (2022). NMR Metabonomic Profile of Preterm Human Milk in the First Month of Lactation: From Extreme to Moderate Prematurity. *Foods*, *11*(3), 345. https://doi.org/10.3390/foods11030345

Pereira, R., Oliveira, J., & Sousa, M. (2020). Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *Journal of Clinical Medicine*, *9*(1), 132. https://doi.org/10.3390/jcm9010132

Petras, D., Koester, I., Da Silva, R., Stephens, B. M., Haas, A. F., Nelson, C. E., Kelly, L. W., Aluwihare, L. I., & Dorrestein, P. C. (2017). High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter. *Frontiers in Marine Science*, *4*. https://doi.org/10.3389/fmars.2017.00405

Pitt, J. J. (2009). Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. *The Clinical Biochemist Reviews*, *30*(1), 19–34.

Poulsen, C. S., Kaas, R. S., Aarestrup, F. M., & Pamp, S. J. (2021). Standard Sample Storage Conditions Have an Impact on Inferred Microbiome Composition and Antimicrobial Resistance Patterns. *Microbiology Spectrum*, *9*(2), e01387-21. https://doi.org/10.1128/Spectrum.01387-21

Poulsen, K. O., & Sundekilde, U. K. (2021). The Metabolomic Analysis of Human Milk Offers Unique Insights into Potential Child Health Benefits. *Current Nutrition Reports*, *10*(1), 12–29. https://doi.org/10.1007/s13668-020-00345-x

Prescott, S. L. (2017). History of medicine: Origin of the term microbiome and why it matters. *Human Microbiome Journal*, *4*, 24–25. https://doi.org/10.1016/j.humic.2017.05.004

Pribyl, A. L., Parks, D. H., Angel, N. Z., Boyd, J. A., Hasson, A. G., Fang, L., MacDonald, S. L., Wills, B. A., Wood, D. L. A., Krause, L., Tyson, G. W., & Hugenholtz, P. (2021). Critical evaluation of faecal microbiome preservation using metagenomic analysis. *ISME Communications*, *1*(1), 1–10. https://doi.org/10.1038/s43705-021-00014-2

Pundir, S., Wall, C. R., Mitchell, C. J., Thorstensen, E. B., Lai, C. T., Geddes, D. T., & Cameron-Smith, D. (2017). Variation of Human Milk Glucocorticoids over 24 hour Period. *Journal of Mammary Gland Biology and Neoplasia*, *22*(1), 85–92. https://doi.org/10.1007/s10911-017-9375-x

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(Database issue), D590-596. https://doi.org/10.1093/nar/gks1219

Ramamoorthy, S., Levy, S., Mohamed, M., Abdelghani, A., Evans, A. M., Miller, L. A. D., Mehta, L., Moore, S., Freinkman, E., & Hourigan, S. K. (2021). An ambient-temperature storage and stabilization device performs comparably to flash-frozen collection for stool metabolomics in infants. *BMC Microbiology*, *21*(1), 59. https://doi.org/10.1186/s12866-021-02104-6

Ramos Meyers, G., Samouda, H., & Bohn, T. (2022). Short Chain Fatty Acid Metabolism in Relation to Gut Microbiota and Genetic Variability. *Nutrients*, *14*(24), Article 24. https://doi.org/10.3390/nu14245361

Ray, B., Speck, M. L., & Dobrogosz, W. J. (1976). Cell wall lipopolysaccharide damage in *Escherichia coli* due to freezing. *Cryobiology*, *13*(2), 153–160. https://doi.org/10.1016/0011-2240(76)90127-9

Regueira-Iglesias, A., Balsa-Castro, C., Blanco-Pintos, T., & Tomás, I. (2023). Critical review of 16S rRNA gene sequencing workflow in microbiome studies: From primer selection to advanced data analysis. *Molecular Oral Microbiology*, *38*(5), 347–399. https://doi.org/10.1111/omi.12434

Resneck, J. S., Jr. (2025). Revisions to the Declaration of Helsinki on Its 60th Anniversary: A Modernized Set of Ethical Principles to Promote and Ensure Respect for Participants in a Rapidly Innovating Medical Research Ecosystem. *JAMA*, *333*(1), 15–17. https://doi.org/10.1001/jama.2024.21902

Reyman, M., van Houten, M. A., van Baarle, D., Bosch, A. A. T. M., Man, W. H., Chu, M. L. J. N., Arp, K., Watson, R. L., Sanders, E. A. M., Fuentes, S., & Bogaert, D. (2019). Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life. *Nature Communications*, *10*(1), 4997. https://doi.org/10.1038/s41467-019-13014-7

Ribo, S., Sánchez-Infantes, D., Martinez-Guino, L., García-Mantrana, I., Ramon-Krauel, M., Tondo, M., Arning, E., Nofrarías, M., Osorio-Conles, Ó., Fernández-Pérez, A., González-Torres, P., Cebrià, J., Gavaldà-Navarro, A., Chenoll, E., Isganaitis, E., Villarroya, F., Vallejo, M., Segalés, J., Jiménez-Chillarón, J. C., … Lerin, C. (2021). Increasing breast milk betaine modulates Akkermansia abundance in mammalian neonates and improves long-term metabolic health. *Science Translational Medicine*, *13*(587), eabb0322. https://doi.org/10.1126/scitranslmed.abb0322

Ridlon, J. M., Kang, D.-J., & Hylemon, P. B. (2006). Bile salt biotransformations by human intestinal bacteria. *Journal of Lipid Research*, *47*(2), 241–259. https://doi.org/10.1194/jlr.R500013-JLR200

Rintala, A., Pietilä, S., Munukka, E., Eerola, E., Pursiheimo, J.-P., Laiho, A., Pekkala, S., & Huovinen, P. (2017). Gut Microbiota Analysis Results Are Highly Dependent on the 16S rRNA Gene Target Region, Whereas the Impact of DNA Extraction Is Minor. *Journal of Biomolecular Techniques: JBT*, *28*(1), 19–30. https://doi.org/10.7171/jbt.17-2801-003

Rios-Covian, D., González, S., Nogacka, A. M., Arboleya, S., Salazar, N., Gueimonde, M., & de los Reyes-Gavilán, C. G. (2020). An Overview on Fecal Branched Short-Chain Fatty Acids Along Human Life and as Related With Body Mass Index: Associated Dietary and Anthropometric Factors. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00973

Roager, H. M., Stanton, C., & Hall, L. J. (2023). Microbial metabolites as modulators of the infant gut microbiome and host-microbial interactions in early life. *Gut Microbes*, *15*(1), 2192151. https://doi.org/10.1080/19490976.2023.2192151

Robe, P., Nalin, R., Capellano, C., Vogel, T. M., & Simonet, P. (2003). Extraction of DNA from soil. *European Journal of Soil Biology*, *39*(4), 183–190. https://doi.org/10.1016/S1164-5563(03)00033-5

Roesch, L. F. W., Casella, G., Simell, O., Krischer, J., Wasserfall, C. H., Schatz, D., Atkinson, M. A., Neu, J., & Triplett, E. W. (2009). Influence of Fecal Sample Storage on Bacterial Community Diversity. *The Open Microbiology Journal*, *3*, 40–46. https://doi.org/10.2174/1874285800903010040

Romano, K. A., Vivas, E. I., Amador-Noguez, D., & Rey, F. E. (2015). Intestinal Microbiota Composition Modulates Choline Bioavailability from Diet and Accumulation of the Proatherogenic Metabolite Trimethylamine-N-Oxide. *mBio*, *6*(2), 10.1128/mbio.02481-14. https://doi.org/10.1128/mbio.02481-14

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P. I., Godneva, A., Kalka, I. N., Bar, N., Shilo, S., Lador, D., Vila, A. V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B. C., … Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature*, *555*(7695), 210–215. https://doi.org/10.1038/nature25973

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., & Tuohy, K. (2018). Gut microbiota functions: Metabolism of nutrients and other food components. *European Journal of Nutrition*, *57*(1), 1–24. https://doi.org/10.1007/s00394-017-1445-8

Roy Sarkar, S., & Banerjee, S. (2019). Gut microbiota in neurodegenerative disorders. *Journal of Neuroimmunology*, *328*, 98–104. https://doi.org/10.1016/j.jneuroim.2019.01.004

Ruiz, L., Moles, L., Gueimonde, M., & Rodriguez, J. M. (2016). Perinatal Microbiomes' Influence on Preterm Birth and Preterms' Health. *Journal of Pediatric Gastroenterology and Nutrition*, *63*(6), e193–e203. https://doi.org/10.1097/MPG.0000000000001196

Sabater, C., Iglesias-Gutiérrez, E., Ruiz, L., & Margolles, A. (2023). Next-generation sequencing of the athletic gut microbiota: A systematic review. *Microbiome Research Reports*, *2*(1), 5. https://doi.org/10.20517/mrr.2022.16

Sakata, T. (2019). Pitfalls in short-chain fatty acid research: A methodological review. *Animal Science Journal*, *90*(1), 3–13. https://doi.org/10.1111/asj.13118

Salamone, M., & Nardo, V. D. (2020). Effects of human milk oligosaccharides (HMOs) on gastrointestinal health. *Frontiers in Bioscience-Elite*, *12*(1), Article 1. https://doi.org/10.2741/E866

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., & Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*, 87. https://doi.org/10.1186/s12915-014-0087-z

Samuel, T. M., Binia, A., de Castro, C. A., Thakkar, S. K., Billeaud, C., Agosti, M., Al-Jashi, I., Costeira, M. J., Marchini, G., Martínez-Costa, C., Picaud, J.-C., Stiris, T., Stoicescu, S.-M., Vanpeé, M., Domellöf, M., Austin, S., & Sprenger, N. (2019). Impact of maternal characteristics on human milk oligosaccharide composition over the first 4 months of lactation in a cohort of

healthy European mothers. *Scientific Reports*, *9*(1), 11767. https://doi.org/10.1038/s41598-019-48337-4

Sanmiguel, C., Gupta, A., & Mayer, E. A. (2015). Gut Microbiome and Obesity: A Plausible Explanation for Obesity. *Current Obesity Reports*, *4*(2), 250–261. https://doi.org/10.1007/s13679-015-0152-0

Sarkar, A., Yoo, J. Y., Valeria Ozorio Dutra, S., Morgan, K. H., & Groer, M. (2021). The Association between Early-Life Gut Microbiota and Long-Term Health and Diseases. *Journal of Clinical Medicine*, *10*(3), Article 3. https://doi.org/10.3390/jcm10030459

Sasaki, M., Suaini, N. H. A., Afghani, J., Heye, K. N., O'Mahony, L., Venter, C., Lauener, R., Frei, R., & Roduit, C. (2024). Systematic review of the association between short-chain fatty acids and allergic diseases. *Allergy*, *79*(7), 1789–1811. https://doi.org/10.1111/all.16065

Sato, M. P., Ogura, Y., Nakamura, K., Nishida, R., Gotoh, Y., Hayashi, M., Hisatsune, J., Sugai, M., Takehiko, I., & Hayashi, T. (2019). Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *26*(5), 391–398. https://doi.org/10.1093/dnares/dsz017

Sayin, S. I., Wahlström, A., Felin, J., Jäntti, S., Marschall, H.-U., Bamberg, K., Angelin, B., Hyötyläinen, T., Orešič, M., & Bäckhed, F. (2013). Gut Microbiota Regulates Bile Acid Metabolism by Reducing the Levels of Tauro-beta-muricholic Acid, a Naturally Occurring FXR Antagonist. *Cell Metabolism*, *17*(2), 225–235. https://doi.org/10.1016/j.cmet.2013.01.003

Schoeler, M., & Caesar, R. (2019). Dietary lipids, gut microbiota and lipid metabolism. *Reviews in Endocrine and Metabolic Disorders*, *20*(4), 461–472. https://doi.org/10.1007/s11154-019-09512-0

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biology*, *14*(8), e1002533. https://doi.org/10.1371/journal.pbio.1002533

Shakeri, A., Yousefi, H., Jarad, N. A., Kullab, S., Al-Mfarej, D., Rottman, M., & Didar, T. F. (2022). Contamination and carryover free handling of complex fluids using lubricant-infused pipette tips. *Scientific Reports*, *12*(1), 14486. https://doi.org/10.1038/s41598-022-18756-x

Sheng, W., Ji, G., & Zhang, L. (2022). The Effect of Lithocholic Acid on the Gut-Liver Axis. *Frontiers in Pharmacology*, *13*, 910493. https://doi.org/10.3389/fphar.2022.910493

Shulman, S. T. (2004). The History of Pediatric Infectious Diseases. *Pediatric Research*, *55*(1), 163–176. https://doi.org/10.1203/01.PDR.0000101756.93542.09

Siezen, R. J., & Kleerebezem, M. (2011). The human gut microbiome: Are we our enterotypes? *Microbial Biotechnology*, *4*(5), 550–553. https://doi.org/10.1111/j.1751-7915.2011.00290.x

Sinha, R., Abu-Ali, G., Vogtmann, E., Fodor, A. A., Ren, B., Amir, A., Schwager, E., Crabtree, J., Ma, S., Microbiome Quality Control Project Consortium, Abnet, C. C., Knight, R., White, O., & Huttenhower, C. (2017). Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nature Biotechnology*, *35*(11), 1077–1086. https://doi.org/10.1038/nbt.3981

Sommer, F., & Bäckhed, F. (2013). The gut microbiota—Masters of host development and physiology. *Nature Reviews Microbiology*, *11*(4), 227–238. https://doi.org/10.1038/nrmicro2974

Song, S. J., Amir, A., Metcalf, J. L., Amato, K. R., Xu, Z. Z., Humphrey, G., & Knight, R. (2016). Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies. *mSystems*, *1*(3), 10.1128/msystems.00021-16. https://doi.org/10.1128/msystems.00021-16

Song, X., Sun, X., Oh, S. F., Wu, M., Zhang, Y., Zheng, W., Geva-Zatorsky, N., Jupp, R., Mathis, D., Benoist, C., & Kasper, D. L. (2020). Microbial bile acid metabolites modulate gut RORγ+ regulatory T cell homeostasis. *Nature*, *577*(7790), 410–415. https://doi.org/10.1038/s41586-019-1865-0

Soyyılmaz, B., Mikš, M. H., Röhrig, C. H., Matwiejuk, M., Meszaros-Matwiejuk, A., & Vigsnæs, L. K. (2021). The Mean of Milk: A Review of Human Milk Oligosaccharide Concentrations throughout Lactation. *Nutrients*, *13*(8), 2737. https://doi.org/10.3390/nu13082737

Stern, S., Powers, T., Changchien, L.-M., & Noller, H. F. (1989). RNA-Protein Interactions in 30S Ribosomal Subunits: Folding and Function of 16S rRNA. *Science*, *244*(4906), 783–790. https://doi.org/10.1126/science.2658053

Stevens, V. L., Hoover, E., Wang, Y., & Zanetti, K. A. (2019). Pre-Analytical Factors that Affect Metabolite Stability in Human Urine, Plasma, and Serum: A Review. *Metabolites*, *9*(8), 156. https://doi.org/10.3390/metabo9080156

Stewart, C. J., Ajami, N. J., O'Brien, J. L., Hutchinson, D. S., Smith, D. P., Wong, M. C., Ross, M. C., Lloyd, R. E., Doddapaneni, H., Metcalf, G. A., Muzny, D., Gibbs, R. A., Vatanen, T., Huttenhower, C., Xavier, R. J., Rewers, M., Hagopian, W., Toppari, J., Ziegler, A.-G., … Petrosino, J. F. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature*, *562*(7728), Article 7728. https://doi.org/10.1038/s41586-018-0617-x

Stiemsma, L. T., & Turvey, S. E. (2017). Asthma and the microbiome: Defining the critical window in early life. *Allergy, Asthma & Clinical Immunology*, *13*(1), 3. https://doi.org/10.1186/s13223-016-0173-6

Stinson, L. F., Boyce, M. C., Payne, M. S., & Keelan, J. A. (2019). The Not-so-Sterile Womb: Evidence That the Human Fetus Is Exposed to Bacteria Prior to Birth. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.01124

Stulberg, E., Fravel, D., Proctor, L. M., Murray, D. M., LoTempio, J., Chrisey, L., Garland, J., Goodwin, K., Graber, J., Harris, M. C., Jackson, S., Mishkind, M., Porterfield, D. M., & Records, A. (2016). An assessment of US microbiome research. *Nature Microbiology*, *1*(1), Article 1. https://doi.org/10.1038/nmicrobiol.2015.15

Suárez-Martínez, C., Santaella-Pascual, M., Yagüe-Guirao, G., & Martínez-Graciá, C. (2023). Infant gut microbiota colonization: Influence of prenatal and postnatal factors, focusing on diet. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1236254

Sun, S., Wang, H., Howard, A. G., Zhang, J., Su, C., Wang, Z., Du, S., Fodor, A. A., Gordon-Larsen, P., & Zhang, B. (2022). Loss of Novel Diversity in Human Gut Microbiota Associated with Ongoing Urbanization in China. *mSystems*, *7*(4), e00200-22. https://doi.org/10.1128/msystems.00200-22

Sundekilde, U. K., Downey, E., O'Mahony, J. A., O'Shea, C.-A., Ryan, C. A., Kelly, A. L., & Bertram, H. C. (2016). The Effect of Gestational and Lactational Age on the Human Milk Metabolome. *Nutrients*, *8*(5), Article 5. https://doi.org/10.3390/nu8050304

Tamburini, S., Shen, N., Wu, H. C., & Clemente, J. C. (2016). The microbiome in early life: Implications for health outcomes. *Nature Medicine*, *22*(7), 713–722. https://doi.org/10.1038/nm.4142

Tang, W. H. W., & Hazen, S. L. (2014). The contributory role of gut microbiota in cardiovascular disease. *The Journal of Clinical Investigation*, *124*(10), 4204–4211. https://doi.org/10.1172/JCI72331

Tang, W. H. W., Li, D. Y., & Hazen, S. L. (2019). Dietary metabolism, the gut microbiome, and heart failure. *Nature Reviews Cardiology*, *16*(3), 137–154. https://doi.org/10.1038/s41569-018-0108-7

Tedjo, D. I., Jonkers, D. M. A. E., Savelkoul, P. H., Masclee, A. A., van Best, N., Pierik, M. J., & Penders, J. (2015). The Effect of Sampling and Storage on the Fecal Microbiota Composition in Healthy and Diseased Subjects. *PLoS ONE*, *10*(5), e0126685. https://doi.org/10.1371/journal.pone.0126685

Theodoridis, G. A., Gika, H. G., Want, E. J., & Wilson, I. D. (2012). Liquid chromatography–mass spectrometry based global metabolite profiling: A review. *Analytica Chimica Acta*, *711*, 7–16. https://doi.org/10.1016/j.aca.2011.09.042

Thomas, V., Clark, J., & Doré, J. (2015). Fecal Microbiota Analysis: An Overview of Sample Collection Methods and Sequencing Strategies. *Future Microbiology*, *10*(9), 1485–1504. https://doi.org/10.2217/fmb.15.87

Thorman, A. W., Adkins, G., Conrey, S. C., Burrell, A. R., Yu, Y., White, B., Burke, R., Haslam, D., Payne, D. C., Staat, M. A., Morrow, A. L., & Newburg, D. S. (2023). Gut Microbiome

Composition and Metabolic Capacity Differ by FUT2 Secretor Status in Exclusively Breastfed Infants. *Nutrients*, *15*(2), 471. https://doi.org/10.3390/nu15020471

Tomasello, G., Mazzola, M., Jurjus, A., Cappello, F., Carini, F., Damiani, P., Gerges Geagea, A., Zeenny, M. N., & Leone, A. (2017). The fingerprint of the human gastrointestinal tract microbiota: A hypothesis of molecular mapping. *Journal of Biological Regulators and Homeostatic Agents*, *31*(1), 245–249.

Tourlousse, D. M., Narita, K., Miura, T., Sakamoto, M., Ohashi, A., Shiina, K., Matsuda, M., Miura, D., Shimamura, M., Ohyama, Y., Yamazoe, A., Uchino, Y., Kameyama, K., Arioka, S., Kataoka, J., Hisada, T., Fujii, K., Takahashi, S., Kuroiwa, M., … Terauchi, J. (2021). Validation and standardization of DNA extraction and library construction methods for metagenomics-based human fecal microbiome measurements. *Microbiome*, *9*(1), 95. https://doi.org/10.1186/s40168-021-01048-3

Tözün, N., & Vardareli, E. (2016). Gut Microbiome and Gastrointestinal Cancer: Les liaisons Dangereuses. *Journal of Clinical Gastroenterology*, *50*, S191. https://doi.org/10.1097/MCG.0000000000000714

Tremaroli, V., & Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature*, *489*(7415), 242–249. https://doi.org/10.1038/nature11552

Trimigno, A., Khakimov, B., Mejia, J. L. C., Mikkelsen, M. S., Kristensen, M., Jespersen, B. M., & Engelsen, S. B. (2017). Identification of weak and gender specific effects in a short 3 weeks intervention study using barley and oat mixed linkage β-glucan dietary supplements: A human fecal metabolome study by GC-MS. *Metabolomics: Official Journal of the Metabolomic Society*, *13*(10), 108. https://doi.org/10.1007/s11306-017-1247-2

Tsukuda, N., Yahagi, K., Hara, T., Watanabe, Y., Matsumoto, H., Mori, H., Higashi, K., Tsuji, H., Matsumoto, S., Kurokawa, K., & Matsuki, T. (2021). Key bacterial taxa and metabolic pathways affecting gut short-chain fatty acid profiles in early life. *The ISME Journal*, *15*(9), 2574–2590. https://doi.org/10.1038/s41396-021-00937-7

van Best, N., Rolle-Kampczyk, U., Schaap, F. G., Basic, M., Olde Damink, S. W. M., Bleich, A., Savelkoul, P. H. M., von Bergen, M., Penders, J., & Hornef, M. W. (2020). Bile acids drive the newborn's gut microbiota maturation. *Nature Communications*, *11*(1), 3692. https://doi.org/10.1038/s41467-020-17183-8

Van de Peer, Y., Chapelle, S., & De Wachter, R. (1996). A Quantitative Map of Nucleotide Substitution Rates in Bacterial rRNA. *Nucleic Acids Research*, *24*(17), 3381–3391. https://doi.org/10.1093/nar/24.17.3381

Vatanen, T., Jabbar, K. S., Ruohtula, T., Honkanen, J., Avila-Pacheco, J., Siljander, H., Stražar, M., Oikarinen, S., Hyöty, H., Ilonen, J., Mitchell, C. M., Yassour, M., Virtanen, S. M., Clish, C. B., Plichta, D. R., Vlamakis, H., Knip, M., & Xavier, R. J. (2022). Mobile genetic elements from the maternal microbiome shape infant gut microbial assembly and metabolism. *Cell*, *185*(26), 4921-4936.e15. https://doi.org/10.1016/j.cell.2022.11.023

Větrovský, T., & Baldrian, P. (2013). The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLOS ONE*, *8*(2), e57923. https://doi.org/10.1371/journal.pone.0057923

Victora, C. G., Bahl, R., Barros, A. J. D., França, G. V. A., Horton, S., Krasevec, J., Murch, S., Sankar, M. J., Walker, N., & Rollins, N. C. (2016). Breastfeeding in the 21st century: Epidemiology, mechanisms, and lifelong effect. *The Lancet*, *387*(10017), 475–490. https://doi.org/10.1016/S0140-6736(15)01024-7

Vijayvargiya, P., Busciglio, I., Burton, D., Donato, L., Lueke, A., & Camilleri, M. (2018). Bile Acid Deficiency in a Subgroup of Patients With Irritable Bowel Syndrome With Constipation Based on Biomarkers in Serum and Fecal Samples. *Clinical Gastroenterology and Hepatology*, *16*(4), 522–527. https://doi.org/10.1016/j.cgh.2017.06.039

Visconti, A., Le Roy, C. I., Rosa, F., Rossi, N., Martin, T. C., Mohney, R. P., Li, W., de Rinaldis, E., Bell, J. T., Venter, J. C., Nelson, K. E., Spector, T. D., & Falchi, M. (2019). Interplay between the

human gut microbiome and host metabolism. *Nature Communications*, *10*(1), 4505. https://doi.org/10.1038/s41467-019-12476-z

Visekruna, A., & Luu, M. (2021). The Role of Short-Chain Fatty Acids and Bile Acids in Intestinal and Liver Function, Inflammation, and Carcinogenesis. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/fcell.2021.703218

Vivarelli, S., Salemi, R., Candido, S., Falzone, L., Santagati, M., Stefani, S., Torino, F., Banna, G. L., Tonini, G., & Libra, M. (2019). Gut Microbiota and Cancer: From Pathogenesis to Therapy. *Cancers*, *11*(1), Article 1. https://doi.org/10.3390/cancers11010038

Vogtmann, E., Chen, J., Amir, A., Shi, J., Abnet, C. C., Nelson, H., Knight, R., Chia, N., & Sinha, R. (2017). Comparison of Collection Methods for Fecal Samples in Microbiome Studies. *American Journal of Epidemiology*, *185*(2), 115–123. https://doi.org/10.1093/aje/kww177

Vuorela, N., Saha, M.-T., & Salo, M. K. (2011). Change in prevalence of overweight and obesity in Finnish children – comparison between 1974 and 2001. *Acta Paediatrica*, *100*(1), 109–115. https://doi.org/10.1111/j.1651-2227.2010.01980.x

Wahlström, A., Sayin, S. I., Marschall, H.-U., & Bäckhed, F. (2016). Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. *Cell Metabolism*, *24*(1), 41–50. https://doi.org/10.1016/j.cmet.2016.05.005

Walker, A. W., & Hoyles, L. (2023). Human microbiome myths and misconceptions. *Nature Microbiology*, *8*(8), 1392–1396. https://doi.org/10.1038/s41564-023-01426-7

Walker, A. W., Martin, J. C., Scott, P., Parkhill, J., Flint, H. J., & Scott, K. P. (2015). 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome*, *3*, 26. https://doi.org/10.1186/s40168-015-0087-4

Wang, Y., & LêCao, K.-A. (2020). Managing batch effects in microbiome data. *Briefings in Bioinformatics*, *21*(6), 1954–1970. https://doi.org/10.1093/bib/bbz105

Wang, Z., Zolnik, C. P., Qiu, Y., Usyk, M., Wang, T., Strickler, H. D., Isasi, C. R., Kaplan, R. C., Kurland, I. J., Qi, Q., & Burk, R. D. (2018). Comparison of Fecal Collection Methods for Microbiome and Metabolomics Studies. *Frontiers in Cellular and Infection Microbiology*, *8*, 301. https://doi.org/10.3389/fcimb.2018.00301

Watson, E.-J., Giles, J., Scherer, B. L., & Blatchford, P. (2019). Human faecal collection methods demonstrate a bias in microbiome composition by cell wall structure. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-53183-5

Weersma, R. K., Zhernakova, A., & Fu, J. (2020). *Interaction between drugs and the gut microbiome*. https://doi.org/10.1136/gutjnl-2019-320204

Wensel, C. R., Pluznick, J. L., Salzberg, S. L., & Sears, C. L. (2022). Next-generation sequencing: Insights to advance clinical investigations of the microbiome. *The Journal of Clinical Investigation*, *132*(7). https://doi.org/10.1172/JCI154944

Widjaja, F., & Rietjens, I. M. C. M. (2023). From-Toilet-to-Freezer: A Review on Requirements for an Automatic Protocol to Collect and Store Human Fecal Samples for Research Purposes. *Biomedicines*, *11*(10), Article 10. https://doi.org/10.3390/biomedicines11102658

Williams, G. M., Leary, S. D., Ajami, N. J., Chipper Keating, S., Petrosin, J. F., Hamilton-Shield, J. P., & Gillespie, K. M. (2019). Gut microbiome analysis by post: Evaluation of the optimal method to collect stool samples from infants within a national cohort study. *PloS One*, *14*(6), e0216557. https://doi.org/10.1371/journal.pone.0216557

Wilson, I. D., & Nicholson, J. K. (2017). Gut Microbiome Interactions with Drug Metabolism, Efficacy and Toxicity. *Translational Research : The Journal of Laboratory and Clinical Medicine*, *179*, 204–222. https://doi.org/10.1016/j.trsl.2016.08.002

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., Bewtra, M., Knights, D., Walters, W. A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F. D., & Lewis, J. D. (2011). Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science*, *334*(6052), 105–108. https://doi.org/10.1126/science.1208344

Wu, W.-K., Chen, C.-C., Panyod, S., Chen, R.-A., Wu, M.-S., Sheen, L.-Y., & Chang, S.-C. (2019). Optimization of fecal sample processing for microbiome study—The journey from bathroom to bench. *Journal of the Formosan Medical Association*, *118*(2), 545–555. https://doi.org/10.1016/j.jfma.2018.02.005

Xiong, J., Hu, H., Xu, C., Yin, J., Liu, M., Zhang, L., Duan, Y., & Huang, Y. (2022). Development of gut microbiota along with its metabolites of preschool children. *BMC Pediatrics*, *22*(1), 25. https://doi.org/10.1186/s12887-021-03099-9

Xu, F., Zou, L., & Ong, C. N. (2009). Multiorigination of Chromatographic Peaks in Derivatized GC/MS Metabolomics: A Confounder That Influences Metabolic Pathway Interpretation. *Journal of Proteome Research*, *8*(12), 5657–5665. https://doi.org/10.1021/pr900738b

Xu, Z., Malmer, D., Langille, M. G. I., Way, S. F., & Knight, R. (2014). Which is more important for classifying microbial communities: Who's there or what they can do? *The ISME Journal*, *8*(12), 2357–2359. https://doi.org/10.1038/ismej.2014.157

Yadav, H., Lee, J.-H., Lloyd, J., Walter, P., & Rane, S. G. (2013). Beneficial Metabolic Effects of a Probiotic via Butyrate-induced GLP-1 Hormone Secretion*. *Journal of Biological Chemistry*, *288*(35), 25088–25097. https://doi.org/10.1074/jbc.M113.452516

Yang, F., Sun, J., Luo, H., Ren, H., Zhou, H., Lin, Y., Han, M., Chen, B., Liao, H., Brix, S., Li, J., Yang, H., Kristiansen, K., & Zhong, H. (2020). Assessment of fecal DNA extraction protocols for metagenomic studies. *GigaScience*, *9*(7), giaa071. https://doi.org/10.1093/gigascience/giaa071

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A., Lauber, C., Clemente, J. C., Knights, D., … Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature*, *486*(7402), 222–227. https://doi.org/10.1038/nature11053

Ye, S. H., Siddle, K. J., Park, D. J., & Sabeti, P. C. (2019). Benchmarking Metagenomics Tools for Taxonomic Classification. *Cell*, *178*(4), 779–794. https://doi.org/10.1016/j.cell.2019.07.010

Yelverton, C. A., Killeen, S. L., Feehily, C., Moore, R. L., Callaghan, S. L., Geraghty, A. A., Byrne, D. F., Walsh, C. J., Lawton, E. M., Murphy, E. F., Van Sinderen, D., Cotter, P. D., & McAuliffe, F. M. (2023). Maternal breastfeeding is associated with offspring microbiome diversity; a secondary analysis of the MicrobeMom randomized control trial. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1154114

Yokota, A., Fukiya, S., Islam, K. B. M. S., Ooka, T., Ogura, Y., Hayashi, T., Hagio, M., & Ishizuka, S. (2012). Is bile acid a determinant of the gut microbiota on a high-fat diet? *Gut Microbes*, *3*(5), 455–459. https://doi.org/10.4161/gmic.21216

Zhang, S., Paul, S., & Kundu, P. (2022). NF-κB Regulation by Gut Microbiota Decides Homeostasis or Disease Outcome During Ageing. *Frontiers in Cell and Developmental Biology*, *10*. https://doi.org/10.3389/fcell.2022.874940

Zheng, D., Liwinski, T., & Elinav, E. (2020). Interaction between microbiota and immunity in health and disease. *Cell Research*, *30*(6), 492–506. https://doi.org/10.1038/s41422-020-0332-7

Zhernakova, A., Kurilshikov, A., Bonder, M. J., Tigchelaar, E. F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A. V., Falony, G., Vieira-Silva, S., Wang, J., Imhann, F., Brandsma, E., Jankipersadsing, S. A., Joossens, M., Cenit, M. C., Deelen, P., Swertz, M. A., LifeLines cohort study, … Fu, J. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science*, *352*(6285), 565–569. https://doi.org/10.1126/science.aad3369

Zierer, J., Jackson, M. A., Kastenmüller, G., Mangino, M., Long, T., Telenti, A., Mohney, R. P., Small, K. S., Bell, J. T., Steves, C. J., Valdes, A. M., Spector, T. D., & Menni, C. (2018). The fecal metabolome as a functional readout of the gut microbiome. *Nat Genet*, *50*(6), 790–795. https://doi.org/10.1038/s41588-018-0135-7

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., & Goodman, A. L. (2019). Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature*, *570*(7762), 462–467. https://doi.org/10.1038/s41586-019-1291-3

Zmora, N., Suez, J., & Elinav, E. (2019). You are what you eat: Diet, health and the gut microbiota. *Nature Reviews Gastroenterology & Hepatology*, *16*(1), 35–56. https://doi.org/10.1038/s41575-018-0061-2

Zoetendal, E. G., von Wright, A., Vilpponen-Salmela, T., Ben-Amor, K., Akkermans, A. D. L., & de Vos, W. M. (2002). Mucosa-Associated Bacteria in the Human Gastrointestinal Tract Are Uniformly Distributed along the Colon and Differ from the Community Recovered from Feces. *Applied and Environmental Microbiology*, *68*(7), 3401–3407. https://doi.org/10.1128/AEM.68.7.3401-3407.2002

Original Publications

Isokääntä H, Tomnikov N, Vanhatalo S, Munukka E, Huovinen P, Hakanen AJ, Kallonen T. (2024). High-throughput DNA extraction strategy for faecal microbiome studies.
Microbiology Spectrum

I

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Spectrum

| Bacteriology | Methods and Protocols

High-throughput DNA extraction strategy for fecal microbiome studies

Heidi Isokääntä,[1,2] Natalie Tomnikov,[3] Sanja Vanhatalo,[1] Eveliina Munukka,[4,5] Pentti Huovinen,[1] Antti J. Hakanen,[1,3,4] Teemu Kallonen[1,3,4]

AUTHOR AFFILIATIONS See affiliation list on p. 15.

ABSTRACT Microbiome studies are becoming larger in size to detect the potentially small effect that environmental factors have on our gut microbiomes, or that the microbiome has on our health. Therefore, fast and reproducible DNA isolation methods are needed to handle thousands of fecal samples. We used the Chemagic 360 chemistry and Magnetic Separation Module I (MSMI) instrument to compare two sample preservatives and four different pre-treatment protocols to find an optimal method for DNA isolation from thousands of fecal samples. The pre-treatments included bead beating, sample handling in tube and plate format, and proteinase K incubation. The optimal method offers a sufficient yield of high-quality DNA without contamination. Three human fecal samples (adult, senior, and infant) with technical replicates were extracted. The extraction included negative controls (OMNIgeneGUT, DNA/RNA shield fluid, and Chemagic Lysis Buffer 1) to detect cross-contamination and ZymoBIOMICS Gut Microbiome Standard as a positive control to mimic the human gut microbiome and assess sensitivity of the extraction method. All samples were extracted using Chemagic DNA Stool 200 H96 kit (PerkinElmer, Finland). The samples were collected in two preservatives, OMNIgeneGUT and DNA/RNA shield fluid. DNA quantity was measured using Qubit-fluorometer, DNA purity and quality using gel electrophoresis, and taxonomic signatures with 16S rRNA gene-based sequencing with V3V4 and V4 regions. Bead beating increased bacterial diversity. The largest increase was detected in gram-positive genera *Blautia,* Bifidobacterium, and *Ruminococcus*. Preservatives showed minor differences in bacterial abundances. The profiles between the V3V4 and V4 regions differed considerably with lower diversity samples. Negative controls showed signs from genera abundant in fecal samples. Technical replicates of the Gut Standard and stool samples showed low variation. The selected isolation protocol included recommended steps from manufacturer as well as bead beating. Bead beating was found to be necessary to detect hard-to-lyse bacteria. The protocol was reproducible in terms of DNA yield among different stool replicates and the ZymoBIOMICS Gut Microbiome Standard. The MSM1 instrument and pre-treatment in a 96-format offered the possibility of automation and handling of large sample collections. Both preservatives were feasible in terms of sample handling and had low variation in taxonomic signatures. The 16S rRNA target region had a high impact on the composition of the bacterial profile.

IMPORTANCE Next-generation sequencing (NGS) is a widely used method for determining the composition of the gut microbiota. Due to the differences in the gut microbiota composition between individuals, microbiome studies have expanded into large population studies to maximize detection of small effects on microbe–host interactions. Thus, the demand for a rapid and reliable microbial profiling is continuously increasing, making the optimization of high-throughput 96-format DNA extraction integral for NGS-based downstream applications. However, experimental protocols are prone to bias and errors from sample collection and storage, to DNA extraction,

Editor Paul A. Jensen, University of Michigan-Ann Arbor, Ann Arbor, Michigan, USA

Address correspondence to Teemu Kallonen, teemu.kallonen@tyks.fi, or Heidi Isokääntä, hejokun@utu.fi.

E.M. is currently working as Medical Advisor for BioCodex Nordics. A.J.H. reports receiving personal fees for lectures from BioCodex, Merck, and Pfizer. All other authors declare no competing interests.

See the funding table on p. 15.

Received 9 August 2023
Accepted 19 April 2024
Published 15 May 2024

primer selection and sequencing, and bioinformatics analyses. Methodological bias can contribute to differences in microbiome profiles, causing variability across studies and laboratories using different protocols. To improve consistency and confidence of the measurements, the standardization of microbiome analysis methods has been recognized in many fields.

KEYWORDS DNA extraction, high throughput, gut microbiome, fecal sample, method development, sample preservative

Due to better understanding of the microbial communities, new links are constantly discovered between the human microbiome and health (1–7). Host and gut microbiota form a complex and active ecosystem, which harbors an enormous variety of microbes forming a community, playing important roles, e.g., in our metabolism and immune system (8–12). During the early stages of life, the gut microbiome is typically low in diversity. In adulthood, the gut microbiome is relatively stable, yet still susceptible to changes caused by life events. It can be stated that a stable and diverse microbiome is a vital factor of an individual's health and well-being. At an older age, the diversity of microbiome decreases and becomes unstable (13–15). Furthermore, due to the high variation in gut microbiome between individuals, defining a global, unequivocal healthy gut microbiome is challenging (6).

Next-generation sequencing (NGS) is a widely used method for determining the genetic composition of the gut microbiota. Due to the great differences in the composition of the gut microbiota between individuals, microbiome studies have expanded into large populations to maximize detection of these small effects of microbe–host interactions. Thus, the demand for a rapid, efficient, and reliable microbial profiling is continuously increasing, making the optimization of high-throughput 96-format DNA extraction integral prior to NGS-based downstream applications (16–19).

Metagenomics has an indispensable role at many stages of microbiome-based product development, identification of microbial targets and clinical trials. However, workflows and experimental protocols are complex and prone to bias and errors at all steps, from sample collection and storage (20, 21) to DNA extraction (22–24), primer selection (25–27) and sequencing, and bioinformatics analyses (28–31). Previous studies have shown that multiple preservatives can be feasible in sample collection and compatible for DNA isolation (32, 33). Some of the challenges may occur due to differences in bacterial species characteristics, such as persistent cell walls or envelopes. Gram-positive bacteria are typically considered hard-to-lyse because of the thicker peptidoglycan cell wall. Mechanical lysis has been found to increase DNA yield but predispose to DNA shearing. With an ideal DNA extraction method, all bacteria should be recovered equally well (34–36).

Methodological bias can contribute to remarkable differences in observed microbiome profiles, causing variability in results across studies and laboratories with different protocols (37–39). To improve consistency and confidence in the accuracy of the measurements, the standardization of microbiome analysis methods has been recognized as an urgent need by academic, diagnostic, industrial, and regulatory sectors (40–43).

Study aims

This study aimed to investigate what is the best practice for automated DNA extraction for downstream sequencing applications and handling of large sample collections in microbiome research and clinical microbiology. The best practice needs to offer sufficient yield of high-quality DNA, minimized contamination and human error and low variability in the results. This study focuses on the impact of human fecal sample collection (preservative), pre-treatment of DNA extraction, and further, choice of 16S rRNA gene hypervariable region to microbiome results.

MATERIALS AND METHODS

The optimization workflow is represented in Fig. 1. Two commercial sample preservatives, namely OMNIgeneGUT (DNA Genotek, Canada) and DNA/RNA shield fluid (Zymo Research, USA), were tested in fecal sample collection. In addition, four different pre-treatment protocols were tested with all samples including bead beating (mechanical lysis) and proteinase K incubation. Comparison of two 16S rRNA target regions (V3V4 and V4) was also carried out.

Test samples

Human fecal samples from healthy infant, adult, and senior volunteers (n = 3) were simultaneously collected in both OMNIgeneGUT and DNA/RNA shield fluid tubes. After 3 d of storage at room temperature, simulating transport from home to the laboratory by mail, the samples were frozen at −80°C. DNA extraction was performed 2 wk after freezing. The ZymoBIOMICS Gut Microbiome Standard (Zymo Research, USA) was used as a positive extraction control to mimic a human gut microbiome with a known microbial composition. The commercial Gut Standard was used to investigate whether the results were close to the theoretical composition, and all taxa were equally extracted. Manufacturer uses the V3V4 region of the 16S gene to assign reference abundances for the standard. A kit lysis buffer (Chemagic Lysis Buffer 1) and OMNIgene fluid and RNA/DNA shield fluid were used as negative controls. Altogether, 192 samples were sequenced (40 repeats of adult sample, 32 repeats of infant sample, 30 repeats of senior sample, 24 repeats of Gut Standard, 60 negative controls, and 6 PCR controls).

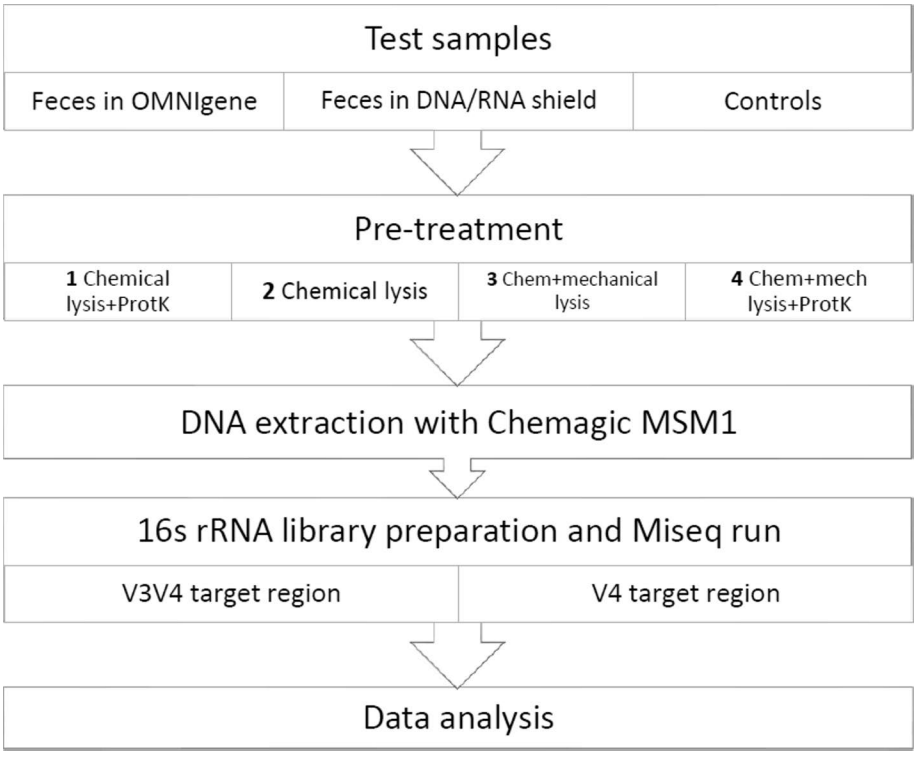


FIG 1 An overview of the study design. Controls included negative controls of extraction (lysis buffer) and preservatives (OMNIgeneGUT and DNA/RNA shield fluid), and Gut Standard as positive extraction controls.

TABLE 1 Pre-treatment groups of DNA extraction

Group	Lysis	Pre-treatment
1	Chemical	Manufacturer's protocol; incubation with proteinase K
2	Chemical	Manufacturer's protocol; incubation without proteinase K
3	Chemical + mechanical	Bead plate + tissue lyser 15 Hz; 2 × 5 min
4	Chemical + mechanical	Bead plate + tissue lyser 15 Hz; 2 × 5 min + proteinase K incubation

Pre-treatment and DNA extraction

Microbial DNA was extracted from adult and infant fecal samples as well as the ZymoBIOMICS Gut Microbiome Standard and negative controls using a DNA Stool 200 H96 kit (PerkinElmer, Finland) with Magnetic Separation Module I (MSM I) extraction robot (PerkinElmer, Finland). Different sample preparation procedures, including bead beating and chemical lysis, were applied to the samples to assess the impact of the sample lysis and homogenization. Three technical replicates of the adult and senior samples and two technical replicates of the infant sample were extracted. The same extraction also included negative controls and the ZymoBIOMICS Gut Standard. Negative controls were placed between fecal samples to detect cross-contamination. The Gut Standard was used to assess the sensitivity of the extraction method.

Pre-treatment procedures were modified from the manufacturer's protocol "Purification Protocol for Human Feces Material Using the Chemagic Magnetic Separation Module I." The following volumes of reagents and samples were used in every pre-treatment group. Lysis Buffer 1 (800 µL) was added to 200 µL of the fecal sample. Subsequently, 925 µL of Lysis Buffer 1 was added to 75 µL of the ZymoBIOMICS Gut Standard. For negative controls (OMNIgene and DNA/RNA shield fluid), the volume of 800 µL of Lysis Buffer 1 was added to 200 µL of each collection fluid. Finally, for Lysis Buffer 1 extraction control, 1 mL of buffer was used. The pre-treatment was divided into four groups (Table 1).

Group 1 included MSM I manufacturer's original protocol with proteinase K incubations. After the addition of the lysis buffer, the tubes were vortexed and 15 µL of proteinase K was added, incubated in a thermo shaker at 70°C for 10 min, followed by incubation at 95°C for 5 min. Samples were centrifuged at a high speed for 5 min. The lysate (800 µL) was then transferred to a sample plate, and the extraction proceeded according to the manufacturer's protocol.

Group 2 included MSM I manufacturer's original protocol without proteinase K incubations.

Group 3 included bead beating with a PowerBead Pro Plates (Glass beads 0.1 mm) and a TissueLyser II (Qiagen, USA). The fecal samples, the Gut Standard, negative controls, and lysis buffer were added to the bead plates, and the bead plate was sealed with a sealing film. The plate was shaken in the TissueLyser II at 15 Hz for 5 min twice. Next, the plate was centrifuged at 4,500 × *g* for 6 min, and 800 µL of the lysate was transferred to the sample plate; the extraction proceeded according to the manufacturer's protocol.

Group 4: bead beating with a bead plate and the TissueLyser II was combined with proteinase K incubations. The plate was shaken in the same way as mentioned in group 3. After shaking, the plate was centrifuged, and 800 µL of the lysate was transferred to 2-mL screw cap tubes with 15 µL of proteinase K. The tubes were vortexed and incubated as mentioned in group 1. The tubes were briefly placed in a spinner, and the lysate was transferred to the sample plate.

After the extraction step, the DNA concentration was measured with a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA) using a Qubit dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific, USA). DNA integrity was evaluated by 1% TBE agarose gel electrophoresis. The DNA was divided into two 100-µL aliquots and stored at −80°C. DNA extraction and a downstream analysis were performed with DNase-/RNase-free plastics.

Library preparation and 16S sequencing

Microbial composition was determined by sequencing both V3V4 and V4 regions of the 16S ribosomal gene using a MiSeq platform (Illumina, USA). The sequence library was constructed according to the Illumina library preparation protocol (44) and V4 library with an in-house protocol (45).

For V4, two replicates of the fecal samples were sequenced. In the V4 library preparation, amplicon PCR and index PCR were combined (45). The DNA was diluted in PCR-grade water to 10 ng/µL prior to PCR. PCR was performed with KAPA HiFi High Fidelity PCR kit with dNTPs (Roche, USA). The desired concentration of each component was the following: 1x for 5x KAPA HiFi Fidelity Buffer, 0.3 mM for dNTP mix, 0.5 U for KAPA HiFi DNA polymerase and PCR-grade water. Sequences of the forward and reverse primers (0.3 µM) were 5′- AATGATACGGCGACCACCGAGATCTACAC -i5- TATGGTAATT-GTGT GCCAGCMGCCGCGGTAA-3′ (forward) and 5′- CAAGCAGAAGACGGCATACGAGAT -i7- AGTC AGTCAG-GCGGACTACHVGGGTWTCTAAT-3′ (reverse), where i5 and i7 indicate the sample specific indices. The concentration of template DNA was 50 ng. The final volume of the reaction was 25 µL. Combined PCR had following conditions: initial denaturation at 98°C for 4 min, followed by 30 cycles consisting of denaturation at 98°C for 20 s, annealing at 65°C for 20 s and extension at 72°C for 35 s, and with a final extension at 72°C for 10 min.

V3V4 sequencing included all replicates of the fecal samples. Sequencing also included positive and negative controls. The V3V4 protocol differed from Illumina's recommendation for the final volumes of the PCR reaction and the DNA visualization procedures. Prior to PCR, the DNA samples were diluted to 2.5 ng/µL in PCR-grade water. Briefly, amplicon PCR included 2x KAPA HiFi HotStart ReadyMix (Roche, USA), Illumina amplicon forward and reverse primers (6.6 µM), PCR-grade water, and microbial DNA (16.5 ng). The final volume of the amplicon PCR reaction was 33 µL. Index PCR was performed according to Illumina's instructions.

After PCR, 8 µL of the product was analyzed with 1.5% TAE agarose gel (120 V, 1 h). The concentration of the library samples was measured with a Qubit Fluorometer using a Qubit dsDNA High Sensitivity Assay kit. The 4 nM library pool was denatured, diluted to a concentration of 4 pM, and an 8% denaturized PhiX control (Illumina, USA) was added. The library samples were sequenced with a MiSeq Reagent kit v3, 600 cycles (Illumina, USA) on a MiSeq system with 2 × 300 base pair (bp) paired ends following the manufacturer's instructions. The library samples were sequenced with an Illumina MiSeq Reagent kit v3 (600 cycles) on a MiSeq system with 2 × 250 bp paired ends following the manufacturer's instructions. A positive plasmid control (DNA 7-mock) and a negative control (PCR-grade water) were included in library preparation to control the PCR.

Bioinformatic methods and data visualization

The raw sequence data for both libraries were processed and analyzed with a CLC Microbial Genomics Module (CLC Genomics Workbench 21.0.3, Qiagen, USA). The workflows "Data quality control (QC) and operational taxonomic unit (OTU) clustering" including read trim and "Alpha and beta diversities" were used to analyze the data using default settings. The 16S read pairs were merged. The cutoff for the number of reads of the fecal samples was 200,000 sequences; however, negative controls were not filtered based on the number of reads. One infant sample from V3V4 sequencing was excluded due to the low number of reads (reads after trim, 193,762). An index and adapter trims from the 5′ end were performed for both libraries. Sequences were mapped using a SILVA 16S version 132 with a 97% similarity for OTU clustering. In the diversity analysis, low abundance OTUs were filtered (<100 reads), and OTUs were aligned using MUSCLE (MUltiple Sequence Comparison by Log-Expectation). A neighbor-joining tree was used to calculate alpha diversity. Chao1 and the total OTU number (observed OTUs) were selected to represent alpha diversity. Alpha diversity of the total OTU number represents richness; the number of species observed in each sample and Chao1 estimates the total richness that accounts for unobserved species (46). Two-sample *t*-test was used to test statistically significant difference in alpha diversity between pre-treatment

groups. A rarefaction level of 78,948 was used for the alpha diversity. Beta diversity calculations were performed using the principal coordinate analysis (PCoA) with Jaccard and Bray-Curtis. OTU tables and alpha and beta diversities were exported to a RStudio 4.1.1 and GraphPad Prism 9.0.1 for data visualization. Differential abundances (DA) were analyzed with CLC Microbial Genomics Module and Deseq2 (47) in R environment. The reported DA results are consensus between the two methods mentioned above.

RESULTS

DNA yields

All pre-treatment methods yielded sufficient amount of DNA (>7 ng/µL) in the fecal samples. The non-bead-beating group 2 with both stabilizers produced the highest concentration for the adult samples. For senior samples, collected in OMNIgeneGUT, the highest concentration was observed in group 4 (bead beating and prot K). In infant samples, the highest concentration was achieved without bead beating (groups 1 and 2). DNA concentrations are shown in Table S1.

Adult fecal samples seemed to have the lowest deviation in concentrations across pre-treatment groups, whereas infant samples had the highest standard deviation. OMNIgeneGUT offered a little higher concentration with senior and infant samples than DNA/RNA shield, but with the adult sample, there was no marked difference. The expected concentration of the Gut Standard was 5 ng/µL according to the manufacturer. Different pre-treatments yielded, on average, 5.35 ng/µL of the standard (4.0–6.4 ng/µL) (Table S1).

The integrity of the DNA isolates was detected with gel electrophoresis, which showed that all fecal samples had a visible amount of DNA (Fig. S1). Senior samples in OMNI ("SO") had some fragmentation. Correspondingly, those senior samples also had the highest DNA concentrations of all samples. Extraction controls ("EC") were all pure.

Gut Standard

In relation to the ZymoBIOMICS Gut Standard, the results were relatively similar across the pre-treatment groups within the same sequencing target. However, the V3V4 sequencing produced more similar results to the Gut Standard than V4, which notably differed from the manufacturer's expected abundances (Fig. 2).

The V3V4 sequencing produced higher relative abundances of the genera *Fusobacterium, Clostridioides, Akkermansia, Bacteroides, Veillonella*, and *Prevotella* than the manufacturer stated. On the other hand, the genera *Bifidobacterium* and *Lactobacillus* were less abundant than expected. With the V3V4 sequencing, differences between the pre-treatments were minor and differences in relative abundances fluctuated between 1% and 2%.

V4 sequencing favored the genera *Veillonella and Prevotella* in comparison with the manufacturer's reported values (Fig. 2). V4 sequencing detected five to six times higher relative abundance of *Prevotella* and a twofold higher abundance of *Veillonella* than the manufacturer's reported abundances. The genera *Faecalibacterium, Roseburia, Bacteroides, Fusobacterium, Clostridioides, Akkermansia, Bifidobacterium*, and *Lactobacillus* were lower in abundance in comparison with the Gut Standard. The V4 sequencing results had more variation between the pre-treatment groups (1%–10%) than the V3V4 sequencing. The genera *Veillonella* and *Prevotella* decreased in abundance from the non-bead-beating groups (1 and 2) to the bead-beating groups (3 and 4). Subsequently, the genera *Bacteroides, Faecalibacterium, Roseburia, Lactobacillus*, and *Bifidobacterium* increased in relative abundance with bead beating (groups 3 and 4). The relative abundance of *Escherichia* stayed consistent across the pre-treatment groups and sequencing targets.

Both sequencing target areas favored the genera *Veillonella* and *Prevotella* in relation to the manufacturer's result, whereas the genera *Bifidobacterium* and *Lactobacillus* were lower in abundance. *Enterococcus* was detected in all samples in the V3V4 sequencing. With the V4 sequencing, *Enterococcus* was seen only in pre-treatment groups 3 and 4.

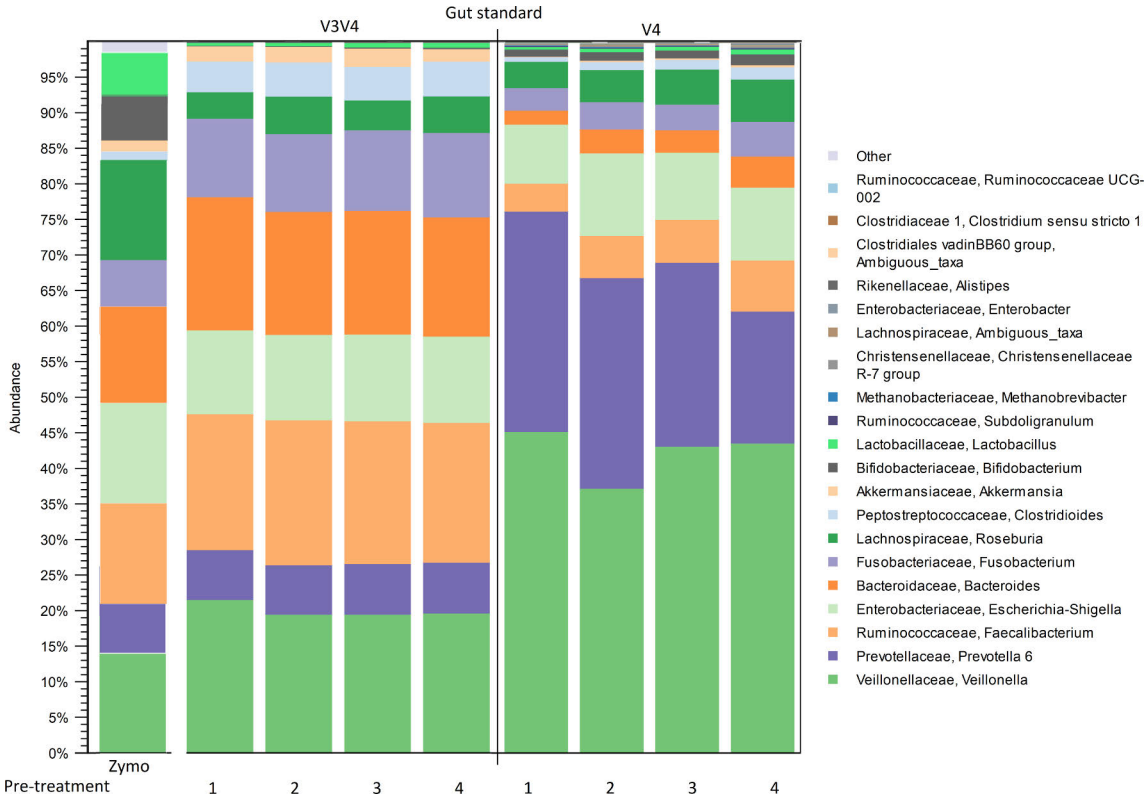


FIG 2 Relative abundance of Gut Standard across different pre-treatments with V3V4 and V4 sequencing. Numbers indicate pre-treatment groups (1–4). The manufacturer's (Zymo) expected abundances are shown on the left of the figure. Legend shows the 20 most abundant genera.

Differences between replicates in relative abundances fluctuated between 0% and 1% with V3V4, and 1% and 4% with V4. In addition to the expected genera, V4 detected 121 and V3V4 detected 31 additional genera, with an abundance of under 1%.

Fecal samples: relative and differential abundances

The adult fecal samples were dominated by the phyla *Firmicutes*, followed by *Bacteroidetes* and *Actinobacteria*. The profiles were relatively similar for both sequencing targets. Bead beating (groups 3 and 4) added signatures from hard-to-lyse gram-positive genera such as *Blautia, Bifidobacterium*, and *Ruminococcus torques* group (Fig. S2). Samples with DNA/RNA shield fluid had a higher abundance of *Faecalibacterium*. Samples with OMNIgeneGUT showed a higher abundance of Bacteroides. The effect of bead beating was not as visible when DNA/RNA shield fluid was utilized.

For senior fecal samples, the most abundant phyla were *Firmicutes*, *Bacteroidetes*, *and Proteobacteria* (Fig. S3). The profiles were relatively similar within the same sequencing target; however, V3V4 and V4 differed from each other. Genus *Klebsiella* from family *Enterobacteriaceae* was missing in V4 sequenced samples, and genus *Enterobacter* was missing in V3V4 sequenced samples.

The most abundant phyla in the infant samples were *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*. V3V4 sequencing favored genera *Bacteroides* and *Klebsiella*, whereas *Veillonella* and *Enterobacter* were more abundant with the V4 sequencing (Fig. S4). There were minor increases in the abundances of *Enterobacter and Pantoea* genera in bead-beating groups 3 and 4 in the V4 sequencing. The profiles were relatively similar

within the same sequencing target. The effect of the target region had a high impact on the infant sample. The profiles were dominated with the same bacteria but the ratios differed greatly.

Taken together, differential abundance analysis (DAA) of fecal samples (Fig. 3) shows that bead beating increased the abundances of several gram-positive bacteria. These differences were significant (FDR-$P \leq 0.05$); however, DAA has limitations, particularly in the context of high inter-individual variation in the microbiome, which may explain the high variability in abundances and prevalence. Figure 3 shows the genera that were both found by the CLC and Deseq2.

Alpha diversities

The alpha diversity indexes were calculated based on observed OTUs and Chao1 (Fig. 4). Both indexes showed similar results. In adult and senior samples, the bead-beating groups 3 and 4 had higher alpha diversity levels than the groups without bead beating (adult $P = 0.0003$, senior $P = 0.0017$). The highest alpha diversity in the adult samples was in group 3, and in the senior samples in group 4. Both bead-beating groups had higher variation compared to the non-bead-beating groups (1 and 2) in adult samples. Senior samples had less variation in the bead-beating groups in Chao1 metric.

With the infant samples, on average, the bead-beating group 3 produced the highest diversity. The variation between the replicates was relatively high. Overall, the effect

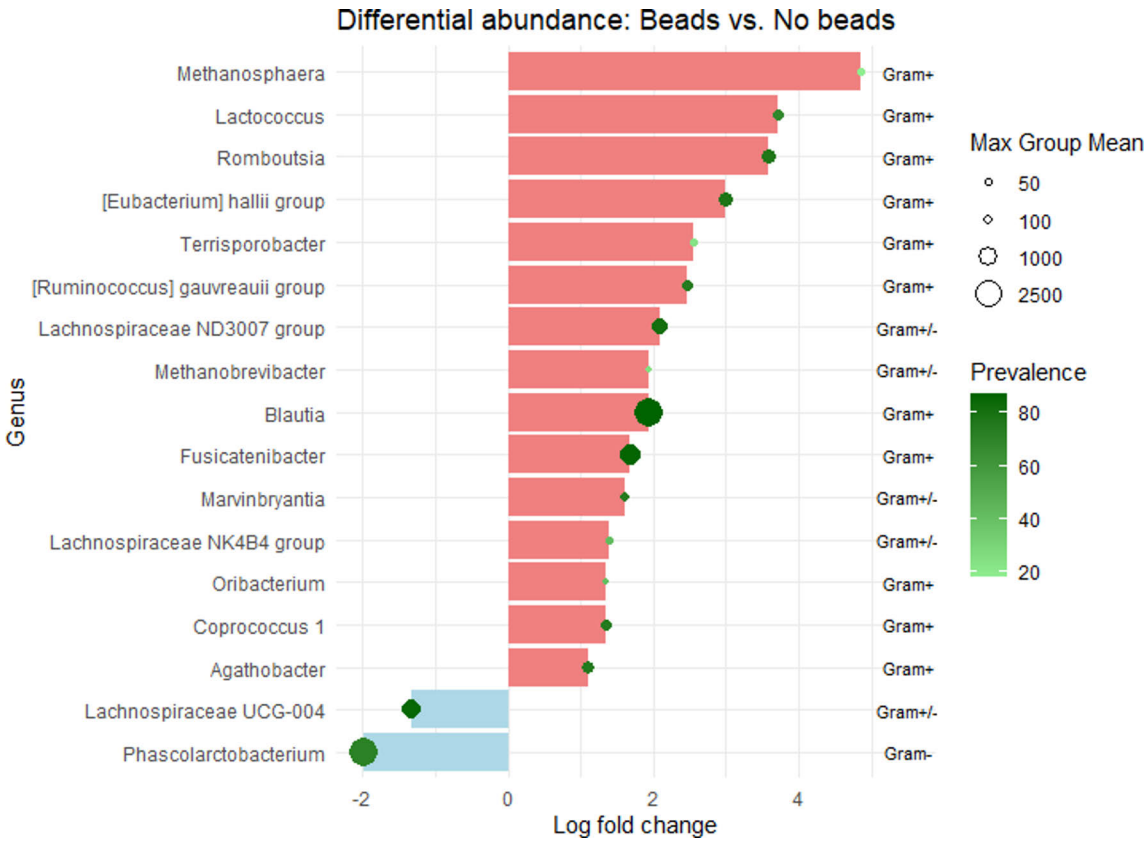


FIG 3 The effect of bead beating. Differential abundance analysis with fecal samples, comparison of bead beating versus no bead beating. X-axis shows log2 fold change. Maximum group means (abundances) are illustrated with different dot sizes, and prevalence of genus by color scale (green). The figure summarizes the consensus of CLC and Deseq2 results. These differences were significant (FDR-$P \leq 0.05$).

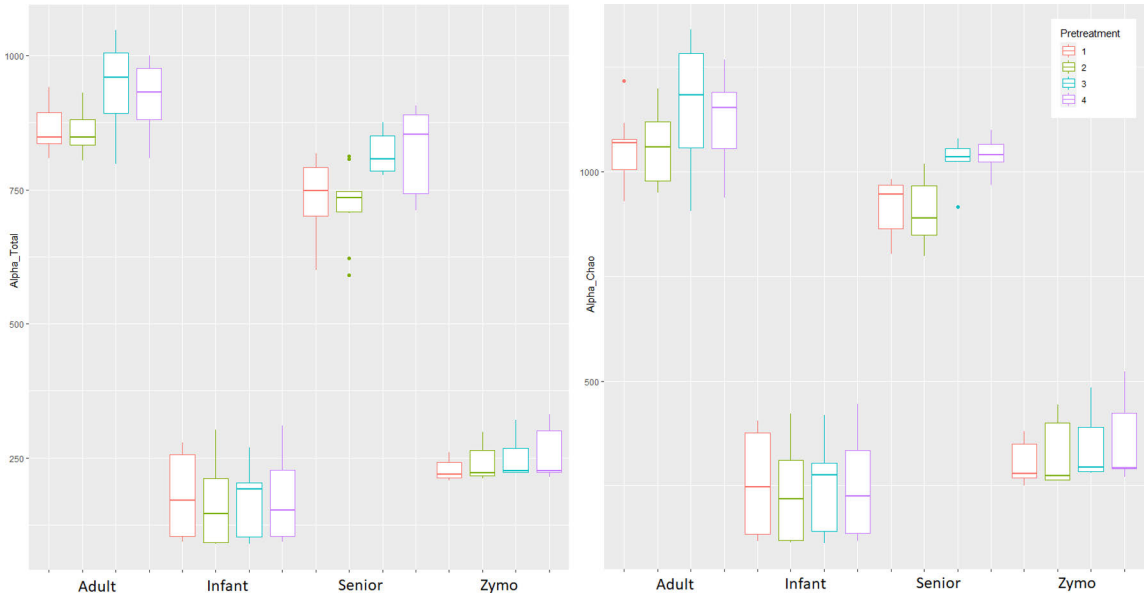


FIG 4 Alpha diversity indexes in boxplots by observed OTUs (left) and Chao1 (right) across sample types and pre-treatment groups. Boxplots show minimum, first quartile (Q1), median, third quartile (Q3), and maximum and outliers.

of pre-treatment was smaller in the infant samples compared to the adult and senior samples. Similarly, with the Gut Standard, the effect of pre-treatment was minor. Group 3 and 4 produced a slightly higher alpha diversity. In the Gut Standard and infant samples, the difference between bead-beating and non-bead-beating groups was not statistically significant.

Beta diversities

Figure 5 summarizes the beta diversities between all the sequenced samples using Bray-Curtis (A) and Jaccard (B) distance metrics. All the sequenced samples (total = 192) clustered in their own groups in exception of seven negative controls, which were located in clusters of fecal samples or the Gut Standard. There were no shifts from one fecal sample group to another. Bead beating did not increase the read count ($P = 0.36$). Moreover, grouping by the 16S target region (V3V4 and V4) can be seen within the subject clusters (Fig. S5).

Bead-beating samples without proteinase K incubation (3) and with proteinase K incubation (4) are loosely grouped in the adult and senior fecal samples (Fig. 6). Similarly, the chemical lysis samples with proteinase K incubation (1) and without proteinase K (2) are loosely grouped. Both sequencing methods exhibit similar trends in the pre-treatment grouping. However, the V4 sequenced senior samples seemed to have more variation within groups 1 and 2. Infant fecal samples did not exhibit specific grouping across the pre-treatment groups.

Negative controls

Negative controls were included in analysis to detect possible cross-contamination (Fig. 7). The negative controls with V3V4 sequencing were dominated by genera present in the fecal samples, such as *Bacteroides*, *Faecalibacterium*, and *Klebsiella*. Approximately half of the V3V4 sequenced negative controls were dominated by the genus *Pseudomonas*, and those profiles were similar to the PCR-0-control, which contained only PCR-grade water. The PCR-0 controls in V3V4 were dominated by the genera of *Pseudomonas*,

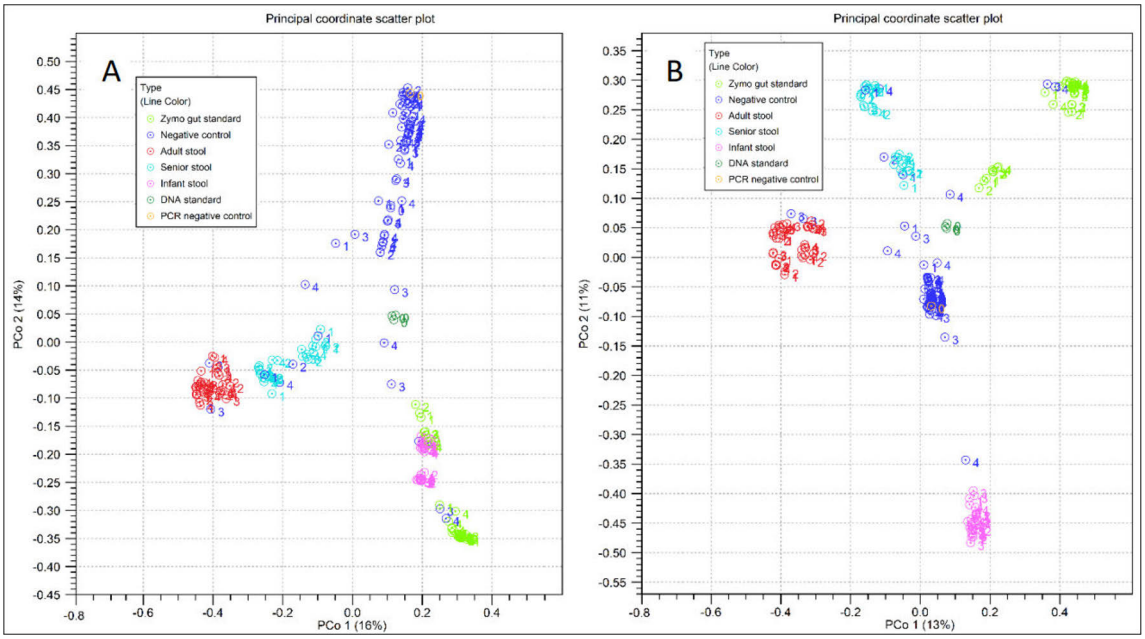


FIG 5 Beta diversity by Bray-Curtis (A) and Jaccard (B) among all sample types. Sample types are listed in legend. Numbers indicate pre-treatment groups 1–4.

Serratia, *and Delftia*, which were not commonly detected in the fecal samples. Accordingly, the negative controls within V4 sequencing were dominated by genera present in the fecal samples, such as *Bacteroides*, *Veillonella*, and *Alloprevotella*. The V3V4 sequenced negative controls had more variation in their profiles, whereas the V4 results were more uniform.

The V3V4 sequenced negative controls had, on average, 53,203 reads of trimmed pairs. The V4 sequenced controls had, on average, 27,387 reads of trimmed pairs. The individual read counts of the negative and extraction controls are shown in Table 2. The V3V4 sequenced controls had a higher read count and more variation than those sequenced by V4.

The number of reads in OTUs across the different sample types can be seen in Figure 8. Although there were a relatively high number of reads in the negative controls, the level was still notably lower compared to the actual samples. If a threshold is needed in quality control, those should be based on relative fraction in read counts (reads in OTUs or total read count). Based on the reads in OTUs, the QC fraction (arithmetic mean of read counts in negative per positive) was 12%, but it should be noted that negative samples had a skewed read count distribution.

DISCUSSION

In this study, we applied several methodological approaches to explore the optimization of a high-throughput stool pre-treatment and DNA isolation method suitable for downstream microbiome sequencing analyses. We established a pre-treatment protocol, which includes bead-beating step followed by automated DNA extraction with Chemagic DNA Stool kit, both in 96-format.

Bead beating yielded a higher microbial diversity and the increase of harder-to-lyse gram-positive bacteria in the fecal sample profiles, as previously reported (22, 48). The senior and adult fecal samples showed the highest alpha diversities with bead beating. Infant fecal samples showed variable microbial diversities across the pre-treatment groups, indicating that infant samples do not benefit from bead beating as much as

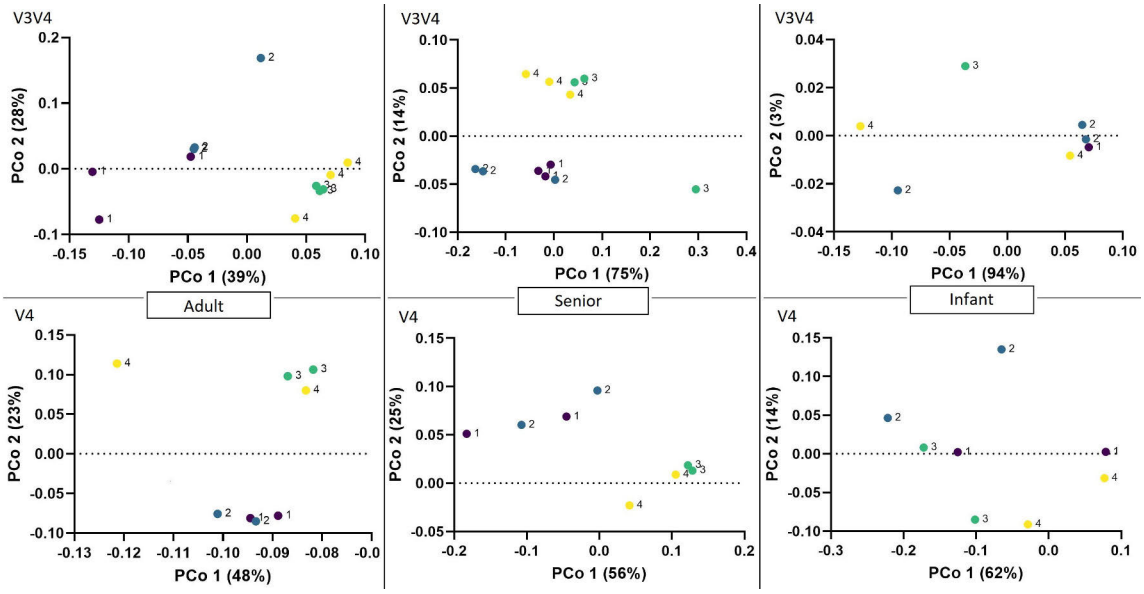


FIG 6 Beta diversities of fecal samples in V3V4 and V4 across different pre-treatment groups with Bray-Curtis measure. Pre-treatments groups are indicated with differently colored dots (1, purple; 2, blue; 3, green; 4, yellow) and numbers adjacent to the dots.

senior and adult samples, because infant gut microbiota is not as rich. It seems that the higher the diversity, the higher the impact of the bead beating. The isolation protocol was also evaluated by using the Gut Microbiome Standard. The observed compositions of the Gut Standard samples differed considerably from the expected theoretical compositions in all the methods. Furthermore, the compositions of Gut Standard showed minor variability between the pre-treatment methods, as well as considerable variation between the sequenced 16S variable regions.

Both preservatives, OMNIgeneGUT and DNA/RNA shield, were usable for DNA extraction and sequencing. The replicates from DNA/RNA shield fluid were more comparable and this might be caused by the 1:10 ratio of sample and preservative, whereas OMNIgeneGUT had a ratio of 1:4. The microbiome profiles showed minor differences between preservatives. The DNA shield had higher abundances of *Faecalibacterium* while OMNIgeneGUT showed higher abundances of *Bacteroides*. These results are in line with Chen et al. (25). OMNIgeneGUT tubes were more practical in terms of sample collection; the correct sample volume was easy to collect, and the mixing was convenient with metal bead in the tube. OMNIgeneGUT can be convenient in field studies with at-home-collected samples and it has been used successfully in several studies (49–51).

The 16S variable target region had a high impact on the microbial compositional profiles as has already been reported (45, 52–54). However, differences in family and genus abundances occurred in similar proportions in the different pre-treatment groups. The impact of the variable region was highest with the infant sample and lowest with the adult sample. Different sequencing targets seemed to favor different bacterial genera. Indeed, the sequencing region seemed to have more impact on the compositional profiles than the pre-treatment. It seemed that the lower the diversity, the higher the impact of the V-target region. It was expected that sequencing with V3V4 would produce more results similar to the Gut Standard because the same target area was used by the manufacturer (55). To reach more rigorous quality control, the use of spike-in controls could be beneficial when controlling the resolution of the DNA extracts and the quality of the libraries. Detection thresholds are challenging to set, and they should be

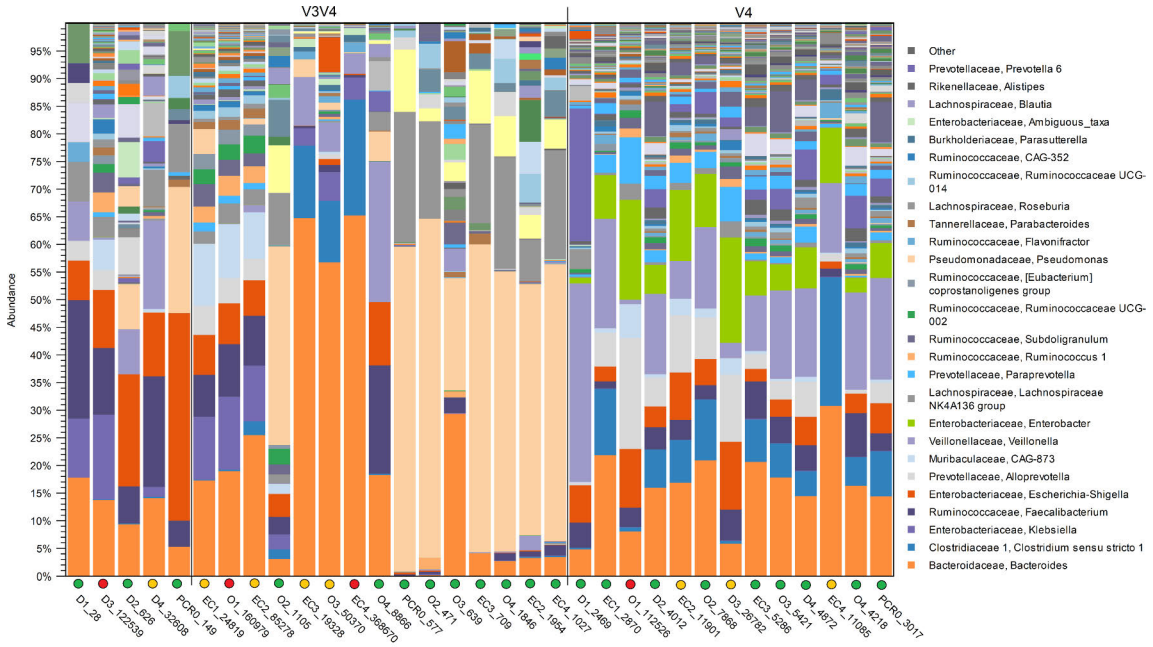


FIG 7 Relative abundances of V3V4 and V4 sequenced negative controls, DNA/RNA shield fluid (D), OMNIgeneGUT (O), Lysis Buffer 1 as extraction controls (EC), and PCR blanks (PCR0). Extraction groups are indicated with numbers 1–4. Numbers after the underscore are reads in OTUs. Colored circles indicate the quantity of reads in OTUs (green < 10,000, yellow 10,000–100,000, red > 100,000).

calculated based on individual runs and sample type and can be highly variable between laboratories. Because the NGS methods are prone to contamination, the possibility of contamination and a need of reruns should be considered, if the read count in negative samples rises above the relative QC fraction (~10%).

The selected protocol included OMNIgeneGUT tubes for sample collection, pre-treatment with bead beating and proteinase K incubation (group 4), a Chemagic stool kit, and an MSM1 extraction robot. The turnaround time for the selected pre-treatment was long, because pre-treatment 4 included bead beating and proteinase K incubation. Moreover, pre-treatments 3 and 4 were more expensive, because bead plates were not included in the extraction kit. Proteinase K was estimated to be necessary to avoid DNases and inhibitors when DNA is stored for a longer period. Manual extraction kits were not tested or compared with Chemagic, because the aim was to an automated method with reduced hands-on time. Based on its DNA yield, efficient recovery of DNA from gram-positives, and overall small bias, as well as its reasonable turnaround time and cost, we selected pre-treatment 4 as the basis for our standard operating procedure (SOP) for DNA extraction. To reduce hands-on time, adjustable multi-channel pipettes and a plate format thermomixer are utilized in further extractions for upcoming studies with large sample collections. To assist the pipetting of fecal material, OMNIgene Liquefaction (DNA Genotek, Canada) can be utilized to make samples less viscous. Chemagic 360 instrument (PerkinElmer, Germany), the next version of MSM1, can also be used for DNA extractions with identical principal and the same stool kit. Bioanalyzer, Tapestation, or a similar device should be utilized in quality control, especially when approaching shotgun sequencing.

The relatively high read count of the negative controls can be explained by the 96-well plate format; the samples are close to each other, and aerosols cannot be fully

TABLE 2 Read counts of negative controls with V4 and V3V4[a]

Negative controls, V4, read counts (paired, trimmed pairs)					
ID	**Reads**	**ID**	**Reads**	**ID**	**Reads**
DNA shield 1_1	✓ 2,863	Omni 1_1	✓15,641	EC 1_1	✓5,415
DNA shield 1_2	✓ 3,271	Omni 1_2	‼ 123,738	EC 1_2	✓5,864
DNA shield 2_1	! 26,052	Omni 2_1	✓15,595	EC 2_1	! 34,811
DNA shield 2_2	! 34,934	Omni 2_2	! 25,796	EC 2_2	! 33,879
DNA shield 3_1	! 40,892	Omni 3_1	! 39,777	EC 3_1	! 57,900
DNA shield 3_2	! 54,439	Omni 3_2	! 36,932	EC 3_2	✓15,211
DNA shield 4_1	✓21,053	Omni 4_1	✓19,867	EC 4_1	✓2,427
DNA shield 4_2	! 31,600	Omni 4_2	✓6,395	EC 4_2	✓16,672
PCR-0_V4	✓13,657				
Negative controls, V3V4, read counts (paired, trimmed pairs)					
ID	**Reads**	**ID**	**Reads**	**ID**	**Reads**
DNA shield 1_1	! 85,290	Omni 1_1	✓5,857	EC 1_1	✓4,367
DNA shield 1_2	✓2,188	Omni 1_2	‼ 261,396	EC 1_2	! 41,043
DNA shield 2_1	✓150	Omni 2_1	✓4,526	EC 2_1	✓2,956
DNA shield 2_2	✓108	Omni 2_2	✓2,576	EC 2_2	‼ 148,071
DNA shield 3_1	‼ 200,210	Omni 3_1	‼ 238,062	EC 3_1	‼ 239,797
DNA shield 3_2	✓856	Omni 3_2	! 77,678	EC 3_2	! 29,192
DNA shield 4_1	✓565	Omni 4_1	✓4,286	EC 4_1	✓2,511
DNA shield 4_2	! 64,789	Omni 4_2	✓14,545	EC 4_2	‼ 553,171
PCR-0_V3V4_1	✓1,427				
PCR-0_V3V4_2	✓ 888				

[a]DNA shield, DNA/RNA shield fluid; Omni, OMNIgeneGUT; EC, extraction control/lysis buffer; first number, pre-treatment group; second number, number of replicate, e.g., EC1_2 is an extraction control with pre-treatment 1 and second replicate. The symbols indicate the number of reads: ✓ < 25,000; 25,000 < ! < 100,000; ‼ > 100,000 reads.

avoided. Cross-contamination was observed from the negative control read counts, the relative abundances of the controls, and beta diversity resembling those of the fecal samples. V3V4 can be more problematic due to two PCR steps and more library preparation steps where contamination might be introduced. Amplicon PCR product purification with V3V4 is particularly susceptible to contamination, because sample-specific indexes have not yet been added. Correspondingly, V4 has only one PCR step, and the sample-specific indexes have been added to the primers.

Due to the variable consistency of the fecal samples, it can be challenging to obtain an equal amount of sample for all replicates. Fecal matter is proven to have a heterogenous consistency of a semi-solid mixture of endogenous and exogenous material (56), and these together might explain the variation within replicates. This variation relates to the resolution of the protocol (fluctuation in relative abundance: V3V4 0%–1% and V4 1%–4%), and it also indicates error rate, which should be considered when analyzing low abundances. Technical replicates are needed to estimate not only the degree of variation but also the existence of contamination. This was a pilot study for a DNA extraction in a large human cohort. Due to the low sample size, statistical analyses were limited. In further studies, a larger sample size is recommended for the statistical power that is needed to observe the effect of different methodological choices.

Conclusion

The applied 96-format extraction system, including the sample pre-treatment steps, proved to be a functional workflow in stool DNA extraction in clinical microbiome research. The sequencing target region effect was larger than the effect of the pre-treatment method. It is notable that the choice of the sample preservative, bead beating, and 16S target region have a varying effect on microbiome profiles. These results indicate the need for standardized methods for microbial profiling. Therefore, we propose the

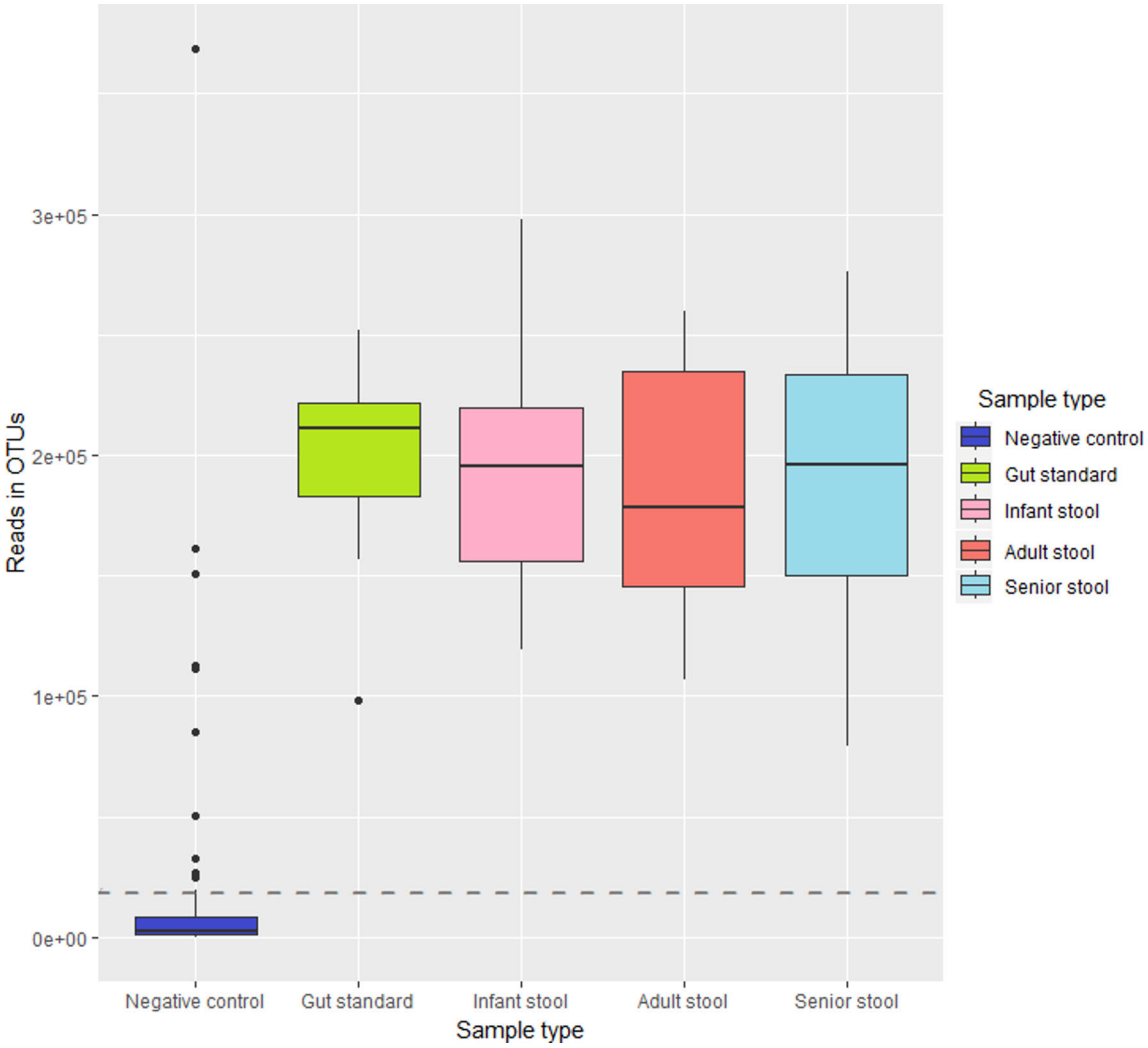


FIG 8 Number of reads in OTUs in different sample types. The horizontal lines inside boxplots indicate the average in a sample type. The horizontal dotted lines reflect the QC fraction.

following points to achieve best practice: consider using a well-established preservative to maintain microbial integrity during storage, utilize a 96-format extraction system coupled with bead beating for efficient and high-throughput stool DNA extraction, and control the level of contamination and bacterial coverage of chosen methods.

ACKNOWLEDGMENTS

We thank TYKS laboratories and Anna-Katariina Aatsinki for the assistance and resources in the analyses.

This study was supported by research grants from The Finnish Foundation for Cardiovascular Research, The Diabetes Research Foundation, the Finnish Cultural Foundation, Signe and Ane Gyllenberg Foundation, the state research funding of Turku University Hospital, and Doctoral Program in Clinical Research at the University of Turku.

The funders had no role in the study design, data collection, analysis, interpretation, or writing of the report.

H.I. drafted the manuscript, which was refined by N.T., S.V., E.M., P.H., A.J.H., and T.K. S.V., and T.K. designed the experiments. N.T., S.V., and H.I. performed the laboratory analyses, and H.I., N.T., S.V., and T.K. carried out the data analysis. All authors approved the final version and had final responsibility for the decision to submit for publication. All authors consent to the publication of the manuscript and have approved the final version.

AUTHOR AFFILIATIONS

[1]Infections and Immunity Unit, Institute of Biomedicine, University of Turku, Turku, Finland
[2]Centre for Population Health Research, University of Turku, Turku, Finland
[3]Department of Clinical Microbiology, Tyks Laboratories, Turku University Hospital, Turku, Finland
[4]Clinical Microbiome Bank, Microbe Center, Turku University Hospital and University of Turku, Turku, Finland
[5]Division of Digestive Surgery and Urology, Turku University Hospital, Turku, Finland

AUTHOR ORCIDs

Heidi Isokääntä http://orcid.org/0000-0002-2443-575X
Teemu Kallonen http://orcid.org/0000-0003-1741-6486

FUNDING

Funder	Grant(s)	Author(s)
Sydäntutkimussäätiö (Finnish Foundation for Cardiovascular Research)		Heidi Isokääntä
Diabetesliitto (Finnish Diabetes Association)		Heidi Isokääntä
Signe ja Ane Gyllenbergin Säätiö (Signe and Ane Gyllenberg Foundation)		Heidi Isokääntä

AUTHOR CONTRIBUTIONS

Heidi Isokääntä, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing | Natalie Tomnikov, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft | Sanja Vanhatalo, Investigation, Methodology, Writing – review and editing | Eveliina Munukka, Conceptualization, Validation, Writing – review and editing | Pentti Huovinen, Project administration, Resources, Supervision, Validation, Writing – review and editing | Antti J. Hakanen, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing | Teemu Kallonen, Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

Sequence data (FastQ-files) and metadata have been deposited into the SRA database under the BioProject PRJNA955433. Supplementary figures and tables are available in a Word document and Excel files accompanying the manuscript. STORMS checklist is available at: https://figshare.com/s/a14e81de02317b2b7580.

ETHICS APPROVAL

Fecal samples were obtained from anonymous fecal donations of three healthy individuals; informed consent was obtained from all donors. Research using anonymized

biological material is not restricted by the Finnish legislation and ethical committee in University of Turku on research involving human beings. Consequently, an ethical review statement was not necessary.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental figures and table (Spectrum02932-23-s0001.docx). Fig. S1–S5; Table S1.
Table S2 (Spectrum02932-23-s0002.xlsx). 6,000 genera (used for negative controls).
Table S3 (Spectrum02932-23-s0003.xlsx). 450 genera (used for fecal samples and positive controls).
Table S4 (Spectrum02932-23-s0004.xlsx). OTU table of relative abundances.
Table S5 (Spectrum02932-23-s0005.xlsx). Table of sample metadata and read counts.

REFERENCES

1. Kelly CR, Ihunnah C, Fischer M, Khoruts A, Surawicz C, Afzali A, Aroniadis O, Barto A, Borody T, Giovanelli A, et al. 2014. Fecal microbiota transplant for treatment of *Clostridium difficile* infection in immunocompromised patients. Am J Gastroenterol 109:1065–1071. https://doi.org/10.1038/ajg.2014.133
2. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, D Lieber A, Wu F, Perez-Perez GI, Chen Y, Schweizer W, Zheng X, Contreras M, Dominguez-Bello MG, Blaser MJ. 2016. Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci Transl Med 8:343ra82. https://doi.org/10.1126/scitranslmed.aad7121
3. Dominguez-Bello MG, Godoy-Vitorino F, Knight R, Blaser MJ. 2019. Role of the microbiome in human development. Gut 68:1108–1114. https://doi.org/10.1136/gutjnl-2018-317503
4. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, Kosciolek T, Janssen S, Metcalf J, Song SJ, Kanbar J, Miller-Montgomery S, Heaton R, Mckay R, Patel SP, Swafford AD, Knight R. 2020. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature 579:567–574. https://doi.org/10.1038/s41586-020-2095-1
5. Taylor BC, Lejzerowicz F, Poirel M, Shaffer JP, Jiang L, Aksenov A, Litwin N, Humphrey G, Martino C, Miller-Montgomery S, Dorrestein PC, Veiga P, Song SJ, McDonald D, Derrien M, Knight R. 2020. Consumption of fermented foods is associated with systematic differences in the gut microbiome and metabolome. mSystems 5:e00901-19. https://doi.org/10.1128/mSystems.00901-19
6. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. 2018. Current understanding of the human microbiome. Nat Med 24:392–400. https://doi.org/10.1038/nm.4517
7. Shreiner AB, Kao JY, Young VB. 2015. The gut microbiome in health and in disease. Curr Opin Gastroenterol 31:69–75. https://doi.org/10.1097/MOG.0000000000000139
8. Rooks MG, Garrett WS. 2016. Gut microbiota, metabolites and host immunity. Nat Rev Immunol 16:341–352. https://doi.org/10.1038/nri.2016.42
9. Lamichhane S, Sen P, Dickens AM, Orešič M, Bertram HC. 2018. Gut metabolome meets microbiome: a methodological perspective to understand the relationship between host and microbe. Methods 149:3–12. https://doi.org/10.1016/j.ymeth.2018.04.029
10. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, Ross MC, Lloyd RE, Doddapaneni H, Metcalf GA, et al. 2018. Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature 562:583–588. https://doi.org/10.1038/s41586-018-0617-x
11. Schmidt TSB, Raes J, Bork P. 2018. The human gut microbiome: from association to modulation. Cell 172:1198–1215. https://doi.org/10.1016/j.cell.2018.02.044
12. de Vos WM, Tilg H, Van Hul M, Cani PD. 2022. Gut microbiome and health: mechanistic insights. Gut 71:1020–1032. https://doi.org/10.1136/gutjnl-2021-326789
13. Salazar N, Valdés-Varela L, González S, Gueimonde M, de Los Reyes-Gavilán CG. 2017. Nutrition and the gut microbiome in the elderly. Gut Microbes 8:82–97. https://doi.org/10.1080/19490976.2016.1256525
14. Thursby E, Juge N. 2017. Introduction to the human gut microbiota. Biochem J 474:1823–1836. https://doi.org/10.1042/BCJ20160510
15. Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, Mele MC. 2019. What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. Microorganisms 7:14. https://doi.org/10.3390/microorganisms7010014
16. Shaffer JP, Marotz C, Belda-Ferre P, Martino C, Wandro S, Estaki M, Salido RA, Carpenter CS, Zaramela LS, Minich JJ, Bryant M, Sanders K, Fraraccio S, Ackermann G, Humphrey G, Swafford AD, Miller-Montgomery S, Knight R. 2021. A comparison of DNA/RNA extraction protocols for high-throughput sequencing of microbial communities. Biotechniques 70:149–159. https://doi.org/10.2144/btn-2020-0153
17. Wagner J, Coupland P, Browne HP, Lawley TD, Francis SC, Parkhill J. 2016. Evaluation of PacBio sequencing for full-length bacterial 16S rRNA gene classification. BMC Microbiol 16:274. https://doi.org/10.1186/s12866-016-0891-4
18. Myer PR, Kim M, Freetly HC, Smith TPL. 2016. Evaluation of 16S rRNA amplicon sequencing using two next-generation sequencing technologies for phylogenetic analysis of the rumen bacterial community in steers. J Microbiol Methods 127:132–140. https://doi.org/10.1016/j.mimet.2016.06.004
19. Shin J, Lee S, Go M-J, Lee SY, Kim SC, Lee C-H, Cho B-K. 2016. Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. Sci Rep 6:29681. https://doi.org/10.1038/srep29681
20. Choo JM, Leong LEX, Rogers GB. 2015. Sample storage conditions significantly influence faecal microbiome profiles. Sci Rep 5:16350. https://doi.org/10.1038/srep16350
21. Watson E-J, Giles J, Scherer BL, Blatchford P. 2019. Human faecal collection methods demonstrate a bias in microbiome composition by cell wall structure. Sci Rep 9:16831. https://doi.org/10.1038/s41598-019-53183-5
22. Lim MY, Song E-J, Kim SH, Lee J, Nam Y-D. 2018. Comparison of DNA extraction methods for human gut microbial community profiling. Syst Appl Microbiol 41:151–157. https://doi.org/10.1016/j.syapm.2017.11.008
23. Ezzy AC, Hagstrom AD, George C, Hamlin AS, Pereg L, Murphy AJ, Winter G. 2019. Storage and handling of human faecal samples affect the gut microbiome composition: a feasibility study. J Microbiol Methods 164:105668. https://doi.org/10.1016/j.mimet.2019.105668
24. Yang F, Sun J, Luo H, Ren H, Zhou H, Lin Y, Han M, Chen B, Liao H, Brix S, Li J, Yang H, Kristiansen K, Zhong H. 2020. Assessment of fecal DNA extraction protocols for metagenomic studies. Gigascience 9:giaa071. https://doi.org/10.1093/gigascience/giaa071
25. Chen Z, Hui PC, Hui M, Yeoh YK, Wong PY, Chan MCW, Wong MCS, Ng SC, Chan FKL, Chan PKS. 2019. Impact of preservation method and 16S rRNA

Hypervariable region on gut microbiota profiling. mSystems 4:e00271-18. https://doi.org/10.1128/mSystems.00271-18
26. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41:e1. https://doi.org/10.1093/nar/gks808
27. Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP. 2015. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. Microbiome 3:26. https://doi.org/10.1186/s40168-015-0087-4
28. Clooney AG, Fouhy F, Sleator RD, O' Driscoll A, Stanton C, Cotter PD, Claesson MJ. 2016. Comparing apples and oranges?: Next generation sequencing and its impact on microbiome analysis. PLoS One 11:e0148028. https://doi.org/10.1371/journal.pone.0148028
29. Ye SH, Siddle KJ, Park DJ, Sabeti PC. 2019. Benchmarking metagenomics tools for taxonomic classification. Cell 178:779–794. https://doi.org/10.1016/j.cell.2019.07.010
30. Wang Y, LêCao K-A. 2020. Managing batch effects in microbiome data. Brief Bioinform 21:1954–1970. https://doi.org/10.1093/bib/bbz105
31. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. 2021. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol 39:1348–1365. https://doi.org/10.1038/s41587-021-01108-x
32. Vogtmann E, Chen J, Amir A, Shi J, Abnet CC, Nelson H, Knight R, Chia N, Sinha R. 2017. Comparison of collection methods for fecal samples in microbiome studies. Am J Epidemiol 185:115–123. https://doi.org/10.1093/aje/kww177
33. Natarajan A, Han A, Zlitni S, Brooks EF, Vance SE, Wolfe M, Singh U, Jagannathan P, Pinsky BA, Boehm A, Bhatt AS. 2021. Publisher Correction: standardized preservation, extraction and quantification techniques for detection of fecal SARS-CoV-2 RNA. Nat Commun 12:7100. https://doi.org/10.1038/s41467-021-27392-4
34. Li X, Bosch-Tijhof CJ, Wei X, de Soet JJ, Crielaard W, Loveren C van, Deng DM. 2020. Efficiency of chemical versus mechanical disruption methods of DNA extraction for the identification of oral Gram-positive and Gram-negative bacteria. J Int Med Res 48:300060520925594. https://doi.org/10.1177/0300060520925594
35. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, Tramontano M, Driessen M, Hercog R, Jung F-E, et al. 2017. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol 35:1069–1076. https://doi.org/10.1038/nbt.3960
36. Robe P, Nalin R, Capellano C, Vogel TM, Simonet P. 2003. Extraction of DNA from soil. Eur J Soil Biol 39:183–190. https://doi.org/10.1016/S1164-5563(03)00033-5
37. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 12:87. https://doi.org/10.1186/s12915-014-0087-z
38. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S, Abnet CC, Knight R, White O, Huttenhower C, Microbiome Quality Control Project Consortium. 2017. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol 35:1077–1086. https://doi.org/10.1038/nbt.3981
39. Han D, Gao P, Li R, Tan P, Xie J, Zhang R, Li J. 2020. Multicenter assessment of microbial community profiling using 16S rRNA gene sequencing and shotgun metagenomic sequencing. J Adv Res 26:111–121. https://doi.org/10.1016/j.jare.2020.07.010
40. Tourlousse DM, Narita K, Miura T, Sakamoto M, Ohashi A, Shiina K, Matsuda M, Miura D, Shimamura M, Ohyama Y, et al. 2021. Validation and standardization of DNA extraction and library construction methods for metagenomics-based human fecal microbiome measurements. Microbiome 9:95. https://doi.org/10.1186/s40168-021-01048-3
41. Stulberg E, Fravel D, Proctor LM, Murray DM, LoTempio J, Chrisey L, Garland J, Goodwin K, Graber J, Harris MC, Jackson S, Mishkind M, Porterfield DM, Records A. 2016. An assessment of US microbiome research. Nat Microbiol 1:15015. https://doi.org/10.1038/nmicrobiol.2015.15
42. Wu W-K, Chen C-C, Panyod S, Chen R-A, Wu M-S, Sheen L-Y, Chang S-C. 2019. Optimization of fecal sample processing for microbiome study — The journey from bathroom to bench. J Formos Med Assoc 118:545–555. https://doi.org/10.1016/j.jfma.2018.02.005
43. Greathouse KL, Sinha R, Vogtmann E. 2019. DNA extraction for human microbiome studies: the issue of standardization. Genome Biol 20:212. https://doi.org/10.1186/s13059-019-1843-8
44. Site search. Available from: https://emea.illumina.com/search.html?q=protocol&filter=all&p=1. Retrieved 06 Oct 2022.
45. Rintala A, Pietilä S, Munukka E, Eerola E, Pursiheimo J-P, Laiho A, Pekkala S, Huovinen P. 2017. 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 28:19–30. https://doi.org/10.7171/jbt.17-2801-003
46. Willis AD. 2019. Alpha diversity, and statistics. Front Microbiol 10:2407. https://doi.org/10.3389/fmicb.2019.02407
47. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8
48. Zhang B, Brock M, Arana C, Dende C, van Oers NS, Hooper LV, Raj P. 2021. Impact of bead-beating intensity on the genus- and species-level characterization of the gut microbiome using amplicon and complete 16S rRNA gene sequencing. Front Cell Infect Microbiol 11:678522. https://doi.org/10.3389/fcimb.2021.678522
49. de Goffau MC, Jallow AT, Sanyang C, Prentice AM, Meagher N, Price DJ, Revill PA, Parkhill J, Pereira DIA, Wagner J. 2022. Gut microbiomes from Gambian infants reveal the development of a non-industrialized Prevotella-based trophic network. Nat Microbiol 7:132–144. https://doi.org/10.1038/s41564-021-01023-6
50. Williams GM, Leary SD, Ajami NJ, Chipper Keating S, Petrosin JF, Hamilton-Shield JP, Gillespie KM. 2019. Gut microbiome analysis by post: evaluation of the optimal method to collect stool samples from infants within a national cohort study. PLoS One 14:e0216557. https://doi.org/10.1371/journal.pone.0216557
51. Keskitalo A, Munukka E, Aatsinki A, Saleem W, Kartiosuo N, Lahti L, Huovinen P, Elo LL, Pietilä S, Rovio SP, Niinikoski H, Viikari J, Rönnemaa T, Lagström H, Jula A, Raitakari O, Pahkala K. 2022. An infancy-onset 20-year dietary counselling intervention and gut microbiota composition in adulthood. Nutrients 14:2667. https://doi.org/10.3390/nu14132667
52. Johnson JS, Spakowicz DJ, Hong B-Y, Petersen LM, Demkowicz P, Chen L, Leopold SR, Hanson BM, Agresta HO, Gerstein M, Sodergren E, Weinstock GM. 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun 10:5029. https://doi.org/10.1038/s41467-019-13036-1
53. Fuks G, Elgart M, Amir A, Zeisel A, Turnbaugh PJ, Soen Y, Shental N. 2018. Combining 16S rRNA gene variable regions enables high-resolution microbial community profiling. Microbiome 6:17. https://doi.org/10.1186/s40168-017-0396-x
54. Soriano-Lerma A, Pérez-Carrasco V, Sánchez-Marañón M, Ortiz-González M, Sánchez-Martín V, Gijón J, Navarro-Mari JM, García-Salcedo JA, Soriano M. 2020. Influence of 16S rRNA target region on the outcome of microbiome studies in soil and saliva samples. Sci Rep 10:13637. https://doi.org/10.1038/s41598-020-70141-8
55. ZymoBIOMICS Gut Microbiome Standard. Zymo research. Available from: https://www.zymoresearch.com/products/zymobiomics-gut-microbiome-standard. Retrieved 05 Oct 2022.
56. Karu N, Deng L, Slae M, Guo AC, Sajed T, Huynh H, Wine E, Wishart DS. 2018. A review on human fecal metabolomics: methods, applications and the human fecal metabolome database. Anal Chim Acta 1030:1–24. https://doi.org/10.1016/j.aca.2018.05.031

Heidi Isokääntä, Lucas Pinto da Silva, Naama Karu, Teemu Kallonen, Anna-Katariina Aatsinki, Thomas Hankemeier, Leyla Schimmel, Edgar Diaz, Tuulia Hyötyläinen, Pieter C. Dorrestein, Rob Knight, Matej Orešič, Rima Kaddurah-Daouk, Alex M. Dickens, Santosh Lamichhane (2024). Comparative metabolomics and microbiome analysis of Ethanol vs. OMNImet/gene®•GUT faecal stabilization.
Analytical Chemistry

II

analytical chemistry

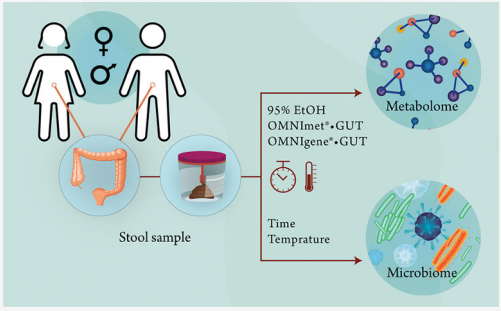
pubs.acs.org/ac Article

Comparative Metabolomics and Microbiome Analysis of Ethanol versus OMNImet/gene•GUT Fecal Stabilization

Heidi Isokääntä, Lucas Pinto da Silva, Naama Karu, Teemu Kallonen, Anna-Katariina Aatsinki, Thomas Hankemeier, Leyla Schimmel, Edgar Diaz, Tuulia Hyötyläinen, Pieter C. Dorrestein, Rob Knight, Matej Orešič, Rima Kaddurah-Daouk,* Alex M. Dickens,* and Santosh Lamichhane*

Cite This: *Anal. Chem.* 2024, 96, 8893−8904

ABSTRACT: Metabolites from feces provide important insights into the functionality of the gut microbiome. As immediate freezing is not always feasible in gut microbiome studies, there is a need for sampling protocols that provide the stability of the fecal metabolome and microbiome at room temperature (RT). Here, we investigated the stability of various metabolites and the microbiome (16S rRNA) in feces collected in 95% ethanol (EtOH) and commercially available sample collection kits with specific preservatives OMNImet•GUT/OMNIgene•GUT. To simulate field-collection scenarios, the samples were stored at different temperatures at varying durations (24 h + 4 °C, 24 h RT, 36 h RT, 48 h RT, and 7 days RT) and compared to aliquots immediately frozen at −80 °C. We applied several targeted and untargeted metabolomics platforms to measure lipids, polar metabolites, endocannabinoids, short-chain fatty acids (SCFAs), and bile acids (BAs). We found that SCFAs in the nonstabilized samples increased over time, while a stable profile was recorded in sample aliquots stored in 95% EtOH and OMNImet•GUT. When comparing the metabolite levels between aliquots stored at room temperature and at +4 °C, we detected several changes in microbial metabolites, including multiple BAs and SCFAs. Taken together, we found that storing samples at RT and stabilizing them in 95% EtOH yielded metabolomic results comparable to those from flash freezing. We also found that the overall composition of the microbiome did not vary significantly between different storage types. However, notable differences were observed in the α diversity. Altogether, the stability of the metabolome and microbiome in 95% EtOH provided results similar to those of the validated commercial collection kits OMNImet•GUT and OMNIgene•GUT, respectively.



■ THEORETICAL BACKGROUND

The gut microbiome is considered an "essential organ" that contributes to the regulation of host development and physiology and facilitates host metabolism[1,2] and is often linked to various human health conditions,[3,4] including inflammatory bowel disease,[5] obesity,[6] and multiple neurological disorders.[7,8] The interaction and dynamics between the host and gut microbiome are mediated by metabolites, which serve as vital signaling molecules.[9] Integrated microbiome and metabolome analyses have emerged as the foremost promising approach to unveil host−microbiota interactions in the context of disease risk.[10]

Over the past decade, fecal metabolomics has received increasing attention as fecal metabolites offer important insights into the functional aspects of the gut microbiome. The molecules associated with the gut microbiome also have the potential to be used in therapeutic strategies and biomarker discovery.[11] Despite growing interest, among the limiting factors in this field are practical challenges involved in the collection of human fecal samples. Albeit logistically impractical, the immediate homogenization and freezing of fecal samples at −80 °C is considered the gold standard for metabolite preservation because it halts enzymatic activity, hydrolysis, oxidation, and other degradation processes.[12] Recently, studies of large human cohorts increased the demand for home collection of human fecal specimens, with the aim of reducing cost as well as improving practicality, donor privacy, and convenience. Although home collection is a convenient option, it includes multiple steps that can involve temperature fluctuations. Moreover, storing and shipping frozen fecal samples can be inconvenient for participants and prohibitively expensive for researchers. This emphasizes the need for sampling systems that can be stored at room temperature.[13]

Received: October 2, 2023
Revised: April 12, 2024
Accepted: April 18, 2024
Published: May 23, 2024

https://doi.org/10.1021/acs.analchem.3c04436
Anal. Chem. 2024, 96, 8893−8904

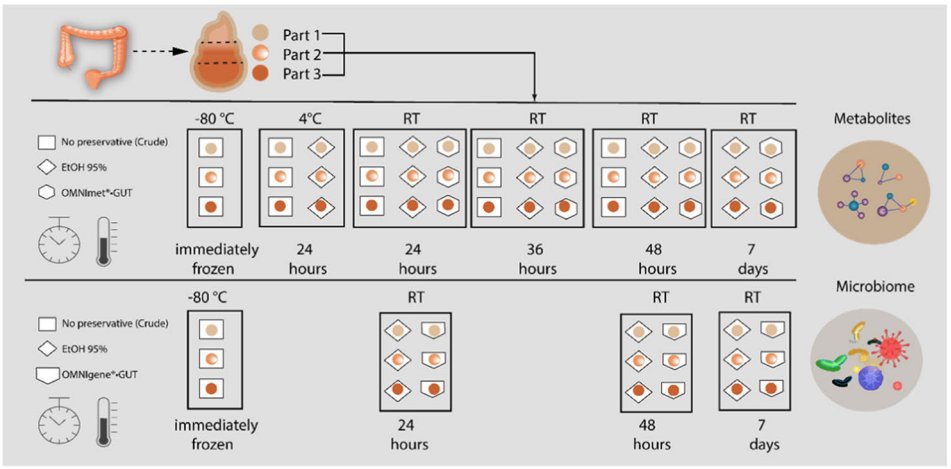


Figure 1. An overview of the study design is presented, illustrating the feces samples collected for metabolite measurement as well as the number of matched feces samples for 16S rRNA gene sequencing at each time point. In this study, one set of samples was frozen immediately (−80 °C) while other aliquots were either stored as crude feces (without preservatives), in 95% ethanol or in the OMNIgene•GUT/OMNImet•GUT at room temperature (RT) for 24 h, 48 h, and 7 days.

To address this need specifically for metabolomics analysis, available sampling devices such as DNA stabilization tubes and fecal immunochemical test tubes were critically assessed.[14,15] These studies found that the numerous detergents, buffers, salts, and other additives in the examined collection tubes deemed them inferior or unsuitable for analysis by liquid chromatography (LC) and mass spectrometry (MS). In addition, some of these collection kits can significantly distort the metabolic profile of fecal samples compared to the gold standard of flash freezing.[14,15] A few studies tested 95% ethanol (EtOH) as a fecal sample preservative and found it suitable for metabolomics.[10,15] Ethanol prevents microbial growth and stabilizes the microbiome until profiling while partly stabilizing the metabolome, as it prevents enzymatic metabolism by the fecal microbiota and affects chemical degradation processes.[16] Another stability contributor effect of 95% EtOH is its nonfrozen state at −80 °C; hence, no freeze−thaw cycles occur and disrupt the sample profile. Fecal collection and storage tubes containing 95% EtOH (OMNImet•GUT, DNA Genotek, Canada) have been introduced as a kit tailored specifically for metabolomics analysis. A corresponding kit for the microbiome (OMNIgene•GUT) has been used in several studies.[17,18] To our knowledge, only a few studies compared feces collection methods for simultaneous gut microbiota profiling and fecal metabolomics.[10] Notably, studies comparing collection with that of 95% EtOH and commercial fecal collection kits are lacking.

Here, we aim to fill the current knowledge gaps and establish the validity of an off-site fecal collection method that preserves the integrity of both the metabolome and the microbiome during the freight and until processing. For this goal, 95% ethanol was selected as a preservative and compared with flash freezing (crude feces without solvent), OMNImet•GUT, and OMNIgene•GUT, which are validated commercial collection kits for fecal metabolome and microbiome profiling, respectively. We also tested the effect of different storage times and temperatures on metabolite coverage with EtOH-containing matrices in terms of their sensitivity, robustness, and throughput.

EXPERIMENTAL SECTION

Study Design. Stool samples were collected from four healthy human volunteers (n = 4) for metabolomics and microbiome profiling (Figure 1); informed consent was obtained from all donors. Fecal samples were collected from the four subjects in the morning next to the operating laboratory. After defecation, the sample was divided into three parts, which were homogenized with a spatula. These represent biological replicates, and they were further divided into aliquots.

For the metabolomics analyses, the aliquot tubes were spiked with stability standards (sodium butyrate-13C4 5 ppm, cholic acid-24−13C 2,5 ppb, palmitic acid (1−13C, 99%) 5 ppm, hippuric acid-d5 5 ppm, indole-2,4,5,6,7-d5−3-acetic acid 5 ppm, nicotinamide-d4 5 ppm, sucrose-1−13Cfru 5 ppm, L-tyrosine-d4 5 ppm, and cortisol-1,2-d2 5 ppm) and dried using a SpeedVac (Thermo Fisher Scientific) beforehand. For the aliquots, 150 ± 10 mg of stool was weighed to each prepared tube. Next, storage fluid was added at a ratio of 1:4 (600 μL of OMNImet•GUT solvent or EtOH 95%) to the tubes. We added 100 μL of ultrapure water to make homogenization easier for samples with no preservative. All aliquots were homogenized with a bullet homogenizer (Next Advance) at a speed of 6 for 2 min. During sample processing, fast freezing of one aliquot per biological replicate was prioritized. Those aliquots were frozen within 15−20 min from defecation, and hereafter, this refers to the golden standard.

After the initial sample processing, the aliquots with no preservative (the crude sample), with OMNImet•GUT solvent, and with EtOH 95% were stored at room temperature (RT) for 24, 36, 48 h, and 7 days. Additionally, the aliquots with EtOH 95% were kept at +4 °C for 24, 36, 48, and 7 days. The crude samples were kept at +4 for 24 h.

For the microbiome profiling, the sample material left from the spatula-mixed biological replicates were divided into aliquots of three sample types: feces with no preservative (crude feces), feces in EtOH 95%, and feces with OMNIgene.

https://doi.org/10.1021/acs.analchem.3c04436
Anal. Chem. 2024, 96, 8893−8904

GUT kit was in the same ratio as above. The aliquots of crude feces were frozen at −80 °C within 2 h (kept on ice) from defecation. The aliquots with EtOH 95% and OMNIgene.-GUT were stored at RT for 2 h, 48 h, and 7 days. After the incubation time, these aliquots were stored at −80 °C until sample preparation for DNA extraction.

Metabolite Extractions. Here, three distinct extraction procedures were performed. Prior to extraction, we balanced the sample weights with liquid volumes by adding 100 μL of water to EtOH/Omni-tubes and 600 μL of EtOH to the crude samples and those labeled as "immediately frozen samples."

First, we combined extraction for SCFA, bile acids (BAs), and untargeted metabolites assay. Glycoursodeoxycholic acid (GUDCA) -d4, glycocholic acid (GCA)-d4, cholic acid (CA) -d4, ursodeoxycholic acid (UDCA) -d4, glycochenodeoxycholic acid (GCDCA) -d4, chenodeoxycholic acid (CDCA) -d4, deoxycholic acid (DCA) -d4, glycolithocholic acid (GLCA) -d4, heptadecanoic acid, deuterium-labeled valine, deterium-labeled succinic acid, deterium-labeled glutamic acid, mass-labeled PFCAs and PFASs solution/mixture (MPFAC-MXA), and d4-androsterone were added as internal standards. After vortexing, samples were filtered through 96-well protein precipitation plates (Supelco/Sigma-Aldrich) with vacuum. Filtrates were divided into 3 parts (20 μL for BAs, 20 μL for untargeted metabolites, and 50 μL for SCFA). Vials for BA analysis were dried under nitrogen, resuspended in 20 μL of methanol in water (4:6 v/v), and stored at −80 °C until analysis. Vials for untargeted metabolites were dried and stored at −80C before analysis. Filtrates for SCFA were stored in Waters deep well plates at −80 °C.

The second extraction was for endocannabinoids (ECCs), and it included 200 μL of fecal slurry and 400ul of crash solvent (95% Ethanol). To avoid contact with plastic, glass vials and syringes were utilized during extraction. Samples were vortexed and incubated for 30 min at −20 °C and then filtered through the protein precipitation plate. The filtrates were evaporated under a gentle steam of nitrogen at +35 °C and reconstituted with the final solvent 50 μL/sample (40% water, 30% ACN, and 30% IPA) before the run. This crash solvent included the following internal standards: THC−COOH-d9, 2-AG-d5, NADA-d8, AEA-d8, and AA-d8.

The third extraction was for lipids done using liquid−liquid extraction, a method based on the Folch procedure, as described previously.[19] Here, 10 μL of 0.9% NaCl and 120 μL of crash solvent CHCl3:MeOH (2:1, v/v) containing internal standards solution were added to 10 μL of fecal sample homogenate. The crash solved included the following internal standards: PE (17:0/17:0), SM(d18:1/17:0), Cer (d18:1/17:0), PC (17:0/17:0), LPC (17:0), PC (16:0/d31/18:1), and TG (17:0/17:0/17:0). After vortexing, samples were allowed to stand on ice for 30 min. Then, the samples were centrifuged (9400*g*, 5 min, 4 °C). Next, 60 μL of the lower layer was collected and transferred to an LC vial with an insert, and 60 μL of CHCl3: EtOH (2:1, v/v) was added. The extracts were stored at −80 °C until analysis.

Metabolite Analysis. *Targeted Analysis.* Targeted methods were used for SCFA, BAs, and ECCs.

Bile Acid Analysis. For BA analysis, the chromatographic separation was carried out using an Acquity Premiere HSS T3 column (100 mm × 2.1 mm i.d., 1.8 μm particle size), fitted with a C18 precolumn (Acquity UPLC HSS T3 1.8 μm, 2.1 mm × 5 mm, Waters Corporation, Wexford, Ireland). Mobile phase A consisted of water:methanol (v/v 70:30), and mobile phase B consisted of methanol, with both phases containing 2 mM ammonium acetate as an ionization agent. The flow rate was set at 0.4 mL/min, with the elution gradient as follows: 0−1.5 min, mobile phase B was increased from 5 to 30%; 1.5−4.5 min, mobile phase B was increased to 70%; and 4.5−7.5 min, mobile phase B was increased to 100% and held for 5.5 min. A post-time of 5 min was used to regain the initial conditions for the next analysis. The total run time per sample was 18 min. The injection volume used was 5 μL. The analyses were performed in negative ion mode, and Analyst v. 1.7.3 (AB SCIEX) was used for all data acquisition. For other details of the method, see the Supporting Information method M1.

SCFA Analysis. For SCFAs, the sample aliquot was derivatized by adding 50 μL of 50 mM 3-nitrophenylhydrazine (3-NPH) in 3:7 H2O: MeOH, followed by addition of 50 μL of 50 mM ethylene dichloride (EDC) in 3:7 H2O: MeOH, 50 μL of pyridine (7% v/v in 3:7 H2O: MeOH). The mixture was then incubated for 60 min at room temperature, after which 100 μL of formic acid (0.2% in 3:7 H2O: MeOH) was added to the mixture to quench the reaction. The analysis of derivatized SCFAs was carried out on an Acquity UPLC BEH C18 column (2.1 mm × 100 mm, 1.7 μm; Waters, Milford) using as mobile phase (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile. Samples were eluted at 0.5 mL min-1, starting with 10% B and increasing to 100% B in 10 min, then holding at 100% B for 2.1 min, returning to 10% B and holding for 2 min. Column temperature was maintained at 50 °C, while the autosampler was maintained at 10 °C during analysis. The injection volume was 5 μL. The analyses were performed in negative ion mode, and Analyst v. 1.7.3 (AB SCIEX) was used for all data acquisition. For other details of the method, see Supporting Information method M2.

Endocannabinoid Analysis. The ECCs analysis was carried out on an XBridge BEH C18 column (2.1 mm × 150 mm, 2.5 μm; Waters, Milford) using as mobile phase (A) 1 mM ammonium acetate and 0.1% formic acid in water and (B) 1 mM ammonium acetate and 0.1% formic acid in ACN: IPA (1:1). Samples were eluted at 0.4 mL/min, starting with 60% B and holding for 0.3 min, then increasing to 100% for 5 min, holding at 100% B for 2.5 min, returning to 10% in 0.1 min, and holding for 3 min. Column temperature was maintained at 40 °C, while the autosampler was maintained at 15 °C during analysis. The injection volume was 10 μL. The analyses were performed in positive and negative ion mode, and SCIEX OS v3.0 (AB SCIEX) was used for all data acquisition. For other details of the method, see Supporting Information method M3 and related publications.[19]

SCFAs and BAs were analyzed with an Exion LC system coupled to a QTrap 5500 MS interfaced with a Turbo V electrospray ion source (SCIEX, Framingham, MA). The ECCs were analyzed with an Exion LC system coupled with a QTrap 7500 MS interfaced with an OptiFlow Pro electrospray ion source (SCIEX, Framingham, MA). The lipids were analyzed with an Exion LC system coupled with TripleTOF 6600 MS interfaced with a DuoSpray electrospray ion source (SCIEX, Framingham, MA).

Untargeted Analysis. Lipids and polar metabolites were analyzed using a combination of untargeted and semitargeted assays. Lipids were quantified using class-based internal standards and authentic standards from each class. Polar metabolites were quantified using a set of internal standards, as described below.

https://doi.org/10.1021/acs.analchem.3c04436
Anal. Chem. 2024, 96, 8893−8904

Lipidomic Analysis. The lipidomic analysis was carried out on an ACQUITY UPLC BEH C18 column (2.1 mm × 100 mm, particle size of 1.7 μm) by Waters (Milford). The eluent system consisted of (A) 10 mM ammonium acetate in H2O and 0.1% formic acid and (B) 10 mM ammonium acetate in ACN: IPA (1:1) and 0.1% formic acid. The gradient was as follows: an increase from 35% B to 80% B over 2 min and then an increase to 100% B over 5 min. The gradient was held at 100% B for 7 min, followed by a re-equilibration at 35% over 7 min. The injection volume was 1 μL. The column compartment and autosampler temperatures were 40 and 10 °C, respectively. All analyses were performed in positive ion mode, and Analyst v1.8.1 (AB SCIEX) was used for all data acquisition. The RSD% for QC samples (n = 8) was, on average, 18.95%.

Data was processed using MZmine 2.[20] Mass spectrometry data processing was performed using the open source software package MZmine 2.53.[20] The following steps were applied in this processing: (i) mass detection with a noise level of 1000, (ii) chromatogram builder with a minimum time span of 0.08 min, minimum height of 1000 and a m/z tolerance of 0.006 m/z or 10.0 ppm, (iii) chromatogram deconvolution using the local minimum search algorithm with a 70% chromatographic threshold, 0.05 min minimum RT range, 5% minimum relative height, 1200 minimum absolute height, a minimum ration of peak top/edge of 1.2, and a peak duration range of 0.08−5.0, (iv), isotopic peak grouper with a m/z tolerance of 5.0 ppm, RT tolerance of 0.05 min, maximum charge of 2 and with the most intense isotope set as the representative isotope, (v) Join aligner with a m/z tolerance of 0.009 or 10.0 ppm and a weight for of 2, a RT tolerance of 0.15 min and a weight of 1 and with no requirement of charge state or ID and no comparison of isotope pattern, (vi) peak list row filter with a minimum of 10% of the samples, (vii) gap filling using the same RT and m/z range gap filler algorithm with an m/z tolerance of 0.009 m/z or 11.0 ppm, and (vii) identification of lipids using a custom database search with an m/z tolerance of 0.008 m/z or 8.0 ppm and a RT tolerance of 0.25 min. Identification of lipids was based on in-house data on the LC-MS/MS retention time and mass spectra. The identification was done with a custom database, with identification levels 1 and 2, i.e., based on authentic standard compounds (level 1) and based on MS/MS identification (level 2) based on Metabolomics Standards Initiative.

Quantification of lipids was performed using a 7-point internal calibration curve (0.1−5 μg/mL) using the following lipid-class specific authentic standards: using 1-hexadecyl-2-(9Z-octadecenoyl)-*sn*-glycero-3-phosphocholine (PC(16:0*e*/18:1(9Z))), 1-(1Z-octadecenyl)-2-(9Z-octadecenoyl)-*sn*-glycero-3-phosphocholine (PC(18:0*p*/18:1(9Z))), 1-stearoyl-2-hydroxy-*sn*-glycero-3-phosphocholine (LPC(18:0)), 1-oleoyl-2-hydroxy-*sn*-glycero-3-phosphocholine (LPC(18:1)), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (PE(16:0/18:1)), 1-(1Z-octadecenyl)-2-docosahexaenoyl-*sn*-glycero-3-phosphocholine (PC(18:0*p*/22:6)) and 1-stearoyl-2-linoleoyl-*sn*-glycerol (DG(18:0/18:2)), 1-(9Z-octadecenoyl)-*sn*-glycero-3-phosphoethanolamine (LPE(18:1)), N-(9Z-octadecenoyl)-sphinganine (Cer(d18:0/18:1(9Z))), 1-hexadecyl-2-(9Z-octadecenoyl)-*sn*-glycero-3-phosphoethanolamine (PE(16:0/18:1)) from Avanti Polar Lipids, 1-Palmitoyl-2-Hydroxy-*sn*-Glycero-3-Phosphatidylcholine (LPC(16:0)), 1,2,3 trihexadecanoalglycerol (TG(16:0/16:0/16:0)), 1,2,3-trioctadecanoylglycerol (TG(18:0/18:0/18:)) and 3β-hydroxy-5-cholestene-3-stearate (ChoE(18:0)), 3β-Hydroxy-5-cholestene-3-linoleate (ChoE(18:2)) from Larodan, were prepared to the following concentration levels: 100, 500, 1000, 1500, 2000, and 2500 ng/mL (in CHCl3:MeOH, 2:1, v/v), including 1250 ng/mL of each internal standard.

Polar Metabolite Analysis. For polar metabolites, aliquots of 10 μL of samples were injected into the Acquity UPLC BEH C18 2.1 mm × 100 mm, 1.7-μm column (Waters Corporation), fitted with a C18 precolumn (Waters Corporation, Wexford, Ireland). The mobile phases consisted of (A) 2 mM NH_4Ac in H_2O: MeOH (7:3) and (B) 2 mM NH_4Ac in MeOH. The flow rate was set at 0.4 mL min-1 with the elution gradient as follows: 0−1.5 min, mobile phase B was increased from 5 to 30%; 1.5−4.5 min, mobile phase B was increased to 70%; 4.5−7.5 min, mobile phase B was increased to 100% and held for 5.5 min. A post-time of 5 min was used to regain the initial conditions for the next analysis. The total run time per sample was 20 min. The dual ESI ionization source settings were as follows: capillary voltage was 4.5 kV, nozzle voltage was 1500 V, N2 pressure in the nebulized was 21 psi, and the N2 flow rate and temperature as sheath gas were 11 Lmin-1 and 379 °C, respectively. In order to obtain accurate mass spectra in the MS scan, the m/z range was set to 100−1700 in negative ion mode. MassHunter B.06.01 software (Agilent Technologies, Santa Clara, CA) was used for all data acquisition. The RSD% for QC samples (n = 8) was, on average, 18.76%.

Quantitation was done using 6-point calibration (bile acids, ca. 20−640 ng/mL; polar metabolites, ca. 0.1 to 80 μg/mL). Quantification was performed using the following compounds: chenodeoxycholic acid (CDCA), cholic acid (CA), deoxycholic acid (DCA), glycochenodeoxycholic acid (GCDCA), glycocholic acid (GCA), glycodehydrocholic acid (GDCA), glycodeoxycholic acid (GDCA), glycohyocholic acid (GHCA), glycohyodeoxycholic acid (GHDCA), glycolitocholic acid (GLCA), glycoursodeoxycholic acid (GUDCA), hyocholic acid (HCA), hyodeoxycholic acid (HDCA), litocholic acid (LCA), α-Muricholic acid (αMCA), tauro-α-muricholic acid (T-α-MCA), tauro-β-muricholic acid(T-β-MCA), taurochenodeoxycholic acid (TCDCA), taurocholic acid (TCA), taurodehydrocholic acid (THCA), taurodeoxycholic acid (TDCA), taurohyodeoxycholic acid (THDCA), taurolitocholic acid (TLCA), tauro-omega-muricholic acid (TωMCA) and tauroursodeoxycholic acid (TDCA), alanine, citric acid, fumaric acid, glutamic acid, glycine, lactic acid, malic acid, 2-hydroxybutyric acid, 3-hydroxybutyric acid, linoleic acid, oleic acid, palmitic acid, stearic acid, cholesterol, fructose, glutamine, indole-3-propionic acid, isoleucine, leucine, proline, succinic acid, valine, asparagine, aspartic acid, arachidonic acid, glycerol-3-phosphate, lysine, methionine, ornithine, phenylalanine, serine, and threonine.

Quality control was accomplished both for lipidomics and for polar metabolites by including blanks, pure standard samples, extracted standard samples, pooled quality control samples, and control plasma samples. The pooled samples were prepared by taking an aliquot (10 μL) of each extract separately for lipidomic and polar metabolite methods, and pooling them and aliquoting the pool into separate vials. In lipidomic and metabolomic analyses, lipids and metabolites that had >30% RSD in the pooled QC samples (an equal aliquot of each sample pooled together) or that were present at high concentrations in the extracted blank samples (ratio

https://doi.org/10.1021/acs.analchem.3c04436
Anal. Chem. 2024, 96, 8893−8904

between samples to blanks <5) were excluded from the data analyses.

Gut Microbiome Analysis. For microbiome profiling, the same fecal samples as above (n = 4) were tested (including 3 biological replicates) with three sample collection types: immediately frozen crude feces, feces in EtOH 95%, and OMNIgene. GUT kit in ratio 1:4 (150 mg +600 μL); samples in 95% EtOH and OMNIgene. GUT were kept at RT for 24, 48, and 7 days and then stored at −80 °C until DNA extraction. The sample volume was 50 mg per DNA extraction, either as crude or dissolved in a preservative. Microbial DNA was extracted using a DNA Stool 200 Kit special H96 (PerkinElmer, Turku, Finland) kit with a corresponding Chemagic with Magnetic Separation Module I (MSM I) extraction robot after bead-beating and proteinase K incubation. Microbial composition was determined by sequencing the V3 V4 region of the 16S ribosomal gene using a MiSeq platform (Illumina). The sequence library was constructed according to the Illumina library preparation protocol, with minor differences from the V3 V4 protocol. After PCR, 8 μL of the integrity of the product was analyzed with 1.5% TAE agarose gel (120 V, 1 h). The concentration of the library samples was measured with a Qubit Fluorometer by using a Qubit dsDNA High Sensitivity Assay kit. The 4 nM library pool was denatured and diluted to a concentration of 4pM, and an 8% denaturalized PhiX control (Illumina) was added. The library samples were sequenced with a MiSeq Reagent kit v3, 600 cycles (Illumina) on a Miseq system with 2× 300 base pair (bp) paired ends following the manufacturer's instructions. A positive control, Zymobiomics Microbial community DNA standard (Zymo Research), and a negative control (PCR-grade water) were included in library preparation to control the PCR. Lysis buffer was used as the negative control in DNA extraction to control contamination.

Statistical Analysis. For metabolomics data, all multivariate statistical analyses were based on log2-transformed intensity data. The transformed data were auto-scaled prior to multivariate analysis to improve the global interpretability. To account for the inconsistency in the fecal water content, the measured metabolites were normalized to the dry weight in the stool. The multivariate analysis was done using the PLS Toolbox 8.2.1 (eigenvector Research Inc., Manson, WA) in MATLAB 2017b (Mathworks, Inc., Natick, MA). ANOVA-simultaneous component analysis (ASCA), a multivariate extension of ANOVA analysis, was performed to allow interpretation of the variation induced by the different factors, including time, individual, and collection matrix. Subsequently, for univariate analysis, the level of each metabolite in each sample storage matrix (i.e., crude feces without any solvent, feces in 95% EtOH, and feces in OMNImet•GUT) was divided by the level of the same metabolite species in the paired immediately frozen sample (golden standard sample). For instance, the concentration of butyrate in the 95% EtOH was divided by the concentration of butyrate in the immediately frozen sample (golden standard). The fold difference was calculated by dividing the mean concentration of a given metabolite species in one group by another. The difference in the metabolites between the groups was tested using a multivariate linear model using the MaAsLin2 package in R, taking into account random effects within an individual sample or subject. The resulting nominal p-values were corrected for multiple comparisons using the Benjamini and Hochberg approach. The adjusted p-values <0.25 (q-values) were considered significantly different among the group of hypotheses tested.

Microbiome data analyses were performed in R Bioconductor ecosystem (R version 4.2.3) and with a CLC Microbial Genomics Module (CLC Genomics Workbench 21.0.3, Qiagen), which complies with QIIME2.[21] Differences in β diversity between methodological settings were evaluated with distance-based redundancy analysis and PERMANOVA. We used the Bray−Curtis dissimilarity. Moreover, β diversity by Jaccard, UniFrac, and Unweighted UniFrac were plotted. We calculated the Shannon index, number of observed OTUs and Chao1, and the Simpson Index. All of the indices were used in the ICC testing. Differences in the Shannon index were assessed with two sample t tests assuming equal variance (Levene's test).

■ RESULTS AND DISCUSSION

Fecal Metabolome and Lipidome Profiles. Fecal metabolites and lipids were analyzed from a total of 168 fecal samples (aliquoted from four individuals) simulating different conditions of the sample storage matrix: (i) crude feces without any solvent, (ii) feces in 95% EtOH, and (iii) feces in OMNImet•GUT solvent. For each sampling condition, we obtained samples from three different parts of the bulk fecal specimen (Figure 1, Part 1−3). To investigate the effects of storage time and temperature, initially, we obtained one aliquot of the homogenized fecal sample and froze it immediately at −80 °C, which was used as a reference (golden standard) sample (Figure 1). Other corresponding aliquots were tested for varying durations at different temperatures: 24 h at +4 °C (except OMNImet•GUT), 24 h at room temperature (RT), 36 h at RT, 48 h at RT, 48 h at +4 °C (except OMNImet•GUT), and 7 days at RT, as shown in Figure 1.

The metabolomics analysis included targeted short-chain fatty acids (SCFA, n = 7) and bile acids (BAs, n = 33), encompassing both primary (glycine/taurine conjugates) and secondary BAs, as well as endocannabinoids (ECCs, n = 9). These ECCs included palmitoylethanolamide (PEA), arachidonoylglycerol (AG), 2-arachidonoylglycerol ether (2-AGE), arachidonoylethanolamide (AEA), oleoylethanolamide (OEA), stearoylethanolamide (SEA), docosahexaenoylethanolamide (DEA), α-linolenoyl ethanolamide (aLEA), and arachidonic acid (AA). Lipidomics analysis provided coverage of the following lipid classes: acylcarnitines (AC), cholesterol esters (CE), ceramides (Cer), diacylglycerols (DG), lysophosphatidylcholines (LPC), phosphatidylcholines (PC), sphingomyelins (SM), and triacylglycerols (TG). Untargeted polar metabolomics analysis provided coverage of the following classes: amino acids, bile acids, carboxylic acids (mainly free fatty acids and other organic acids), hydroxy acids, phenolic compounds, alcohols, and sugar derivatives.

Fecal Metabolic Profile. *Multifactorial Analysis.* To understand the contributions of different sampling factors to the fecal metabolome, we performed analysis of variance (ANOVA)-simultaneous component analysis (ASCA) with the following factors: subjects; sample storage matrix (crude, 95% EtOH, OMNImet•GUT, immediately frozen); time (24, 36, 48 h, 7 days); and temperature (RT, + 4 °C). ASCA is a multivariate extension of ANOVA analysis that allows interpretation of the variation induced by the different factors, including individual, sample storage, and time. We found that interindividual differences (Model Effect (%) 71.15, 65.10,

https://doi.org/10.1021/acs.analchem.3c04436
Anal. Chem. 2024, 96, 8893−8904

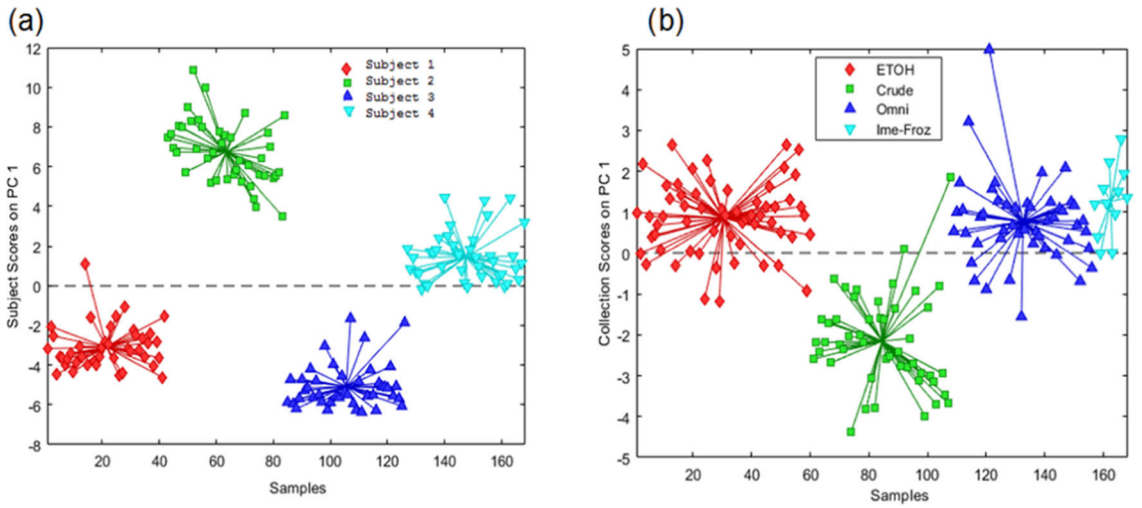


Figure 2. Principal component analysis (PCA) score plots based on ANOVA-simultaneous component analysis (ASCA). (a) Principal component (PC1) score plot obtained based on interindividual score in ASCA analysis. This figure represents the bile acid data set arranged according to interindividual samples in the PCA score plot. Here, each sample is represented by a point and colored according to the individual (red diamond: Subject1, green square: Subject 2, blue triangle up: Subject 3, cyan triangle down: Subject4). (b) Principal component (PC1) score plot obtained based on sample storage matrix score in ASCA analysis. This figure represents the bile acid profile arranged according to the sample storage matrix in the PCA score plot. Here, each sample is represented by a point and colored according to the sample storage matrix (red diamond: 95% EtOH, green square: crude, blue triangle up: OMNImet•GUT, cyan triangle down: immediately frozen).

77.70, p = 0.0010) and subsequent sample storage matrix (Model Effect (%) 8.2, 9.6, 2.5, p = 0.0010) had the strongest effect on BAs, SCFAs, and ECCs fecal profiles. In comparison, the ASCA could not find significant effects of the duration of storage (Model Effect (%) 2.1, 2.0, 1.7, $p > 0.05$) and storage temperature (Model Effect (%) 1.60, 1.09, 1.34, $p > 0.05$). Figure 2 illustrates the clustering differences in the subject (Figure 2a) and sample storage matrix (Figure 2b) explained by the levels of fecal BAs. Similar trends were observed for SCFAs and ECCs (Supporting Information Figures S1 and S2). Similar analysis utilizing untargeted lipidomics and polar metabolites (see Supporting Information Figure S3) found only a minor intersubject significant effect (Model Effect (%) 15.3, 19.9, p = 0.0010), with no contribution of the sample storage matrix, storage time, and temperature ($p > 0.05$). This may be attributed to larger within-group variance.

Metabolic Changes throughout Storage at Room Temperature. To better understand the effects of sample storage duration (24, 36, 48 h, 7 days) at room temperature on the fecal metabolome, we analyzed the stability of metabolites in aliquots stored without solvent (crude), with 95% EtOH, and with an OMNImet•GUT tube. For each sample matrix, we calculated the relative change of each metabolite compared with the gold standard (immediately frozen sample).

We found that SCFAs in the crude samples increased over time, while a stable profile was recorded in sample aliquots stored in a 95% EtOH or OMNImet•GUT tube. In particular, butyric acid, isobutyric acid, and valeric acid increased over 1.5 fold in raw feces left at RT from 24 h until 48 h (Figure 3a,b, p.adj <0.25, Supporting Information Figure S4 and Table S1), suggesting continuous microbial activity over time. We also observed an increase in the levels of butyric acid, isobutyric acid, and valeric acid by 1.72, 1.73, and 1.47-fold (FC), respectively, within the first 24 h compared to the sample that was frozen immediately. However, we observed a stable pattern of these SCFAs in fecal samples collected with 95% EtOH and/or OMNImet GUT solvent (Figure 3 and Supporting Informationy Figure S4 and Tables S2−S3). Taken together, crude feces (stored without any solvent) at room temperature showed at least a 50% increase in SCFA over 24−48 h. SCFAs are primarily produced by gut microbes through the saccharolytic fermentation of complex microbiota-accessible carbohydrates (MACs).[22,23] Therefore, our results suggest that crude samples at room temperature are susceptible to microbial metabolism by specific microorganisms encoding MAC-degrading enzymes. This is corroborated by the increased SCFA levels we recorded at room temperature when compared to samples stored at 4 °C. These results suggest that 95% EtOH can inhibit microbial activity such as saccharolytic fermentation even at room temperature, which, in turn, can prevent and reduce metabolite degradation. However, there were exceptions of conjugated BAs such as GLCA, GCDCA, GDCA, GCA, and several unknown metabolites (Supporting Information Tables S4−S5) that appeared to be affected by temperature (nominal p-value <0.05) and require further investigation.

No consistent time-dependent pattern was found during 7 days of storage at room temperature in all sample storage matrices (crude, 95% EtOH, OMNImet•GUT, immediately frozen), with the exception of SCFA in crude feces. We also analyzed the concentration differences of fecal BAs, lipids, and metabolites over time in these sampling groups. Tauro- and/or glycoconjugated bile acids (GLCA, THDCA, GCDCA), lipids (mainly TGs), and unknown polar features increased over time ($p < 0.05$) and differed in at least one of the three sample storage matrices. However, none of these metabolites exceeded the significance level at the selected FDR threshold of 0.25 (Figure 3c,d, Supporting Information Figures S5−S7, and

https://doi.org/10.1021/acs.analchem.3c04436
Anal. Chem. 2024, 96, 8893−8904

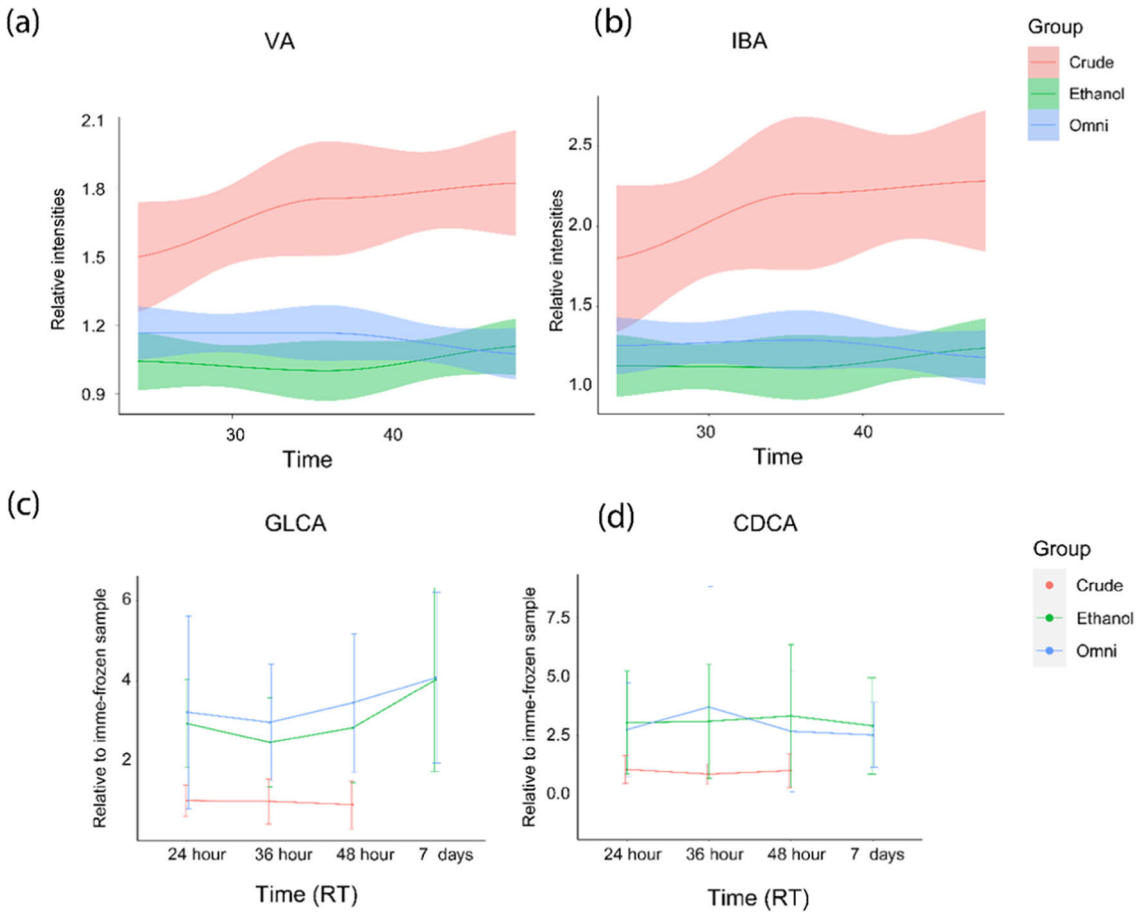


Figure 3. Alterations in metabolites during storage at room temperature. A Loess curve plot showing the changes in the levels of two SCFA [(a) Valeric acid and (b) Isobutyric acid] over time (24, 36, 48 h) in feces samples collected as crude, in 95% EtOH, and with OMNImet•GUT solvent. The *X*-axis shows the sample storage duration (24, 36, and 48 h). The *Y*-axis shows the relative change of each metabolite compared to the gold standard. VA is valeric acid, and IBA is isobutyric acid. The changes in bile acids (BAs) over time (24, 36, 48, and 7 days) were examined in feces samples collected in three different ways: crude, in 95% EtOH, and in OMNImet•GUT solvent. (c) Glycolithocholic acid (GLCA) and (d) Chenodeoxycholic acid (CDCA).

Tables S1−S3). Research investigating the gut microbiome−metabolome relationship has predominantly concentrated on water-soluble polar metabolites, with microbe-linked lipids receiving less emphasis.[24−26] To aid in this effort, apart from the methodological comparison, our results also highlight that fecal lipids can serve as a functional indicator of gut microbial metabolism.

Fecal Metabolic Changes in 95% Ethanol at Different Temperatures. We compared the differences of individual metabolite levels between fecal aliquots stored at room temperature and at +4 °C in a 24 h sample set. We found significant changes in microbial metabolites, including conjugated bile acids and SCFAs. GLCA content was 1.31 FC higher in samples stored at room temperature than in fecal aliquots stored at +4 °C in 95% EtOH (Figure 4). In crude feces, two BAs [7-oxo DCA (FC = 4.1), DHCA (FC = 0.45)] and five SCFAs, including acetic acid (FC = 1.51), butyric acid (FC = 1.64), isobutyric acid (FC = 1.37), valeric acid (FC = 1.28), propionic acid (FC = 1.32), and the internal stability standard Butyric acid-$^{13}C_4$(FC = 0.51) changed when stored at RT (Figure 4, Supporting Information Tables S4−S5). In contrast, the corresponding metabolites (e.g., SCFAs) remained stable in fecal samples collected in 95% EtOH, which may be partly due to active enzymatic metabolism by the gut microbiota at room temperature but not when deactivated by 95% EtOH (Supporting Information Tables S4−S5).

We generally observed good metabolite stability in the 95% EtOH and OMNImet•GUT kits, irrespective of the storage temperature. Our findings are consistent with previous studies suggesting that fecal samples in 95% EtOH or OMNImet•GUT have comparable metabolite profiles to samples that were frozen shortly after collection.[10,13,15,27] In addition, 95% EtOH was reported to be the most suitable matrix for preserving the fecal metabolite profile in comparison to clinically used fecal collection kits that do not target metabolites (RNAlater, OMNIgene•GUT, fecal occult blood test (FOBT) cards).[15,27] Feces collected without any solvent was suitable for

https://doi.org/10.1021/acs.analchem.3c04436
Anal. Chem. 2024, 96, 8893−8904

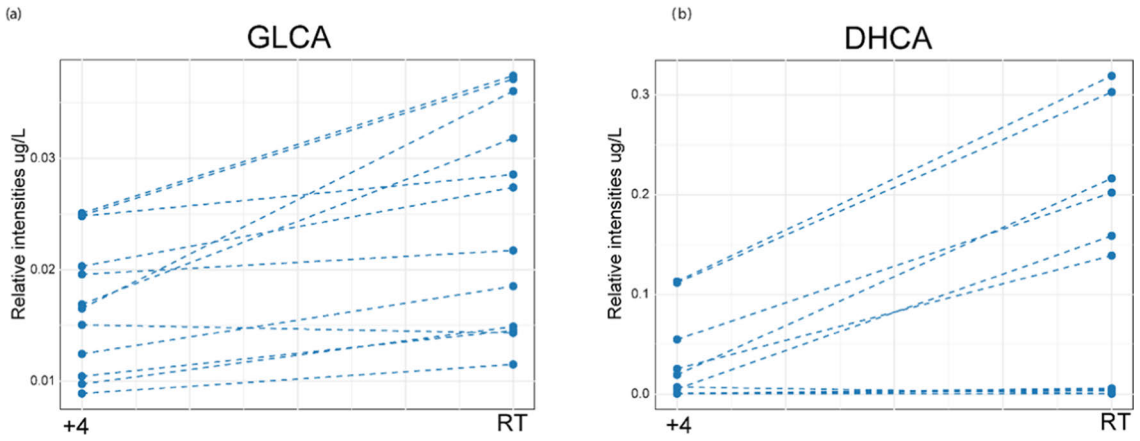


Figure 4. Stability of fecal metabolites at room temperature (RT) compared to +4 °C for (a) 95% EtOH samples and (b) crude samples. The *Y*-axis denotes concentration of metabolites, and the *X*-axis denotes the different temperatures (+4 °C and RT). GLCA is glycolithocholic acid, and DHCA is dehydrocholic acid.

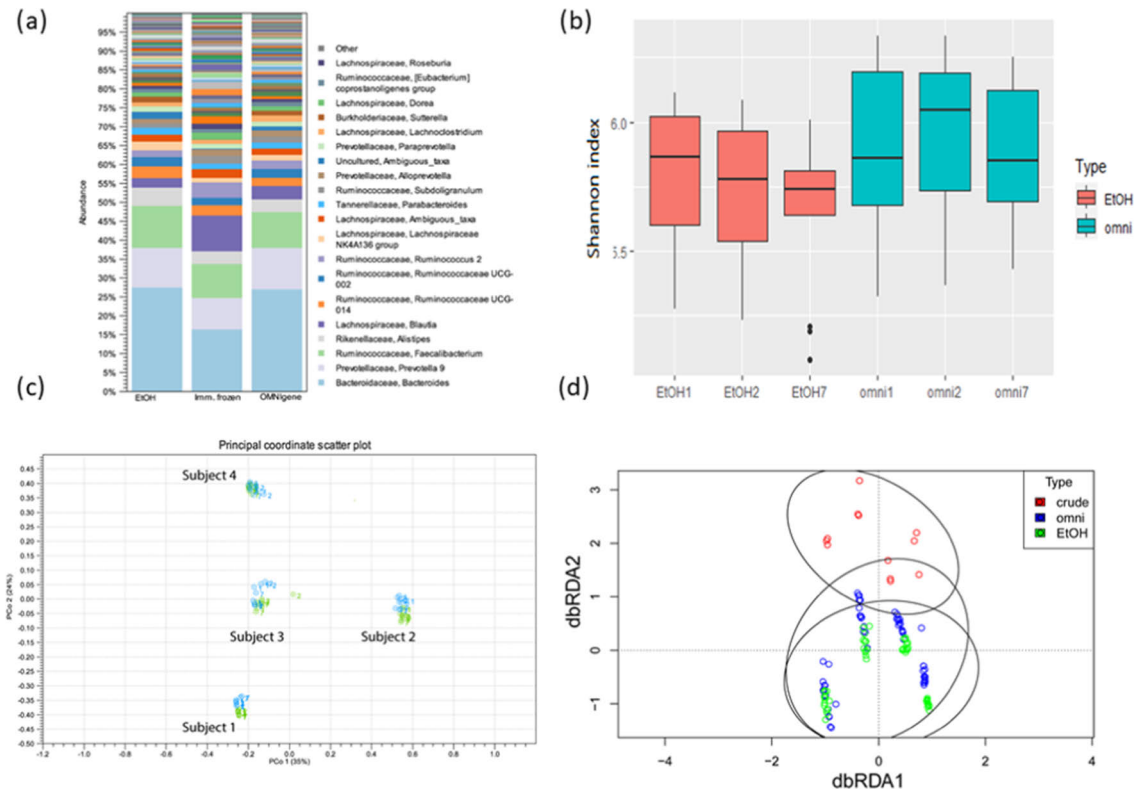


Figure 5. Stability of the fecal microbiome in 95% EtOH stored at RT. (a) Microbiome profiles by relative abundances in different storage types. Legend shows 20 most abundant genera. Main genera are the same; abundances differ between immediately frozen samples and samples in preservatives. (b) α diversity (Shannon index) by storage type and time. Numbers indicate days of storage. (c) β diversity by Principal Coordinates analysis with Bray−Curtis dissimilarity metrics. Colors indicate storage types with the number of storage days. Green and blue circles represent 95% EtOH and OMNIgene•GUT, respectively. (d) Distance-based redundancy analysis with Bray−Curtis representing dissimilarity between storage types with ellipses of 95% confidence interval.

metabolomics analysis when frozen immediately after collection.[15] However, it is worth noting that this approach is inconvenient, expensive, and not feasible for population-based studies on a large scale. Liu et al. demonstrated that

https://doi.org/10.1021/acs.analchem.3c04436
Anal. Chem. 2024, 96, 8893−8904

metabolite measures obtained from OMNIgene•GUT were comparable to those obtained from samples that were immediately frozen after collection for up to 21 days. These findings suggest that OMNIgene•GUT is sufficient for obtaining data on the gut microbiome and gut metabolome.[14] On the contrary, other studies report that OMNIgene•GUT may not be optimal for collecting fecal samples for metabolomics profiles.[10,27]

To our knowledge, this is the first study to directly compare the performance of the OMNImet•GUT kit, which is designed to preserve fecal metabolites at room temperature, to flash freezing feces or using 95% ethanol. Unlike similar studies, we also utilized internal standards and biological replicates to assess the stability of the detected metabolites at room temperature. We applied a wide range of metabolomics platforms: targeted metabolomics to measure SCFA, BAs, and endocannabinoids, as well as untargeted metabolomics to detect lipids and polar metabolites. Various studies have shown that both the human gut microbiota and metabolome are highly heterogeneous between individuals.[28−30] Indeed, our work showed that variance between fecal samples was mainly attributed to interindividual differences and affected by the sample storage matrix, while less affected by the duration at room temperature when stored in 95% EtOH and OMNImet•GUT liquid. This is consistent with similar work, where metabolite concentrations have been shown to vary considerably between individuals.[31,32] Besides that, we also found that the metabolome pattern varied across different parts of the fecal specimen (Supporting Information Figure S8). This is in line with previous studies suggesting that metabolites may not be evenly distributed within the fecal samples.[31,33] Trost et al. studied the variation in metabolites from four sampling areas of cryogenically collected fecal specimens and found that fecal metabolites are not homogeneously distributed within the specimens.[32] Similarly, Jones et al. also showed that SCFA concentrations vary profoundly across a single stool.[31] Pooling of the samples from different parts of the stool could be considered to minimize variation arising from the sampling.[32,34]

Fecal Microbiome Profile. We aimed to determine the stability of the fecal microbiome in 95% EtOH stored at RT. The composition of the fecal microbiota was analyzed by 16S rRNA gene sequencing using the V3−V4 hypervariable region (n = 118). These samples correspond to aliquots obtained from the fecal sample that were frozen straight after processing at −80 °C while other aliquots were stored for 24 h, 48 h, and 7 days at RT in 95% EtOH and OMNIgene•GUT (Figure 1). The bacterial profiles with relative abundances from these collections are shown in Figure 5a and are summarized for all subjects and time points. We found that the bacterial profiles of the fecal samples stored in 95% EtOH and OMNIgene•GUT were similar. However, the immediately frozen samples differed from those stored in 95% EtOH and OMNIgene•GUT, mainly by Bacteroides and Blautia. Positive controls included in the NGS-protocol indicated high reproducibility, and the accuracy was sufficient based on the theoretical and identified abundances (Supporting Information Figure S9).

Similarly, we analyzed the differences in the α diversity among the three study groups. We found that interindividual differences had a dominant effect on microbiome α diversity (Supporting Information Figure S10). Lower α diversity was observed in both storage solvents compared with immediately frozen samples. We also found a significant decrease in the α diversity in the longitudinal series of samples collected in 95% EtOH. Figure 5b shows that α diversity was lower over time in 95% EtOH compared with OMNIgene•GUT (p = 0.007). We also compared similarities between storage types using intraclass correlation coefficients with immediately frozen samples used as reference (ICC, see Supporting Information Figure S11). We compared similarity in α diversity metrics (Shannon, Simpson, Chao1, number of observed OTUs) and three most prevalent genera (*Bacteroides, Bifidobacterium, and Faecalibacterium*) between storage types. OMNIgene•GUT showed higher α diversity intraclass similarity with the immediately frozen samples than in the 95% EtOH samples.

Additionally, we also analyzed the differences in overall gut microbiota composition, i.e., β diversity between samples collected in 95% EtOH and OMNIgene•GUT. PcoA with Bray−Curtis dissimilarity showed little difference between storage types and time points (Figure 5c). PcoA with Jaccard, UniFrac, and Unweighted UniFrac showed parallel results; however, there were fewer differences with UniFrac metrics (Supporting Information Figures S12−S13). However, the score plot revealed that interindividual variability is the key factor shaping the fecal microbiome (Figure 5d). There were no significant differences in β diversity between storage types (distance-based redundancy analysis with Bray−Curtis dissimilarity, PERMANOVA, p = 0.3). However, interindividual differences were significant (p = 0.001), as expected. Meanwhile, the variation among biological replicates of a single individual was low (Figure S14).

Limitations. We acknowledge that there are some limitations that must be considered. The main limitation is that instead of collecting directly within each inspected matrix, we only simulated such collection; however, samples were processed immediately. This is an inherent technical limitation of the study design, which requires aliquots of the same biological sample to be compared across various collection matrices and storage conditions. Our study suggests that the metabolomes of different portions of whole feces vary profoundly. We acknowledge that aliquots of feces may not represent the entire specimen accurately. However, pooled sampling across the whole specimen may be a strategy to account for the intrasample variance. Additionally, we recognize a discrepancy in the trend of fecal butyric acid compared to labeled butyric acid, which was added prior to sample collection as stability standards. This inconsistency could likely arise due to the adsorption of the spiked standards (including other undetected spiked standards) onto the walls of the vials during drying. This may make it potentially difficult to redissolve them, especially considering that the vials are filled with fecal homogenates. Another potential limitation is the collection by trained staff (volunteers). For instance, using noncommercial 95% ethanol kits for sampling may present practical challenges compared to using commercial kits such as OMNIgene•GUT. In the current study, trained staff conducted the sampling; however, the effectiveness of using 95% EtOH for lay study participants for home collection needs to be tested. Notwithstanding this, the ongoing Alzheimer's Gut Microbiome Project (https://alzheimergut.org/) has already tested the feasibility of using 95% EtOH. It is also worth noting that all subjects in this study were adults, and the metabolite and microbiome content of feces differ in different age ranges, such as in children. Therefore, further validation may be necessary to test fecal samples collected from a wider range of age groups. Another factor that limits the reliability of

https://doi.org/10.1021/acs.analchem.3c04436
Anal. Chem. 2024, 96, 8893−8904

fecal sample collection is the variability in the volume and weight of the feces (water and fiber content). To address this, in the current setting, we collected fecal samples volumetrically, using a weight-to-volume ratio of 1:4 (weight of feces to volume of ethanol 95%) to match the W:V ratio in the commercially available kit OMNIgene•GUT. Notwithstanding that, two steps may be taken to address this issue. First, the collection tube (95% ethanol kits) should be weighed before it is sent to the participant to obtain an accurate measurement of the fecal sample weight. Second, participants report their stool consistency, which is strongly associated with the composition of the gut microbiota and metabolite content. By considering dry weight, we may be able to account for the physicochemical bias during the sample collection process. In terms of microbiome profiling, our study was limited to metataxonomics (i.e., 16S rRNA gene sequencing) analyses, and future studies should consider metagenomics (whole shotgun metagenomic sequencing) and study if 95% EtOH collection is suitable for metatranscriptomics (gene expression study).

■ CONCLUSIONS

Overall, our study found that storing feces samples at room temperature and stabilizing them in OMNImet•GUT or 95% EtOH yielded metabolomic results generally comparable to flash freezing. Specifically, we observed similar identities and abundances of detected biochemicals as well as comparable metabolic profiles of the study subjects. Moreover, we characterized metabolic changes in crude feces over time, which could be attributed to microbiota activity and nonenzymatic reactions such as oxidation−reduction. Therefore, samples could be reasonably stored in these examined preservatives at room temperature for up to 7 days. Utilizing 95% EtOH as a fecal collection matrix can offer a more convenient and cost-effective way to collect and store feces samples at home. Individual differences in microbiome's overall composition dominated those of the storage type. However, OMNIgene•GUT was slightly better than 95% EtOH at preserving microbiota based on α diversity. Further exploration of an existing commercial kit is ongoing and will expand the microbiome and metabolome assessment.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.3c04436.

List of targeted metabolites including GLCA, GCDCA, GDCA, GCA, and several unknown metabolites (ZIP)

Additional Methods section; ASCA results from endocannabinoids, lipids, and untargeted metabolomics data; plots showing concentration levels of endocannabinoids and lipids over time (24, 36, 48 h, 7 days); and α and β diversity results (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Rima Kaddurah-Daouk − *Department of Psychiatry and Behavioral Sciences, Duke University, Durham, North Carolina 27708-0187, United States*; Email: rima.kaddurahdaouk@duke.edu

Alex M. Dickens − *Turku Bioscience Centre, University of Turku, 20520 Turku, Finland; Department of Chemistry, University of Turku, 20500 Turku, Finland*; orcid.org/0000-0002-3178-8449; Email: alex.dickens@utu.fi

Santosh Lamichhane − *Turku Bioscience Centre, University of Turku, 20520 Turku, Finland*; orcid.org/0000-0002-9292-3595; Email: santosh.lamichhane@utu.fi

Authors

Heidi Isokääntä − *Research Center for Infections and Immunity, Institute of Biomedicine, University of Turku, 20520 Turku, Finland*

Lucas Pinto da Silva − *Turku Bioscience Centre, University of Turku, 20520 Turku, Finland*

Naama Karu − *Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden 2333 CC, The Netherlands*

Teemu Kallonen − *Department of Clinical Microbiology, Laboratory Division, Turku University Hospital, 20520 Turku, Finland; Clinical Microbiome Bank, Microbe Center, University Hospital and University of Turku, 20520 Turku, Finland*

Anna-Katariina Aatsinki − *Centre for Population Health Research, University of Turku, 20520 Turku, Finland*

Thomas Hankemeier − *Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden 2333 CC, The Netherlands*

Leyla Schimmel − *Department of Psychiatry and Behavioral Sciences, Duke University, Durham, North Carolina 27708-0187, United States*

Edgar Diaz − *Department of Psychiatry and Behavioral Sciences, Duke University, Durham, North Carolina 27708-0187, United States*

Tuulia Hyötyläinen − *School of Science and Technology, Örebro University, 70281 Örebro, Sweden*; orcid.org/0000-0002-1389-8302

Pieter C. Dorrestein − *Center for Microbiome Innovation, University of California, San Diego, La Jolla, California 92093-6607, United States; Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California 92093-0657, United States*; orcid.org/0000-0002-3003-1030

Rob Knight − *Center for Microbiome Innovation, University of California, San Diego, La Jolla, California 92093-6607, United States*

Matej Orešič − *Turku Bioscience Centre, University of Turku, 20520 Turku, Finland; School of Medical Sciences, Faculty of Medicine and Health, Örebro University, 702 81 Örebro, Sweden*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.3c04436

Notes

The authors declare the following competing financial interest(s): Dr. Kaddurah-Daouk is an inventor on a series of patents on the use of metabolomics for the diagnosis and treatment of CNS diseases and holds equity in Metabolon Inc., Chymia LLC and PsyProtix.

■ ACKNOWLEDGMENTS

The authors thank the Turku Metabolomics Center for the assistance and resources in the analysis of fecal metabolome and lipidome. The authors thank Matilda Kråkström for her excellent support in the technical details of the metabolomics

analysis and Päivi Haaranen for her technical support in NGS library preparation. The authors also thank Henok Karvonen for his assistance with the graphical abstract. This study was supported by the National Institute on Health grant (U19AG063744; PI: Kaddurdah-Daouk) and the Academy of Finland project grant, (No. 323171 to S.L.), (No. 333981 to M.O.), Swedish Research Council (Grant No. 2016-05176 to T.H. and M.O.), Formas (Grant No. 2019-00869 to T.H. and M.O.), and the Novo Nordisk Foundation (Grant No. NNF20OC0063971 to T.H. and M.O.). Further support was received from "Inflammation in human early life: targeting impacts on life-course health" (INITIALISE) consortium funded by the Horizon Europe Program of the European Union under Grant Agreement 101094099 (to M.O. and T.H.) and Alzheimer's Gut Microbiome Project (https://alzheimergut.org/). A.A. and H.I. were supported by the Signe and Ane Gyllenberg Foundation (grant no. 6273). H.I. received funding from the Finnish Cultural Foundation (grant no. 00230482) and further supported by the Doctoral Program in Clinical Research at the University of Turku. Funding sources had no role in study design, the collection, analysis, and interpretation of data, the writing of the report, or in the decision to submit the article for publication. All authors approved the final version and had final responsibility for the decision to submit for publication.

■ REFERENCES

(1) Visconti, A.; Le Roy, C. I.; Rosa, F.; Rossi, N.; Martin, T. C.; Mohney, R. P.; Li, W.; de Rinaldis, E.; Bell, J. T.; Venter, J. C.; Nelson, K. E.; Spector, T. D.; Falchi, M. *Nat. Commun.* **2019**, *10* (1), No. 4505.

(2) Sommer, F.; Bäckhed, F. *Nat. Rev. Microbiol.* **2013**, *11* (4), 227−238.

(3) Gomaa, E. Z. *Antonie van Leeuwenhoek* **2020**, *113* (12), 2019−2040.

(4) Fan, Y.; Pedersen, O. *Nat. Rev. Microbiol.* **2021**, *19* (1), 55−71.

(5) Franzosa, E. A.; Sirota-Madi, A.; Avila-Pacheco, J.; Fornelos, N.; Haiser, H. J.; Reinker, S.; Vatanen, T.; Hall, A. B.; Mallick, H.; McIver, L. J.; Sauk, J. S.; Wilson, R. G.; Stevens, B. W.; Scott, J. M.; Pierce, K.; Deik, A. A.; Bullock, K.; Imhann, F.; Porter, J. A.; Zhernakova, A.; Fu, J.; Weersma, R. K.; Wijmenga, C.; Clish, C. B.; Vlamakis, H.; Huttenhower, C.; Xavier, R. J. *Nat. Microbiol.* **2019**, *4* (2), 293−305.

(6) Muscogiuri, G.; Cantone, E.; Cassarano, S.; Tuccinardi, D.; Barrea, L.; Savastano, S.; Colao, A. *Int. J. Obes. Suppl.* **2019**, *9* (1), 10−19.

(7) Cryan, J. F.; O'Riordan, K. J.; Cowan, C. S. M.; Sandhu, K. V.; Bastiaanssen, T. F. S.; Boehme, M.; Codagnone, M. G.; Cussotto, S.; Fulling, C.; Golubeva, A. V.; Guzzetta, K. E.; Jaggar, M.; Long-Smith, C. M.; Lyte, J. M.; Martin, J. A.; Molinero-Perez, A.; Moloney, G.; Morelli, E.; Morillas, E.; O'Connor, R.; Cruz-Pereira, J. S.; Peterson, V. L.; Rea, K.; Ritz, N. L.; Sherwin, E.; Spichak, S.; Teichman, E. M.; van de Wouw, M.; Ventura-Silva, A. P.; Wallace-Fitzsimons, S. E.; Hyland, N.; Clarke, G.; Dinan, T. G. *Physiol. Rev.* **2019**, *99* (4), 1877−2013.

(8) Sorboni, S. G.; Moghaddam, H. S.; Jafarzadeh-Esfehani, R.; Soleimanpour, S. *Clin. Microbiol. Rev.* **2022**, *35* (1), No. e0033820.

(9) Ursell, L. K.; Haiser, H. J.; Van Treuren, W.; Garg, N.; Reddivari, L.; Vanamala, J.; Dorrestein, P. C.; Turnbaugh, P. J.; Knight, R. *Gastroenterology* **2014**, *146* (6), 1470−1476.

(10) Wang, Z.; Zolnik, C. P.; Qiu, Y.; Usyk, M.; Wang, T.; Strickler, H. D.; Isasi, C. R.; Kaplan, R. C.; Kurland, I. J.; Qi, Q.; Burk, R. D. *Front. Cell. Infect. Microbiol.* **2018**, *8*, No. 301.

(11) Zierer, J.; Jackson, M. A.; Kastenmüller, G.; Mangino, M.; Long, T.; Telenti, A.; Mohney, R. P.; Small, K. S.; Bell, J. T.; Steves, C. J.; Valdes, A. M.; Spector, T. D.; Menni, C. *Nat. Genet.* **2018**, *50* (6), 790−795.

(12) Stevens, V. L.; Hoover, E.; Wang, Y.; Zanetti, K. A. *Metabolites* **2019**, *9* (8), 156.

(13) Ramamoorthy, S.; Levy, S.; Mohamed, M.; Abdelghani, A.; Evans, A. M.; Miller, L. A. D.; Mehta, L.; Moore, S.; Freinkman, E.; Hourigan, S. K. *BMC Microbiol.* **2021**, *21* (1), No. 59.

(14) Lim, M. Y.; Hong, S.; Kim, B. M.; Ahn, Y.; Kim, H. J.; Nam, Y. D. *Sci. Rep* **2020**, *10* (1), No. 1789.

(15) Loftfield, E.; Vogtmann, E.; Sampson, J. N.; Moore, S. C.; Nelson, H.; Knight, R.; Chia, N.; Sinha, R. *Cancer Epidemiol. Biomarkers Prev.* **2016**, *25* (11), 1483−1490.

(16) Ingram, L. O. *Crit. Rev. Biotechnol.* **1989**, *9* (4), 305−319.

(17) de Goffau, M. C.; Jallow, A. T.; Sanyang, C.; Prentice, A. M.; Meagher, N.; Price, D. J.; Revill, P. A.; Parkhill, J.; Pereira, D. I. A.; Wagner, J. *Nat. Microbiol.* **2022**, *7* (1), 132−144.

(18) Williams, G. M.; Leary, S. D.; Ajami, N. J.; Chipper Keating, S.; Petrosin, J. F.; Hamilton-Shield, J. P.; Gillespie, K. M. *PLoS One* **2019**, *14* (6), No. e0216557.

(19) Kråkström, M.; Dickens, A. M.; Alves, M. A.; Forssten, S. D.; Ouwehand, A. C.; Hyötyläinen, T.; Orešič, M.; Lamichhane, S. *Metabolites* **2023**, *13* (3), No. 355.

(20) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. *BMC Bioinf.* **2010**, *11*, No. 395.

(21) Bolyen, E.; Rideout, J. R.; Dillon, M. R.; Bokulich, N. A.; Abnet, C. C.; Al-Ghalith, G. A.; Alexander, H.; Alm, E. J.; Arumugam, M.; Asnicar, F.; Bai, Y.; Bisanz, J. E.; Bittinger, K.; Brejnrod, A.; Brislawn, C. J.; Brown, C. T.; Callahan, B. J.; Caraballo-Rodríguez, A. M.; Chase, J.; Cope, E. K.; Da Silva, R.; Diener, C.; Dorrestein, P. C.; Douglas, G. M.; Durall, D. M.; Duvallet, C.; Edwardson, C. F.; Ernst, M.; Estaki, M.; Fouquier, J.; Gauglitz, J. M.; Gibbons, S. M.; Gibson, D. L.; Gonzalez, A.; Gorlick, K.; Guo, J.; Hillmann, B.; Holmes, S.; Holste, H.; Huttenhower, C.; Huttley, G. A.; Janssen, S.; Jarmusch, A. K.; Jiang, L.; Kaehler, B. D.; Kang, K. B.; Keefe, C. R.; Keim, P.; Kelley, S. T.; Knights, D.; Koester, I.; Kosciolek, T.; Kreps, J.; Langille, M. G. I.; Lee, J.; Ley, R.; Liu, Y. X.; Loftfield, E.; Lozupone, C.; Maher, M.; Marotz, C.; Martin, B. D.; McDonald, D.; McIver, L. J.; Melnik, A. V.; Metcalf, J. L.; Morgan, S. C.; Morton, J. T.; Naimey, A. T.; Navas-Molina, J. A.; Nothias, L. F.; Orchanian, S. B.; Pearson, T.; Peoples, S. L.; Petras, D.; Preuss, M. L.; Pruesse, E.; Rasmussen, L. B.; Rivers, A.; Robeson, M. S., 2nd; Rosenthal, P.; Segata, N.; Shaffer, M.; Shiffer, A.; Sinha, R.; Song, S. J.; Spear, J. R.; Swafford, A. D.; Thompson, L. R.; Torres, P. J.; Trinh, P.; Tripathi, A.; Turnbaugh, P. J.; Ul-Hasan, S.; van der Hooft, J. J. J.; Vargas, F.; Vázquez-Baeza, Y.; Vogtmann, E.; von Hippel, M.; Walters, W.; Wan, Y.; Wang, M.; Warren, J.; Weber, K. C.; Williamson, C. H. D.; Willis, A. D.; Xu, Z. Z.; Zaneveld, J. R.; Zhang, Y.; Zhu, Q.; Knight, R.; Caporaso, J. G. *Nat. Biotechnol.* **2019**, *37* (8), 852−857.

(22) Lamichhane, S.; Yde, C. C.; Jensen, H. M.; Morovic, W.; Hibberd, A. A.; Ouwehand, A. C.; Saarinen, M. T.; Forssten, S. D.; Wiebe, L.; Marcussen, J.; Bertelsen, K.; Meier, S.; Young, J. F.; Bertram, H. C. *J. Proteome Res.* **2018**, *17* (3), 1041−1053.

(23) Nogal, A.; Valdes, A. M.; Menni, C. *Gut Microbes* **2021**, *13* (1), 1−24.

(24) Lamichhane, S.; Sen, P.; Alves, M. A.; Ribeiro, H. C.; Raunioniemi, P.; Hyötyläinen, T.; Orešič, M. *Metabolites* **2021**, *11* (1), 55.

(25) Brown, E. M.; Ke, X.; Hitchcock, D.; Jeanfavre, S.; Avila-Pacheco, J.; Nakata, T.; Arthur, T. D.; Fornelos, N.; Heim, C.; Franzosa, E. A.; Watson, N.; Huttenhower, C.; Haiser, H. J.; Dillow, G.; Graham, D. B.; Finlay, B. B.; Kostic, A. D.; Porter, J. A.; Vlamakis, H.; Clish, C. B.; Xavier, R. J. *Cell Host Microbe* **2019**, *25* (5), 668−680.e7.

(26) Lee, M. T.; Le, H. H.; Johnson, E. L. *J. Lipid Res.* **2021**, *62*, No. 100034.

(27) Guan, H.; Pu, Y.; Liu, C.; Lou, T.; Tan, S.; Kong, M.; Sun, Z.; Mei, Z.; Qi, Q.; Quan, Z.; Zhao, G.; Zheng, Y. *mSphere* **2021**, *6* (5), No. e0063621.

(28) Chen, L.; Zhernakova, D. V.; Kurilshikov, A.; Andreu-Sánchez, S.; Wang, D.; Augustijn, H. E.; Vich Vila, A.; Weersma, R. K.; Medema, M. H.; Netea, M. G.; Kuipers, F.; Wijmenga, C.; Zhernakova, A.; Fu, J. *Nat. Med.* **2022**, *28* (11), 2333−2343.
(29) Zhu, A.; Sunagawa, S.; Mende, D. R.; Bork, P. *Genome Biol.* **2015**, *16* (1), No. 82.
(30) Gilbert, J. A. *Genome Biol.* **2015**, *16* (1), No. 97.
(31) Jones, J.; Reinke, S. N.; Ali, A.; Palmer, D. J.; Christophersen, C. T. *Sci. Rep.* **2021**, *11* (1), No. 13964.
(32) Trošt, K.; Ahonen, L.; Suvitaival, T.; Christiansen, N.; Nielsen, T.; Thiele, M.; Jacobsen, S.; Krag, A.; Rossing, P.; Hansen, T.; Dragsted, L. O.; Legido-Quigley, C. *Sci. Rep.* **2020**, *10* (1), No. 885.
(33) Millspaugh, J. J.; Washburn, B. E. *Gen. Comp. Endocrinol.* **2003**, *132* (1), 21−26.
(34) Lamichhane, S.; Sundekilde, U. K.; Blædel, T.; Dalsgaard, T. K.; Larsen, L. H.; Dragsted, L. O.; Astrup, A.; Bertram, H. C. *Anal. Methods* **2017**, *9* (30), 4476−4480.

https://doi.org/10.1021/acs.analchem.3c04436
Anal. Chem. 2024, 96, 8893−8904

Anna-Katariina Aatsinki, Santosh Lamichhane, Heidi Isokääntä, Partho Sen, Matilda Kråkström, Marina Amaral Alves, Anniina Keskitalo, Eveliina Munukka, Hasse Karlsson, Laura Perasto, Minna Lukkarinen, Matej Orešič, Henna-Maria Kailanto, Linnea Karlsson, Leo Lahti, Alex M Dickens (2025). Dynamics of Gut Metabolome and Microbiome Maturation during Early Life.
Accepted in iScience

III

Dynamics of Gut Metabolome and Microbiota Maturation during Early Life

Anna-Katariina Aatsinki[1,2]#, Santosh Lamichhane[3,4]#, Heidi Isokääntä[1,2,4]#, Partho Sen[3], Matilda Kråkström[3], Marina Amaral Alves[3,5], Anniina Keskitalo[5], Eveliina Munukka[6], Hasse Karlsson[1,2,7], Laura Perasto[1,2], Minna Lukkarinen[1,2,9], Matej Oresic[3,10,11], Henna-Maria Kailanto[1,2], Linnea Karlsson[1,2,8], Leo Lahti[12], Alex M Dickens[3,13*]

1. Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland
2. FinnBrain Birth Cohort Study, Turku Brain and Mind Center, Department of Clinical Medicine, University of Turku, Turku, Finland.
3. Turku Bioscience Centre, University of Turku and Åbo Akademi University, 20520 Turku, Finland
4. Research Center for Infections and Immunity, Institute of Biomedicine, University of Turku, Turku, Finland
5. Walter Mors Institute of Research on Natural Products, Federal University of Rio de Janeiro, Rio de Janeiro, RJ 21941-902, Brazil
6. Department of Clinical Microbiology, Turku University Hospital, 20520 Turku, Finland
7. Faculty of Medicine, Microbiome Biobank, University of Turku and Turku University Hospital, Turku, Finland.
8. Department of Psychiatry, University of Turku and Turku University Hospital, Turku, Finland.
9. Department of Pediatrics and Adolescent Medicine, Turku University Hospital and University of Turku, Turku, Finland.
10. School of Medical Sciences, Örebro University, 702 81 Örebro, Sweden
11. Department of Life Technologies, University of Turku, 20014 Turku, Finland
12. Department of Computing, University of Turku, 20014 Turku, Finland.
13. Department of Chemistry, University of Turku, 20520 Turku, Finland

Contact information:
Anna Aatsinki - ankaaa@utu.fi
Santosh Lamichhane - santosh.lamichhane@utu.fi
Heidi Isokääntä - heidi.kunnasranta@utu.fi
Partho Sen - partho.sen@utu.fi
Matilda Kranstrom - matilda.v.kranstrom@utu.fi
Marina Amaral Alves - marina.amaral@ippn.ufrj.br
Anniina Keskitalo - Anniina.Johanna.Keskitalo@tyks.fi
Eveliina Munukka - laevmu@utu.fi
Hasse Karlsson - hasse.karlsson@utu.fi
Laura Perasto - laura.e.perasto@utu.fi
Minna Lukkarinen - mimapo@utu.fi
Matej Oresic - matej.oresic@utu.fi
Henna-Maria Kailanto - henna-maria.kailanto@iff.com
Linnea Karlsson linnea.karlsson@utu.fi
Leo Lahti leo.lahti@utu.fi
Alex Dickens
alex.dickens@utu.fialex.dickens@utu.fialex.dickens@utu.fialex.dickens@utu.fi

# These authors contributed equally
* Corresponding author: Alex M. Dickens, alex.dickens@utu.fi

Summary

Early-life gut microbiome-metabolome crosstalk plays a crucial role in maintaining host physiology. The microbially produced metabolites often convey the effects on host health and physiology. This study investigates the gut metabolites, including short-chain fatty acids (SCFAs), bile acids (BAs), and polar metabolites, and their relationship to gut microbiota composition in a birth cohort of 670 children. Samples were collected at 2.5 (n=272), 6 (n=232), 14 (n=289), and 30 months (n=157) of age.

We identified the trajectories of the fecal metabolome that relate to the maturation of the early-life gut microbiota. We found that prevalent gut microbial abundances were associated with microbial metabolite levels, particularly in 2.5-month-old infants. Here, the abundances of early colonizers, *e.g. Bacteroides, Escherichia*, and *Bifidobacterium*, were associated with microbial metabolites, especially secondary BAs, particularly in breastfed infants.

Our results suggest that early-life gut microbiota associates with changes in metabolome composition, particularly BAs, which may have physiological implications.

Key words: gut microbiota, metabolome, early life, breastfeeding, bile acids

Introduction

Human adult gut harbors an estimated average of 500 –1000 species of microbes. The gut microbiome, which includes the by-products of microbes, as amino acids, vitamins, and organic acids, and the host interaction, is considered to be an “essential organ” within human beings [1-3]. The process of gut microbiome colonization after birth has been intensively studied during the last decade [4-6]. It has been established that members of *Bifidobacterium* and *Enterobacteriaceae* are typical in early infancy, whereas *Bacteroides* and *Ruminococcus* increase in abundance during later development, particularly when the diet diversifies [5, 7-10] Recent studies have shown that gut microbiome plays a crucial role in human health and disease, and its disturbances associate with many common diseases including inflammatory bowel disease [11], obesity [12], and various neurological and psychiatric disorders [13, 14].

Crosstalk between the gut microbiome and host metabolism is vital for maintaining human metabolic capacity [17]. Many complex interactions between the gut microbiome and the host occur via enterohepatic circulation between the liver and the intestine, and the capacity for metabolite production begins already prenatally [18, 19]. As such, profiling fecal metabolites can provide an indirect functional readout of the gut microbiome composition. The metabolites can act as an intermediate phenotype mediating host-microbiome interactions[20]. In fact, bidirectional interactions exist between the gut microbiome and metabolome[21]. For example, microbial biotransformation of bile acids (BAs) can regulate human physiology and in turn the overall host BA pool can control the microbial diversity[22]. Intriguingly, a recent rodent study suggests that gut metabolome drives gut microbiota development and maturation[23]. However, our understanding on early-life gut microbiota-metabolome maturation trajectories in humans is limited[5, 24-26]. Previous studies have shown short-chain fatty acid (SCFA) concentration increases by age, of which acetate typically plateaus first and remains relatively stable[27-30]. Additionally, there are individual studies highlighting the developmental patterns in bile acids, untargeted metabolites and

aromatic amino acids [29, 31], or study infant or child gut metabolites cross-sectionally [32]. The existing literature also highlights that preterm birth associates with gut metabolites levels [33], there is age-related changes in the metabolite levels [31], and that the early feeding causes variation in the gut metabolome composition [34-37]. However, few studies have integrated different metabolomic assays and gut microbiota data in a longitudinal fashion in general population-based birth cohort study. Although early feeding has been identified as important factor causing variation in the gut metabolome in several studies [35-37], comprehensive view on how and which early life factors associate with gut metabolome development is missing.

Here, we study how early-life gut metabolome is associated with the maturation of gut microbiota. More specifically we aimed to identify trajectories of fecal metabolome which may drive the maturation of early gut microbiota. We also explore how different early life factors, such as breastfeeding, associate with gut metabolome and microbiota.

Results

The study subjects are described in Table S1. We analyzed longitudinal metabolome, using both mass spectrometry based targeted and untargeted techniques, and microbiota in stool samples collected at 2.5 (n=444 for microbiota, n=272 for metabolome), 6 (n=256 for microbiota, n=232 for metabolome), 14 (n=302 for microbiota and n=289 metabolome), and 30 (n=207 for microbiota and n=157 for metabolome) months (mo) of age. The metabolomics dataset used for the analysis included identified metabolites from the following classes: short chain fatty acids (SCFA), bile acids (including taurine (tauro) and glycine (glyco) conjugated BAs), amino acids, carboxylic acids (mainly free fatty acids and other organic acids), hydroxy acids, phenolic compounds, alcohols, and sugar derivatives. There was no complete overlap between the timepoints: 37 children had microbiota data from all the timepoints, whereas two or three samples were available from 208 or 110 children, respectively.

Fecal Metabolites Change During Early Development

First, we explored how the gut metabolome changes by age. As expected, age-related variation displayed the major effect on the gut metabolome. Most of the SCFAs, except for acetic acid, increased with age (Fig. 1A.). Individual BAs and polar metabolites showed no clear age-related patterns (Fig 1.). Secondary BAs were positively, whilst, primary and tauroconjugated BAs remained negatively associated with age (Fig. 1B., C.) Glycoconjugated BAs were positively associated with age, however, this association attenuated when adjusting for breastfeeding (Figure S1). Some of the metabolites including 5-Hydroxyindoleacetate, 4-Hydroxyphenylacetic acid and multiple unidentified polar metabolites that had a significant age trend were attenuated when adjusting for breastfeeding associated with breastfeeding (Figure S1). Next, we also sought to explore the SCFA and BA trends in the subsample that had all the timepoints available (n=37, Figure S2). We found that the trends were similar to those in the whole sample set.

Breastfed children have lower concentrations of bile acids

In order to understand the overall contributions of various factors to gut metabolome, we performed variance analysis using variables previously shown to associate with gut microbiota maturation, *i.e.* breastfeeding, delivery mode, antibiotics intake, prenatal birth, biological sex assigned at birth, pet ownership and having siblings. In general, demographic exposures explained on average <1 % of variance in polar metabolites, SCFAs, and BA concentrations (Fig.1D.,E.,F.).

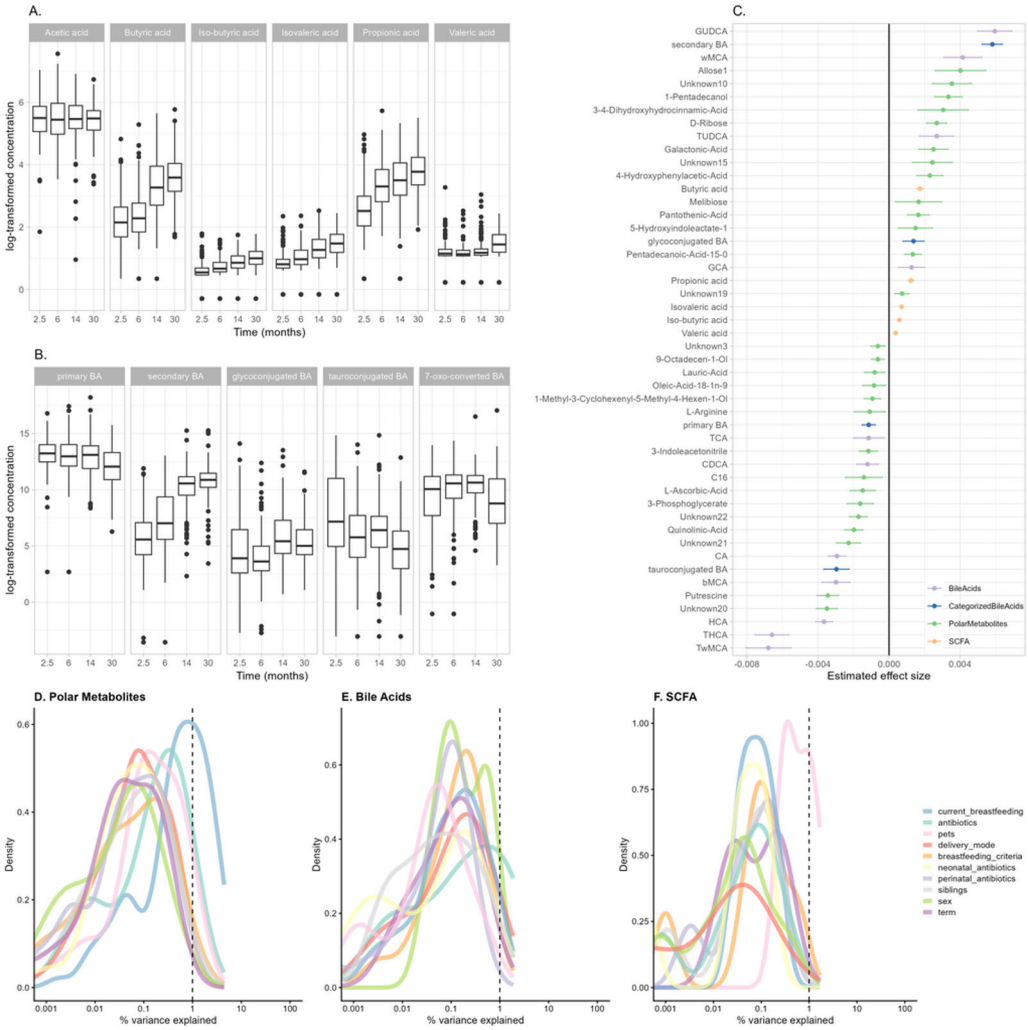


Figure 1. Metabolites varied in their age trends. SCFA tended to increase, while conjugated BA tended to decrease. A, B. The average changes in SCFAs and BAs concentrations observed across different age groups. Each box in the plot shows the median (horizontal line), the interquartile range (box spanning the 25th to 75th percentiles), and whiskers extending to data points within 1.5 times the interquartile range from the quartiles. C. Fixed effect effect-size (age coefficients) for individual metabolites as estimated from the linear mixed models, with metabolite concentration as the response variable, age as the fixed effect and child as the random effect. Error bars represent 95% confidence interval. Lighter colours indicate lower concentration. Density plots showing explained variances (%) by total D. polar metabolites, E. BAs and F. SCFAs associated with the clinical and demographic factors.

Next, to study in more detail how gut metabolites related to demographic exposures, we implemented a linear mixed-effect model with metabolite concentration as the response variable, age and demographic variable as fixed effects, and child as random effect. We found that breastfed infants had lower concentration of secondary and individual tauro- and glycoconjugated BAs especially at an early age (Fig. 2A,B, Figure S3, Figure S4).

Further, we investigated if duration of exclusive breastfeeding associated with metabolite concentrations in the 6, 14, and 30-month-olds (median 4.5 months, mean = 3.96, SD 1.95). We modelled the metabolite concentration by duration of exclusive breastfeeding, age, and any current breastfeeding (fixed effects), and child identity (random effect). Specifically, the duration of exclusive breastfeeding negatively associated with pinitol, lauric acid, ribonic acid, 1,2,3,4,5,6-hexatrimethylsilylinositol, 7-oxo-HDCA, propionic acid and iso-butyric acid, whereas succinic acid was positively associated ($q < 0.05$).

Vaginal delivery was related to lower concentration of hydroxyindoleacetate (Fig. 1C), and exposure to intravenous antibiotics in the neonatal period was associated with higher butyric acid concentration (Fig.1D). In the cross-sectional group comparison , vaginally born infants

had lower concentration of 7-oxo-converted BA at 14 mo. The primary BAs at 14 mo, and tauroconjugated BAs at 2.5 mo were also lower. Likewise, breastfed infants had lower concentration of secondary and primary BAs at 2.5 months. Having pets was positively associated with tauroconjugated BAs concentration at 14 mo, whereas having siblings was positively associated with secondary BAs concentration at 6 and 14 months. It seems that factors related to optimal microbiota development, especially breastfeeding, associated with fecal metabolite concentrations.

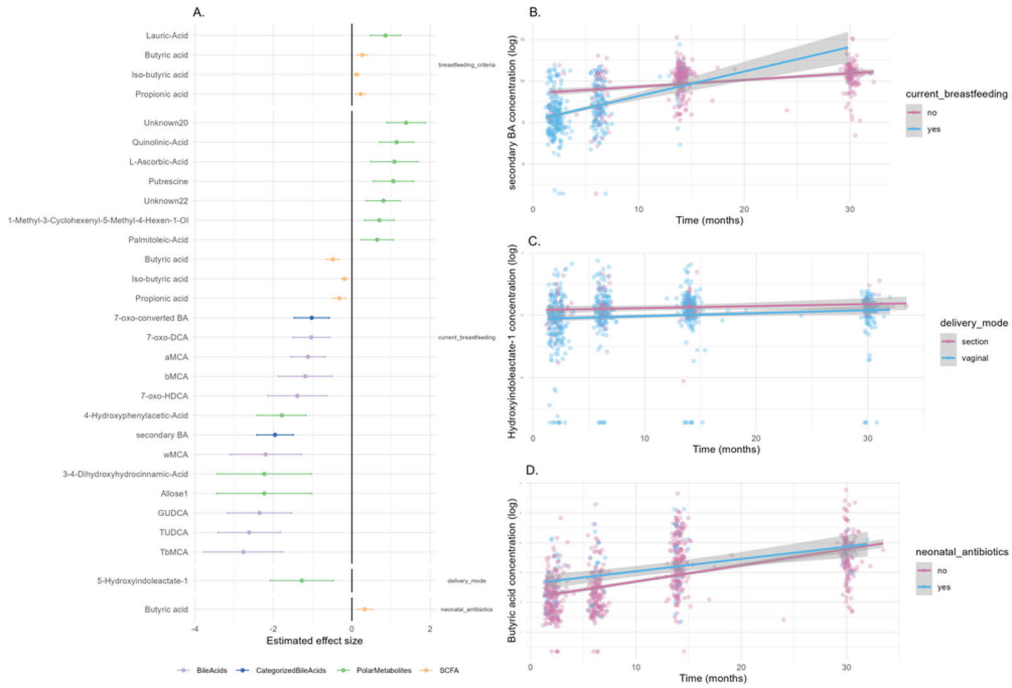


Figure 2. Out of comprehensive list of background factors, breastfeeding associated with concentrations of multiple metabolites from different assays. A. Estimates for each demographic variable from a mixed model with the metabolite concentration as the dependent variable, demographic variable and age as fixed effects, and child identity as random effect. Error bars represent 95% confidence intervals. B. Secondary BA concentrations were lower among breastfed infants in the 2.5-6-month timepoints. C. Vaginally born infants had

consistently lower concentration of hydroxyindoleacetate-1 across all timepoints. D. Concentration of butyric acid was higher in infants who received antibiotic treatment in the neonatal period. B-D. Grey area depicts 95% confidence interval.

Infant microbiota shows more diverse microbiota community types compared with toddlers

Previous literature has suggested a successional development of infant gut microbiota taxonomic composition, and we wanted to confirm these patterns in our data [5, 38]. To examine the patterns of gut microbiota succession during early-life we performed Dirichlet Multinomial Mixture (DMM) model to identify gut microbiota community types and stratify the individuals accordingly. We identified 7 community types according to Laplace criteria when jointly analysing the samples from all time points. The first timepoint was dominated by three community types, that were driven by the abundances of *Bacteroides* and *Bifidobacterium* (C1), *Escherichia* (C2), *Veillonella,* and an unidentified genus in *Enterobacriaceae* (C3), (Figure 2 and Figure S5). The majority of the later timepoints were dominated by a single community type that were driven by *Bacteroides*, *Clostridium* or *Veillonella* with differing proportions (C4-7, Figure 3, Figure S5).

Consistently with previous reports, the gut microbe community differed according to the background factors, including delivery mode and breastfeeding (Figure 3). Some additional trends were consistent with earlier reports but did not reach statistical significance. On the other hand, infant sex, having pets, overall duration of exclusive breastfeeding and intravenous neonatal or recent antibiotic intake was not associated with gut microbe community membership. When stratified by timepoint, delivery mode at 2.5 months (C1 3.3%, C2 15.1%, C3 29.4% of C-section born infants, X^2 $q < 0.005$, Figure S6) and preterm delivery at 6 months (C1 50%, C2 89%, C3 67%, C4 98%, Figure S6) were enriched in a specific

community type (Supplementary tables 2-35-6), whereas perinatal (2.5 mo, 30 mo) and recent antibiotic treatments (6 mo, 30 mo), siblings (2.5 mo, 14 mo) were not significant (Table S5). In a mixed model with the community type membership as the dependent variable, demographic variable and age as fixed effects, and child identity as random effect, the vaginal delivery and current breastfeeding were negatively related to community type progression (Fig 2C,D).

Figure 3. We stratified gut microbiota community composition into distinct community types with Dirichlet Multinomial Mixture model, and linked the community types with breastfeeding and delivery mode. A. We identified seven community types. The breaks between rows are derived from hierarchical clustering of clr-transformed abundances of genera with prevalence over 10 % and abundance over 0.1 %. Here, 50 % height of the maximum dendrogram branch height is used to visualize clusters of taxa in the heatmap. B. The community membership is indicated on the PCoA ordination. It seems that C7 was the most homogenous as indicated

by DMM theta (Table S4). The data points with non-transparent color belong to the timepoint indicated above the figures, and the partially transparent points belong to other time points. The color represents the community type. C,D. In a mixed model with the community type membership as the dependent variable, demographic variable and age as fixed effects, and child identity as random effect, C. Current breastfeeding and D. delivery mode explained transition between community types.

Microbiota alpha diversity and genera abundances are associated with fecal metabolites

Next, we sought to determine whether the microbiota composition associated with metabolome profiles. We found microbiota alpha diversity was correlated with multiple metabolite classes (Fig. 5), in particular, SCFA concentration showed consistently positive associations with alpha diversity. The observed richness was also associated positively with SCFA concentration. Linear mixed model showed that THCA, TωMCA and several polar metabolites (Arachidonic acid, 2-Methylpentadecanoic acid, Putrescine) were negatively associated with Shannon Index, adjusted for age (fixed effect) and child identify (random effect) ($q < 0.015$, Fig. 4), whereas butyric, propionic, isovaleric and iso-butyric acid, ωMCA, αMCA, UDCA, and other polar metabolite concentrations were positively associated with Shannon index when adjusted for age ($p<0.039$, Fig. 4). In addition, we found Shannon index was positively associated with 7-oxo-converted BA concentrations (estimate = 0.3, 95%-CI 0.3-0.56, $q = 0.047$). In differential abundance testing, *Clostridium* and *Bifidobacterium* showed associations with butyric acid P-hydrophenyllactic acid, and conjugated BAs in opposing directions (Figure S7). In 30-month-olds, unidentified genera in the Oscillospirales order associated negatively with BA, such as the 7-oxo-converted BA (Figure S8)

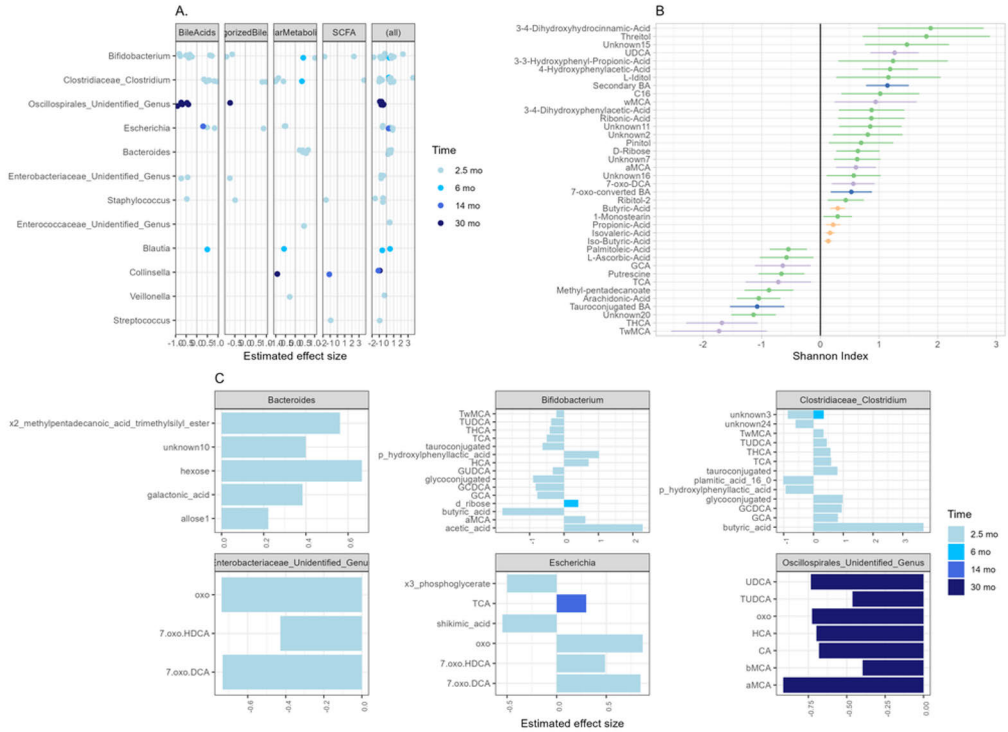


Figure 4. Gut microbiota composition associated with fecal metabolite concentrations, and the associations with genera differed based on age. A. Differential abundance analysis showed multiple associations between genus abundances and metabolites (ALDEx2). Only significant associations ($q < 0.05$) are visualized. *Bifidobacterium* (n=22), *Clostridium* (n=18), unidentified genus in Oscillaspirales (n=11), *Bacteroides* (n=9), *Escherichia* (n=9) had most significant associations. B. SCFA tended to positively correlate with alpha diversity, whereas individual polar metabolites and BAs correlated both negatively and positively. Error bars represent 95% confidence intervals. C. Most significant associations between genera and metabolites were at 2.5 mo timepoint. *Bifidobacterium* at 2.5 mo associated negatively and *Clostridium* at 2.5 mo associated positively with conjugated BAs and butyric acid. *Streptococcus* was associated negatively with propionic acid. Unidentified genus in Oscillospiroles at 30 months associated negatively with multiple BAs, especially 7-oxo-converted and tauroconjugated BAs. Data is represented as the effect size derived from ALDEx2.

We performed a network analysis for each timepoint. There, the node and edge numbers were higher in the 30 months, and the density was highest in the first timepoint (Fig. 5). Additionally, the degree distribution was more left-skewed in the 30 months compared with 6 and 14 months (Kolmogorov-Smirnov test, p-value 0.046 for 6 month, p-value 0.024 for 14 months). There were most sub-communities in the 14 months, and it had the highest modularity score (Sub-communities 2.5 months n= 5, 6 months n=7, 14 months n=13, 30 months n=10, 2.5 months modularity = 0.55, 6 months modularity = 0.53, 14 months modularity =0.57, 30 months modularity = 0.37). Metabolites and genera with highest degree and betweenness were different depending on the timepoint. For instance, *Bifidobacterium* and *Clostridium* had high degree and betweenness in the 2.5-month-olds, whereas Oscillaspirales and Ruminococcaceae had high degree and betweenness in the 30-month-olds (Fig. 5). Bile acids also had high score, and CA was among the most connected metabolites in 6, 14 and 30 months (Fig. 5).

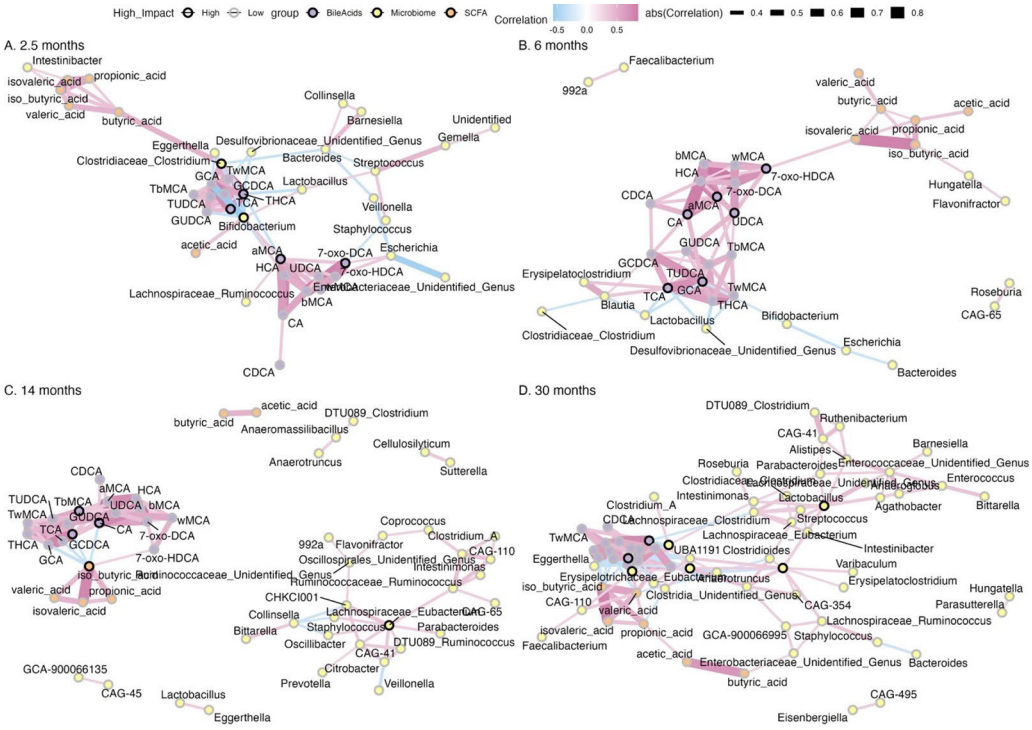


Figure 5. Networks of microbiome and metabolite inter-correlations dependent on the age. Spearman correlation was used here. Visually, the 30 months was dominated by microbial inter-correlations whereas correlations with metabolites were limited. At 2.5 months, *Bifidobacterium* and Clostridium both associated with bile acids and short-chain fatty acids. As expected, both SCFA and BA had strong correlations within the group. Black outer circle indicates ”high impact”, i.e. nodes that have high degree and betweenness, and are in the top 25 % in both. The color of the line indicates Spearman correlation coefficient and width of the line indicates absolute strength of the correlation. A. 2.5 months, B. 6 months, C. 14 months, D. 30 months.

Microbiota community types associate with different levels of metabolites

Furthermore, we examined whether metabolite concentrations were different between community types. Community types showed different levels of fecal metabolites per timepoint, and largest effect sizes were for TwMCA, TCA, THCA, GCA as well as succinic acid and an unknown polar metabolite at 2.5 months. For all the above-mentioned BAs, C1 had a lower concentration compared with C2 and/or C3. Additionally, both glucoconjugated and tauroconjugated BA concentrations were lower in C1 at 2.5 months . At 14 months, butyric acid concentration was higher in C6 compared with C5. Additionally, at 30 months, C7 had higher concentrations of valeric acid, βMCA, succinic acid and βMCA with moderate effect size.

The BAs TωMCA, THCA, TCA, GCA and arachidonic acid showed positive association with community type membership. Whereas multiple polar metabolites, UDCA, propionic acid, and branched SCFA showed negative associations with community type membership (Fig. 6). Likewise, glycoconjugated and tauroconjugated were both positively associated with community types C2-C6 and C2, C3 and C6, respectively (FDR < 0.05, C1 as reference, Fig.

6). In addition, between-community type differences in SCFA and BA concentrations were similar in the subsample of subjects with the whole timeseries available (Figure S9)

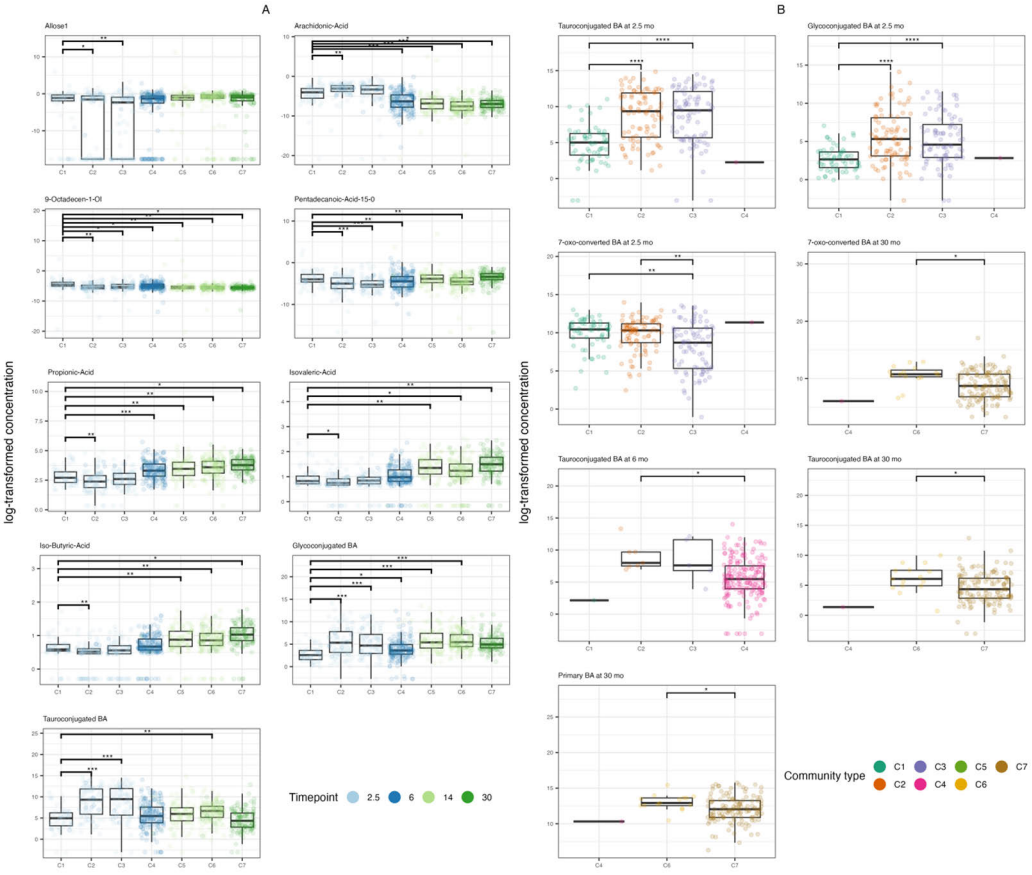


Figure 6. Community type showed different levels of metabolites. A. Several associations remained after adjusting for breastfeeding in the mixed model with the metabolite concentration as the dependent variable, community type, child age and current breastfeeding as the fixed effects, and child identity as random effect, color indicating the timepoint. B. Community types at 2.5, 6 and 30 months had different levels of BA based on cross-sectional group comparison and post hoc testing. * q < 0.05 & q > 0.01, ** q <= 0.01 & q > 0.001, ***q <= 0.001. Each box in the plot shows the median (horizontal line), the interquartile range (box

spanning the 25th to 75th percentiles), and whiskers extending to data points within 1.5 times the interquartile range from the quartiles.

Interactions Between Breastfeeding, Gut Microbiota, and Metabolites

Breastfeeding drives the microbiota maturation. We observed that breastfeeding showed the strongest associations with metabolite levels. Thus, we wanted to further explore the interactions between gut microbes abundances, metabolite levels and breastfeeding. As the prevalent genera were driving the community types and they showed the most associations metabolite levels, we studied how the interaction between prevalent genera and breastfeeding status associated with metabolite levels with metabolite concentration as the dependent variable, age and the interaction between any current breastfeeding and rclr-transformed abundance of genus as the fixed effects and child identity as the random effect.

We observed that *Bifidobacterium* abundances were associated negatively with tauroconjugated BA concentration only in breastfed infants (Fig. 7). On the other hand, *Bacteroides* was positively associated with secondary BA in the breastfed infants (Fig. 7). Moreover, the less there is *Escherichia* in breastfed infants gut microbiota, the less there is 7-oxo-HDCA (Fig. 7). Of the polar metabolites, *Bacteroides* abundances were positively associated with pinitol concentrations in the breastfed infants (Fig. 7).

We further wanted to test if cumulative exclusive breastfeeding duration interacts similarly with prevalent genera abundances, with a model that had with metabolite concentration as the dependent variable, age, any current breastfeeding and the interaction between duration of breastfeeding and rclr-transformed concentration of genus abundance as the fixed effects and child identity as the random effect. The interaction between exclusive breastfeeding duration and *Bacteroides* associated with 7-oxo-converted BA and particularly 7-oxo-DCA (Figure S10). The interaction between exclusive breastfeeding duration and *Escherichia* abundances associated with 7-oxo-HDCA (Figure S10).

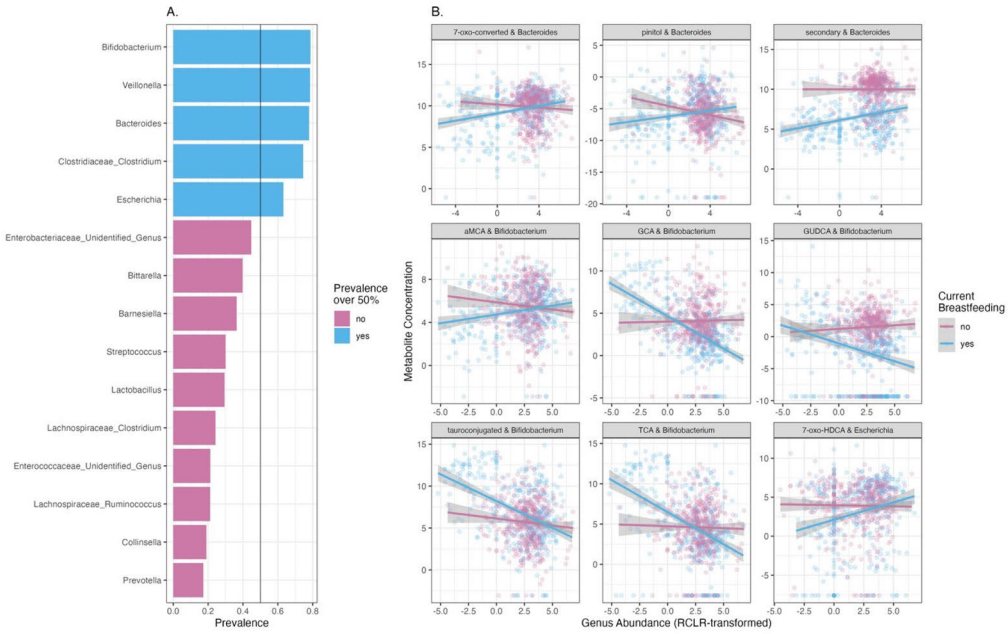


Figure 7. Prevalent genera abundances interaction with breastfeeding status associated with microbially metabolized metabolite concentrations. A. Only the five most prevalent taxa were observed in >50 % of the study subjects, and those were selected for the interaction analyses. B. *Escherichia, Bifidobacterium* and *Bacteroides* showed significant interaction with breastfeeding. Scatterplots for significant interaction models. Grey areas depict 95 % confidence intervals.

Discussion

Gut microbiota undergoes successional development in early life[38], which is affected by factors such as breastfeeding and delivery mode[5]. However, less is known about development of fecal metabolites, which are important mediators of physiological effects of the gut microbiota. Here, we showed in our population-based cohort that the fecal metabolome develops alongside the gut microbiota, and individual variation in microbiota is associated with the metabolome composition. Additionally, our observations suggest that breastfeeding, an important microbiota-modulating factor, is related to metabolite concentration depending on gut microbiota composition. This not only shows that metabolome is related to microbiota development, but that common exposures may have individualized effects based on microbiota composition. We studied not only SCFA trends, but also BA and untargeted polar metabolites in our study in addition to showing links between gut metabolites and multiple early life exposures thus extending the current understanding on early life gut metabolome and associating factors.

SCFAs, except acetic acid, were systematically increased by age and this might be explained by more complex microbiota and increased intake of indigestible fiber by age. This is in agreement with earlier studies, which suggest an increasing stool SCFA trend after birth[39] with exception for acetic acid. We observed no significant age-related increase for acetic acid, which might relate to the lack of very early sampling in our study. On the other hand, developmental patterns of BAs were more nuanced. Secondary BAs increased by age, whereas primary and tauroconjugated BAs decreased by age, which is partially in line with our previous findings[8]. The decrease in BAs could be related to increased bile salt hydrolase (BSH) activity, potentially driven by increasing abundances of *Clostridium* and *Bacteroides[40]*. Interestingly, glycoconjugated BAs were not increased by age when adjusting for breastfeeding. This could be explained by the observation that most of the infants in the first time point were breastfed, and thus harbored more *bifidobacteria*, which often have

BSH enzymes with preference for glycine as a substrate over taurine[41]. Thus, it may be that the *Bifidobacterium*-dominated microbiota is already capable of deconjugating glycine in earlier phases, which is further supported by our observation that *Bifidobacterium* was negatively associated with glycoconjugated BA concentration.

We observed that breastfeeding was related to lower abundances of butyric, iso-butyric and propionic acid, which is in contrast to Brink et al. report[41]. However, we noted a negative association between *Bifidobacterium* and butyric acid, which corroborates a finding by Nguyen et al. As noted by them, certain *Bifidobacterium* strains can compete for the same substrates as butyrate producers[24, 42], and thus the strain-level variation between the studies may underline the discrepancies in the reports. Our data would suggest that if a breastfed infants has a lower *Bifidobacterium* or higher *Bacteroides* abundance, there is a concomitant higher concentration of microbially modified BAs. Both *Bacteroides* and *Bifidobacterium* are hallmark genera of breastfed infants gut ecosystem, and those harbour differential capacity for BA metabolism. We acknowledge that strain level information is missing in our study. Notwithstanding that, we corroborate that secondary BA concentration was lower in breastfed infants[43], which might reflect slower acquisition of microbiota with BA metabolizing capacity.

In line with previous reports, we observed community typing in the gut microbiota that mostly aligned with age reflecting typical colonization patterns in early life. Notable exception was the 2.5 months' timepoint when most infants were breastfed, where three community types with either *Bifidobacterium* and *Bacteroides*, *Veillonella* and *Enterobacteriaceae* or *Escherichia* dominance were observed. *Bifidobacterium* and *Bacteroides* dominated community type was related to lower rate of C-section, which aligns with existing literature[5, 44]. Moreover, in addition to vaginal delivery, breastfeeding was related to a slower community type progression, which may indicate slower maturation of gut microbiota[5]. Thus, our data would

support the observation that cessation of breastfeeding would result in faster maturation of gut microbiota.

The differences in metabolite concentrations between community types in the 2.5-month-olds further elucidates the interaction between microbiota, factors affecting colonization and metabolites. The *Bifidobacterium* and *Bacteroides* dominated community type, which had a higher proportion of vaginally born infants, was associated with lower concentration of conjugated BAs than the two other major clusters in the first time point, most likely reflecting differences in BSH enzymatic activity. On the other hand, *Bifidobacterium* and *Bacteroides* dominated community type had higher concentration of propionic acid and branched SCFA iso-butyric and iso-valeric acids than the *Escherichia*-dominated community type, which may indicate increased availability of protein for microbial fermentation. The difference may relate to variation in human milk composition[45], as no difference in breastfeeding was observed between community types dominating the first timepoint.

It is evident that breastfeeding is an essential factor in determining the microbiota. However, there is variation in the individual colonization patterns also in the breastfed infants, and we wanted to explore how the interaction between breastfeeding and prevalent taxa is associated with metabolite concentrations. Not surprisingly, conjugated BA concentrations were lower in the breastfed infants the more they had *Bifidobacteria*. On the other hand, breastfed infants with high *Bacteroides* abundances had higher concentrations of secondary BAs. This indicates a complex interaction between early nutrition, early life microbiota and microbially metabolized products. Thus, future focus on human milk components that potentially relate to the colonization patterns in microbiota when serving as substrates for microbial fermentation is warranted.

BAs participate to regulation of inflammatory and metabolic processes via farnesoid X receptor and other bile acid-responsive receptors. For instance, secondary BAs, more abundant in breastfed infants with high *Bacteroides*-levels, may inhibit pro-inflammatory processes in microglia[46], and they are also required to activate vitamin D receptor to support optimal growth and development of adaptive immunity[47, 48]. Early-life microbiota-bile acid crosstalk may then participate in programming of growth and later brain health. However, it is uncertain how exactly the complex feedback systems affect the physiological outcomes, since gut metabolites shape the postnatal gut microbiota composition[23], and for instance tauroconjugated BAs metabolized by gut bacteria may in feedback inhibit BA synthesis via FXR antagonism[49].

Conclusions:

First, we showed that SCFA concentrations, except acetic acid, and secondary BA increase , whereas tauroconjugated BA decrease within the first 30 months. Second, breastfeeding, among the background factors known to influence gut microbiota maturation, associated with multiple metabolites. Interestingly, the secondary BA concentrations were lower in the breastfed infants. Third, we corroborated that gut microbiota shows successional maturation during the first 30 months of life. Fourth, we showed that prevalent gut microbe abundances are associated with metabolite levels, especially in the 2.5-month-olds. Finally, we demonstrate that the prevalent early colonizers *Bacteroides*, *Escherichia* and *Bifidobacterium* abundances associate with the microbial metabolized BAs especially in the breastfed infants. Alterations in early-life bile acid-microbiota crosstalk may in future studies prove important mechanism in developmental programming of health. Breastfeeding and human milk composition are likely to be important moderators in the process.

Limitations of The Study

The main limitation of the study is the lack of longitudinal samples from all the participants, and we have partially distinct participants in the different timepoints. Although we can study group-level differences between developmental stages, the small sample size of participants with the full time series prevents us from detecting nuanced intra-individual dynamics of microbiota and metabolome. Moreover, although our study benefits from a large sample of children and a representative variation in breastfeeding and delivery mode, our sample collection time points do not extend to the neonatal time nor was the sampling dense. This may have limited us to detect more nuanced patterns in the colonization and metabolome development. Additionally, our sample consisted mostly of infants and children who received some breastmilk, and we do not have adequate sample size of exclusively formula-fed infants. The utilized 16s rRNA sequencing data provided important information on the overall microbiota profiles, but the results call for future studies focusing on gene-level differences in gut microbiota. Leveraging metagenomic sequencing in future studies will help to disentangle the role of BA metabolizing capacity in the developing gut microbiome. Moreover, more detailed data on early diet, such as analysis of human milk composition, may also help to describe the differences in microbiota composition and the functional output especially in breastfed infants. Future integration of the reported exploratory findings to mechanistic models will help to elucidate the clinical potential related to inflammation[47, 50] and metabolic programming[40].

Authors' Contributions

Conception or design of the work: AKA, SL, AD, LL. Acquisition, analysis, or interpretation of data: LK, HK, EM, HMK, HI, AK, LP, ML, MO, AD, AKA, SL, LL, MAA. Drafting or substantial revision of the work: AKA, SL, HI, AD, LL. All authors have approved the submitted revision.

Acknowledgements

We want to thank all the participating families and the FinnBrain staff and assisting personnel. Turku Metabolomics Centre and Biocenter Finland is acknowledged for the collaboration regarding fecal sample metabolomics. This work was supported by the “Inflammation in human early life: targeting impacts on life-course health” (INITIALISE) consortium funded by the Horizon Europe Program of the European Union under Grant Agreement 101094099.

Finnbrain Birth cohort Study (HK) has been funded by Research Council of Finland (grant numbers 253270, 134950), Jane and Aatos Erkko Foundation, as well as Signe and Ane Gyllenberg Foundation. LK was funded by the Research Council of Finland (grant number 308176 and 325292), Yrjö Jahnsson Foundation (6847, 6976), Signe and Ane Gyllenberg Foundation, Finnish State Grants for Clinical Research (P3654), Jalmari and Rauha Ahokas Foundation, and Waterloo Foundation (2110-3601). AKA was supported by Yrjö Jahnsson Foundation, Psychiatry Research Foundation, Emil Aaltonen Foundation, Brain Foundation, Instrumentarium Science Foundation, Signe and Ane Gyllenberg Foundation, Duodecim Finnish Medical Society, Juho Vainio Foundation and Research Council of Finland (grant number 347640). HI had grant from Finnish cultural foundation [no 00230482]. LL was supported by Research Council of Finland (grant number 330887). EM was supported by the government research grant awarded to Turku University Hospital. AD has been funded by the Waterloo foundation and Research Council of Finland (347924). “Inflammation in human early life: targeting impacts on life-course health” (INITIALISE) consortium funded by the Horizon Europe Program of the European Union under Grant Agreement 101094099 (to MO HK AD).

Declaration of interests

HMK is employee International Food and Fragrances. EM was employee of the Biocodex Finland. Other authors report no conflicts of interest.

Figure Titles and Legends

Figure 1. Metabolites varied in their age trends. SCFA tended to increase, while conjugated BA tended to decrease. A, B. The average changes in SCFAs and BAs concentrations

observed across different age groups. Each box in the plot shows the median (horizontal line), the interquartile range (box spanning the 25th to 75th percentiles), and whiskers extending to data points within 1.5 times the interquartile range from the quartiles. C. Fixed effect effect-size (age coefficients) for individual metabolites as estimated from the linear mixed models, with metabolite concentration as the response variable, age as the fixed effect and child as the random effect. Error bars represent 95% confidence interval- Lighter colours indicate lower concentration. Density plots showing explained variances (%) by total D. polar metabolites, E. BAs and F. SCFAs associated with the clinical and demographic factors.

Figure 2. Out of comprehensive list of background factors, breastfeeding associated with concentrations of multiple metabolites from different assays. A. Estimates for each demographic variable from a mixed model with the metabolite concentration as the dependent variable, demographic variable and age as fixed effects, and child identity as random effect. Error bars represent 95% confidence intervals. B. Secondary BA concentrations were lower among breastfed infants in the 2.5-6-month timepoints. C. Vaginally born infants had consistently lower concentration of hydroxyindoleacetate-1 across all timepoints. D. Concentration of butyric acid was higher in infants who received antibiotic treatment in the neonatal period. B-D. Grey area depicts 95% confidence interval.

Figure 3. We stratified gut microbiota community composition into distinct community types with Dirichlet Multinomial Mixture model, and linked the community types with breastfeeding and delivery mode. A. We identified seven community types. The breaks between rows are derived from hierarchical clustering of clr-transformed abundances of genera with prevalence over 10 % and abundance over 0.1 %. Here, 50 % height of the maximum dendrogram branch height is used to visualize clusters of taxa in the heatmap. B. The community membership is indicated on the PCoA ordination. It seems that C7 was the most homogenous as indicated by DMM theta (Table S4). The data points with non-transparent color belong to the timepoint indicated above the figures, and the partially transparent points belong to other time points.

The color represents the community type. C,D. In a mixed model with the community type membership as the dependent variable, demographic variable and age as fixed effects, and child identity as random effect, C. current breastfeeding and D. delivery mode explained transition between community types.

Figure 4. Gut microbiota composition associated with fecal metabolite concentrations, and the associations with genera differed based on age. A. Differential abundance analysis showed multiple associations between genus abundances and metabolites (ALDEx2). Only significant associations ($q < 0.05$) are visualized. *Bifidobacterium* (n=22), *Clostridium* (n=18), unidentified genus in Oscillaspirales (n=11), *Bacteroides* (n=9), *Escherichia* (n=9) had most significant associations. B. SCFA tended to positively correlate with alpha diversity, whereas individual polar metabolites and BAs correlated both negatively and positively. Error bars represent 95% confidence intervals. C. Most significant associations between genera and metabolites were at 2.5 mo timepoint. *Bifidobacterium* at 2.5 mo associated negatively and *Clostridium* at 2.5 mo associated positively with conjugated BAs and butyric acid. *Streptococcus* was associated negatively with propionic acid. Unidentified genus in Oscillospiroles at 30 months associated negatively with multiple BAs, especially 7-oxo-converted and tauroconjugated BAs. Data is represented as the effect size derived from ALDEx2.

Figure 5. Networks of microbiome and metabolite inter-correlations dependent on the age. Spearman correlation was used here. Visually, the 30 months was dominated by microbial inter-correlations whereas correlations with metabolites were limited. At 2.5 months, *Bifidobacterium* and Clostridium both associated with bile acids and short-chain fatty acids. As expected, both SCFA and BA had strong correlations within the group. Black outer circle indicates ”high impact”, i.e. nodes that have high degree and betweenness, and are in the top 25 % in both. The color of the line indicates Spearman correlation coefficient and width of the line indicates absolute strength of the correlation. A. 2.5 months, B. 6 months, C. 14 months, D. 30 months.

Figure 6. Community type showed different levels of metabolites. A. Several associations remained after adjusting for breastfeeding in the mixed model with the metabolite concentration as the dependent variable, community type, child age and current breastfeeding as the fixed effects, and child identity as random effect, color indicating the timepoint. B. Community types at 2.5, 6 and 30 months had different levels of BA based on cross-sectional group comparison and post hoc testing. * q < 0.05 & q > 0.01, ** q <= 0.01 & q > 0.001, ***q <= 0.001. Each box in the plot shows the median (horizontal line), the interquartile range (box spanning the 25th to 75th percentiles), and whiskers extending to data points within 1.5 times the interquartile range from the quartiles.

Figure 7. Prevalent genera abundances interaction with breastfeeding status associated with microbially metabolized metabolite concentrations. A. Only the five most prevalent taxa were observed in >50 % of the study subjects, and those were selected for the interaction analyses. B. *Escherichia, Bifidobacterium* and *Bacteroides* showed significant interaction with breastfeeding. Scatterplots for significant interaction models. Grey areas depict 95 % confidence intervals.

STAR Methods

Resource Availability

Due to national legislation on personal data protection and the rights of the study participants, the individual-level data cannot be made available online. The study subjects have given their consent after being informed that research data may be shared with research partners, that these partners are bound by confidentiality obligations, and that the participants will be informed of these partners on the research project website.

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Linnea Karlsson (linnea.karlsson@utu.fi).

Materials Availability

Data can be shared with Research Agreement as part of research collaboration. Requests for collaboration can be sent to the Board of the FinnBrain Birth Cohort Study; please contact the Lead Contact mentioned above.

Data and Code Availability

- The individual-level data cannot be shared openly due to national legislation and the rights of the study participants.
- The R scripts for data analyses can be found in Zenodo (DOI 10.5281/zenodo.14967319).
- The key resource table (supplement) presents the reagents and other items used in the study. However, this study did not generate new unique reagents.

Experimental Model and Subject Details

The study subjects are children from the FinnBrain Cohort Study[51] that is a general population birth cohort study located in the southwestern Finland. The FinnBrain Birth Cohort Study recruited families with sufficient fluency in Finnish or Swedish, and normal 1st trimester ultrasound examination. A subset of the cohort participated in the study visits, and there were no exclusion criteria for the collection of fecal samples. The initial recruitment took place between December 2011 and April 2015, and fecal samples were collected from May 2013 to May 2018. The fecal samples were collected from the children by the parents according to written and oral instructions at 2.5, 6, 14 and 30 months postpartum. The samples were collected in plastic tubes, and parents were instructed to store the sample in a refrigerator, and bring the sample to the laboratory within 24 h. The samples were processed in the Medical Microbiology laboratory of the Research Center for Infections and Immunity, University of Turku. The sample collection time was reported.

Clinical data used in the study were collected with parental reports during and after pregnancy at 14, 24, 34 gestational weeks, 3, 6, 12, and 24 months postpartum and during study visits (2.5, 6, 14, and 30 months). Likewise, the data on maternal pre-pregnancy body mass index (BMI; kg/m2), duration of gestation as well as mode of delivery (caesarian section vs. vaginal) were collected from National Birth Registry provided by the National Institute for Health and Welfare of Finland (www.thl.fi). The information on maternal perinatal and infant neonatal intravenous antibiotic intake was collected from the hospital records. Breastfeeding was categorized in two ways: 1) any current breastfeeding (yes vs. no); 2) exclusive breastfeeding at least 4 months and partial breastfeeding for at least 6 months (breastfeeding_criteria, yes vs. no).

Ethical issues have been considered and there is a research permit for the project. FinnBrain has a permit from the Ethics Committee of the the wellbeing services county of Southwest

Finland (ETMK: 57/180/2011), which has approved Cohort profile and research protocol (Karlsson et al. 2018). FinnBrain parents have signed a consent form about their children's participation in research and given permission to use their samples for scientific purposes. Samples went through the laboratory process anonymously with research code to protect participants' privacy. STORMS guideline was used for reporting the methods and materials (Table S5).

Method Details

Metabolome analysis

The BAs were measured in fecal samples as described previously[21]. Only samples frozen within 24 h of sample collection were included in the metabolome analyses. The order of the samples was randomized before sample preparation. Two aliquots (50 mg) of each fecal sample were weighed. An aliquot was freeze-dried prior to extraction to determine the dry weight. The second aliquot was homogenized by adding homogenizer beads and 20 µL of water for each mg of dry weight in the fecal sample, followed by samples freezing to at least -70 °C and homogenizing them for five minutes using a bead beater. The BAs analysed were Litocholic acid (LCA), 12-oxo-litocholic acid(12-oxo-LCA), Chenodeoxycholic acid (CDCA), Deoxycholic acid (DCA), Hyodeoxycholic acid (HDCA), Ursodeoxycholic acid (UDCA), Dihydroxycholestanoic acid (DHCA), 7-oxo-deoxycholic acid (7-oxo-DCA), 7-oxo-hyocholic acid (7-oxo-HCA), Hyocholic acid(HCA), β-Muricholic acid (b-MCA), Cholic acid (CA), Ω/α-Muricholic acid (w/a-MCA), Glycolitocholic acid (GLCA), Glycochenodeoxycholic acid (GCDCA), Glycodeoxycholic acid (GDCA), Glycohyodeoxycholic acid (GHDCA), Glycoursodeoxycholic acid (GUDCA), Glycodehydrocholic acid (GDHCA), Glycocholic acid (GCA), Glycohyocholic acid (GHCA), Taurolitocholic acid (TLCA), Taurochenodeoxycholic acid (TCDCA), Taurodeoxycholic acid (TDCA), Taurohyodeoxycholic acid (THDCA), Tauroursodeoxycholic acid (TUDCA), Taurodehydrocholic acid (TDHCA), Tauro-α-muricholic

acid (TaMCA), Tauro-β-muricholic acid (TbMCA), Taurocholic acid (TCA), Trihydroxycholestanoic acid (THCA) and Tauro-Ω-muricholic acid (TwMCA). BAs were extracted by adding 40 µL fecal homogenate to 400 µL crash solvent (methanol containing 62,5 ppb each of the internal standards LCA-d4, TCA-d4, GUDCA-d4, GCA-d4, CA-d4, UDCA-d4, GCDCA-d4, CDCA-d4, DCA-d4 and GLCA-d4) and filtering them using a Supelco protein precipitation filter plate. The samples were dried under a gentle flow of nitrogen and resuspended using 20 µL resuspenstion solution (Methanol:water (40:60) with 5 ppb Perfluoro-n-[13C9]nonanoic acid as in injection standard). Quality control (QC) samples were prepared by combining an aliquot of every sample into a tube, vortexing it and preparing QC samples in the same way as the other samples. Blank samples were prepared by pipetting 400 µL crash solvent into a 96-well plate, then drying and resuspending them the same way as the other samples. Calibration curves were prepared by pipetting 40 µL of standard dilution into vials, adding 400 µL crash solution and drying and resuspending them in the same way as the other samples. The concentrations of the standard dilutions were between 0.0025 and 600 ppb.

The LC separation was performed on a Sciex Exion AD 30 (AB Sciex Inc., Framingham, MA) LC system consisting of a binary pump, an autosampler set to 15 °C and a column oven set to 35 °C. A waters Aquity UPLC HSS T3 (1.8µm, 2.1x100mm) column with a precolumn with the same material was used. Eluent A was 0.1 % formic acid in water and eluent B was 0.1 % formic acid in methanol. The gradient started from 15 % B and increased to 30 % B over 1 minute. The gradient further increased to 70 % B over 15 minutes. The gradient was further increased to 100 % over 2 minutes. The gradient was held at 100 % B for 4 minutes then decreased to 15 % B over 0.1 minutes and re-equilibrated for 7.5 minutes. The flow rate was 0.5 mL/min and the injection volume was 5 µL.

The mass spectrometer used for this method was a Sciex 5500 QTrap mass spectrometer operating in scheduled multiple reaction monitoring mode in negative mode. The ion source

gas1 and 2 were both 40 psi. The curtain gas was 25 psi, the CAD gas was 12 and the temperature was 650 °C. The spray voltage was 4500 V. Data processing was performed on Sciex MultiQuant.

Quantification of SCFA

We adapted and modified the targeted SCFA analysis from previous work[52]. Fecal samples were homogenized by adding water (10 µL per mg of dry weight as determined for the BA analysis) to wet feces, the samples were homogenized using a bead beater. Analysis of SCFA was performed on fecal homogenate (50 µL) crashed with 500 µL methanol containing internal standard (propionic acid-d6 and hexanoic acid-d3 at 10 ppm). Samples were vortexed for 1 min, followed by filtration using 96-Well protein precipitation filter plate (Sigma-Aldrich, 55263-U). Retention index (RI, 8 ppm C10-C30 alkanes and 4 ppm 4,4-Dibromooctafluorobiphenyl in hexane) was added to the samples. Gas chromatography (GC) separation was performed on an agilent 5890B GC system equipped with a Phenomenex Zebron ZB-WAXplus (30 m × 250 µm × 0.25 µm) column a short blank pre-column (2 m) of the same dimensions was also added. A sample volume of 1 µL was injected into a split/splitless inlet at 285°C using split mode at 2:1 split ratio using a PAL LSI 85 sampler. Septum purge flow and split flow were set to 13 mL/min and 3.2 mL/min, respectively. Helium was used as carrier gas, at a constant flow rate of 1.6 mL/min. The GC oven program was as follows: initial temperature 50°C, equilibration time 1 min, heat up to 150°C at the rate of 10°C/min, then heat at the rate of 40°C/min until 230°C and hold for 2 min. Mass spectrometry was performed on an Agilent 5977A MSD. Mass spectra were recorded in Selected Ion Monitoring (SIM) mode. The detector was switched off during the 1 min of solvent delay time. The transfer line, ion source and quadrupole temperatures were set to 230, 230 and 150°C, respectively. Dilution series of SCFA standards of acetic, propionic, butyric, valeric, hexanoic acid, isobutyric, and iso-valeric acid were prepared in concentrations of 0.1, 0.5, 1, 2, 5, 10, 20, 40, and 100 ppm for the construction of standard curves for quantification.

Analysis of polar metabolites

Polar metabolites were extracted in methanol. The method was adapted from the method used by Lamichhane et al.[8]. Fecal homogenates (60 µL) were diluted with 600 µL methanol crash solvent containing internal standards (heptadecanoic acid (5 ppm) valine-d8 (1 ppm) and glutamic acid-d5 (1 ppm)). After precipitation the samples were filtered using Supelco protein precipitation filter plates. One aliquot (50 µL) was transferred to a shallow 96-well plate to create a QC sample. The rest of the sample volume was dried under a gentle stream of nitrogen and stored in -80 °C until analysis. After thawing the samples were again dried to remove any traces of water. Derivatization was carried out on a Gerstel MPS MultiPurpoe Sampler using the following protocol: 25 µL methoxamine (20 mg/mL) was added to the sample followed by incubation on a shaker heated to 45 °C for 60 minutes. N-Methyl-N-(trimethylsilyl) trifluoroacetamide (25 µL) was added followed by incubation (60 min). After that, 25 µL retention index was added, the sample was allowed to mix for one min followed by injection. The automatic derivatization was carried out using the Gerstel maestro 1 software (version 1.4).

Gas chromatographic (GC) separation was carried out on an Agilent 7890B GC system equipped with an Agilent DB-5MS (20 m x 0,18 mm (0,18 µm)) column. A sample volume of 1 µl was injected into a split/splitless inlet at 250°C using splitless mode. The system was guarded by a retention gap column of deactivated silica (internal dimensions 1.7 m, 0.18 mm, PreColumn FS, Ultimate Plus Deact; Agilent Technologies, CA, USA). Helium was used as carrier gas at a flow rate of 1.2 ml/min for 16 min followed by 2 mL/min for 5.75 min. The temperature programme started at 50°C (5 min), then a gradient of 20°C/min up to 270°C was applied and then finally a gradient of 40°/min to 300°C, where it was held stable for 7 min. The mass spectrometry was carried out on a LECO Pegasus BT system (LECO). The acquisition delay was 420 sec. The acquisition rate was 16 spectra/sec. The mass range was 50 – 500 m/z and the extraction frequency was 30 kHz. The ion source was held at 250 °C and the transferline heater temperature was 230 °C. ChromaTOF software (version 5.51) was used for data aquisition. The samples were run in 9 batches, each consisting of 100 samples and a

calibration curve. In order to monitor the run a blank, a QC and a standard sample with a known concentration run between every 10 samples. Between every batch the septum and liner on the GC were replaced, the precolumn was cut if necessary and the instrument was tuned.

The retention index was determined with ChromaTOF using the reference method function. For every batch a reference file was created. The reference file contained the spectras and approximate retention times of the alkanes from C10 to C30 as determined manually). A reference method was implemented for every sample in order to determine the exact retention time of the alkanes. Text files with the names and retention times of the alkanes were then exported and converted to the correct format for MSDIAL using an in-house R script. The samples were exported from ChromaTOF using the netCDF format. After this they were converted to abf files using the abfConverter software (Reifycs). Untargeted data processing was carried out using MSDIAL (version 4.7). The minimum peak height was set to an amplitude of 1000, the sigma window value was 0.7 and the EI spectra cut off was 10. The identification was carried out using retention index with the help of the GCMS DB-Public-kovatsRI-VS3 library provided on the MSDIAL webpage. A separate RI file was used for each sample. The RI tolerance was 20 and the m/z tolerance was 0.5 Da. the EI similarly cut off was 70 %. The identification score cut off was 70 % and retention information was used for scoring. Alignment was carried out using the RI with an RI tolerance of 10. The EI similarity tolerance was 60 %. The RI factor was 0.7 and the EI similarity factor was 0.5. The results were exported as peak areas and further processed with excel. In excel the results were normalized using heptadecanoic acid as internal standard and the features with a coefficient of variance of less than 30 % in QC samples were selected. Further filtering was carried out to remove alkanes and duplicate features. The IDs of the features which passed the CV check were further checked using the Golm Metabolome Database.

Microbiota analysis

DNA extraction and sample processing

The samples were divided into cryotubes and frozen in -80C within 2 days after arriving at the laboratory. Samples were kept at +4C before freezing. Only samples that were frozen within 48 h of sample collection were sequenced. Sample volume for DNA extraction was approximately 100 mg. Lysis buffer was added 1 ml, and the samples were homogenized with glass beads 1000 rpm / 3 min. The samples were centrifuged at high speed (> 13000 rpm) for 5 min. The lysate (800µL) was then transferred to tubes and the extraction proceeded according to the manufacturer's protocol. DNA was extracted using a semi-automatic extraction instrument Genoxtract with DNA stool kit (HAIN life science, Germany). DNA yields were measured with Qubit fluorometer using Qubit dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific, USA). The DNA extraction and sequencing was performed at the University of Turku.

16S ribosomal RNA (rRNA) amplicon sequencing

Bacterial community composition was determined by sequencing the V4 region of 16S rRNA gene using Illumina MiSeq platform (Illumina, USA). The sequence library was constructed with an in-house developed protocol where amplicon PCR and index PCR were combined.

The DNA samples were diluted in PCR grade water to 10 ng/µL concentration prior to library PCR. PCR was performed with KAPA HiFi High Fidelity PCR kit with dNTPs (Roche, USA). Reverse and forward primers included in-house modifications verified by Rintala et al.[53]. The forward and reverse primer sequences were 5'-AATGAT-ACGGCGACCACCGAGATCTACAC -i5- TATGGTAATT -GT-GTGCCAGCMGCCGCGGTAA-3' and 5'-CAAGCAGAAGACGGCATACGAGAT -i7-AGTCAGTCAG-GC-GGACTACHVGGGTWTCTAAT-3', respectively, where i5 and i7 indicate the sample specific indexes. After PCR, 5µl of the product was analyzed with 1,5% TBE

agarose gel (100V, 1h15min). PCR products were purified with AMPure XP magnetic beads (Becman Coulter, USA). The DNA concentrations of the purified samples were measured with Qubit fluorometer using Qubit dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific, USA), after which the samples were mixed in equimolar concentration into a 4 nM library pool. The library pool was denatured, diluted to a concentration of 4 pM and a denaturized PhiX control (Illumina, USA) was added. The sequencing was performed with Illumina MiSeq Reagent kit v3 (600 cycles) on MiSeq system with 2x 250 base pair (bp) paired ends following the manufacturer's instructions. Positive control (DNA 7-mock standard) and negative control (PCR grade water) were included in library preparation and sequencing runs (Supplementary figures 11-14).

DADA2-pipeline (version 1.14) was used to preprocess the 16S rRNA gene sequencing data to infer exact amplicon sequence variants (ASVs)[54]. The reads were truncated to length 225 and reads with more than two expected errors were discarded (maxEE = 2). SILVA taxonomy database (version 138)[55, 56] and RDP Naive Bayesian Classifier algorithm[57] were used for the taxonomic assignments of the ASVs. Library sizes for all timepoints are shown in the Figure S15.

Quantification and Statistical analysis

The data analyses were performed with R version 4.2.0 with packages including *phyloseq*, *mia*, *vegan, DirichletMultinomial* and *lme*. Heatmaps were created with the *pheatmap* R package. Shannon Index and observed richness were used as alpha diversity indices and those were calculated with *mia* package from the untransformed ASV-table, i.e. count assay. Metabolite concentrations were log-transformed with a pseudocount (minimum value / 2). Dirichlet Multinomial Mixture Model (DMM) with the rarified (minimum read count 10000), genus-level count data were used to identify community types in the microbiota data [58]. The optimal number of community types was determined by the Laplace criteria.

Factor analysis, the relative contribution of a clinical/demographic factor towards the total variance of the metabolite classes were estimated by fitting a linear regression model. The total metabolite concentrations of a particular class were regressed to a clinical/demographic factor of interest, and median marginal coefficient of determination (R2) and % of explained variance were estimated. Factor analysis was performed using the *scater* package in R.

Wilcoxon test was used to test metabolite concentration difference between groups (such as breastfed and non-breastfed). Chi-square test was used to test difference in community type proportions between timepoints and groups (such as breastfed and non-breastfed). Kruskal-Wallis test with Dunn's posthoc test were used to test metabolite concentrations differences between timepoints. Linear mixed models with child ID as random effect and sampling age as fixed effect were used to study i. metabolite age-trends, ii. association between metabolite concentrations and demographic factors, iii., association between microbiota community type membership and demographic factors, iv. associations between metabolite concentrations and microbiota community type membership, and v. association between metabolite concentrations and the interaction with breastfeeding and rclr-transformed prevalent genus abundances as breastfeeding has been shown to drive the microbiota maturation[5]. Genera observed in >50 % of the study subjects were categorized as prevalent. Package lme4 was used to check for model singularity, and nlme was used for running the mixed model. The clr-module from the ALDEx2 was used for the differential abundance analysis[59]. Variance explained in the metabolome assays by demographic factors was calculated with the package *scater[60]*. p-values were adjusted for multiple testing with Benjamini-Hochberg procedure.

Supplemental Information Titles and Legends

Document S1.

Table S 1. Sample characteristics, related to Figure 1.

Figure S 1Estimates and 95 % confidence intervals for metabolites associated with age when adjusting for current breastfeeding, related to Figure 1. Error bars represent 95% confidence intervals.

Figure S 2 Metabolite age-trends in the subsample of children with the whole timeseries, related to Figure 1. A., B. Each box in the plot shows the median (horizontal line), and the interquartile range (box spanning the 25th to 75th percentiles). The upper whisker extends from the hinge to the highest value that is within 1.5 * IQR of the hinge, where IQR is the inter-quartile range. The lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge. C. Error bars represent 95% confidence intervals.

Figure S 3 Secondary bile acid concentrations in breastfed and non-breastfed children per timepoint, related to Figure 2. Each box in the plot shows the median (horizontal line), and the interquartile range (box spanning the 25th to 75th percentiles). The upper whisker extends from the hinge to the highest value that is within 1.5 * IQR of the hinge, where IQR is the inter-quartile range. The lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge.

Figure S 4. The age-trends in SCFA and BA levels were similar in the subset of children with the whole timeseries available, and breastfeeding covaried with the age, related to Figure 1.

Figure S 5. Contribution of taxonomic groups to the DMM clusters, related to Figure 3.

Figure S 6 Differences in demographic factors in clusters per timepoint. The bars represent number of individuals. Findings with p-value <0.05 are visualized, related to Figure 3.

Table S 2. Demographic factors associated with cluster membership per timepoint, related to Figure 3.

Table S 3. Dunn's posthoc test for significant cluster differences in timepoint-wise group comparison, related to Figure 3

Table S 4. Cluster pi and theta values. Clusters differed by the variability (theta). Lower theta values indicate higher variance, and the C5 and C7 have the most variance in the taxonomic contributions. Moreover, the C1 seems to have higher variance compared with C2 and C3, related to Figure 3.

Figure S 7. ALDEx2 indicated multiple associations between bile acids concentrations and Bifidobacterium and Clostridium abundances at 2.5 mo. Interestingly, they were in opposing directions. Bile acid data is log-transformed whereas abundance data is robust centered log-transformed (rclr) for the visualization, related to Figure 4.

Figure S 8. At 30 months unidentified genus in Oscillospirales order associated negative with bile acids. Bile acid data is log-transformed whereas abundance data is robust centered log-transformed (rclr) for the visualization, related to Figure 4.

Figure S 9. The BA and SCFA level differences between clusters were similar in the subset of children with the whole timeseries available, related to Figure 6. A., B. Each box in the plot shows the median (horizontal line), and the interquartile range (box spanning the 25th to 75th percentiles). The upper whisker extends from the hinge to the highest value that is within 1.5 * IQR of the hinge, where IQR is the inter-quartile range. The lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge

Figure S 10. Duration of exclusive breastfeeding and interaction with Bacteroides and Escherichia interaction associated with 7-oxo-converted BA. Mean split for the duration of exclusive breastfeeding is used here for illustration purposes, although continuous variable was used in the analyses. Model formula was metabolite concentration ~ duration of exclusive breastfeeding + age + any current breastfeeding + 1|ID, related to Figure 7. Grey areas depict 95 % confidence intervals.

Figure S 11. Positive and negative control samples in 2.5 monts time point. A. read counts across control samples. B. Relative abundances of core genera in positive control samples, related to STAR Methods, Method Details.

Figure S 12. Positive and negative control samples in 6 months time point. A. read counts across control samples. B. Relative abundances of core genera in positive control samples, related to STAR Methods, Method Details.

Figure S 13. Positive and negative control samples in 14 months time point. A. read counts across control samples per control sample type. B. Read counts in all individual control samples. C. Relative abundances of core genera in positive control samples, related to STAR Methods, Method Details.

Figure S 14. Positive and negative control samples in 30 months time point. A. read counts across control samples. B. Relative abundances of core genera in positive control samples, related to STAR Methods, Method Details.

Figure S 15. Library sizes in all timepoints and the whole sample, related to STAR Methods, Method Details. Each box in the plot shows the median (horizontal line), and the interquartile range (box spanning the 25th to 75th percentiles). The upper whisker extends from the hinge to the highest value that is within 1.5 * IQR of the hinge, where IQR is the inter-quartile range. The lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge.

Table S 5. STORMS checklist

References

1. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449(7164):804-10; doi: 10.1038/nature06244.
2. O'Hara AM, Shanahan F. The gut flora as a forgotten organ. EMBO Rep. 2006;7(7):688-93; doi: 10.1038/sj.embor.7400731.
3. Joos R, Boucher K, Lavelle A, Arumugam M, Blaser MJ, Claesson MJ, et al. Examining the healthy human microbiome concept. Nat Rev Microbiol. 2025;23(3):192-205; doi: 10.1038/s41579-024-01107-0.
4. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe. 2015;17(6):852; doi: 10.1016/j.chom.2015.05.012.
5. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature. 2018;562(7728):583-8; doi: 10.1038/s41586-018-0617-x.
6. Korpela K, de Vos WM. Early life colonization of the human gut: microbes matter everywhere. Curr Opin Microbiol. 2018;44:70-8; doi: 10.1016/j.mib.2018.06.003.
7. Mercer EM, Ramay HR, Moossavi S, Laforest-Lapointe I, Reyna ME, Becker AB, et al. Divergent maturational patterns of the infant bacterial and fungal gut microbiome in the first year of life are associated with inter-kingdom community dynamics and infant nutrition. Microbiome. 2024;12(1):22; doi: 10.1186/s40168-023-01735-3.
8. Lamichhane S, Sen P, Dickens AM, Alves MA, Harkonen T, Honkanen J, et al. Dysregulation of secondary bile acid metabolism precedes islet autoimmunity and type 1 diabetes. Cell Rep Med. 2022;3(10):100762; doi: 10.1016/j.xcrm.2022.100762.
9. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe. 2015;17(5):690-703; doi: 10.1016/j.chom.2015.04.004.
10. Roswall J, Olsson LM, Kovatcheva-Datchary P, Nilsson S, Tremaroli V, Simon MC, et al. Developmental trajectory of the healthy human gut microbiota during the first 5 years of life. Cell Host Microbe. 2021;29(5):765-76 e3; doi: 10.1016/j.chom.2021.02.021.
11. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Microbiol. 2019;4(2):293-305; doi: 10.1038/s41564-018-0306-4.
12. Muscogiuri G, Cantone E, Cassarano S, Tuccinardi D, Barrea L, Savastano S, et al. Gut microbiota: a new path to treat obesity. Int J Obes Suppl. 2019;9(1):10-9; doi: 10.1038/s41367-019-0011-7.
13. Cryan JF, O'Riordan KJ, Cowan CSM, Sandhu KV, Bastiaanssen TFS, Boehme M, et al. The Microbiota-Gut-Brain Axis. Physiol Rev. 2019;99(4):1877-2013; doi: 10.1152/physrev.00018.2018.
14. Andrioaie IM, Duhaniuc A, Nastase EV, Iancu LS, Lunca C, Trofin F, et al. The Role of the Gut Microbiome in Psychiatric Disorders. Microorganisms. 2022;10(12); doi: 10.3390/microorganisms10122436.
15. Robertson RC, Manges AR, Finlay BB, Prendergast AJ. The Human Microbiome and Child Growth - First 1000 Days and Beyond. Trends Microbiol. 2019;27(2):131-47; doi: 10.1016/j.tim.2018.09.008.
16. Tamburini S, Shen N, Wu HC, Clemente JC. The microbiome in early life: implications for health outcomes. Nat Med. 2016;22(7):713-22; doi: 10.1038/nm.4142.

17. Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. Nat Rev Immunol. 2016;16(6):341-52; doi: 10.1038/nri.2016.42.
18. Vuong HE, Pronovost GN, Williams DW, Coley EJL, Siegler EL, Qiu A, et al. The maternal microbiome modulates fetal neurodevelopment in mice. Nature. 2020;586(7828):281-6; doi: 10.1038/s41586-020-2745-3.
19. Sprockett D, Fukami T, Relman DA. Role of priority effects in the early-life assembly of the gut microbiota. Nat Rev Gastroenterol Hepatol. 2018;15(4):197-205; doi: 10.1038/nrgastro.2017.173.
20. Zierer J, Jackson MA, Kastenmuller G, Mangino M, Long T, Telenti A, et al. The fecal metabolome as a functional readout of the gut microbiome. Nat Genet. 2018;50(6):790-5; doi: 10.1038/s41588-018-0135-7.
21. Lamichhane S, Sen P, Dickens AM, Oresic M, Bertram HC. Gut metabolome meets microbiome: A methodological perspective to understand the relationship between host and microbe. Methods. 2018;149:3-12; doi: 10.1016/j.ymeth.2018.04.029.
22. Guzior DV, Quinn RA. Review: microbial transformations of human bile acids. Microbiome. 2021;9(1):140; doi: 10.1186/s40168-021-01101-1.
23. van Best N, Rolle-Kampczyk U, Schaap FG, Basic M, Olde Damink SWM, Bleich A, et al. Bile acids drive the newborn's gut microbiota maturation. Nat Commun. 2020;11(1):3692; doi: 10.1038/s41467-020-17183-8.
24. Nguyen QP, Karagas MR, Madan JC, Dade E, Palys TJ, Morrison HG, et al. Associations between the gut microbiome and metabolome in early life. BMC Microbiol. 2021;21(1):238; doi: 10.1186/s12866-021-02282-3.
25. Jian C, Carpen N, Helve O, de Vos WM, Korpela K, Salonen A. Early-life gut microbiota and its connection to metabolic health in children: Perspective on ecological drivers and need for quantitative approach. EBioMedicine. 2021;69:103475; doi: 10.1016/j.ebiom.2021.103475.
26. Matharu D, Ponsero AJ, Dikareva E, Korpela K, Kolho KL, de Vos WM, et al. Bacteroides abundance drives birth mode dependent infant gut microbiota developmental trajectories. Front Microbiol. 2022;13:953475; doi: 10.3389/fmicb.2022.953475.
27. Wu S, Ren L, Li J, Shen X, Zhou Q, Miao Z, et al. Breastfeeding might partially contribute to gut microbiota construction and stabilization of propionate metabolism in cesarean-section infants. Eur J Nutr. 2023;62(2):615-31; doi: 10.1007/s00394-022-03020-9.
28. Loniewska B, Fraszczyk-Tousty M, Tousty P, Skonieczna-Zydecka K, Maciejewska-Markiewicz D, Loniewski I. Analysis of Fecal Short-Chain Fatty Acids (SCFAs) in Healthy Children during the First Two Years of Life: An Observational Prospective Cohort Study. Nutrients. 2023;15(2); doi: 10.3390/nu15020367.
29. Laursen MF, Sinha AK, Pedersen M, Roager HM. Key bacterial taxa determine longitudinal dynamics of aromatic amino acid catabolism in infants' gut. Gut Microbes. 2023;15(1):2221426; doi: 10.1080/19490976.2023.2221426.
30. Tsukuda N, Yahagi K, Hara T, Watanabe Y, Matsumoto H, Mori H, et al. Key bacterial taxa and metabolic pathways affecting gut short-chain fatty acid profiles in early life. ISME J. 2021;15(9):2574-90; doi: 10.1038/s41396-021-00937-7.
31. Holzhausen EA, Shen N, Chalifour B, Tran V, Li Z, Sarnat JA, et al. Longitudinal profiles of the fecal metabolome during the first 2 years of life. Sci Rep. 2023;13(1):1886; doi: 10.1038/s41598-023-28862-z.

32. Barker-Tejeda TC, Zubeldia-Varela E, Macias-Camero A, Alonso L, Martin-Antoniano IA, Rey-Stolle MF, et al. Comparative characterization of the infant gut microbiome and their maternal lineage by a multi-omics approach. Nat Commun. 2024;15(1):3004; doi: 10.1038/s41467-024-47182-y.
33. Frau A, Lett L, Slater R, Young GR, Stewart CJ, Berrington J, et al. The Stool Volatile Metabolome of Pre-Term Babies. Molecules. 2021;26(11); doi: 10.3390/molecules26113341.
34. Chalifour B, Holzhausen EA, Lim JJ, Yeo EN, Shen N, Jones DP, et al. The potential role of early life feeding patterns in shaping the infant fecal metabolome: implications for neurodevelopmental outcomes. NPJ Metab Health Dis. 2023;1; doi: 10.1038/s44324-023-00001-2.
35. Rodriguez-Herrera A, Tims S, Polman J, Porcel Rubio R, Munoz Hoyos A, Agosti M, et al. Early-life fecal microbiome and metabolome dynamics in response to an intervention with infant formula containing specific prebiotics and postbiotics. Am J Physiol Gastrointest Liver Physiol. 2022;322(6):G571-G82; doi: 10.1152/ajpgi.00079.2021.
36. He X, Parenti M, Grip T, Lonnerdal B, Timby N, Domellof M, et al. Fecal microbiome and metabolome of infants fed bovine MFGM supplemented formula or standard formula with breast-fed infants as reference: a randomized controlled trial. Sci Rep. 2019;9(1):11589; doi: 10.1038/s41598-019-47953-4.
37. Sillner N, Walker A, Lucio M, Maier TV, Bazanella M, Rychlik M, et al. Longitudinal Profiles of Dietary and Microbial Metabolites in Formula- and Breastfed Infants. Front Mol Biosci. 2021;8:660456; doi: 10.3389/fmolb.2021.660456.
38. Beller L, Deboutte W, Falony G, Vieira-Silva S, Tito RY, Valles-Colomer M, et al. Successional Stages in Infant Gut Microbiota Maturation. mBio. 2021;12(6):e0185721; doi: 10.1128/mBio.01857-21.
39. Xiong J, Hu H, Xu C, Yin J, Liu M, Zhang L, et al. Development of gut microbiota along with its metabolites of preschool children. BMC Pediatr. 2022;22(1):25; doi: 10.1186/s12887-021-03099-9.
40. Wahlstrom A, Sayin SI, Marschall HU, Backhed F. Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. Cell Metab. 2016;24(1):41-50; doi: 10.1016/j.cmet.2016.05.005.
41. Brink LR, Mercer KE, Piccolo BD, Chintapalli SV, Elolimy A, Bowlin AK, et al. Neonatal diet alters fecal microbiota and metabolome profiles at different ages in infants fed breast milk or formula. Am J Clin Nutr. 2020;111(6):1190-202; doi: 10.1093/ajcn/nqaa076.
42. Moens F, Weckx S, De Vuyst L. Bifidobacterial inulin-type fructan degradation capacity determines cross-feeding interactions between bifidobacteria and Faecalibacterium prausnitzii. Int J Food Microbiol. 2016;231:76-85; doi: 10.1016/j.ijfoodmicro.2016.05.015.
43. Khine WWT, Rahayu ES, See TY, Kuah S, Salminen S, Nakayama J, et al. Indonesian children fecal microbiome from birth until weaning was different from microbiomes of their mothers. Gut Microbes. 2020;12(1):1761240; doi: 10.1080/19490976.2020.1761240.
44. Reyman M, van Houten MA, van Baarle D, Bosch A, Man WH, Chu M, et al. Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life. Nat Commun. 2019;10(1):4997; doi: 10.1038/s41467-019-13014-7.

45. Borewicz K, Gu F, Saccenti E, Hechler C, Beijers R, de Weerth C, et al. The association between breastmilk oligosaccharides and faecal microbiota in healthy breastfed infants at two, six, and twelve weeks of age. Sci Rep. 2020;10(1):4270; doi: 10.1038/s41598-020-61024-z.
46. Joo SS, Kang HC, Won TJ, Lee DI. Ursodeoxycholic acid inhibits pro-inflammatory repertoires, IL-1 beta and nitric oxide in rat microglia. Arch Pharm Res. 2003;26(12):1067-73; doi: 10.1007/BF02994760.
47. Song X, Sun X, Oh SF, Wu M, Zhang Y, Zheng W, et al. Microbial bile acid metabolites modulate gut RORgamma(+) regulatory T cell homeostasis. Nature. 2020;577(7790):410-5; doi: 10.1038/s41586-019-1865-0.
48. Ahmad O, Nogueira J, Heubi JE, Setchell KDR, Ashraf AP. Bile Acid Synthesis Disorder Masquerading as Intractable Vitamin D-Deficiency Rickets. J Endocr Soc. 2019;3(2):397-402; doi: 10.1210/js.2018-00314.
49. Sayin SI, Wahlstrom A, Felin J, Jantti S, Marschall HU, Bamberg K, et al. Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. Cell Metab. 2013;17(2):225-35; doi: 10.1016/j.cmet.2013.01.003.
50. Devkota S, Wang Y, Musch MW, Leone V, Fehlner-Peach H, Nadimpalli A, et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. Nature. 2012;487(7405):104-8; doi: 10.1038/nature11225.
51. Karlsson L, Tolvanen M, Scheinin NM, Uusitupa HM, Korja R, Ekholm E, et al. Cohort Profile: The FinnBrain Birth Cohort Study (FinnBrain). Int J Epidemiol. 2018;47(1):15-6j; doi: 10.1093/ije/dyx173.
52. Trimigno A, Khakimov B, Mejia JLC, Mikkelsen MS, Kristensen M, Jespersen BM, et al. Identification of weak and gender specific effects in a short 3 weeks intervention study using barley and oat mixed linkage beta-glucan dietary supplements: a human fecal metabolome study by GC-MS. Metabolomics. 2017;13(10):108; doi: 10.1007/s11306-017-1247-2.
53. Rintala A, Pietila S, Munukka E, Eerola E, Pursiheimo JP, Laiho 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):19-30; doi: 10.7171/jbt.17-2801-003.
54. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3; doi: 10.1038/nmeth.3869.
55. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6; doi: 10.1093/nar/gks1219.
56. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 2014;42(Database issue):D643-8; doi: 10.1093/nar/gkt1209.
57. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73(16):5261-7; doi: 10.1128/AEM.00062-07.
58. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS One. 2012;7(2):e30126; doi: 10.1371/journal.pone.0030126.

59. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. PLoS One. 2013;8(7):e67019; doi: 10.1371/journal.pone.0067019.
60. McCarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics. 2017;33(8):1179-86; doi: 10.1093/bioinformatics/btw777.

Heidi Isokääntä, Laura Perasto, Santosh Lamichhane, Minka Ovaska, Teemu Kallonen, Eveliina Munukka, Hasse Karlsson, Linnea Karlsson, Henna-Maria Kailanto, Ulrik Sundekilde, Matej Orešič, Alex M. Dickens and Anna-Katariina Aatsinki (2025). Human milk metabolites association with infant gut microbiome and metabolome development.
Manuscript

IV

Human milk metabolites associate with infant gut microbiome and metabolome development

Heidi Isokääntä[1,2,3,4], Laura Perasto[1,2], Santosh Lamichhane[4], Minka Ovaska[1], Teemu Kallonen[3,5,6], Eveliina Munukka[3,6], Matilda Kråkström[4], Hasse Karlsson[1,2,7], Linnea Karlsson[1,2,8], Henna-Maria Kailanto[2], Katri Kantojärvi[9], Tiina Paunio[9], Matej Orešič[4,10], Ulrik Sundekilde[11]*, Alex M. Dickens[4,12]* and Anna-Katariina Aatsinki[1,2]*

[1] Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland
[2] FinnBrain Birth Cohort Study, Turku Brain and Mind Center, Department of Clinical Medicine, University of Turku, Turku, Finland.
[3] Research Center for Infections and Immunity, Institute of Biomedicine, University of Turku, Turku, Finland
[4] Turku Bioscience Centre, University of Turku, Turku, Finland
[5] Clinical Microbiology, Turku University Hospital, The wellbeing services county of Southwest Finland, Turku, Finland
[6] Clinical Microbiome Bank, University of Turku and Turku University Hospital, Turku, Finland
[7] Department of Psychiatry, University of Turku and Turku University Hospital, Turku, Finland.
[8] Department of Pediatrics and Adolescent Medicine, Turku University Hospital and University of Turku, Turku, Finland.
[9] Finnish Institute for Health and Welfare (THL)
[10] School of Medical Sciences, Örebro University
[11] Department of food science, Aarhus University, Aarhus, Denmark
[12] Department of Chemistry, University of Turku, Turku, Finland
*Shared last authorship
corresponding author: Heidi Isokääntä

Introduction: Human milk is vital in establishing a healthy infant gut microbiome. Gut microbes and their metabolites are important for host development. However, our understanding of how milk metabolites are connected to the gut microbiome and metabolome in early-life remains limited. We aimed to investigate associations between the human milk metabolome and the gut microbiome in infancy.

Methods: The fecal and milk samples were collected at 2.5 (n=283), 6 (n=129) and 14 (n=65) months of age. Gut microbiota was analyzed with amplicon sequencing, gut metabolome with GC-TOF-MS and LC-MS and milk metabolites with ^{1}H-NMR.

Results: Bifidobacterium and other butyrate producers were associated with milk metabolites, such as ethanolamine, in a time-dependent manner. After the introduction of solid foods, higher milk LNFP-I and caprylate were positively associated with secondary bile acids. Additionally, 3-SL and 6-SL were associated with long-chain carbohydrates positively before the solid foods and afterwards negatively.

Conclusion: Our results highlight that milk metabolites associate with biologically relevant gut microbial metabolites. There were limited associations between milk composition and microbiome, and those differed in early- versus late-infancy, indicating that milk metabolome may have differential effect on microbial metabolism depending on the diversity of diet and/or maturity of the microbiome.

Key words: early life, infant gut, gut microbiome, gut metabolome, fecal metabolites, human milk, milk metabolites, secretor status, FUT2 gene

Importance

It's essential to raise awareness about breastfeeding and its variety of health benefits. Expanding our understanding of how human milk metabolites interact with the gut microbiome can help highlight these positive effects. Exploratory studies of human milk composition with gut microbial metabolites may support achieving this goal. Furthermore, understanding of biological effects of different components in human milk is important, and it can help to develop strategies to ensure optimal growth and development of an infant regardless of the breastfeeding status of the child.

1 Introduction

The infant gut microbiome undergoes dynamic developmental stages and is shaped by various early-life factors, including gestational age, delivery mode, and feeding practices[1–3]. Among these, human milk (HM) plays a central role in fostering a gut environment dominated by *Bifidobacterium* species. HM is rich in bioactive components—such as energy-yielding metabolites, enzymes, immunoglobulins, vitamins, and human milk oligosaccharides (HMOs)—that support microbial colonization and infant health [4, 5]. HMOs, whose composition is partly dictated by the maternal secretor genotype (FUT2), are indigestible by the infant but serve as essential substrates for specific gut bacteria. These bacteria, in turn, generate microbial metabolites that act as signalling molecules, influencing immune function, neurodevelopment, and metabolic processes. [6–9] Another cohort study found associations between breastfeeding and bacterial genera in a longitudinal manner when Bifidobacterium and Bacteroides appear as early keystone organisms, directing microbiota development and consistently predicting positive health outcomes. [10, 11] In addition, they saw specific patterns in the associations between HMOs and infant gut microbiota depending on the delivery mode, however the temporal changes in HMO profiles were unobserved [12].

Although it is relatively well-established how breastfeeding and HMOs are related to gut microbes, less is known about the metabolites produced by gut microbes. Microbial metabolites produced by gut microbes are often key mediators of health effects [13–15]. We and others have examined how feeding mode affects the composition of infant fecal metabolites, focusing on short-chain fatty acids (SCFAs) and secondary bile acids [16, 17]. Although, the shifts in the early-life fecal metabolome are driven largely by dietary alterations up to the first 2 years of life [18], there is still a lack of longitudinal research that simultaneously considers human milk composition, gut microbiota and fecal metabolite profiles. Understanding how natural variation in human milk influences microbiota assembly and function could guide strategies to promote healthy microbial development, support metabolic health, and mitigate disruptions such as those caused by antimicrobial exposures. To address this gap, we conducted a longitudinal cohort study involving infants to investigate the interconnected development of human milk metabolites, gut microbiota, and fecal metabolome across infancy.

In this study, we aimed to elucidate how HM metabolites, known to shape the infant microbiota, relate to fecal microbiota and metabolome at 2.5, 6 and 14 months. We hypothesized that variation in HM metabolite composition and secretor status of mother and child may drive differences in gut microbial communities among breastfed infants and that these microbial shifts are reflected in the fecal metabolite profile. We speculate that both maternal and infant FUT2 secretor status may have independent contribution to infant gut microbiota driven by difference in HMO composition and fucosylated mucin production, respectively. We expect to observe links between milk HMOs and aromatic lactic acids [19], as well as lipid-class metabolites and fecal bile acids. Moreover, we expect that associations between milk, fecal microbiota and metabolome may change during the first year of life. Results may reveal the variation in healthy microbiome development. This may be important for the foundation of gut microbiome and how the seeding effect of microbiome for later life health is formed.

2 Materials and Methods

2.1 Study participants and sample collection

Data was collected in FinnBrain birth cohort study [20] from the region of southwestern Finland and Åland in 2013-2016. Infant fecal samples (total 1002) were collected at 2.5 (n=444), 6 (n=256) and 14 (n=302) months of age in a preservative-free tube at home. They were delivered to a study visit and laboratory at +4C. The fecal samples were frozen at -80°C within 48h after sample collection. For fecal metabolomics, only samples that were frozen within 24h were included.

Human milk (HM) samples were collected at 2 (n=406), 6 (n=176) and 14 (n=80) months of age during study visits in the presence of a study nurse. Mothers were instructed to feed their infant from their right breast 1.5–2 h prior to the study visit, but breastfeeding from the left breast was allowed based on infant needs during the same day. While wearing latex gloves, mothers were instructed to express manually 10 ml of foremilk into a sterile cup from their right breast. Samples with lower volumes were not excluded. After the sample collection, the sample was stirred and then divided into aliquots, stored at 4°C, and frozen at −80° during the same day.

Mothers filled questionnaires on background factors at 14 GW (gestational weeks), 24GW, 36 GW and on breastfeeding habits after delivery at 2.5, 6 and 14 months of infant age. Health records including information on delivery mode and perinatal antibiotic treatments were collected in from the hospital records of the Wellbeing Services of the County of Southwestern Finland.

2.2 Milk metabolomics

The milk samples were analyzed in Aarhus University, Department of Food Sciences. The samples intended for ^{1}H nuclear magnetic resonance (NMR)-based metabolomics were processed in a random order as followed using standard protocol for milk-based metabolomics [21]. Samples were thawed in water bath and kept on ice while Amplicon Ultra 0.5-ml 10-kDa spin filters (Millipore, Billerica, MA, United States) were being washed three times. The samples were skimmed by centrifugation at 4,000 g, at 4°C for 10 min, fat layer removed, and 500 µl of the skimmed milk transferred to individual Amplicon Ultra 0.5-ml 10 kDa spin filters. Next, the skimmed milk was filtered by centrifugation at 10,000 g at 4°C, for 30 min and 400 µl of filtered milk from each sample was transferred to an individual 5-mm NMR tube. In each tube, 200 µl D2O with 0.05% 3-(trimethylsilyl)propanoic acid (TSP, Sigma-Aldrich, Saint-Louis, MO, USA) was added. Spectra acquisition was acquired according to the study of Sundekilde et al. [21]. Using a Bruker Avance III 600 spectrometer equipped with a 5-mm 1H TXI probe (Bruker BioSpin, Rheinstetten, Germany), ^{1}H-NMR spectra were acquired at 298 K and a 1H frequency of 600.13 MHz. A single 90° pulse experiment (Bruker pulse sequence: noesypr1d) was run to acquire one-dimensional spectra with a relaxation delay of 5 s. During the relaxation delay, water suppression was performed, and a total of 64 scans were comprised of 32.768 data points with a spectral width of 12.15 ppm. The resulting ^{1}H-NMR spectra were all referenced to TSP signal at 0 ppm. A line-broadening function by 0.3 Hz was applied to each ^{1}H NMR spectra, followed by a Fourier transformation. Preprocessing of ^{1}H-NMR spectra was subsequently conducted by phase and baseline corrections, both automatically and manually using Topspin 3.2 (Bruker Biospin, Rheinstetten, Germany).

Milk metabolite data consists of 44 compounds. Milk metabolites were grouped into five groups based on their chemical classification: HMOs, Energy metabolism, Amino acids & Protect nutrients, Bacterial fermentation and Lipids & fatty acids (Table S1). Milk samples give a snapshot of compound

concentrations and therefore normalization was needed. Normalization for milk metabolites was done by lactose content since the lactose is relatively stable in mature milk and is correlated with milk volume [22] and has previously been used for normalization [21]. Data was log-transformed.

2.2.1 Pheno- and genotypes of secretor status (FUT2)

Mother´s secretor status of 2-Fucosyllactose (FUT2) was determined by the concentration of 2'-FL in milk. Infant secretor status (FUT2) was obtained from genetic SNPs from blood samples as in previous study [23]. Briefly, DNA from cord blood was extracted as instructed by standard procedures at the Finnish Institute for Health and Welfare (THL). Extracted DNAs were genotyped with Illumina Infinium PsychArray BeadChip (Illumina, San Diego, CA) comprising 603132 SNPs at Estonian Genome Centre (Tartu, Estonia), and quality control was carried out with PLINK (www.cog-genomics.org/plink/1.9/). Markers were deleted for missingness (> 5%) and Hardy–Weinberg equilibrium ($P < 1 \times 10–6$). Subjects were checked for missing genotypes (> 5%), relatedness (identical by descent calculation, PI_HAT > 0.2) and population stratification (scaled multidimensionally). Two SNPs of FUT2 gene rs3894326_T (non-functional FUT2 gene) and rs812936_G (non-functional FUT2 gene) were investigated to determine the secretor status of infant. The SNPs were coded as minor/minor=2, major/minor=1 and major/major=0, meaning that subjects with zero values have normally functioning FUT2 gene (secretors) and other combinations were considered as non-secretors.

2.3 Gut microbiome

Only samples that were frozen within 48 h of sample collection were sequenced. For DNA extraction, 1 ml of lysis buffer was added and the samples (~100mg) were homogenized with glass beads 1000 rpm / 3 min. The samples were centrifuged at high speed (> 13000 rpm) for 5 min. The lysate (800µL) was then transferred to clean tubes. DNA was extracted using a semiautomatic extraction instrument Genoxtract with DNA stool kit (HAIN life science, Germany) and the extraction proceeded according to the manufacturer's protocol. For all procedures related to microbiota analysis, DNA- and RNAse-free plastics were employed to prevent nucleic acid degradation.

DNA yields were measured with Qubit fluorometer using Qubit dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific, USA). The DNA extraction and sequencing were performed in the University of Turku. Bacterial community composition was determined by sequencing the V4 region of 16S rRNA gene using Illumina MiSeq platform (Illumina, USA). The sequence library was constructed with an in-house developed protocol where amplicon PCR and index PCR were combined [24]. Positive control (DNA 7-mock standard) and negative control (PCR grade water) were included in library preparation and sequencing runs. DADA2-pipeline (version 1.14) was used to preprocess the 16S rRNA gene sequencing data to infer exact amplicon sequence variants (ASVs)[25]. The reads were truncated to length 225 and reads with more than two expected errors were discarded (maxEE = 2). SILVA taxonomy database (version 138) [26, 27] and RDP Naive Bayesian Classifier algorithm [28] were used for the taxonomic assignments of the ASVs.

2.4 Fecal metabolomics

Fecal metabolites were already measured as described in the previous study [29]. The order of the samples was randomized before sample preparation. An aliquot was freeze-dried prior to extraction

to determine the dry weight. The second aliquot (~50mg) was homogenized with beads and 20 µL of water for each mg of dry weight in the fecal sample.

Bile acids were extracted by adding 40 µL fecal homogenate to 400 µL crash solvent (methanol containing 62,5 parts per billion ppb each of the internal standards LCA-d4, TCA-d4, GUDCA-d4, GCA-d4, CA-d4, UDCA-d4, GCDCA-d4, CDCA-d4, DCA-d4 and GLCA-d4) and filtering them using a Supelco protein precipitation filter plate. The samples were dried under a gentle flow of nitrogen and resuspended using 20 µL resuspension solution (Methanol:water (40:60) with 5 ppb Perfluoro-n-[13C9] nonanoic acid as in injection standard). Quality control (QC) samples were prepared by combining an aliquot of every sample into a tube, vortexing it and preparing QC samples in the same way as the other samples. Blank samples were prepared by pipetting 400 µL crash solvent into a 96-well plate, then dried and resuspended as the other samples. Calibration curves were prepared by pipetting 40 µL of standard dilution into vials, adding 400 µL crash solution and drying and resuspending them in the same way as the other samples. The concentrations of the standard dilutions were between 0.0025 and 600 ppb. The LC separation was performed on a Sciex Exion AD 30 (AB Sciex Inc., Framingham, MA) LC system consisting of a binary pump, an autosampler set to 15 °C and a column oven set to 35 °C. A waters Aquity UPLC HSS T3 column with a precolumn with the same material was used. The flow rate was 0.5 mL/min, and the injection volume was 5 µL.

The mass spectrometer used for this method was a Sciex 5500 QTrap mass spectrometer operating in scheduled multiple reaction monitoring mode in negative mode. Data processing was performed on Sciex MultiQuant.

2.4.1 Quantification of SCFA

We adapted and modified the targeted SCFA analysis from previous work [30]. Fecal samples were homogenized by adding water (10 µL per mg of dry weight as determined for the BA analysis) to wet feces, the samples were homogenized using a bead beater. Analysis of SCFA was performed on fecal homogenate (50 µL) crashed with 500 µL methanol containing internal standard (propionic acid-d6 and hexanoic acid-d3 at 10 parts per million (ppm)). Samples were vortexed for 1 min and filtrated using 96-Well protein precipitation filter plate (Sigma-Aldrich, 55263-U). Retention index (RI, 8 ppm C10-C30 alkanes and 4 ppm 4,4-Dibromooctafluorobiphenyl in hexane) was added to the samples. Gas chromatography (GC) separation was performed on an Agilent 5890B GC system equipped with a Phenomenex Zebron ZB-WAXplus column; a short blank pre-column of the same dimensions was also added. A sample volume of 1 µL was injected into a split/splitless inlet at 285°C using split mode at 2:1 split ratio using a PAL LSI 85 sampler. Mass spectrometry was performed on an Agilent 5977A MSD. Mass spectra were recorded in Selected Ion Monitoring (SIM) mode. The detector was switched off during the 1 min solvent delay time.

2.4.2. Analysis of polar metabolites

Polar metabolites were extracted in methanol. The method was adapted from the method used by Lamichhane et al.[31]. Fecal homogenate (60 µL) was diluted with 600 µL methanol crash solvent containing internal standards (heptadecanoic acid (5 ppm) valine-d8 (1 ppm) and glutamic acid-d5 (1 ppm)). After precipitation the samples were filtered using the same filter plates as above. One aliquot (50 µL) was transferred to a shallow 96-well plate to create a QC sample. The rest of the sample volume

was dried under a gentle stream of nitrogen and stored at -80 °C until analysis. After thawing the samples were again dried to remove any traces of water. Derivatization was carried out on a Gerstel MPS MultiPurpoe Sampler followed by injection. The automatic derivatization was carried out using the Gerstel maestro 1 software (version 1.4).

Gas chromatographic (GC) separation was carried out on an Agilent 7890B GC system equipped with an Agilent DB-5MS column. A sample volume of 1 µl was injected into a split/splitless inlet at 250°C using splitless mode. The mass spectrometry was carried out on a LECO Pegasus BT system (LECO). ChromaTOF software (version 5.51) was used for data aquisition. The samples were run in 9 batches, each consisting of 100 samples and a calibration curve. To monitor the run a blank, a QC and a standard sample with a known concentration run between every 10 samples. Between every batch the septum and liner on the GC were replaced, the precolumn was cut if necessary and the instrument was tuned.

The retention index was determined with ChromaTOF using the reference method function. The reference file contained the spectras and approximate retention times of the alkanes from C10 to C30 as determined manually. A reference method was implemented for every sample to determine the exact retention time of the alkanes. Untargeted data processing was carried out using MSDIAL (version 4.7). The identification was carried out using retention index with the help of the GCMS DB-Public-kovatsRI-VS3 library provided on the MSDIAL webpage. The results were exported as peak areas and further processed with excel where the results were normalized using heptadecanoic acid as internal standard and the features with a coefficient of variance of less than 30 % in QC samples were selected. Further filtering removed alkanes and duplicate features. The IDs of the features which passed the CV check were further checked using the Golm Metabolome Database.

2.5 Statistical analyses

The analyses were performed using R [32] environment, version 4.2.2, 2022. Libraries factoextra [33], mia [34], vegan [35], boot [36, 37], MOFA2 [38–40] and ggplot2 [41] were used. P-values (two-tailed) smaller than 0.05 were interpreted as statistically significant. While multiple testing, adjusted p-values (false discovery rate FDR) smaller than 0.15 were interpreted as statistically significant.

2.5.1 Milk composition, community composition and fecal metabolome

Beta diversity calculations were based on the Bray-Curtis dissimilarity using relative abundances of the detected bacterial genera. Dissimilarity of milk metabolome was analyzed with Euclidean. PERMANOVA analysis was performed using the adonis2 function with 999 permutations. Both marginal and by terms effects were calculated. To observe the association of overall milk metabolome and to reduce dimensionality, beta diversity Principal components analysis (PCA) was conducted for milk metabolites. The first three principal components (PC1-PC3) were used as independent variables in the PERMANOVA. PCA was conducted using the function prcomp.

Dissimilarity of infant fecal metabolome was studied with Euclidean in a principal component analysis by infant FUT2 status and mother secretor status. Group difference was tested with Adonis (an R implementation of PERMANOVA). Taxonomical diversity (Shannon index) was compared between secretors and non-secretors (two sample t-test) in each timepoint separately.

2.5.2 Differential abundance analysis

Differential abundance analysis (DAA) was conducted using the Wilcoxon signed-rank test between the genus level presence-absence data and milk metabolites. The genera that had at least ten occurrences in both groups (presence/absence) were included in the analyses. P-values were adjusted for multiple testing using the FDR method. Analysis was stratified by timepoints (2.5mo, 6mo and 14mo). Additionally, DAA was conducted with Aldex2 [42] which uses probabilistic modelling and considers the count-compositional nature of the data by including scale modelling by default both unadjusted and adjusted for covariates… The DAA was completed for each timepoint.

Covariates that were used in the models as categorical variables were breastfeeding (partial, full), parity (primipara, multipara), delivery mode (vaginal, section) and maternal education (low, mid, high). Continuous variables were maternal BMI, child's birth month, gestational age and maternal distress. Maternal distress was calculated as the sum of scaled mean EPDS (14 gw, 24, gw, 34 gw, 3 mo, 6 mo and 1y) and scaled mean SCL (14 gw, 24, gw, 34 gw, 3 mo and 6 mo). In cases of missing values, the mean sum score was calculated based on the available measurement points. Age terms refer to the segments of the piecewise linear function used to model the (continuous) child's age at the sampling time. The covariates were selected a priori based on expected causal relationships (Fig. S1). The breakpoint, that is, where the line was allowed to turn (break), was at 6 months. This was selected since Finnish recommendations state that solid foods should be started at 4-6 months [43]. Same covariates were used in mixed model.

2.5.3 Associations between milk and fecal metabolites

The associations between each milk metabolite and each fecal metabolite were examined using Spearman's correlation. Fecal metabolites were log2 transformed with pseudocount of half minimum values. P-values were adjusted for multiple testing using the FDR method. The 95% bias-corrected and accelerated (BCa) bootstrap confidence intervals were calculated for the correlation coefficient (based on 1000 bootstrap samples). Analysis was stratified by timepoints (2.5mo, 6mo and 14mo).

The following linear mixed effects models were used to examine longitudinally the association between fecal metabolite and milk metabolites:

M: fecal metabolites ~ intercept + maternal BMI + birth month + parity+ gestational age + maternal distress + delivery mode + maternal education + age terms x milk metabolites,

where fecal metabolites (Polar metabolites, SCFAs, BAs) and milk metabolites were used in the model one at a time. Interaction between each milk metabolite and child's age at the sampling time was included in the models. Covariates are described above in previous section.

Milk and fecal metabolites were scaled (mean=0, sd=1). P-values were adjusted for multiple testing using the FDR method.

2.5.4 Multi-omics factor analysis

Multi-omics factor analysis (MOFA) was conducted using the MOFA2 function. MOFA discovers the principal sources of variation in multi-omics datasets. Milk and fecal metabolites and genera level microbiome datasets were used in MOFA2 at once. Only the prevalent genera were used (detection=0.01, prevalence=0.1). Genera was examined using the centered log ratio-transformation with pseudo count one. Timepoints were treated as groups to enable comparing the sources of variability that drive each timepoint i.e., group. Fecal metabolites were log2 transformed.

3 Results

In the studied sample, 78% of infants were fully breastfed in the 2-month timepoint. In the 6-month timepoint, 71% were partially breastfed (Fig. 1A). In the 14-month-old, 14% were receiving breast milk. The overlap of milk and stool samples got lower by age due to decreasing breastfeeding (Fig. 1B). Approximately, 84% of mothers were secretors in the whole cohort, which is slightly higher than previously observed in other cohorts globally (75-80%) [44, 45]. Infant FUT2 genotype mostly matched mother's secretor type (Fig. 1C). The background characteristics of study population have been listed in each timepoint (Table S2.)

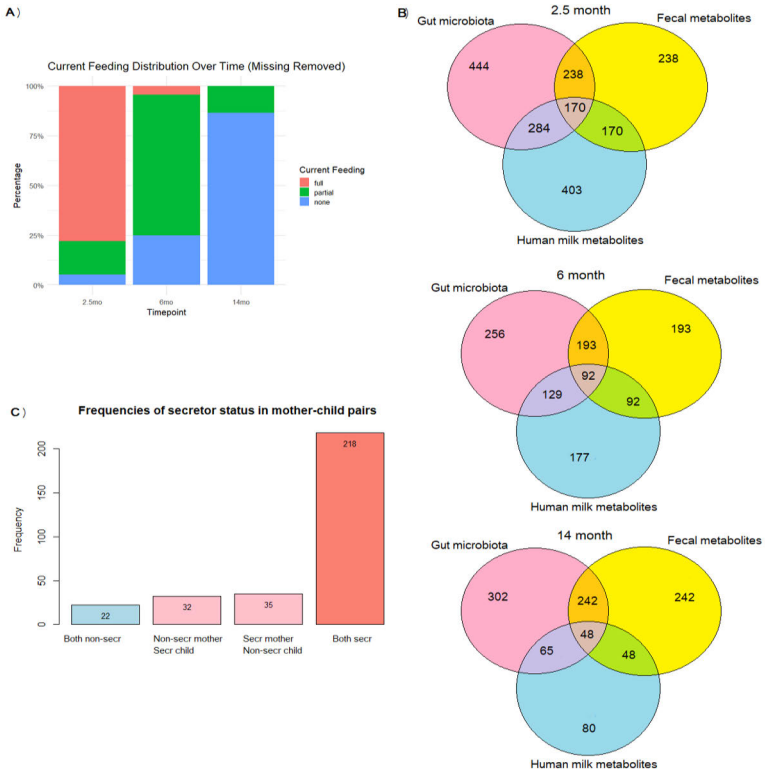


Figure 1. Overview of study samples. A) Breast feeding status in three time points. B) Overlap of stool and milk samples in multiomic setting. C) Frequencies of secretor status of mother-child dyads. 71% of dyads were both secretors.

Diversity and dissimilarities in gut microbiome and milk metabolome

The gut microbiome got more uniform by time (Fig. 2A) and milk metabolites clustered according to mothers' secretor status (Fig. 2B). Although, the milk composition was clustered separately between secretors and non-secretors, the average amount of total HMOs was similar between the groups. Total HMO concentration differed only in 2-month timepoint (mean concentration Secr. 16.9, Non-secr. 14.9, p=0.0002) by secretor status. HMO-profiles (Fig. S3) showed visually that non-secretors had more 3-FL while they were lacking 2'-FL, LNFP I, LDFT, and LNDFH I.

The milk composition did not associate the differences observed in the gut microbiome beta diversity in the infants, except PC1 of milk metabolome in 6 months explained a part of microbiome beta

diversity (PERMANOVA, R^2 = 0.02 p=0.02, Table S3). The taxonomic diversity (Shannon index) was not associated with either the secretor status of mother or the infant (t-test, p>0.4 in all timepoints) or the milk metabolome (Table S4). This indicates that the milk metabolome was not related to gut microbiota in the overall community composition level. Additionally, we examined whether fecal metabolome is driven by secretor status. Although fecal metabolome showed visually time-dependence (Fig. S4), no grouping of fecal metabolome by secretor status of mother or child were found (adonis p>0.05, Table S5).

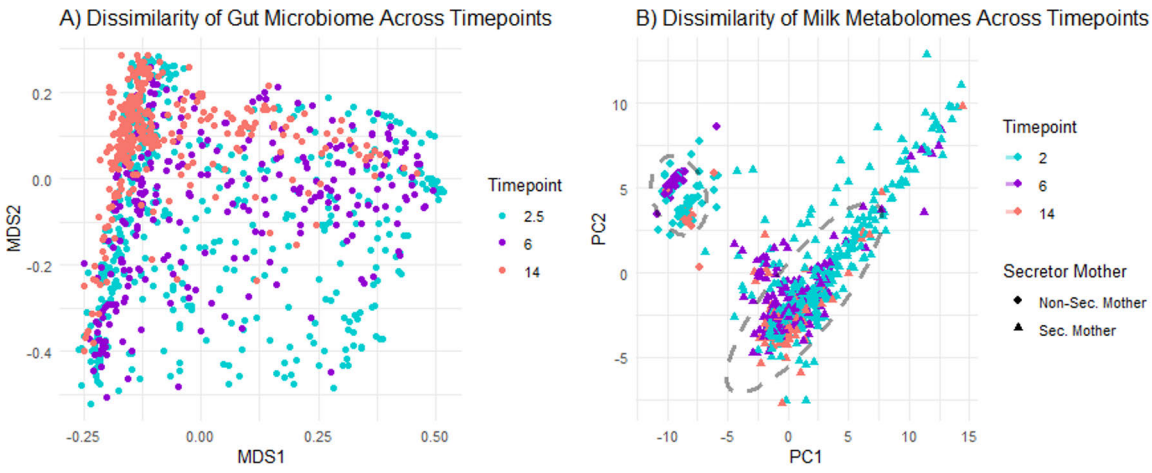


Figure 2. Dissimilarities. A) Infant gut microbiome beta diversity by Bray Curtis dissimilarity across three timepoints with multidimensional scaling. Microbiome got more uniform by time. B) Milk metabolomes by Euclidean and principal components by secretor status across timepoints. Non-secretors shaped as square and secretors as triangle. Grey ellipses represent the two clusters determined by secretor status.

Milk metabolites and microbes in genus level correlate

In presence/absence analysis of microbes, there were many nominal significant differences in milk metabolite concentrations, but none withstood FDR correction (Table S6-S11). With ALDEx2, only two metabolite-microbe links remained after FDR correction; negative correlations between *Bifidobacterium* and ethanolamine/methionine. Moreover, LNFP I was associated with *Bifidobacterium* in 6-month timepoint (Fig. 3). Unidentified genus of *Enterobacteriaceae* associated positively with 3-SL in 2-month timepoint, while 3-FL associated negatively with Bacteroides. Overall, there were more associations in later timepoints showing more diverse microbiome composition. Glutamine had positive correlation with butyrate-producers *Faecalibacterium* and *Roseburia* [46] *in* 14-month timepoint, while 2'-FL correlated negatively with *Ruminococcus, Clostridioides and Citrobacter*. Sn-glycero-3-phosphocholine, a precursor in the synthesis of acetylcholine (a neurotransmitter), was associated in all three timepoints but with different genera. In the last timepoint, it was positively associated with *Bifidobacterium* and *Lactobacillus*.

When adjusting for covariates using ALDEx2 with a linear model matrix, the associations between milk metabolites and microbial genera remained partially consistent (Fig. S5–S7). Notably, at 14 months, fewer genera showed significant correlations after adjustment, yet the effect estimates were stronger. For example, 2'-FL and 3-FL were negatively associated with *Ruminococcus*, while

several milk energy metabolites showed positive associations with *Ruminococcus*. Interestingly, *Bifidobacterium* did not appear in the associations at either 6 or 14 months in the adjusted analysis.

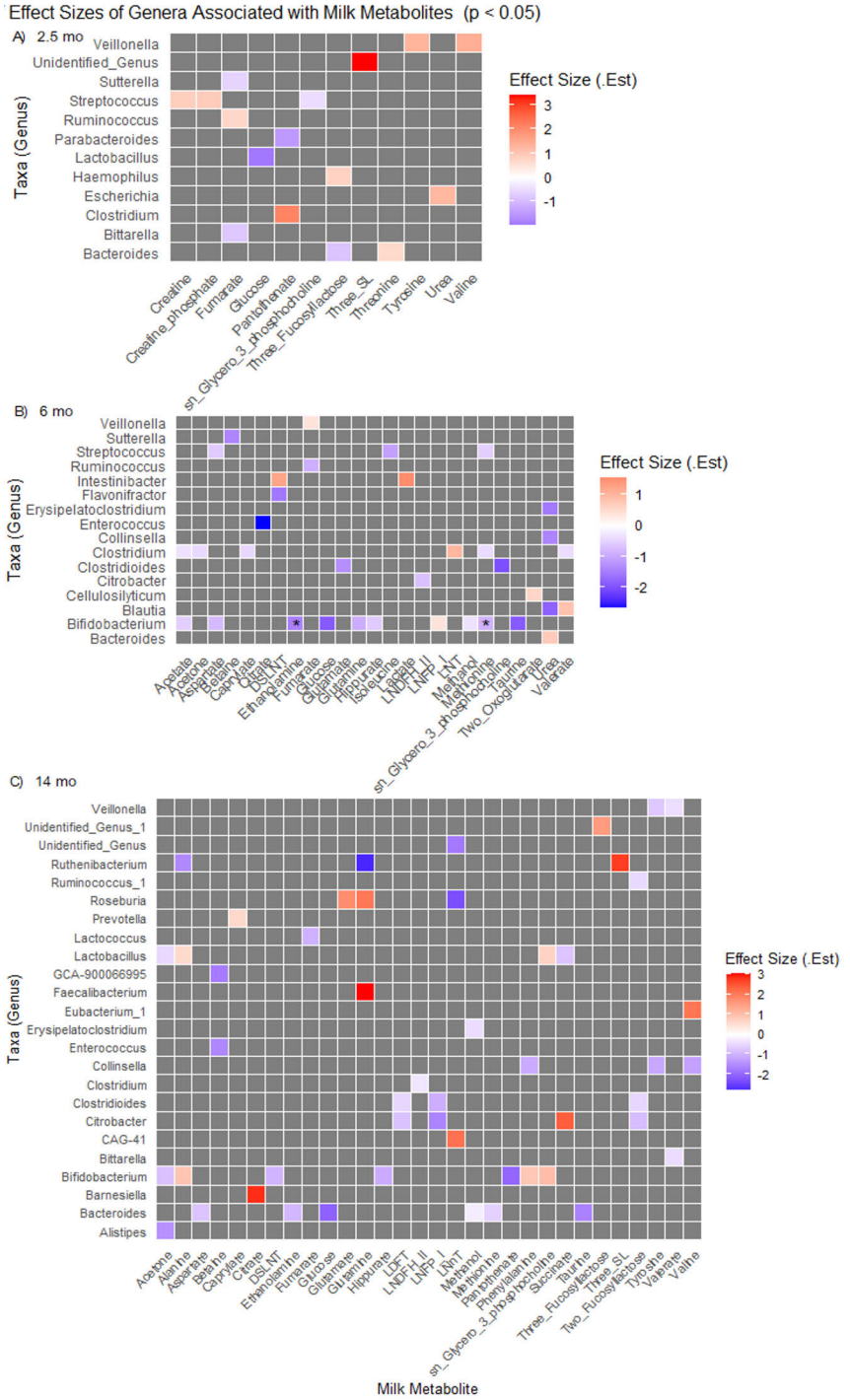


Figure 3. Differential abundances by ALDEx2 with milk metabolites across timepoints. All associations with unadjusted p-value <0.05 are presented here. Colors indicate direction and magnitude of estimates. Asterisks indicate adjusted p-values <0.05. A) estimates in 2.5 month timepoint. Unidentified_Genus is an unknown genus of Enterobacteriaceae B) estimates in 6 month timepoint C) estimates in 14 month timepoint.

3.3 Milk and fecal metabolites

Several associations (Fig. 4) were found between milk metabolites and fecal SCFA, however, the results did not withstand p-value correction. LDFT, which was missing with non-secretors, correlated positively with several fecal SCFAs, especially in 2.5- and 14-month timepoint. Milk amino acids had negative correlations to fecal acetic acid in 6-month timepoint. Milk amino acids associated with branched-chain fatty acids (Isovaleric and Isobutyric acid) in 2- and 6-month timepoints.

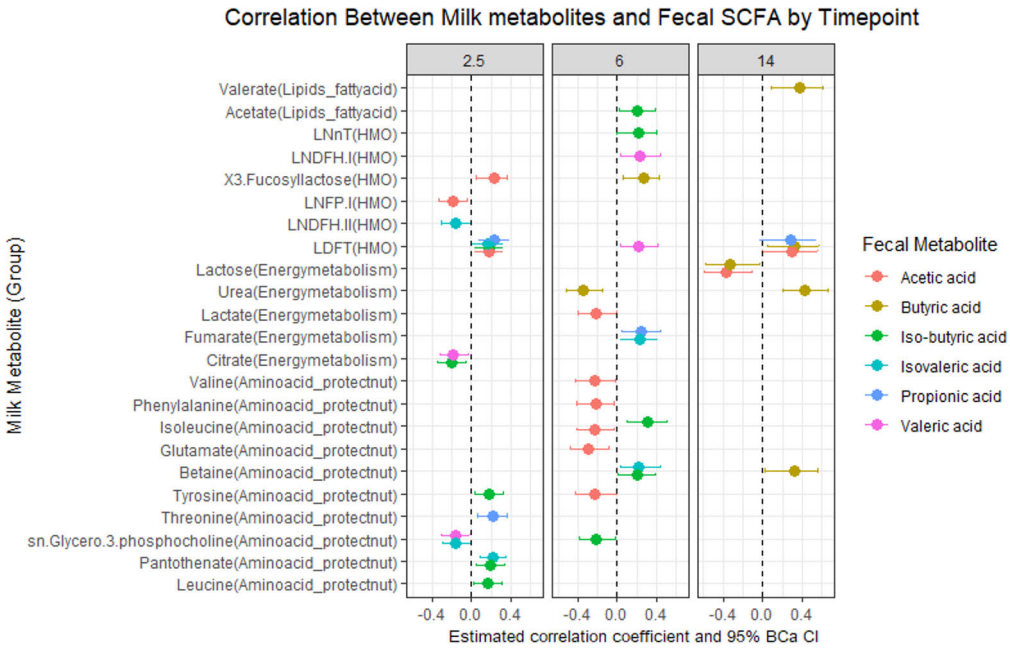


Figure 4. Milk metabolites associated (Spearman) with fecal short chain fatty acids (SCFAs, see colors). These correlations have unadjusted p-values <0.05. The results did not withstand p-value correction (FDR). The 95% bias-corrected and accelerated (BCa) bootstrap confidence intervals were calculated for the correlation coefficient (based on 1000 bootstrap samples). Analysis was stratified by timepoints (2.5mo, 6mo and 14mo in gray panel).

Likewise, bile acids (BAs) were associated with several milk metabolites, but the results did not withstand p-value correction (Fig. 5). In 2.5-month timepoint, milk metabolites had positive associations with primary BAs, while 7-oxo-converted and secondary BAs were negatively associated. Interestingly, 2'-FL and 3-FL had opposite direction in BAs correlation.

In 6-month timepoint, correlations were mainly in glycoconjugated BAs which were not seen in other timepoints. In 14-month timepoint, secondary BAs had several positive correlations. HMOs shifted the directions of correlation from 2.5month to 6month, i.e. 2'-FL and LNFP associated positively with tauroconjugated BAs in 2.5 months and negatively at 6 months, whereas an opposite pattern was observed for 3-FL.

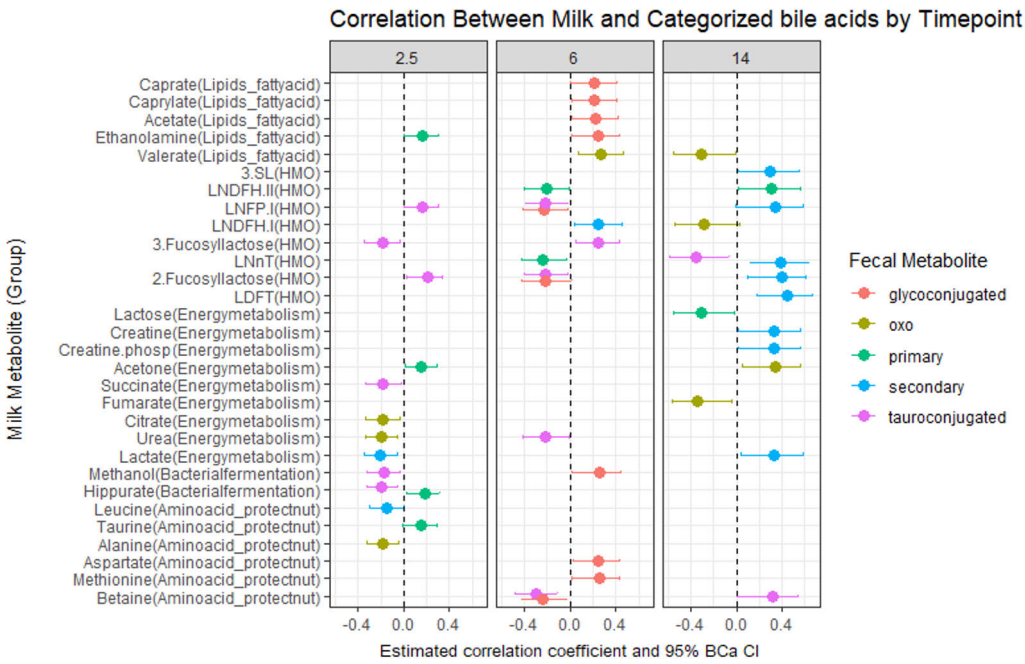


Figure 5. Milk metabolites associated (Spearman) with fecal categorized bile acids (see colors). These correlations have unadjusted p-values <0.05. The results did not withstand p-value correction (FDR). The 95% bias-corrected and accelerated (BCa) bootstrap confidence intervals were calculated for the correlation coefficient (based on 1000 bootstrap samples). Analysis was stratified by timepoints (2.5mo, 6mo and 14mo in gray panel).

There were high number of correlations (>700, p<0.05) between milk metabolites and fecal polar metabolites (Table S12). For instance, in 2.5 months milk caprate and caprylate, antimicrobial medium-chain fatty acids [47], correlated negatively with multiple fecal metabolites such as p-hydroxyphenyl lactic acid. Overall, associations were mainly negative and those were between milk fatty acids and fecal microbial and carbohydrate metabolism. However, in 2.5 months, when the most associations occurred, there were also positive associations between milk fatty acids (valerate, caprylate) and fecal polar metabolites, possibly originating from energy metabolism and xenobiotic sources.

3.5 Mixed model

In mixed model (adjusted for covariates mentioned in 2.5.3) we saw multiple correlations between milk and gut metabolites by continuous age model adjusted for 2 time-intervals by sampling age (Fig.7, Table S13). From fecal polar metabolites, long-chain carbohydrates fucose and ribonic acid had associations with milk 3-SL and 6-SL (Fig. 7C-D). Betaine and urea, which were already noted in DAA and Spearman, were associated with butyric acid in the later time interval (Fig. 7G-H). LDFT, which associated with multiple SCFAs with Spearman, correlated with 3,4-dihydroxyhydrocinnamic acid (likely xenobiotic) in both time intervals. (Fig. 7E). LNFP I and caprylate correlated positively with secondary bile acids in the later time-interval (Fig. 7A-B). Microbiota-derived 5-hydroxyindoleacetic acid, which has seen to alleviate diarrhea [48], associated positively with milk sn-glycero-3-phosphocholine in the first time-interval. Nearly all correlations shifted direction after the first time-interval.

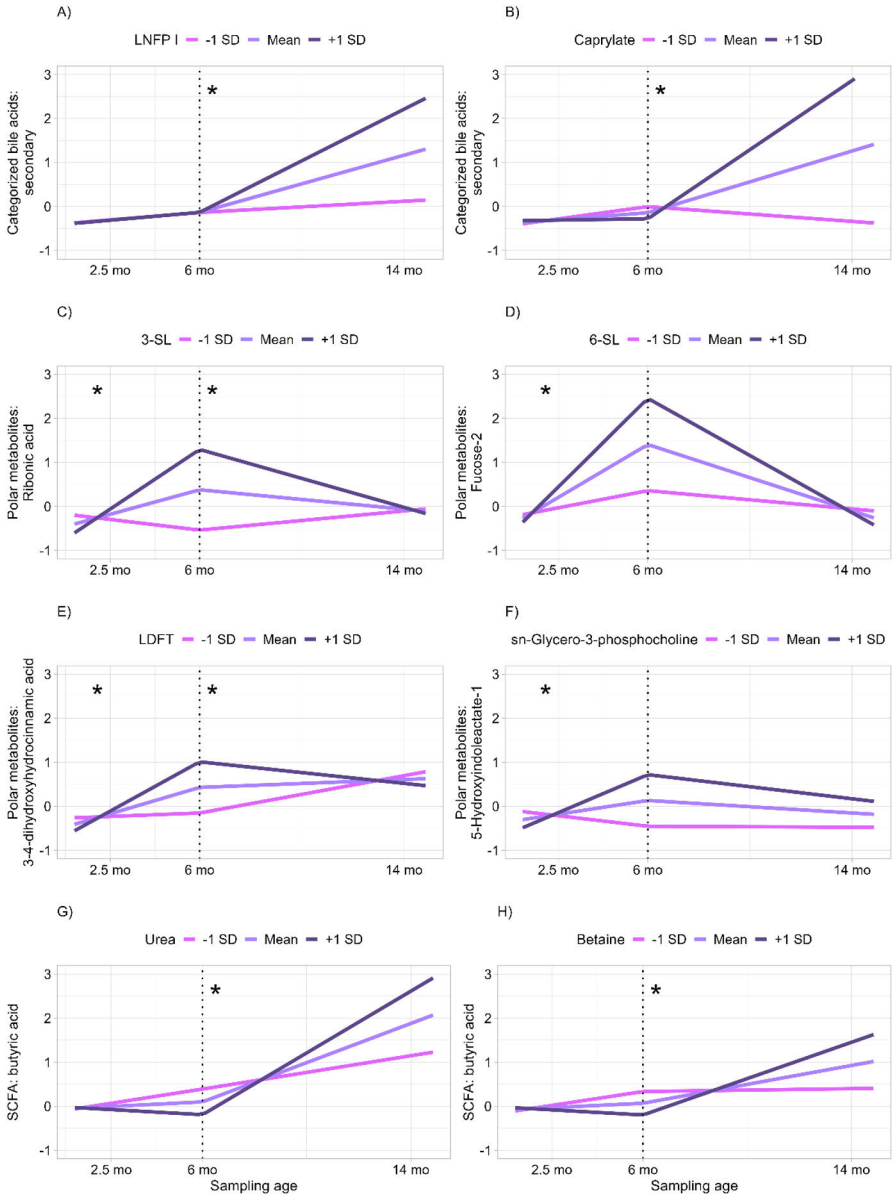


Figure 6. Mixed model by two time-intervals by sampling age. Correlation between milk metabolite (on top of the plots) and fecal metabolites (y-axis) presented with mean values and standard deviations. Asterisks (*) indicate adjusted p-values <0.05.

3.6 Multi-omic factor analysis

Multi-omic factor analysis (MOFA, Fig. 7A.) with all omic-data sets was set out to see how much variance milk, fecal metabolites and gut bacterial genera explain in the data (Table S14) and to reduce features. Eight factors explained the variance and factors were loaded by different features (Factors 1-4 in Fig. 7B, all factors in Fig. S8). Overall, milk metabolites and stool-based omics explained variance

in different latent factors. With milk metabolites, variance was mainly explained by factors 1 (R^2 21.4), 3 (R^2 23.3) and 7 (R^2 10.6) in 2-month timepoint and those remained largest throughout the timepoints. Those factors were loaded by milk energy metabolites, HMOs and fecal propionic acid.

In bile acid data set, variance was explained by latent factor 2 (R^2 20.9) and to some extent by 4 (R^2 5.2) and 5 (R^2 7.0) in 2-month timepoint and those factor values fluctuated by time. Other assays explained variance to lesser degree. Factor 1 was the only shared factor between milk and fecal metabolites, and it explained minor part of SCFAs (2mo: R^2 1.5; 6mo: R^2 1.7).

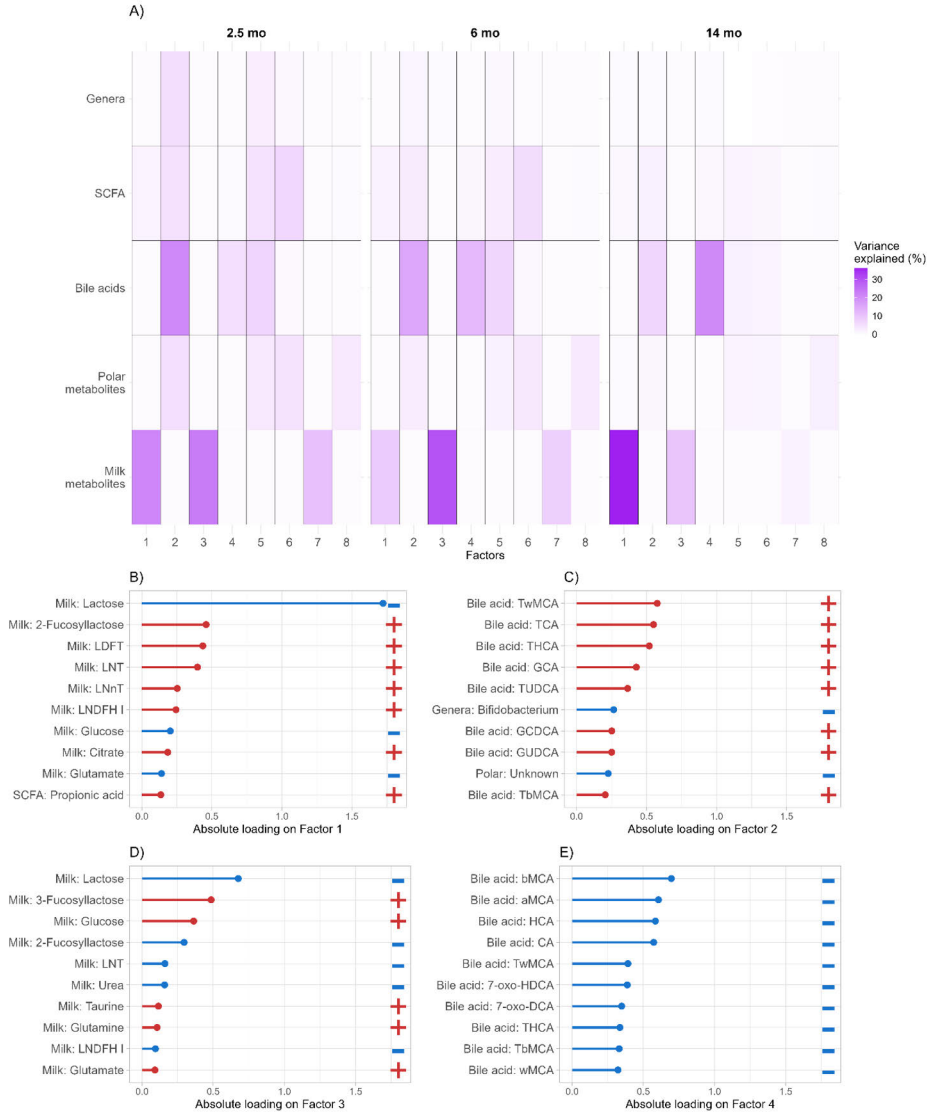


Figure 7. A) Multi-omic factor analysis (MOFA) with five data sets and 8 factors in three timepoints. B)-E) Loadings of factors 1, 2, 3 and 4. Top 4 factors and their top 10 features are presented here. Colors (red/blue) and +/- indicate the direction of the loadings.

Additionally, individual factor points were compared with breastfeeding (BF) variables (Fig. S9).

Overall, categories of current breastfeeding associated with factor 1 and factors 5 to 8 mainly at 2.5 months, while secretor status (mother) associated with factors 1, 3 and 5 (Fig. S9). Specifically, factors 5 and 6 were higher with full-breastfed in 2.5mo timepoint (Fig. S9). Factors 1 and 5 were elevated in secretors at 2.5 and 6 months, respectively, while factor 3 was lower in 2.5 and 6 months (Fig. S9). Although both maternal secretor status and breastfeeding related to factor 5, these were at different timepoints.

Factors 1 and 3 were loaded mainly by HMOs and negatively by lactose (Fig. S8). Factor 1 was positively loaded with fucosylated HMOs, while factor 3 was positively loaded by 3-FL, milk glucose and milk amino acids and negatively by fucosylated HMOs (Fig S8). Factor 5 was negatively loaded by conjugated bile acid (turine), fecal butyrate and positively with Bifidobacterium. Factor 6 was positively loaded with fecal sugars and negatively loaded with propionic acid (Fig S8).

4 Discussion

This study was set out to investigate the association between human milk metabolites and infant gut metabolites and microbiota. It is well known that human milk fosters infant healthy gut microbiome which produce beneficial microbial metabolites. However, there is a lack of longitudinal research that simultaneously considers human milk composition, gut microbiota and fecal metabolite profiles

As expected, the milk HMO composition was influenced by mother's secretor status, primarily due to the absence of specific fucosylated HMOs such as 2'-fucosyllactose (2'-FL), lacto-N-difucohexaose I (LNDFH I), and lacto-difucotetraose (LDFT). In addition, non-secretor mothers had higher concentration of 3-FL. In the mammary gland, FUT2 encodes an α1,2-fucosyltransferase for the synthesis of 2-FL, while FUT3 gene is responsible for 3-FL production [49]. 3-FL has been also found to have prebiotic property, immunomodulatory effect, antiadhesive antimicrobials, antiviral ability, and gastrointestinal protection [50]. However, 2-FL and 3-FL had opposing associations with fecal metabolites, such as secondary BAs. Although 2-FL and 3-FL have been found to have similar properties, our results suggest that they also might differ in their relation to gut microbial metabolism. Hence, it can be speculated that the compensation of higher 3-FL may not address the deficiency of 2'-FL in the milk of non-secretor mothers in relation to gut metabolism.

In our results, neither maternal or infant secretor status (FUT2) were associated with infant gut microbiota alpha and beta diversities. Although they have been considered important for gut health [45, 51, 52], our results align with recent studies indicating modest associations at best [19, 53, 54]. Moreover, in our results, the associations between the overall milk composition and infant gut microbiome diversity and community composition were mostly non-existent. We observed only association between milk first principal component of the milk and gut microbiota beta diversity at 6 months. Previous study showed that the proportion of human milk in the infant diet at 6 months was the prominent determinant of infant gut microbiota diversity [55]. Another study showed energy content in milk negatively correlates with alpha diversity [56]. Although beta diversity has been rarely examined, one animal study reported that mother secretor status differentiate gut microbial beta diversity in one dimension but HMO supplementation did not. [57].

However, individual genera abundances were related to human milk composition. Fucosylated pentasaccharide LNFP-I was positively correlated with *Bifidobacterium* at 6 months and negatively with *Clostridioides* and *Citrobacter* at 14 months. LNFP-I is known to be consumed by *Bifidobacteria*, and can protect from enteropathogens [58] Additionally, in our study LNFP I exhibited a positive correlation with secondary BAs in later time interval, possibly indicating gut maturation [16]. Further, sialylated DSLNT in 2 months associated positively with *Veillonella* and negatively with *Escherichia* and further in 6 months with *Hungatella* and *Citrobacter*. Escherichia may also include possible pathogenic species, and in a rat model of necrotizing enterocolitis, DSLNT was found to provide protection [59]. Interestingly, 2'-FL did not associate with *Bifidobacterium* but fucosylated HMOs LDFT, LNFP-I and 2-FL associated negatively with *Clostridioides*, *Citrobacter* and *Ruminococcus* at 14 months. Hence HMOs appeared to limit potentially opportunistic pathogenic [60–62]. Although the links between Bifidobacteria and HMOs are well-documented [7, 63], previous study showed that also other bacteria have the capacity to metabolize HMOs and the by-products of HMO metabolism [7, 64]. This may underline our observation that HMOs were also related to other bacteria than *Bifidobacterium*, including members of *Bacillota*.

In addition to associations with bacterial genera, HMOs associated with bile acids, sugars and a xenobiotic metabolite. 3-SL and 6-SL, both containing sialic acid bound with lactose, associated with higher fecal ribonic acid and fucose concentrations, respectively. The associations in 6-SL and fucose was apparent before weaning and 3-SL and ribonic acid associated also after weaning. Of note, 3-SL was also positively associated with unidentified genera in *Enterobactreriaceae* at 2.5 months. Previous research has demonstrated that sialylated HMO supplementation can increase transcription of genes related to monosaccharide and carbohydrate metabolism in *E. coli* and *Bacteroides fragilis* [65]. In this light, our finding can reflect that increased simple sialylated HMOs can result in higher abundance of members of Enterobacteriaceae and more extensive saccharolytic metabolism and higher concentration of carbohydrate derivates in stool.

Additionally, milk sn-glycero-3-phosphocholine (GPC), a breakdown product of phosphatidylcholine and a bioavailable choline source, was negatively associated with *Streptococcus* in 2-month-olds and *Clostridioides* in 6-month-olds and positively with *Bifidobacterium* and *Lactobacillus* at 14 months. GPC may shape the infant gut microbiota since some bacteria possess the necessary enzymes to break down GPC into its constituent parts, which can then be further metabolized for growth [66]. However, the effects on host health are uncertain since certain bacteria can utilize unabsorbed choline (e.g. some Clostridium and Escherichia strains) potentially to produce trimethylamine (TMA) which may further end up in host liver as TMAO. [67, 68]

Moreover, in our cohort the higher milk amino acid levels were linked to higher amount of branched short-chain acid (BCFA) isobutyric acid in 2-month olds, but not in later timepoints. Previous studies have linked higher protein intake and amino acid metabolism to increased BCFA concentrations [69, 70]. Although amino acids are typically efficient at absorbing amino acids in small intestine, it may be that some escape to colon for bacterial metabolism if they are abundant in the diet. Interestingly, while isobutyric acid is metabolized from valine and isovaleric acid is produced from leucine [71], we observed positive correlations between isobutyric acid and leucine and tyrosine at 2.5 months as well as isoleucine at 6 months. In addition, milk amino acids, including valine, phenylalanine, isoleucine, tyrosine and glutamate, were negatively correlated fecal acetic acid at 6 months. BCFA are less studied

in comparison to traditional SCFA, and whether our observation related to more complex bacterial cross-feeding in the intestines, remains an open question.

Urea and betaine were linked to higher butyric acid especially after weaning. Urea is a major source of nitrogen in the human milk, which may affect the gut homeostasis and bacterial metabolism. While the human host does not encode urease, which is an enzyme responsible for urea hydrolyzation. However, multiple microbes encode this enzyme, including *Bifidobacteria* and *Lacnospiraceae* [72]. Interestingly, the commensal bacteria with urease activity can use urea to produce SCFA, including butyrate [73]. Human milk betaine, primarily originating from dietary sources, has been associated with normal growth patterns in healthy infants as opposed to accelerated growth [74]. In addition, the betaine supplementation in early life in rodents and higher milk betaine in humans was related to increased *Akkermansia* abundance. Although *Akkermansia* had too low prevalence in our sample, it is known that *Akkermansia* is a SCFA producer [75]. Previous study has associated higher betaine levels in milk with lower adiposity and improved glucose homeostasis in adulthood, as observed in a mouse model [74]. These results highlight that also other human milk components than HMOs, such as urea and betaine, are associated with key microbial metabolites, including butyrate, and this link between human milk and gut microbiome is likely important for infant health [74].

We observed that milk lipids such as caprylate associates positively with glycoconjugated and secondary bile acids, especially after weaning. Caprylate is a medium-chain fatty acid that typically increase during lactation [76] that is absorbed from the small intestines and serves as an energy source for the infant. Often, lipid-containing diet increases the bile production [77, 78]. Host produces bile acids and conjugates them with taurine or glycine, and the conjugated bile acids are deconjugated by bacteria and further metabolized to secondary BAs. The observed associations may reflect increased bile production subsequent microbial metabolism of BAs in response to higher lipid content in the milk. Bile acids are important physiological modulators that often undergo microbial metabolism [16], and our results suggest that milk lipids may be one factor affecting the bile acid enterohepatic circulation in infancy.

Our multi-omic analyses reflected that the data from different samples related to separate multi-omic factors, i.e. fecal data explained data-driven variance different from milk metabolome. This aligns with the observation that overall milk metabolome composition was not related to gut microbiota or metabolome overall composition. On the other hand, the factor loadings were related to breastfeeding and secretor status of the mother. Full-breastfed infants in 2.5 months, compared to partial feeding, had higher score in factors loaded by positively 7-oxo-HDCA, wMCA, butyric acid and Bifidobacterium, which can be considered indications of healthy gut microbiome maturation. On the other hand, there were negative loadings in other BAs, such as tauro- and glycine conjugates. This supports our earlier findings when we demonstrated that Bifidobacterium abundances were associated negatively with tauroconjugated BA concentration only in breastfed infants. Our results also corroborate earlier observation that 7-oxo-converted BAs correlated positively with Bacteroides and Escherichia in breastfed babies [29].

Limitations and strengths

All milk samples were taken in a repeatably manner during study visits during daytime. Limitations of milk data concern the milk sampling (foremilk or hindmilk) and timing which may affect the milk composition. For that reason and the nature of lactation, collected milk samples represents a snap

shot of the milk. Moreover, metabolomics was performed using different analytical platforms — NMR for milk and GC- and LC-MS methods for fecal metabolites. While this approach may limit direct comparability of results, it also enables complementary coverage of distinct metabolite classes and enhances the overall breadth of metabolic profiling suitable for the sample type. Strength of this study is the unique longitudinal sample collection with multiomic approach. However, a part of data suffered from high number of features which may explain high adjusted p-values. However, we observed similar relationships between different analysis methods, including cross-sectional correlations, DAA, multi-omic analyses and mixed models.

5 Conclusions

Our study uncovers dynamic, time-dependent relationships between human milk metabolites, the infant gut microbiome, and fecal metabolites. These associations varied between early and late infancy, suggesting that the influence of milk composition on microbial metabolism shifts with dietary diversification and microbiome maturation. While the overall milk composition was not related to large difference in gut microbiota or metabolome overall, our results highlight the relation between milk HMOs, amino acids, urea, lipids and key fecal metabolites such as butyrate and bile acids. Together, these insights advance our understanding of how human milk relates to the developing gut microbial metabolism and lay the groundwork for strategies to support optimal early-life health.

Ethical evaluation

FinnBrain has a permit from VSSHP ethical committee (ETMK: 57/180/2011), which has approved Cohort profile and research protocol[23]. FinnBrain parents have signed a consent form about their children's participation in research and given permission to use their samples for scientific purposes. Samples went through the laboratory process anonymously with research code to protect participants' privacy. Raw metabolite and microbiome data were shared without personal information and research codes were changed to running numbers.

Conflict of interest/Disclosure statement

HMK is senior researcher at IFF. EM has previously worked at Biocodex. Other authors have no conflict of interest.

Author Contributions

Conception and design of the work: HI, AKA, US, SL and AD. Acquisition, analysis, or interpretation of data: US, LK, HK, HMK, HI, TK, LP, MO, AD, AKA and SL. Drafting and substantial revision of the work: AKA, SL, HI, AD and LP. All authors have approved the submitted manuscript.

Acknowledgements

We acknowledge the staff of FinnBrain project and research group of Alex Dickens, Matej Orešič and Antti Hakanen for the support of the accomplishment of this study. Lastly, heartfelt thanks to FinnBrain families for their active participance in the study.

Funding

We acknowledge several funders for making it possible to carry out this study. HI had grants from Finnish cultural foundation [no 00230482], Juho Vainio's foundation and Doctoral school of clinical

research of university of Turku. FinnBrain project has been funded by Research council of Finland, Finnish State Grants for Clinical Research (ERVA) and Signe and Ane Gyllenberg Foundation.

Data availability statement and deposition

Due to Finnish national legislation and study participant rights, the individual-level data cannot be made available online, but data can potentially be shared with Material Transfer Agreement as part of research collaboration. Requests for collaboration can be sent to the Board of the FinnBrain Birth Cohort Study; please contact Linnea Karlsson (linnea.karlsson@utu.fi).

References

1. Suárez-Martínez C, Santaella-Pascual M, Yagüe-Guirao G, Martínez-Graciá C. Infant gut microbiota colonization: influence of prenatal and postnatal factors, focusing on diet. Front Microbiol. 2023;14. https://doi.org/10.3389/fmicb.2023.1236254.

2. Dominguez-Bello MG, Godoy-Vitorino F, Knight R, Blaser MJ. Role of the microbiome in human development. Gut. 2019;68:1108–14. https://doi.org/10.1136/gutjnl-2018-317503.

3. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature. 2018;562:583–8. https://doi.org/10.1038/s41586-018-0617-x.

4. Azad MB, Konya T, Maughan H, Guttman DS, Field CJ, Chari RS, et al. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. CMAJ. 2013;185:385–94. https://doi.org/10.1503/cmaj.121189.

5. Ames SR, Lotoski LC, Azad MB. Comparing early life nutritional sources and human milk feeding practices: personalized and dynamic nutrition supports infant gut microbiome development and immune system maturation. Gut Microbes. 2023;15:2190305. https://doi.org/10.1080/19490976.2023.2190305.

6. Brink LR, Mercer KE, Piccolo BD, Chintapalli SV, Elolimy A, Bowlin AK, et al. Neonatal diet alters fecal microbiota and metabolome profiles at different ages in infants fed breast milk or formula. The American Journal of Clinical Nutrition. 2020;111:1190–202. https://doi.org/10.1093/ajcn/nqaa076.

7. Lordan C, Roche AK, Delsing D, Nauta A, Groeneveld A, MacSharry J, et al. Linking human milk oligosaccharide metabolism and early life gut microbiota: bifidobacteria and beyond. Microbiology and Molecular Biology Reviews. 2024;88:e00094-23. https://doi.org/10.1128/mmbr.00094-23.

8. Yelverton CA, Killeen SL, Feehily C, Moore RL, Callaghan SL, Geraghty AA, et al. Maternal breastfeeding is associated with offspring microbiome diversity; a secondary analysis of the MicrobeMom randomized control trial. Front Microbiol. 2023;14. https://doi.org/10.3389/fmicb.2023.1154114.

9. Porro M, Kundrotaite E, Mellor DD, Munialo CD. A narrative review of the functional components of human breast milk and their potential to modulate the gut microbiome, the consideration of maternal and child characteristics, and confounders of breastfeeding, and their impact on risk of obesity later in life. Nutrition Reviews. 2023;81:597–609. https://doi.org/10.1093/nutrit/nuac072.

10. Jokela R, Ponsero AJ, Dikareva E, Wei X, Kolho K-L, Korpela K, et al. Sources of gut microbiota variation in a large longitudinal Finnish infant cohort. eBioMedicine. 2023;94:104695. https://doi.org/10.1016/j.ebiom.2023.104695.

11. Hickman B, Salonen A, Ponsero AJ, Jokela R, Kolho K-L, de Vos WM, et al. Gut microbiota wellbeing index predicts overall health in a cohort of 1000 infants. Nat Commun. 2024;15:8323. https://doi.org/10.1038/s41467-024-52561-6.

12. Matharu D, Ponsero AJ, Lengyel M, Meszaros-Matwiejuk A, Kolho K-L, de Vos WM, et al. Human milk oligosaccharide composition is affected by season and parity and associates with infant gut microbiota in a birth mode dependent manner in a Finnish birth cohort. EBioMedicine. 2024;104:105182. https://doi.org/10.1016/j.ebiom.2024.105182.

13. Roager HM, Stanton C, Hall LJ. Microbial metabolites as modulators of the infant gut microbiome and host-microbial interactions in early life. Gut Microbes. 2023;15:2192151. https://doi.org/10.1080/19490976.2023.2192151.

14. Zhang Y, Chen R, Zhang D, Qi S, Liu Y. Metabolite interactions between host and microbiota during health and disease: Which feeds the other? Biomedicine & Pharmacotherapy. 2023;160:114295. https://doi.org/10.1016/j.biopha.2023.114295.

15. Liu J, Tan Y, Cheng H, Zhang D, Feng W, Peng C. Functions of Gut Microbiota Metabolites, Current Status and Future Perspectives. Aging Dis. 2022;13:1106–26. https://doi.org/10.14336/AD.2022.0104.

16. van Best N, Rolle-Kampczyk U, Schaap FG, Basic M, Olde Damink SWM, Bleich A, et al. Bile acids drive the newborn's gut microbiota maturation. Nat Commun. 2020;11:3692. https://doi.org/10.1038/s41467-020-17183-8.

17. Hoen AG, Coker MO, Madan JC, Pathmasiri W, McRitchie S, Dade EF, et al. Association of Cesarean Delivery and Formula Supplementation with the Stool Metabolome of 6-Week-Old Infants. Metabolites. 2021;11:702. https://doi.org/10.3390/metabo11100702.

18. Holzhausen EA, Shen N, Chalifour B, Tran V, Li Z, Sarnat JA, et al. Longitudinal profiles of the fecal metabolome during the first 2 years of life. Sci Rep. 2023;13:1886. https://doi.org/10.1038/s41598-023-28862-z.

19. Laursen MF. Gut Microbiota Development: Influence of Diet from Infancy to Toddlerhood. Ann Nutr Metab. 2021;:1–14. https://doi.org/10.1159/000517912.

20. Karlsson L, Tolvanen M, Scheinin NM, Uusitupa H-M, Korja R, Ekholm E, et al. Cohort Profile: The FinnBrain Birth Cohort Study (FinnBrain). International Journal of Epidemiology. 2018;47:15–16j. https://doi.org/10.1093/ije/dyx173.

21. Sundekilde UK, Downey E, O'Mahony JA, O'Shea C-A, Ryan CA, Kelly AL, et al. The Effect of Gestational and Lactational Age on the Human Milk Metabolome. Nutrients. 2016;8:304. https://doi.org/10.3390/nu8050304.

22. Dror DK, Allen LH. Overview of Nutrients in Human Milk. Advances in Nutrition. 2018;9:278S-294S. https://doi.org/10.1093/advances/nmy022.

23. Korhonen LS, Lukkarinen M, Kantojärvi K, Räty P, Karlsson H, Paunio T, et al. Interactions of genetic variants and prenatal stress in relation to the risk for recurrent respiratory infections in children. Sci Rep. 2021;11:7589. https://doi.org/10.1038/s41598-021-87211-0.

24. Rintala A, Pietilä S, Munukka E, Eerola E, Pursiheimo J-P, Laiho 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;28:19–30. https://doi.org/10.7171/jbt.17-2801-003.

25. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3. https://doi.org/10.1038/nmeth.3869.

26. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41 Database issue:D590-596. https://doi.org/10.1093/nar/gks1219.

27. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Research. 2014;42:D643–8. https://doi.org/10.1093/nar/gkt1209.

28. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Applied and Environmental Microbiology. 2007;73:5261–7. https://doi.org/10.1128/AEM.00062-07.

29. Aatsinki A-K, Lamichhane S, Isokääntä H, Sen P, Kråkström M, Alves MA, et al. Dynamics of Gut Metabolome and Microbiome Maturation during Early Life. 2023;:2023.05.29.23290441. https://doi.org/10.1101/2023.05.29.23290441.

30. Trimigno A, Khakimov B, Mejia JLC, Mikkelsen MS, Kristensen M, Jespersen BM, et al. Identification of weak and gender specific effects in a short 3 weeks intervention study using barley and oat mixed linkage β-glucan dietary supplements: a human fecal metabolome study by GC-MS. Metabolomics. 2017;13:108. https://doi.org/10.1007/s11306-017-1247-2.

31. Lamichhane S, Sen P, Dickens AM, Orešič M, Bertram HC. Gut metabolome meets microbiome: A methodological perspective to understand the relationship between host and microbe. Methods. 2018;149:3–12. https://doi.org/10.1016/j.ymeth.2018.04.029.

32. R: The R Project for Statistical Computing. https://www.r-project.org/. Accessed 14 Oct 2024.

33. Kassambara A, Mundt F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. 2020.

34. microbiome/mia. 2024.

35. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. vegan: Community Ecology Package. 2001;:2.6-8. https://doi.org/10.32614/CRAN.package.vegan.

36. S) AC (author of original code for, R BR (conversion to, maintainer 1999--2022, support) author of parallel, fixes) ARB (minor bug. boot: Bootstrap Functions (Originally by Angelo Canty for S). 2024.

37. Davison AC, Hinkley DV. Bootstrap Methods and their Application. Cambridge: Cambridge University Press; 1997. https://doi.org/10.1017/CBO9780511802843.

38. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018;14:e8124. https://doi.org/10.15252/msb.20178124.

39. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol. 2020;21:111. https://doi.org/10.1186/s13059-020-02015-1.

40. Velten B, Braunger JM, Argelaguet R, Arnol D, Wirbel J, Bredikhin D, et al. Identifying temporal and spatial patterns of variation from multimodal data using MEFISTO. Nat Methods. 2022;19:179–86. https://doi.org/10.1038/s41592-021-01343-9.

41. Wickham H. ggplot2. Cham: Springer International Publishing; 2016. https://doi.org/10.1007/978-3-319-24277-4.

42. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. PLOS ONE. 2013;8:e67019. https://doi.org/10.1371/journal.pone.0067019.

43. Finnish food authority. Ruokavirasto. 2025. https://www.ruokavirasto.fi/elintarvikkeet/terveytta-edistava-ruokavalio/ravitsemus--ja-ruokasuositukset/imevaisikaiset-ja-lapset/. Accessed 11 Apr 2025.

44. Soyyılmaz B, Mikš MH, Röhrig CH, Matwiejuk M, Meszaros-Matwiejuk A, Vigsnæs LK. The Mean of Milk: A Review of Human Milk Oligosaccharide Concentrations throughout Lactation. Nutrients. 2021;13:2737. https://doi.org/10.3390/nu13082737.

45. Azad MB, Wade KH, Timpson NJ. FUT2 secretor genotype and susceptibility to infections and chronic conditions in the ALSPAC cohort. Wellcome Open Res. 2018;3:65. https://doi.org/10.12688/wellcomeopenres.14636.2.

46. Sabater C, Iglesias-Gutiérrez E, Ruiz L, Margolles A. Next-generation sequencing of the athletic gut microbiota: a systematic review. Microbiome Res Rep. 2023;2:5. https://doi.org/10.20517/mrr.2022.16.

47. Nair MKM, Joy J, Vasudevan P, Hinckley L, Hoagland TA, Venkitanarayanan KS. Antibacterial Effect of Caprylic Acid and Monocaprylin on Major Bacterial Mastitis Pathogens. Journal of Dairy Science. 2005;88:3488–95. https://doi.org/10.3168/jds.S0022-0302(05)73033-2.

48. Han Q, Liu R, Wang H, Zhang R, Liu H, Li J, et al. Gut Microbiota-Derived 5-Hydroxyindoleacetic Acid Alleviates Diarrhea in Piglets via the Aryl Hydrocarbon Receptor Pathway. J Agric Food Chem. 2023;71:15132–44. https://doi.org/10.1021/acs.jafc.3c04658.

49. Bode L. Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology. 2012;22:1147–62. https://doi.org/10.1093/glycob/cws074.

50. Li Z, Zhu Y, Ni D, Zhang W, Mu W. Occurrence, functional properties, and preparation of 3-fucosyllactose, one of the smallest human milk oligosaccharides. Crit Rev Food Sci Nutr. 2023;63:9364–78. https://doi.org/10.1080/10408398.2022.2064813.

51. Gazi MA, Fahim SM, Hasan MM, Hossaini F, Alam MA, Hossain MS, et al. Maternal and child FUT2 and FUT3 status demonstrate relationship with gut health, body composition and growth of children in Bangladesh. Sci Rep. 2022;12:18764. https://doi.org/10.1038/s41598-022-23616-9.

52. Lewis ZT, Totten SM, Smilowitz JT, Popovic M, Parker E, Lemay DG, et al. Maternal fucosyltransferase 2 status affects the gut bifidobacterial communities of breastfed infants. Microbiome. 2015;3:13. https://doi.org/10.1186/s40168-015-0071-z.

53. Thorman AW, Adkins G, Conrey SC, Burrell AR, Yu Y, White B, et al. Gut Microbiome Composition and Metabolic Capacity Differ by FUT2 Secretor Status in Exclusively Breastfed Infants. Nutrients. 2023;15:471. https://doi.org/10.3390/nu15020471.

54. Wang A, Diana A, Rahmannia S, Gibson RS, Houghton LA, Slupsky CM. Impact of milk secretor status on the fecal metabolome and microbiota of breastfed infants. Gut Microbes. 2023;15:2257273. https://doi.org/10.1080/19490976.2023.2257273.

55. Sugino KY, Ma T, Kerver JM, Paneth N, Comstock SS. Human Milk Feeding Patterns at 6 Months of Age are a Major Determinant of Fecal Bacterial Diversity in Infants. J Hum Lact. 2021;37:703–13. https://doi.org/10.1177/0890334420957571.

56. Kebbe M, Shankar K, Redman LM, Andres A. Human Milk Components and the Infant Gut Microbiome at 6 Months: Understanding the Interconnected Relationship. The Journal of Nutrition. 2024;154:1200–8. https://doi.org/10.1016/j.tjnut.2024.02.029.

57. Gurung M, Schlegel BT, Rajasundaram D, Fox R, Bode L, Yao T, et al. Microbiota from human infants consuming secretors or non-secretors mothers' milk impacts the gut and immune system in mice. mSystems. 2024;9:e00294-24. https://doi.org/10.1128/msystems.00294-24.

58. Gao X, Wu D, Wen Y, Gao L, Liu D, Zhong R, et al. Antiviral effects of human milk oligosaccharides: A review. International Dairy Journal. 2020;110:104784. https://doi.org/10.1016/j.idairyj.2020.104784.

59. Jantscher-Krenn E, Zherebtsov M, Nissan C, Goth K, Guner YS, Naidu N, et al. The human milk oligosaccharide disialyllacto-N-tetraose prevents necrotising enterocolitis in neonatal rats. Gut. 2012;61:1417–25. https://doi.org/10.1136/gutjnl-2011-301404.

60. Jabeen I, Islam S, Hassan AKMI, Tasnim Z, Shuvo SR. A brief insight into Citrobacter species - a growing threat to public health. Front Antibiot. 2023;2. https://doi.org/10.3389/frabi.2023.1276982.

61. Spigaglia P. Clostridioides difficile and Gut Microbiota: From Colonization to Infection and Treatment. Pathogens. 2024;13:646. https://doi.org/10.3390/pathogens13080646.

62. Zhai L, Huang C, Ning Z, Zhang Y, Zhuang M, Yang W, et al. *Ruminococcus gnavus* plays a pathogenic role in diarrhea-predominant irritable bowel syndrome by increasing serotonin biosynthesis. Cell Host & Microbe. 2023;31:33-44.e5. https://doi.org/10.1016/j.chom.2022.11.006.

63. Thomson P, Medina DA, Garrido D. Human milk oligosaccharides and infant gut bifidobacteria: Molecular strategies for their utilization. Food Microbiology. 2018;75:37–46. https://doi.org/10.1016/j.fm.2017.09.001.

64. Chapman JA, Masi AC, Beck LC, Watson H, Young GR, Browne HP, et al. Human milk oligosaccharide metabolism by Clostridium species suppresses inflammation and pathogen growth. 2025;:2025.01.21.633585. https://doi.org/10.1101/2025.01.21.633585.

65. Charbonneau MR, O'Donnell D, Blanton LV, Totten SM, Davis JCC, Barratt MJ, et al. Sialylated Milk Oligosaccharides Promote Microbiota-Dependent Growth in Models of Infant Undernutrition. Cell. 2016;164:859–71. https://doi.org/10.1016/j.cell.2016.01.024.

66. Lewis ED, Richard C, Goruk S, Wadge E, Curtis JM, Jacobs RL, et al. Feeding a Mixture of Choline Forms during Lactation Improves Offspring Growth and Maternal Lymphocyte Response to Ex Vivo Immune Challenges. Nutrients. 2017;9:713. https://doi.org/10.3390/nu9070713.

67. Romano KA, Vivas EI, Amador-Noguez D, Rey FE. Intestinal Microbiota Composition Modulates Choline Bioavailability from Diet and Accumulation of the Proatherogenic Metabolite Trimethylamine-N-Oxide. mBio. 2015;6:10.1128/mbio.02481-14. https://doi.org/10.1128/mbio.02481-14.

68. Seki D, Errerd T, Hall LJ. The role of human milk fats in shaping neonatal development and the early life gut microbiota. Microbiome Res Rep. 2023;2:8. https://doi.org/10.20517/mrr.2023.09.

69. Macfarlane GT, Macfarlane S. Bacteria, Colonic Fermentation, and Gastrointestinal Health. Journal of AOAC INTERNATIONAL. 2012;95:50–60. https://doi.org/10.5740/jaoacint.SGE_Macfarlane.

70. Rios-Covian D, González S, Nogacka AM, Arboleya S, Salazar N, Gueimonde M, et al. An Overview on Fecal Branched Short-Chain Fatty Acids Along Human Life and as Related With Body Mass Index: Associated Dietary and Anthropometric Factors. Front Microbiol. 2020;11. https://doi.org/10.3389/fmicb.2020.00973.

71. Yang S, Yu X, Zuo Q. Branched- Chain Fatty Acids and Obesity: A Narrative Review. Nutr Rev. 2025;83:1314–26. https://doi.org/10.1093/nutrit/nuaf022.

72. You X, Rani A, Özcan E, Lyu Y, Sela DA. Bifidobacterium longum subsp. infantis utilizes human milk urea to recycle nitrogen within the infant gut microbiome. Gut Microbes. 2023;15:2192546. https://doi.org/10.1080/19490976.2023.2192546.

73. Firth IJ, Sim MAR, Fitzgerald BG, Moore AE, Pittao CR, Gianetto-Hill C, et al. Urease in acetogenic Lachnospiraceae drives urea carbon salvage in SCFA pools. Gut Microbes. 2025;17:2492376. https://doi.org/10.1080/19490976.2025.2492376.

74. Ribo S, Sánchez-Infantes D, Martinez-Guino L, García-Mantrana I, Ramon-Krauel M, Tondo M, et al. Increasing breast milk betaine modulates Akkermansia abundance in mammalian neonates and improves long-term metabolic health. Sci Transl Med. 2021;13:eabb0322. https://doi.org/10.1126/scitranslmed.abb0322.

75. Li Z, Hu G, Zhu L, Sun Z, Jiang Y, Gao M, et al. Study of growth, metabolism, and morphology of Akkermansia muciniphila with an in vitro advanced bionic intestinal reactor. BMC Microbiology. 2021;21:61. https://doi.org/10.1186/s12866-021-02111-7.

76. Poulsen KO, Meng F, Lanfranchi E, Young JF, Stanton C, Ryan CA, et al. Dynamic Changes in the Human Milk Metabolome Over 25 Weeks of Lactation. Front Nutr. 2022;9. https://doi.org/10.3389/fnut.2022.917659.

77. Schoeler M, Caesar R. Dietary lipids, gut microbiota and lipid metabolism. Rev Endocr Metab Disord. 2019;20:461–72. https://doi.org/10.1007/s11154-019-09512-0.

78. Yokota A, Fukiya S, Islam KBMS, Ooka T, Ogura Y, Hayashi T, et al. Is bile acid a determinant of the gut microbiota on a high-fat diet? Gut Microbes. 2012;3:455–9. https://doi.org/10.4161/gmic.21216.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2025

ISBN 978-952-02-0289-7 (PRINT)
ISBN 978-952-02-0290-3 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)