Complementary feeding to nourish the infant gut microbiome

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Nutritional Science

at

Massey University

Palmerston North

New Zealand

Starin Alyssa McKeen

2022

Abstract

The transition from milk-feeding to consuming a range of solid foods during the first 1000 days of life constitutes the most dramatic dietary shift in the human lifespan. This period coincides with the development of host-microbial processes. This PhD dissertation aimed to identify relationships between complementary feeding and characteristics of the developing microbiome of the infant using faecal samples as a proxy of the large intestine. The composition and gene abundances of the microbiome and the metabolites of dietary, host, and/or microbial origins were detected and analysed.

The *Nourish to Flourish* pilot clinical study was conducted to inform subsequent study designs and develop sampling and processing protocols for research in weaning infants. The faecal samples were analysed to characterise the microbiome and metabolome at 4, 9, and 12 months of age. These data revealed greater shifts during the introduction of solid foods between 4 and 9 months of age than during the continued diversification of solid foods in the diet between 9 and 12 months of age. The changes occurring in the microbiome and metabolome in faecal samples were then interpreted in the context of milk-feeding behaviours and associated dietary diversification.

The associations between the faecal microbiome and metabolome and dietary nutrient composition estimates were identified and evaluated. The patterns of association that emerged were varied and showed low prediction values but supported published evidence about relationships between specific nutrients, microbial taxa, and predicted functional pathways based on microbial gene abundance. Correlation networks between dietary food groups and relative abundance of microbes, KEGG pathway abundance, and metabolites revealed patterns that implicate the food matrix by protein-rich foods compared to the relatively weak effect of starchy foods on the luminal environment.

This research demonstrates the window of opportunity for the greatest influence of solid foods on the infant gastrointestinal environment is prior to 9 months of age. However, the impact of continued milk feeding behaviours is likely to be relatively greater than solid foods later in the complementary feeding window. This research provides a basis for selecting food products that can benefit the infant's development by modulating the microbiome and metabolome in the gastrointestinal tract.

Acknowledgements

I would like to extend my deepest thanks to my supervisors and advisors, without whom this thesis would not have been possible. Foremost, Professor Nicole Roy and Dr Jane Mullaney who have offered continued guidance, measured patience, moral support, and countless opportunities for professional and personal development during the course of this PhD dissertation. Second, Dr Karl Fraser has offered the most reliable optimism, enthusiasm, and expertise towards solving the most infuriating problems on the darkest of days and brightest of Friday afternoons alike. Third, Dr Wayne Young has offered broad expertise and excellent criticism that continues to challenge me to prove that I know what I am writing about. Professor Warren McNabb who has brought scientific scope and diplomatic management of what proved to be a complex collaboration.

Additional thanks go to the clinical team in Auckland, who designed and conducted the human side of the study on which this thesis was based. Professor Clare Wall provided her extensive paediatric nutrition and dietetics expertise and project management. Dr Amy Lovell provided extensive age-specific clinical skills, and Hannah Eriksen ensured that study participants and their families were comfortable, supported, and engaged.

I am grateful to the New Zealand Ministry for Business, Innovation, and Enterprise for funding the National Science Challenges and the High-Value Nutrition National Science Challenge to recognise the value of and opportunity to advance our understanding of nutrition and systems biology in weaning infants. I would also like to thank the Riddet Institute Centre of Research Excellence for hosting me as a student at Massey University and the many contributions from the Riddet Institute that enabled the presentation of this research at multiple conferences and symposiums. Finally, I thank the Riddet students, researchers, and administrators for being so welcoming and supportive.

I am endlessly thankful to AgResearch Ltd. for hosting and supporting me throughout my studies and letting me participate in various projects that were launching during my time there. The Manawatu campus felt like a home away from home, thanks in no small part to the welcoming and dedicated administrators, IT professionals, grounds crew, scientists, and technicians from across the organisation. Several people directly contributed to this thesis through technical expertise, advice, and encouragement. For this, I must thank Mr Hedley Stirrat and Dr Heike Schwendel, who were excellent Mass Spectrometry technicians and teachers, and Dr Doug Rosendale and Mrs Helena Stoklosinski, who contributed Branched Chain Fatty Acid and Short Chain Fatty Acid expertise in the early days of this project and continued professional opportunities and support.

Thank you to my overseas family, who have been infinitely supportive of my pursuits in New Zealand and understanding of my extended absence through what has been a strange and trying couple of years. Another important thank you to the many friends I made in Palmy who have become my adopted New Zealand family, now spread throughout this amazing country, and who have made me feel at home from the start with their generosity, shared dinners, weekend adventures, and endless patience and sense of humour while I ponder the best ways to extract baby poo from paper towels.

Finally, the biggest thank you to all the participants of the Nourish to Flourish study and their parents for doing the heavy lifting in the collection of samples and recording of data. Without them, the investigations presented here would not have been possible. I sincerely hope that the lessons learned, hypotheses generated, and questions posed from this study inform further research into how to leverage complementary feeding for long term health and vitality.

Table of Contents

Chapter 1. Introduction 14
Chapter 2: Review of the literature 16
2.1. Complementary feeding **20**
2.2. Digestion in the infant GIT **25**
2.3. GIT barrier development **27**
2.4. GIT mucosa **29**
2.5. Establishment of the microbiome and immune system in the first year of life **31**
2.6. Immune ontogeny **33**
2.7. Microbiome assembly **35**
2.8. Functional transitions during complementary feeding **39**
2.9. Short Chain Fatty Acids **45**
2.10. Vitamins **51**
2.11. Summary **53**
2.12. Aims of the thesis **55**
Chapter 3. Study design, randomisation, and sampling in nutrition and systems biology research in weaning infants 59
3.1. Introduction **60**
3.2. Methods **62**
3.2.1. Clinical study 62
3.2.2. Sample collection 64
3.2.3. Faecal microbiome 64
3.2.4. Faecal metabolome 65
3.2.5. Saliva metabolome 67
3.2.6. Assessment and analysis of interventions 68
3.3. Results **69**
3.3.2. Sample collection: protocols, contamination, and normalisation 73
3.3.3. Faecal samples 75
3.3.4. Saliva contamination 75
3.3.5. Blood, urine, and breastmilk sample collection 77
3.3.6. Comparison of interventions 77
3.4. Discussion **79**
3.5. Conclusions **82**
Chapter 4. Adaptation of the infant GIT microbiome during complementary feeding 84
4.1. Introduction **85**
4.2. Methods **87**

4.2.1. Statistical and bioinformatic analyses 87
4.3. Results **88**
4.3.1. Broad trends in the faecal microbiome composition and function 88
4.3.2. Shifts in the faecal microbiome and metabolome between time points 90
4.3.3. Correlation and integration of the microbiome with metabolites between time points 97
4.4. Discussion **104**
4.5. Conclusions **108**
Chapter 5. Mode of milk-feeding differentially influences microbiome composition and function during complementary feeding 109
5.1. Introduction **110**
5.2. Methods **114**
5.2.1. Diet 114
5.2.2. Statistical analyses 115
5.3. Results **118**
5.3.1. Milk feeding summary 118
5.3.2. Sources of nutrients 118
5.3.3. Diet diversity and microbiome diversity 121
5.3.4. PLS-DA clustering of microbiome variables 123
5.3.5. ANOVA comparisons 125
5.4. Discussion **132**
5.5. Conclusions **136**
Chapter 6. Associations between estimated nutrient consumption and the infant GIT microbiome and metabolome during complementary feeding 137
6.1. Introduction **138**
6.2. Methods **142**
6.2.1. Statistical analysis: normalisation and scaling of datasets 142
6.3. Results **143**
6.3.1. OPLS-R model performance 143
6.3.2. Linear correlations between nutrients and sequences assigned to taxa and KEGG pathways at 9 months of age 147
6.3.3. Linear correlations between nutrients and faecal metabolites at 9 months of age 149
6.3.4. Linear correlations between nutrients and sequences assigned to taxa and KEGG pathways at 12 months of age 152
6.3.5. Linear correlations between nutrients and sequences assigned to taxa and KEGG pathways at 12 months of age 157
6.4. Discussion **160**
6.5. Conclusions **164**
Chapter 7. Complementary food networks in the infant faecal microbiome 165

7.1. Introduction **166**
7. 2. Methods **169**
7.2.1. Correlations and network analyses 169
7.3 Results **170**
7.3.1 Level 1 food groups at 9 months of age 170
7.3.2. Level 2 food groups at 9 months of age 174
7.3.3. Level 1 food groups at 12 months of age 177
7.3.4. Level 2 food groups at 12 months of age 180
7.4. Discussion **183**
Chapter 8. General discussion 189
8.1. Summary of the findings **191**
8.2. Limitations of the study **195**
8.3. Implications **195**
8.4. Future perspectives **197**
References **198**
Appendices **210**

List of Figures

Figure 2. 1. Nutrient requirements met by breastmilk in the first year of life. 21
Figure 2. 2. Core structures of HMOs in breastmilk and formula vs dietary fibres 24
Figure 2. 3. Prebiotic modulation of gut epithelial development and the immune system..... 32
Figure 2. 4. Biosynthesis of Short Chain Fatty Acids by infant gut microbiota..... 46
Figure 2. 5. Overview of the thesis structure 58
Figure 3. 1. CONSORT Flow of the Nourish to Flourish cohort 63
Figure 3. 2. Variance of sequences assigned to taxa and KEGG pathways at baseline by PCA and PLS-DA coloured by environmental variable. 72
Figure 3. 3. Summary figure of sample collections and attempted collections 74
Figure 3. 4. PCA of metabolites detected in saliva with and without known contamination .. 76
Figure 4. 1. Trendlines and Shannon diversity of sequences assigned to taxa and sequences assigned to KEGG pathways at 4, 9, and 12 months of age 89
Figure 4. 2. Ordination analyses of sequences assigned to taxa and sequences assigned to KEGG pathways at 4, 9, and 12 months of age 91
Figure 4. 3. Filtered canonical correlation heatmap of the delta value of taxonomic assignments and metabolites between 4 and 9 months of age..... 100
Figure 4. 4. Filtered canonical correlation heatmap of the delta value of KEGG pathways and metabolites between 4 and 9 months of age 101
Figure 4. 5. Filtered canonical correlation heatmap of the delta value of taxonomic assignments and metabolites between 9 and 12 months of age..... 102
Figure 4. 6. Filtered canonical correlation heatmap of the delta value of KEGG pathways and metabolites between 9 and 12 months of age. 103
Figure 5. 1. Food group classifications of the Nourish to Flourish cohort. 117
Figure 5. 2. Average nutrients obtained from food sources by infants who were breastfed, formula-fed or mixed feeding. 120
Figure 5. 3. Relative abundance of food group contribution to energy intake, diet diversity by mode of milk feeding, and correlation between diet diversity and microbiome diversity at 9 and 12 months of age. 122
Figure 5. 4. Sparse PLS-DA of taxonomic composition, KEGG pathways, lipid metabolites, and aqueous metabolites by mode of milk feeding at 4, 9, and 12 months of age. 124
Figure 6. 1. Filtered heatmap indicating the strength and direction of linear correlation between VIPs (KEGG pathways and taxa) of OPLS-Regressions and estimated nutrient intakes included in models at 9 months of age 146
Figure 6. 2. Filtered heatmap indicating the strength and direction of linear correlation between VIPs (metabolites) of OPLS-Regressions and estimated nutrient intakes included in models at 9 months of age 151
Figure 6. 3. Filtered heatmap indicating the strength and direction of linear correlation between Variables Import to the Projection (VIPs) (KEGG pathways and microbes) of OPLS-Regressions and nutrients included in models at 12 months of age. 156

Figure 6. 4**.** Filtered heatmap indicating the strength and direction of linear correlation between Variable Important to the Projection (VIPs) (metabolites) of OPLS-Regressions and nutrients included in models at 12 months of age.158Figure 7. 1**.** Correlation networks between Level 1 food groups, taxa, KEGG pathways, and metabolites at 9 months of age.....173
Figure 7. 2**.** Correlation networks between Level 2 food groups, taxa, KEGG pathways, and metabolites at 9 months of age.....176
Figure 7. 3**.** Correlation networks between Level 1 food groups, taxa, KEGG pathways, and metabolites at 12 months of age.....179
Figure 7. 4**.** Correlation networks between Level 2 food groups, taxa, KEGG pathways, and metabolites at 12 months of age.....182
Figure 8. 1. Summary figure of findings.....194

List of Tables

Table 3. 1. Characteristics of the Nourish to Flourish cohort. 70
Table 4. 1. Significant changes in sequences assigned to taxa over time, 4 vs 9 months of age, 9 vs 12 months of age, grouped by phylum and direction of change of relative abundance. ..94
Table 4. 2. Significant changes in sequences assigned to KEGG pathways over time, 4 vs 9 months of age, 9 vs 12 months of age, grouped by phylum and direction of change of relative abundance. 95
Table 5. 1. Anthropometric and feeding characteristics of the Nourish to Flourish cohort ..116
Table 5. 2. Significant differences (ANOVA) in sequences assigned to taxonomy and KEGG pathways between breastfed, formula-fed, and mixed fed infants at 4, 9, and 12 months of age. 128
Table 5. 3. Significant differences (ANOVA) in aqueous and lipid metabolites between breastfed, formula-fed, and mixed fed infants at 4, 9, and 12 months of age. 131
Table 6. 1. OPLS-R model performance between nutrients (macronutrient, micronutrients, and vitamins) and sequences assigned to KEGG pathways, taxonomy, lipid metabolites, and aqueous metabolites at 9 and 12 months of age 145

Appendices

Appendix 1. Taxa included in analyses with mean relative abundance at 4, 9, and 12 months of age, as well as Log2 fold change (Log2FC) between time points, permanova p values with FDR correction and associated significance (FDR p <0.05).

Appendix 2. KEGG pathways included in analyses with mean relative abundance at 4, 9, and 12 months of age, as well as Log2 fold change (Log2FC) between time points, permanova p values with FDR correction and associated significance (FDR p <0.05).

Appendix 3. Lipid metabolites included in analyses with annotations, median mass (m/z), median retention time in seconds (rt), and valence.

Appendix 4. HILIC (aqueous) metabolites included in analyses with annotations, median mass (m/z), median retention time in seconds (rt), and valence.

Appendix 5. Metagenome sequence quality data for individual samples collected.

Abbreviations

ANOVA	Analysis of Variance statistical test
Cer	Ceramide
CerG1	Glucosylceramide
CerG2	Diglycosylceramide
DG	Diglyceride
GIT	Gastrointestinal Tract
HILIC	Hydrophilic Interaction Chromatography
HMO	Human Milk Oligosaccharide
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LCMS	Liquid Chromatography Mass Spectrometry
LPC	Lysophosphatidylcholine
MG	Monoglyceride
MGSDG	Monogalactosyldiacylglycerol
NDC	Non-Digestible Carbohydrate
PCA	Principal Component Analysis
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
OPLS-R	Orthogonal Partial Least Squares/Projection to Latent Structure-Regression
PLS-DA	Partial Least Squares/Projection to Latent Structure-Discriminant Analysis
PLS-R	Partial Least Squares/Projection to Latent Structure-Regression
RDI	Recommended Daily Intake
SCFA	Short Chain Fatty Acid
SO	Sphingosine
TG	Triglyceride
TJ	Tight Junction
TLR	Toll Like Receptor
VIP	Variables Important to Projection

Publications

1. **McKeen, S.**, Young, W., Mullaney, J., Fraser, K., McNabb, W. C., & Roy, N. C. (2019). Infant complementary feeding of prebiotics for the microbiome and immunity. *Nutrients*, *11*(2), 364.

2. **McKeen, S.**, Young, W., Fraser, K., Roy, N. C., & McNabb, W. C. (2019). Glycan utilisation and function in the microbiome of weaning infants. *Microorganisms*, *7*(7), 190

3. **McKeen, S.**, Roy, N., Mullaney, J., Eriksen, H., Lovell, A., Kussman, M., Young, W., Fraser, K., Wall, C., McNabb, W. (2022) Adaptation of the infant gut microbiome during the complementary feeding transition. Under review for publication at *PLoS Biology*

Chapter 1. Introduction

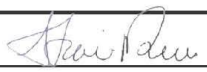
Complementary feeding of solid foods in addition to the milk-based diet of infants constitutes a window of opportunity to influence the establishment of the infant gastrointestinal tract (GIT) microbiome with implications for immune, metabolic, and cognitive health, which are concurrently in a phase of rapid development (Bäckhed et al., 2015; Bridget E Young & Krebs, 2013). The composition of the immature microbiome of the infant prior to the introduction of solid foods is shaped by key environmental exposures, including birth, household environment, and feeding mode (breastmilk or formula) (Fallani et al., 2011). The early colonisers of this new enteric community are key to determining the environment and influence factors such as pH and oxygen saturation. They can adapt to the shift in nutrients from a changing diet due to high levels of functional redundancy in the gene groups harboured in the GIT microbiota (Davis, Wang, & Donovan, 2017). However, once a species has established into ecological niches and trophic networks formed by the solid food diet, changes to the GIT microbiome become transient around a core consortium of species (Faith et al., 2013).

The shifting abundances of these species and their fermentative by-products have been found to play significant roles in supporting the innate immune system of the infant and training the development of the adaptive immune system (Depner et al., 2020; Feehley et al., 2019; Fergusson, Horwood, & Shannon, 1990; Fujimura et al., 2016; Goenka & Kollmann, 2015; Goulet, 2015; Paun & Danska, 2015). However, the role of solid foods in modulating that immune-programming has not been established using a systems biology approach to nutrition research in human infants.

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Starin A. McKeen
Name/title of Primary Supervisor:	Prof Warren C. McNabb
Name of Research Output and full reference:	
McKeen, S., Young, W., Mullaney, J., Fraser, K., McNabb, W. C., & Roy, N. C. (2019). Infant Complementary Feeding of Prebiotics for the Microbiome and Immunity. Nutrients, 11(2), 364.	
In which Chapter is the Manuscript /Published work:	Chapter 2
Please indicate:	
• The percentage of the manuscript/Published Work that was contributed by the candidate:	90
and	
• Describe the contribution that the candidate has made to the Manuscript/Published Work:	
The candidate wrote the literature review, created the figures, and incorporated revisions contributed by the rest of the authors.	
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	
Date:	April 18 2022
Primary Supervisor's Signature:	
Date:	May 09 2022

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

Chapter 2: Review of the literature

Complementary feeding transitions infants from a milk-based diet to solid foods, providing essential nutrients to the infant and the developing GIT microbiome while influencing immune development (Ruiz, Delgado, Ruas-Madiedo, Sánchez, & Margolles, 2017). Some of the earliest microbial colonisers readily ferment specific oligosaccharides, influencing the ongoing establishment of the microbiome. Non-digestible oligosaccharides in prebiotic-supplemented formula and human milk oligosaccharides promote commensal immune-modulating bacteria such as taxa from the *Bifidobacterium* genus, which decrease in abundance during weaning (Koenig et al., 2011; Lawson et al., 2020; D. A. Sela et al., 2011). Incorporating complex, bifidogenic, non-digestible carbohydrates during the transition to solid foods may present an opportunity to feed commensal bacteria and promote balanced concentrations of beneficial organic acids (of which short-chain fatty acids (SCFAs) are part of) and vitamins that support the GIT barrier maturation and immunity throughout the complementary feeding window.

Between 4 and 6 months of age, nutrient demands of growing infants surpass what is provided by breastmilk or formula alone (N. Krebs & Hambidge, 1986; Lönnerdal & Hernell, 2016; Sen, Mardinogulu, & Nielsen, 2017; Bridget E Young & Krebs, 2013). Complementary foods accompany and gradually replace breastmilk and formula throughout the weaning period, providing essential nutrients to the maturing digestive system and modulating microbial colonisation (Bäckhed et al., 2015; Cong et al., 2016; Hill et al., 2017; Koenig et al., 2011; Sen et al., 2017). The young immune system is influenced by the GIT microbiome and supported by metabolites produced during the microbial fermentation of prebiotic compounds, leading to tolerance for commensal microbes and specific responses to pathogens (Amenyogbe, Kollmann, & Ben-Othman, 2017; Clavel, Gomes-Neto, Lagkouvardos, & Ramer-Tait, 2017;

Goulet, 2015; Kaplan, Shi, & Walker, 2011; M. Li, Wang, & Donovan, 2014; Praveen, Jordan, Priami, & Morine, 2015; W. A. Walker, 2013).

Prebiotic compounds in breastmilk and supplemented formula promote commensal immune-modulating bacteria, such as those from genus *Bifidobacterium,* and beneficial metabolites, such as SCFAs and vitamins (Bertelsen, Jensen, & Ringel-Kulka, 2016; Blacher, Levy, Tatirovsky, & Elinav, 2017; Kwak et al., 2016; Sierra et al., 2015; Thomson, Medina, & Garrido, 2018; Vieira & Vinolo, 2019). Introducing non-digestible starches through complementary foods may promote commensal bacteria and support microbial production of beneficial metabolites throughout the complementary feeding window, with lasting effects on health (Maier et al., 2017; Schroeder et al., 2018; Tanaka & Nakayama, 2017).

Prior to weaning, the healthy infant GIT microbiome is shaped by maternal factors, such as mode of delivery, environment, and first foods (breastmilk and infant formula) (Cui et al., 2016; Duranti et al., 2017; Gonzalez-Perez et al., 2016; Hartwig, Diemert, Tolosa, Hecher, & Arck, 2015; Praveen et al., 2015; Rogier et al., 2014a, 2014b; Romano-Keeler & Weitkamp, 2015; Schlinzig et al., 2017; Schwartz et al., 2017). The establishment of microbial species dramatically changes throughout the first two to three years of life before reaching an adult-like composition (Bäckhed et al., 2015). While individual variations in taxonomic composition persist, analogous genes consistently and predictably fill similar functions as new foods are introduced, and formula or breastfeeding ceases (Bäckhed et al., 2015). Commensal species that colonise the immature GIT modulate gene expression of epithelial and immune cells and, in turn, are regulated by adaptive and innate immune responses in the mucosal immune system (Belkaid & Segre, 2014; Brestoff & Artis, 2013; Kabat, Srinivasan, & Maloy, 2014; Kaplan et al., 2011; R. Martin et al., 2010; Paun & Danska, 2015; Rogier et al., 2014b; Romano-Keeler & Weitkamp, 2015; Ruff & Kriegel, 2015; Shi, Li, Duan, & Niu, 2017; Sjögren et al., 2009; Turroni et al., 2016).

Breastmilk and prebiotic-supplemented formulas provide non-digestible oligosaccharides to the GIT microbiome, which exert a strong influence on its composition and metabolism (Triantis, Bode, & van Neerven, 2018). In addition, the introduction of starchy foods such as cereals, porridges, and pureed tubers is common practice due to the neutral tastes, smooth textures, and ease of swallowing as oral coordination develops (Prell & Koletzko, 2016). The role of these starches in the community dynamics of the immature and unstable microbiome has recently been characterised *in vitro* from faecal samples collected from pre-weaned and weaning infants in Malawi (Y. Wang et al., 2019). Wang et al. found the infant faecal microbiome can ferment resistant starch and produce SCFAs both prior to weaning and during weaning, but the increase in bacterial load, acetate concentrations, and decrease in pH was more substantial with the faecal inoculate from the weaning than the pre-weaning group. Additionally, propionate production was only observed with the faecal inoculate from the weaning group, which was associated with increases in *Prevotella*, *Veillonella*, and *Collinsella* genera (Y. Wang et al., 2019). However, these results have not been reported *in vivo* or using faecal microbiota collected from infants consuming western diets.

Based on investigations into human milk oligosaccharides (HMOs), whole prebiotic foods support immunity and immune development through various direct and indirect mechanisms. While poorly characterised compared to oligosaccharides, starches may act as receptor analogues to pathogens, reducing the quantity of enteric pathogens that reach the GIT epithelium and subsequent infection (Shoaf, Mulvey, Armstrong, & Hutkins, 2006). Starches also promote populations of bacteria, of which some strains directly interact with immunomodulatory factors in the GIT mucosa (Abrahamse-Berkeveld et al., 2016). These and other commensal bacteria also ferment starches into metabolites such as SCFAs and vitamins, which benefit GIT barrier integrity, immune regulation, and immune response (Swennen, Courtin, & Delcour, 2006).

406 This review summarises the current body of knowledge on the complementary feeding of
407 prebiotic starches for the microbiome with a focus on the interactions of commensal species,
408 microbial metabolites, and the development of the GIT barrier and immune system.
409
410

2.1. Complementary feeding

Complementary feeding is the necessary inclusion of solid foods alongside the milk-based diet of infants during the transition to adult foods. The inclusion of solid foods is recommended to coincide with sufficient oral maturation and an imbalance between the nutrient requirements of infants and the nutritional provisions of breastmilk and formula, as demonstrated in Figure 2.1 (Prell & Koletzko, 2016). Previously, it was thought that the inclusion of solid foods in the diet was driven by increased demand for energy and protein between 4 and 6 months of age. However, Krebs and Hambidge (1986) found that absorption of zinc from breastmilk is inadequate to meet factorial estimates of requirements based on healthy growth curves (N. Krebs & Hambidge, 1986). Similarly, iron requirements increase with erythrocyte mass and myoglobin in lean tissue from 4 to 12 months of age, surpassing the low concentrations (0.2–0.4 mg/L) of bioavailable (50%) iron in breastmilk at approximately 6 months of age (Dallman, 1988).

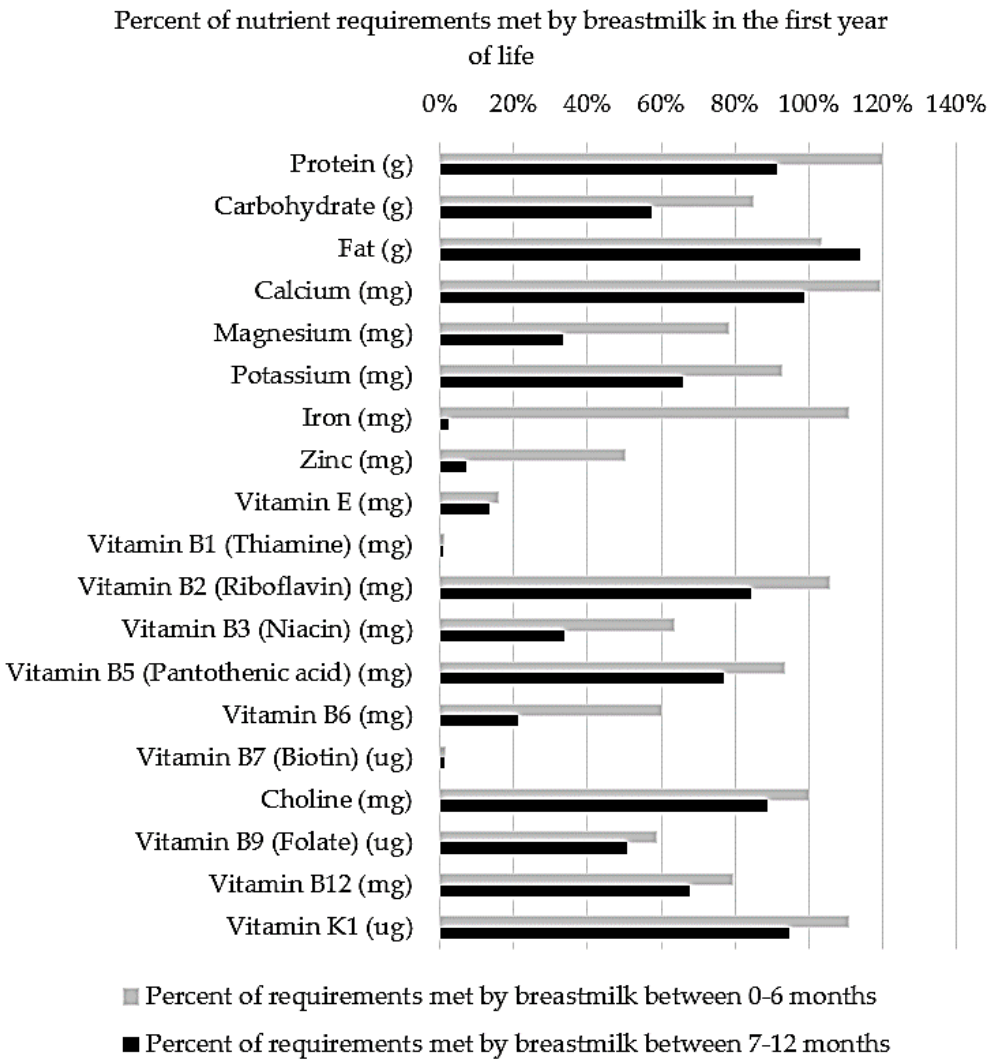


Figure 2. 1. The percentage of nutrient requirements based on the recommended daily intakes (RDIs) (Zealand, 2012) that are met via average daily breastmilk consumption (750 mL from 0 to 6 months and 800 mL from 7 to 12 months) (Allen, 2012).

The timing of the introduction of solid foods has been investigated in both low- and high-income countries. Delaying solids until 6 months of age was thought to be associated with lower body mass index in high-income countries and with lower rates of allergy and decreased water-borne diarrhoeal disease in low- and middle-income countries (Ong et al., 2015; Schack-Nielsen, Sørensen, Mortensen, & Michaelsen, 2010). However, recent studies in larger cohorts have challenged this assertion, proposing that individual oral maturation, nutrient requirements, and environmental disease burden should determine when to introduce solids (Bridget E Young & Krebs, 2013). Results from the PIAMA (Prevention and Incidence of Asthma and Mite Allergy) cohort in the Netherlands suggest that a short duration of breastfeeding (4 months or less) is associated with an increased risk for being overweight during childhood rather than the early introduction of solid foods, and the risk is similar between breastfed and formula-fed infants (Pluymen et al., 2018). However, this study does not report on the types of solid foods that were introduced, the duration of the overlap of breastfeeding and solid feeding, or the potential mechanisms of metabolic programming.

In addition to nutritional provisions, breastmilk also provides non-nutritive and immune-modulatory factors that impart significant benefits, even in partial concentrations or shorter durations (Amenyogbe et al., 2017). The health-promoting properties of breastmilk include varying levels and types of carbohydrates, non-digestible HMOs, immunoglobulins (IgG, IgM, and isoforms of sIgA), amino acids, polyunsaturated fatty acids, monoglycerides, lauric acid, linoleic acid, cytokines, chemokines, soluble receptors, antibacterial proteins/peptides, polyamines, and intact immune cells that are governed by maternal Lewis blood type, secretor status, and phase of lactation (Georgi, Bartke, Wiens, & Stahl, 2013).

HMOs have received significant attention in infant nutrition for their ability to influence various GIT functions: epithelial integrity, mucosal integrity, susceptibility to pathogenic infection, microbial community structure, SCFA production, and vitamin synthesis. Over 2000

distinct HMO structures (Figure 2.2A) have been identified, with significant variation between individuals and phase of lactation: a 9:1 ratio of galactooligosaccharides (GOS):fructooligosaccharides (FOS) is typical (Engfer, Stahl, Finke, Sawatzki, & Daniel, 2000; Oozeer et al., 2013). Infant formulas continue to be developed based on an increasing understanding of the roles of each of these factors in microbiome maturation, brain development, and immunity. Synthetic and plant-derived GOS, FOS (Figure 2.2B), inulin, pectin, and β-glucans, either alone or in comparable ratios, are well characterised and have been primary targets of infant formula research and product development (Boehm & Moro, 2008; Oozeer et al., 2013). Staged and follow-on formulas that vary in composition according to the recommended daily allowances and the introduction of complementary foods are also increasingly recommended (Lönnerdal & Hernell, 2016).

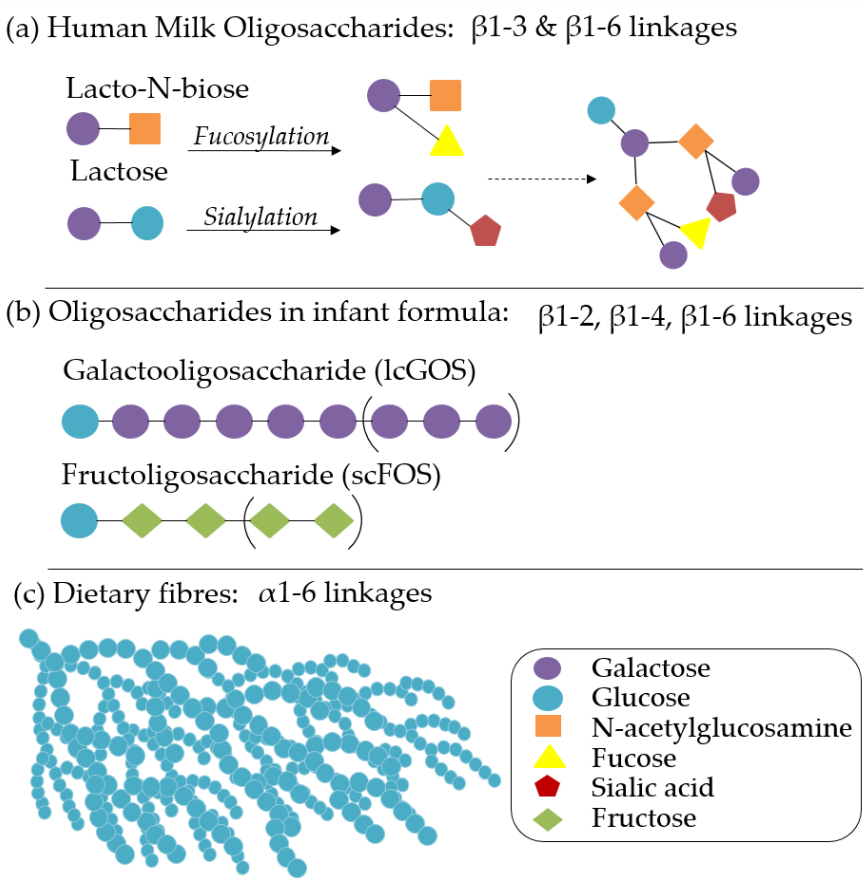


Figure 2. 2. ***A.*** *The core structures of human milk oligosaccharides (HMOs), common modification pathways, and an example of a complex HMO, connected by β1-3 and β 1-6 linkages resistant to enzymatic cleavage by human-derived enzymes.* ***B.*** *The structure of galactooligosaccharide (long-chain) and fructooligosaccharide (short-chain), which are common prebiotic molecules in supplemented infant formulas: β1-2, β1-4, and β1-6 linkages are resistant to enzymatic cleavage by human-derived enzymes.* ***C.*** *A model of dietary starch, characterised by glucose molecules connected by α1-6 linkages in a complex higher structure, which contributes to incomplete enzymatic cleavage by human enzymes.* *Figure generated by author in PowerPoint*.

2.2. Digestion in the infant GIT

The digestive system of the infant is uniquely suited to digest macronutrients provided by breastmilk. The epithelium of neonates has narrow villi and small crypts (Figure 2.3), which duplicate and expand with age, a process that is influenced by components in breastmilk and host-microbe interactions (Cummins & Thompson, 2002). The expansion of the epithelial surface during weaning is necessary to accommodate the increased nutrient load, but dysregulation of this process can lead to hyperplastic crypts, blunted villi, inflammatory responses in the mucosa, and subsequent malabsorption of nutrients (Keusch et al., 2014).

The enzymatic dynamics of infant digestion are poorly characterised due to wide variations between individuals over time and infrequent studies with replicated results (Bourlieu et al., 2014). Lactose, fatty acids, and proteins are the most abundant macronutrients in milk, which are absorbed and utilised predominantly in the small intestine (Thomson et al., 2018). Lipase and trypsin (enzymes for lipid and protein digestion) are present in concentrations comparable to adults and are sufficiently active at the less extreme pH of the infant GIT. However, amylase secretion and activity are distinct in infants. Compared to lipid and protein digestion, the ability to digest carbohydrates is limited to simple carbohydrates such as lactose and sucrose, rather than complex carbohydrates, until weaning.

At weaning, salivary α-amylase and pancreatic α-amylase are present at reduced concentrations compared to that of adults (Sibley, 2004). However, glucoamylase (also referred to as amyloglucosidase), a brush border enzyme in the small intestine capable of cleaving α1,4-glycosidic bonds, is produced at 100–150% of adult concentrations at birth, which may compensate for the otherwise minimal starch hydrolysis (Lebenthal, Lee, & Heitlinger, 1983; Lin & Nichols, 2017). Non-digestible structures such as HMOs and non-digestible carbohydrates (NDCs) resist complete enzymatic degradation and pass to the large intestine, where they become available as a nutrient source for the enteric microbiota (Parracho,

McCartney, & Gibson, 2007). Breastfed infants also receive varying amounts of maternal amylase, and low concentrations of up to 50 other digestive enzymes, through breastmilk (Dewit, Dibba, & Prentice, 1990). During weaning, infants that continue to consume breastmilk may have increased capacity to digest dietary starches compared to those receiving formulas, but this and the subsequent interactions with enteric microbes have yet to be investigated.

2.3. GIT barrier development

The epithelial barrier of the GIT is formed by enterocytes, the primary absorptive cells with crypts and villi, connected by tight junctions (TJs). The absorption of nutrients occurs transepithelially (through) and paracellularly (between TJ pores of approximately 4Å in healthy epithelia), requiring the specificity and structural functionality of TJ proteins (Anderson & Van Itallie, 2009). Compromised barrier integrity is characteristic of inflammation and can lead to aberrant immune responses that have been implicated in the development of allergies (Georas & Rezaee, 2014).

Apart from the histomorphological analysis of tissue biopsies, epithelial integrity is measured in healthy infants by feeding non-digestible sugars, lactulose, and mannitol, and then measuring their ratios in urine over a multi-hour collection to indicate paracellular sugar translocation into the bloodstream and subsequent excretion in urine (Denno et al., 2014). Faecal calprotectin has also been used as an indicator of barrier integrity. Faecal calprotectin levels are higher in infants than in adults, possibly indicative of the immature GIT undergoing cellular division, replication, and differentiation, and are higher in breastfed infants than formula-fed infants (Dorosko, MacKenzie, & Connor, 2008). However, this marker is unreliable and highly variable among and within individuals and populations (Denno et al., 2014; Rugtveit & Fagerhol, 2002).

Clinical investigations with reliable measurement of barrier integrity in infants are rare, particularly those that are sufficiently powered to understand the effects of foods and nutrients. To understand the mechanisms by which nutrients, probiotics, and microbial metabolites may affect epithelial integrity, experiments using single-cell monolayers (for example, Colon Carcinoma cell line 2 (Caco-2)) are common. *Ex vivo* tissue microscopy from porcine and murine models provides further insights. Both studies are limited in their ability to translate directly to humans.

Both HMOs and NDCs support barrier integrity by increasing TJs and promoting crypt and villus differentiation (Figure 2.3). The effects of prebiotics on structural integrity are best understood for GOS, which prevents loss of structure in Caco-2 monolayers when challenged by deoxynivalenol, a mycotoxin that inhibits protein synthesis and increases paracellular permeability (Van De Walle et al., 2010). Additionally, the stabilisation of claudin3, a TJ protein, and suppressed cytokine synthesis have been detected in GOS-treated media (Akbari et al., 2015). Formulations with higher ratios of short-chain oligosaccharides to longer polysaccharides provide the most protection to the epithelium, suggesting that complex resistant starch structures may have indirect, non-microbial mediated benefits on the epithelium (Akbari et al., 2017).

2.4. GIT mucosa

At the luminal surface of the epithelium, the mucous layer provides a physical barrier that provides lubrication and separates the microbiota from epithelial cells while allowing for the transport of nutrients and metabolites. The mucus is a complex heterogeneous suspension matrix with high concentrations of high molecular weight glycoproteins called mucins, which are secreted by goblet cells, and contains antimicrobial peptides, such as defensins (Sjögren et al., 2009). Different types of mucins have specific roles in the lumen: secreted mucins form the mucous layer over the epithelium, transmembrane mucins appear to be involved in signalling pathways, and some species of bacteria rely on mucins as an energy source (Shi et al., 2017).

Several bacterial products, including lipopolysaccharides and flagellin on Gram-negative bacteria and lipoteichoic acids on Gram-positive bacteria, have been found to upregulate mucin gene expression and stimulate mucin secretion (Cornick, Tawiah, & Chadee, 2015). Some probiotics, such as specific strains from *Bifidobacterium* and *Lactobacillus* genera, which colonise different compartments of the gut, successfully adhere to mucins and reach epithelial surfaces using non-flagellar appendages called tight adherence pili, which also influence immune responses (Turroni et al., 2013; A. Walker, 2009). The differences in locations and adherence contribute to differences between the microbiome identified in faecal collections and the microbiome in the mucosa and at the epithelial surface (Haange et al., 2012; Zmora et al., 2018). Probiotic treatment, particularly with bacteria from the *Lactobacillus* genus, has been shown to increase mucin and defensin secretion in murine models and *in vitro* cell monolayers (Caballero-Franco, Keller, De Simone, & Chadee, 2007).

Prebiotics influence the composition of mucus by increasing the concentration of glycans (Schroeder et al., 2018), decreasing the luminal pH, and increasing mucin glycosylation and sulfation (Nofrarías, Martínez-Puig, Pujols, Majó, & Pérez, 2007), which protects mucins from being degraded by host proteases and bacterial glycosidases (Figure 3) (Brockhausen, 2003).

Mucus, specifically secreted MUC-2, has been found to have immune-regulatory activity: by inhibiting gene transcription through nuclear factor NF-κB (the nuclear factor kappa-light-chain-enhancer of activated B cells) in dendritic cells it interferes with the expression of inflammatory cytokines but not tolerogenic cytokines in intestinal tissues (Shan et al., 2013). The mucus layer plays a significant role in microbial signalling, cross-feeding, microbe-host interactions, and enteric immunity but can only be partially simulated using *in vitro* experiments.

2.5. Establishment of the microbiome and immune system in the first year of life

Microbial community composition during the first year of life is dynamic, unstable, and susceptible to perturbations (Koenig et al., 2011; Yassour et al., 2016). The GIT is the largest immune organ in the human body, containing approximately 65% of immunologic tissues and up to 80% of the immunoglobulin-producing tissues of the body (Jeurink, van Esch, Rijnierse, Garssen, & Knippels, 2013). The foetal immune system is downregulated during gestation, making neonates particularly susceptible to infection and aberrant immune responses. The epithelial barrier, mucosa, and environmental conditions, such as pH, provide the majority of the protection against pathogens in the neonatal period (Figure 2.3) (Goenka & Kollmann, 2015; Kollmann, Levy, Montgomery, & Goriely, 2012).

Healthy immune development in infants is characterised by a transition from innate type 1 immunity, dominated by non-specific macrophages and neutrophils, to adaptive type 2 immunity (Duerkop, Vaishnava, & Hooper, 2009). This transition is characterised by specific T cells and B cells, which is fundamental to the establishment of tolerance: the ability to distinguish between beneficial commensal bacteria and harmful pathogens, leading to the appropriate scale and duration of responses to actual threats (Figure 2.3) (Duerkop et al., 2009). Spatial and temporal interactions between the microbiome, microbial metabolites, and epithelial cells in the lumen of the GIT and in the gut-associated lymphoid tissue modulate balanced immune development, immune response, homeostasis, and disease (Figure 2.3) (Fung, Artis, & Sonnenberg, 2014).

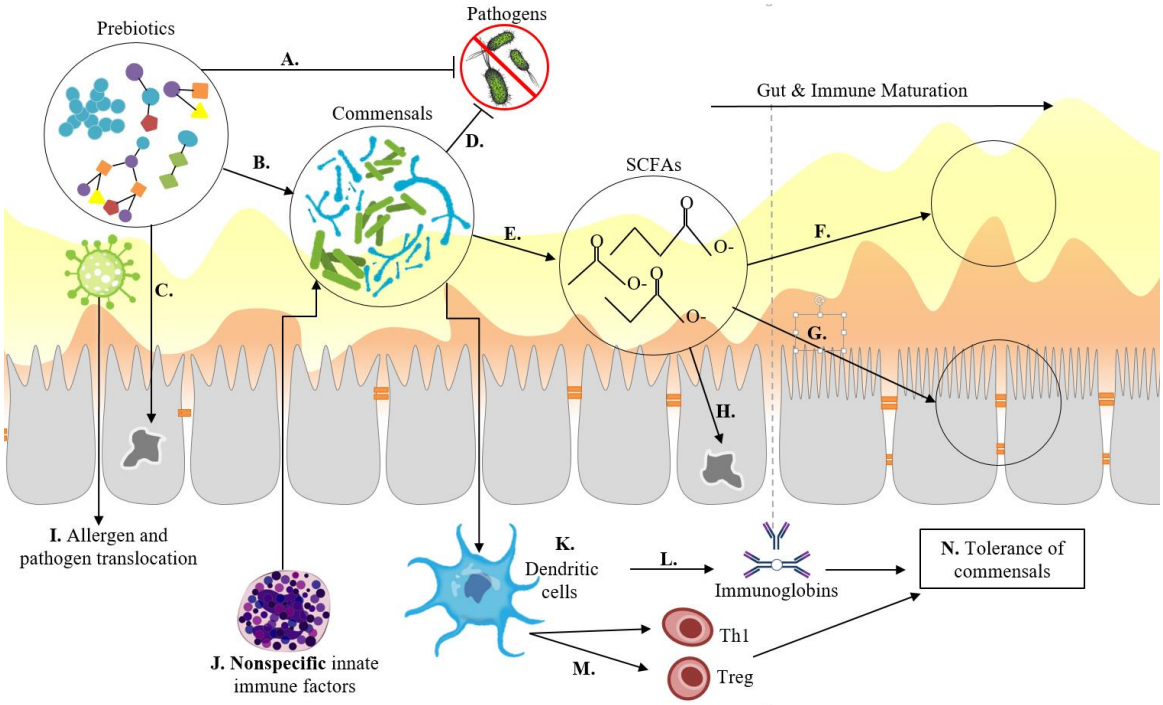


*Figure 2. 3. A schematic of multiple mechanisms by which prebiotics modulate immune and GIT development. **A.** Prebiotics bind to pathogens as receptor analogues, preventing adhesion to the epithelial surface and subsequent infection. **B.** Prebiotics promote populations of commensal microbes, which outcompete pathogens for resources. **C.** reducing infections. **D.** Prebiotics act directly upon the epithelium promoting the mRNA transcription of proteins involved in barrier integrity. **E.** Commensal microbes produce metabolites, such as short-chain fatty acids (SCFAs), that decrease the lumen pH and **F.** increase mucus, **G.** increase tight junction (TJ) proteins and crypt and villi development **H.** and serve as an energy source for enterocytes that form the epithelium. - In infants, the immature GIT is susceptible to allergy and pathogen translocation, **I.** through leaky GIT barrier. **J.** Non-specific immune factors, such as macrophages and neutrophils, attack commensals and pathogens alike in poorly regulated inflammatory responses. During immune development, dendritic cells **K.** sample commensal microbes through toll-Like receptor (TLR) recognition, allowing for antigen-specific immunoglobin production **L.** and promoting the differentiation of T and B cells **M.** resulting in improved tolerance to commensals and targeted response to pathogens **N.** Figure generated by author in PowerPoint.*

2.6. Immune ontogeny

Innate immunity favours Th2 responses and macrophage and neutrophil inflammatory activity, which use specific classes of toll-like receptors (TLRs), such as TLR4, which are capable of recognising structurally conserved molecules on microbes (De Kivit, Tobin, Forsyth, Keshavarzian, & Landay, 2014). As the immune system develops, additional mechanisms of microbe recognition with increased specificity and response cascades develop: C-type lectin receptors, pattern recognition receptors and expression of TLR2, and TLR9 by immune cells, such as DCs, in the mucosa and epithelium (Geijtenbeek & Gringhuis, 2009). These immune cells can be both upregulated and downregulated by exogenous factors, such as microbial metabolites of starch fermentation, and they demonstrate cross-regulatory activity amongst themselves by way of immune factors and regulatory cytokines (Geijtenbeek & Gringhuis, 2009).

Due to the poor specificity of the young immune system in early postnatal life, commensal bacteria are rapidly killed by macrophages. However, dendritic cells can retain small numbers of live commensals for days, protecting them from innate immune responses while selectively inducing a protective IgA response that protects against mucosal penetration by commensals (Macpherson & Uhr, 2004). Mesenteric lymph nodes restrict these commensal-loaded dendritic cells to the mucosal immune compartment, which allows for localised immune responses while preventing more damaging systemic responses (Macpherson & Uhr, 2004). These dendritic cells express TLRs, which have been implicated in GIT homeostasis and inflammatory responses characteristic of food allergies, inflammation, and infections when poorly regulated (Belkaid & Segre, 2014). Insufficient TLR exposure to commensals, as found through antibiotic-mediated dysbiosis in murine models, is also correlated with increased susceptibility to viral infections (Gonzalez-Perez & Lamousé-Smith, 2017). Infant TLR responses to commensal microbes differ from responses in adults, demonstrating the impaired production

of inflammatory mediators and heightened production of inflammatory cytokines, such as interleukin (IL)-10 (Duerkop et al., 2009; Kollmann et al., 2012).

TLRs recognise oligosaccharides, such as 1,3 glucans and mannoses on yeast and fungi, respectively, which initiate immune responses. This capacity also makes TLRs susceptible to modulation by dietary starches, as shown using *in vitro* models. Different starch structures bind differentially to TLRs, activating NF-κB and activator proteins, but the strong immune-stimulating effects may also be attenuated by starch-exposed intestinal epithelial cells (Bermudez-Brito, Rösch, Schols, Faas, & de Vos, 2015). B2→1 fructans and High-Maize 260 mainly stimulate TLR2, whereas Novelose 330 binds to TLR2 and TLR5 (Vogt et al., 2013). High-Maize 260, which has a smaller average particle size of 12.8 µm, smooth surface, and a high degree of molecular order, was found to have a stronger regulatory effect on epithelial cells than Novelose 330, which has a larger average particle size of 46.6 µm and consists of destroyed and convoluted granules due to the retrogradation process. Despite the attenuating activity, TLRs continue to induce the production of Th1 cytokines (Bermudez-Brito et al., 2015). High-maize 260 is also more effective than inulin and sugar pectin in reducing chemokine release in response to *Sphingomonas paucimobilis* infections *in vitro* (Bergström et al., 2014). *In vivo,* the mucosal matrix is expected to alter the exposure of epithelial cells to starch structures, limiting the applicability of these findings to mechanisms occurring *in vivo*.

2.7. Microbiome assembly

Pioneer species in the infant GIT shape the early environment, which influences the dynamic succession of subsequent microbes and immune cascades. Nutrients, digestive processes, gases, and pH gradients throughout the GIT modulate the microbial community, which also influences the characteristics of these attributes. Microbial taxa resemblant of the maternal microbiota may begin to colonise the infant GIT *in utero*, for example, low abundance commensal bacteria such as genus *Prevotella*, *Neisseria*, and *Escherichia Coli*, which have been found through the sequencing of amniotic fluids and placentas of preterm infants. However, this subject continues to be debated due to the inherent challenges of conducting sample collections prior to birth (Aagaard et al., 2014; Gerald W. Tannock & Ercolini, 2021). The mode of delivery is considered the first major event confirmed to seed the infant microbiome with lasting microbial colonisers (Bäckhed et al., 2015).

Vaginally delivered infants are predominantly colonised by taxa from genera *Bacteroides*, *Bifidobacterium*, *Parabacteroides*, and *Escherichia*/*Shigella*, several of which are obligate anaerobes. Infants born by caesarean section are enriched with genera *Enterobacter*, *Haemophilus*, *Staphylococcus*, *Streptococcus*, and *Veillonella*, which are associated with skin, oral, and environment (Bäckhed et al., 2015), a larger proportion of which are aerobic.

The differences in microbial community structure and gene content (i.e., the metagenome) between caesarean- and vaginally-delivered infants gradually decrease over the first year of life. However, the differences in innate and adaptive immunity remain detectable for up to 2 years of age. Caesarean-delivered infants have lower levels of IgA-, IgG-, and IgM-secreting cells, indicating reduced adaptive immune responses, have lower levels of Th1 supporting chemokines, IFNy and IL-8, and have decreased CD4+ T cell responses (Amenyogbe et al., 2017). Caesarean-delivered infants, in particular those who were born by elective caesarean delivery instead of emergency delivery, are at higher risk for asthma, atopy, juvenile arthritis,

and inflammatory bowel disease (Bager, Wohlfahrt, & Westergaard, 2008; Negele et al., 2004; Sevelsted, Stokholm, Bønnelykke, & Bisgaard, 2015). This effect is particularly pronounced for developing obesity where any caesarean delivery has been associated with a 15% increased risk for obesity, but there is a 30% increased risk in elective caesarean-delivered infants (Yuan et al., 2016). The risk for the development of infectious diseases is unclear. Considering these differences, microbiome studies in infants must consider the mode of delivery to account for differences between emergency and elective caesarean-delivered infants.

During the first weeks of life, pioneer facultative anaerobic species, which have metabolic flexibility in the presence of oxygen, shift the environment to favour obligate anaerobic species by utilising oxygen to create an anaerobic environment (Penders et al., 2006). They also reduce luminal substrates through redox (oxygen)-dependent genetic pathways that produce metabolites, such as acetate, which is often required or highly stimulatory for anaerobes (Wormser & Frank, 2005). The meconium of neonates is rich in facultative anaerobes such as *E. Coli*, but the faecal microbiota becomes more diverse with the appearance of obligate anaerobes such as those from the *Bifidobacterium* and *Clostridium* genera within the first week (Jiménez et al., 2008). In a cohort of 19 healthy breastfed full-term Japanese infants, the averaged percentage of obligate anaerobic bacteria in the GIT progressed from 32% (1 day), 37% (7 days), 54% (1 month), 70% (3 months), 64% (6 months), to 99% at 3 years of age. Significant individual variations within this homogenous cohort diminished by 3 years of age (Nagpal, Kurakawa, et al., 2017; Nagpal, Tsuji, et al., 2017). This study did not specify the delivery modes of this cohort and the consequent possibility of significant differences in the colonisation patterns of facultative and obligate anaerobes.

The effects of breastmilk and formula feeding on the infant microbiome and immunity are a popular topic of research. Breastfeeding has been associated with a decreased risk of necrotising enterocolitis, infections, and diarrhoea in early life, and a lower incidence of

inflammatory bowel disease, type 2 diabetes, and obesity later in life compared to formula-fed infants (B. E. Young, 2017). Another meta-analysis found no association between breastmilk consumption and allergy, asthma, high blood pressure, or high cholesterol (Victora et al., 2016). Considering the complexity of the immune-modulating factors of breastmilk, the identifying characteristics of the microbiome that contribute to these benefits is challenging. Taxa from the *Bifidobacterium* genera have consistently been found to exist in higher abundances in exclusively breastfed infants (Bergström et al., 2014; Bode, L., 2012).

In contrast, taxa from the *Lactobacillus* genera have been reported to be in higher abundance in formula-fed infants in some studies (Penders et al., 2007; Penders et al., 2006), while at other times reported to be in higher abundance in breastfed infants (Rinne, Kalliomaki, Arvilommi, Salminen, & Isolauri, 2005). In addition, Bäckhed et al. associated exclusive breastfeeding with lower phylogenetic diversity dominated by taxa from the *Bifidobacterium* and *Lactobacillus* genera and lower relative abundances of bacteria from the *Clostridiales* and *Bacteroides* genera compared to mixed-fed infants (Bäckhed et al., 2015). These differences may persist throughout the weaning phase as breastmilk and formula feeding continues with supplementation of solid foods.

To impart similar bifidogenic effects, the supplementation of infant formula with prebiotics or symbiotic (prebiotics and probiotics) has become common. A 9:1 ratio of synthetic linear polymers of GOS:FOS is standard, but these prebiotics represent a simplistic uniform version of the HMO structures found in breastmilk (Thomson et al., 2018). Abrahamse-Berkeveld et al. (2016) found that a combination of short-chain GOS (scGOS dp of 3–15), long-chain FOS (lcFOS dp of 3–6), and *Bifidobacterium breve* increased the abundance of taxa from the *Bifidobacterium* genus from 48% to 60% of the total bacterial species and reduced the percentage of *Clostridium lituseburense*/*C. histolyticum* from 2.6% to 2.0%. (Abrahamse-Berkeveld et al., 2016). In an *in vitro* study, Leder et al. (1999) found that

734 many different strains of the *Bifidobacterium* genus are capable of utilising scGOS. However,
735 of the species analysed, only *B. adolescentis* can utilise lcFOS, providing evidence of the
736 selectivity between related commensal strains and prebiotic structures (Leder, Hartmeier, &
737 Marx, 1999). These investigations into the utilisation of HMOs and prebiotics in formula offer
738 a starting point for exploring the effects of prebiotics provided by complementary foods.
739 Oligosaccharides also provide additional protection against pathogenic infection by acting
740 as structural mimics of the pathogen binding sites that coat the surface of intestinal epithelial
741 cells. Pathogenic bacteria such as *E. coli* bind to oligosaccharides in the lumen, reducing the
742 pathogen load available for adhesion to intestinal epithelial cells. In Caco-2 and human
743 epithelial type 2 (Hep-2) cell lines, purified GOS reduced adhesion by 70% and 65%,
744 respectively. This effect was dose-dependent and reached a maximum at 16 mg/mL (Shoaf et
745 al., 2006). It is unclear if complex starches, such as resistant starch, have the same effect.
746

2.8. Functional transitions during complementary feeding

Investigations into the functional differences between modes of feeding at the metagenomic and transcriptomic levels are less common. Bäckhed et al. found differences in the relative abundance of genes in the faecal microbiome of the breastfed and formula-fed infants that accounted for approximately 1.30% of the variation in gene abundances, which is substantial considering most microbial genes are involved in basic functions and expected to be constitutively expressed (Bäckhed et al., 2015). This study did not specify the types of formulas used in this comparison. The gene expression differences during this dynamic stage may be more facultative than constitutive due to the inherent instability of the immature infant microbiome.

The community structure and metabolic functions of the infant GIT microbiota are strongly influenced by dietary prebiotics. The bifidogenic nature of breastmilk is well-established and has been attributed to HMOs (Smilowitz, Lebrilla, Mills, German, & Freeman, 2014). However, HMO consumption has only been identified in specific *Bacteroides* and *Lactobacillus* genera. Different species and subspecies have been found to utilise other protein-substrate binding and enzymatic mechanisms to metabolise HMOs (Asakuma et al., 2011; Marcobal & Sonnenburg, 2012).

B. longum subsp. *infantis*, which is enriched in breastfed infants, express an overabundance of proteins that transport HMO substrates into the cell, where they are broken up into their constituent sugars before being metabolised. This breakdown limits the sugars that are available to other species within the microbial community (Garrido, Kim, German, Raybould, & Mills, 2011). *B. bifidum*, however, relies on a set of diverse membrane-associated extracellular glycosyl hydrolases, lacto-*N*-biosidase and endo-*N*-acetylgalactosaminidase (Wada et al., 2008). These enzymes have comparable affinities for HMOs but may release

771 monosaccharides such as lactose, fucose, and sialic acid into the lumen, which become
772 available to other microbes (Garrido et al., 2015).

773 Glycosylation patterns on HMOs influence the enzymatic activity that microbes employ.
774 *B. breve* has been found to prefer sialylated HMOs over neutral HMOs, using enzymes that
775 convert HMOs into multiple intracellular products, but it does not internalise the molecule
776 (Ruiz-Moyano et al., 2013). *B. longum* has numerous genes for carbohydrate utilisation,
777 including 30 glycosyl hydrolases that are involved in HMO degradation, though strains isolated
778 from adults prefer plant polysaccharides (González, Klaassens, Malinen, Vos, & Vaughan,
779 2008). The transcriptomic analysis of *B. longum* SC596 when shifting from a neutral HMO
780 substrate to a fucosylated HMO substrate shows that the gene expression was altered to
781 resemble the intracellular import strategy of *B. infantis*, which may provide an example of the
782 facultative gene expression of infant microbiota in response to dietary factors (Thomson et al.,
783 2018). A meta-transcriptomic analysis of faecal samples from a single breastfed baby from
784 birth to 6 months of age was completed, during which formula, dairy, and solid foods were
785 introduced (Hugenholtz et al., 2017). The authors found that the carbohydrate fermentation
786 activity of taxa from the *Bifidobacterium* genus, based on β-galactosidase activity, decreases
787 during weaning (Hugenholtz et al., 2017). At the same time, that of the resident bacteria from
788 the Firmicutes phylum increases, which corresponds with changes in the relative abundance of
789 major and minor species (Hugenholtz et al., 2017).

790 At approximately 3 months of age, genes implicated in complex carbohydrate utilisation
791 are enriched in faecal samples compared to meconium samples, which favour lactose/galactose
792 and sucrose uptake and utilisation based on a metagenomic analysis (Koenig et al., 2011). Prior
793 to introducing solid foods between 4 and 6 months of age, the GIT microbiome derives energy
794 through the degradation of simple sugars, lactose-specific transport, and carbohydrate uptake,
795 as expected for a milk-based diet. However, genes involved in plant-polysaccharide

metabolism are present prior to the introduction of complementary weaning foods (Koenig et al., 2011).

By 12 months of age, the GIT microbiome is highly enriched with species and genes active in the degradation of complex sugars and starches (Bäckhed et al., 2015). For instance, *Bacteroides thetaiotamicron*, an anaerobic glycan degrading enzyme producer of the Bacteroidetes phylum, can typically be detected at 12 months of age (Koenig et al., 2011). Another study showed that the increased abundance of taxa from the *Bifidobacterium* genus and decreased abundance of those from the *Bacteroides* and *Clostridium* genera in breastfed infants compared to formula-fed or mixed-fed infants persist into the weaning phase (Fallani et al., 2011).

Thompson et al. identified differences before and after the introduction of solid foods between the faecal microbiomes of exclusively breastfed and non-exclusively breastfed infants (Thompson, Monteagudo-Mera, Cadenas, Lampl, & Azcarate-Peril, 2015). The *Veillonella* and *Roseburia* genera and members of the *Lachnospiraceae* family appeared with the introduction of solids in breastfed infants, whereas the *Streptococcus* and *Coprobacillus* genera were identified after the introduction of solids in non-exclusively breastfed infants (Thompson et al., 2015). Most notable of these findings was the increased relative abundance of taxa from the *Bifidobacterium* genus after the inclusion of solids in non-exclusively breastfed infants. In contrast, there was a decreased relative abundance of taxa from the *Bifidobacterium* genus after the inclusion of solids in exclusively breastfed infants, which may reflect differential effects of dietary oligosaccharides and starches during complementary feeding. Metabolic inferences using a PiCRUST analysis (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States predicts metagenome functional content from marker genes and genomes) of this limited 16S dataset showed that only 24 gene clusters encoding enzymes were overrepresented in exclusively breastfed infants after the introduction of solid foods, including

polysaccharide degradation, compared to 230 enzymatic gene clusters overrepresented in the non-exclusively breastfed microbiome, which were primarily involved in signal transduction regulatory systems (Langille et al., 2013; Thompson et al., 2015). This finding indicates differences in metabolic plasticity of the microbiota between exclusively breastfed and non-exclusively breastfed infants. The substantial immune factors in breastmilk may also have a stronger effect on which gene clusters are overrepresented in this small cohort.

Human GIT microbiota may develop the capacity to degrade a specific type of starch (Type III Resistant Starch) at weaning. This characteristic was demonstrated in an *in vitro* fermentation study using faecal inoculum collected from breastfed and formula-fed infants before and during weaning (Scheiwiller, Arrigoni, Brouns, & Amadò, 2006). However, species with the potential capacity to degrade starch have been found to be present at birth (Koenig et al., 2011). From a metagenome perspective, microbial networks of the GIT microbiota of infants at 4 months of age are drastically different to that of infants at 12 months of age.

However, polysaccharide degradation has been found to be more pronounced after the cessation of breastfeeding rather than during the introduction of solid foods in breastfed infants (Bäckhed et al., 2015). It is possible that the cessation of HMO substrates decreases the need for the expression of HMO-degrading genes and reduces the competitive advantage of bacterial species selective for HMOs. These changes will allow species with a preference for polysaccharide substrates to assume a greater ecological niche. However, neither the *in vitro* experiment of Scheiwiller et al. (2006) nor the metagenomic analysis of Thompson et al. (2015) consider the nutrient availability and degradation in the proximal regions of the large intestine prior to analysis of the faecal microbiome (Scheiwiller et al., 2006; Thompson et al., 2015).

Starch degradation in the large intestine is a cooperative process that includes enzymatic starch degradation into glucose, glycolysis leading to organic acid and hydrogen production. Starch binding capacity and enzyme specificity underpin the ability of amylolytic microbes to

metabolise starch structures (Birt et al., 2013). The presence and function of cellulosomes, amylosomes, and starch utilisation system gene clusters have been investigated in keystone species belonging to the Firmicutes and Bacteroidetes phyla. Three broad classes of amylases have been identified in amylolytic bacteria that hydrolyse starch into D-glucose: α-amylase for α-1,4 linkages, type 1 pullalanase for α-1,6 linkages, and amylopullalanases for α-1,4 and α-1,6 linkages (Flint et al., 2012). Stable isotope probing, which allows for the tracking of ^{13}C-isotope labelled carbon utilisation through metabolite production, has identified complex trophic structures that implicate primary starch degraders, such as *Ruminococcus bromii*, in downstream carbon utilisation by microbes found in the infant GIT such as taxa from the *Prevotella* and *Bifidobacterium* genera, and *Eubacterium rectale* (Kovatcheva-Datchary et al., 2009).

The association of amylolytic enzymes with the cell wall and the ability to stabilise large molecules for cleavage may indicate the function of a given microbe within the trophic network (Crittenden et al., 2001; Shipman Joseph, Berleman James, & Salyers Abigail, 2000). For instance, extracellular protein complexes on *Bacteroides thetaiotamicron* import starch molecules for internal degradation, limiting the extracellular release of mono- and di-saccharides, compared to outer membrane protein complexes on *Clostridium butyricum* which degrade starches outside of the cell before importing the mono- and di-saccharides for subsequent metabolism to SCFAs (N.M. Koropatkin & Smith, 2010; J. Shipman , K. Cho, H. Siegel, & A. Salyers, 1999; Swennen et al., 2006). This variety of enzyme structures and systems points to the metabolic flexibility, which may be increased during dietary transitions such as weaning that the microbiome utilises to maximise energy harvest.

Fermentation profiles vary by substrate structure, which changes throughout enzymatic degradation. Short oligosaccharide chains, such as scFOS, are more rapidly fermented than long oligosaccharide chains, such as inulin (M. L. Stewart, Timm, & Slavin, 2008). In a study

by Stewart et al. (2008), SCFA production was highest during the first 4 hours of *in vitro* fermentation of a faecal inoculum with scFOS substrate, whereas long-chain inulin produced the most SCFA between 12–24 h of fermentation (M. L. Stewart et al., 2008).

Warren et al. (2018) expanded upon these findings by comparing digested to non-digested starches from a range of processed, un-processed, digested, and un-digested starch and resistant starch substrates (Warren et al., 2018). They found that faecal microbiota isolated from infants can ferment amorphous and crystalline starches equally well, perhaps attributable to the range of amylolytic enzymes found in the microbiome and found no difference in the fermentation rates of the digested versus undigested substrates (Warren et al., 2018). Both the 16S rRNA gene amplicon sequencing analysis and the SCFA analysis of the inoculum revealed different responses that were dependent on time and class of starch substrate (Warren et al., 2018).

2.9. Short Chain Fatty Acids

SCFAs are the primary class of microbial metabolites of starch degradation and are implicated in immune regulation. Microbes produce SCFAs to dispose of electrons by generating ATP via substrate-level and electron transport linked phosphorylation (Macfarlane & Macfarlane, 2003). However, colonocytes and downstream microbes utilise SCFAs as an energy source, which can alter lipid metabolism, protect against infection, have an anti-inflammatory effect, influence the gut-brain axis, facilitate immune cell metabolic reprogramming, and regulate immune cell transcription through epigenetic modifications (Miller & Wolin, 1996; Wong, de Souza, Kendall, Emam, & Jenkins, 2006). SCFA production varies throughout the colon because of substrate availability, population dynamics, and microbial cross-feeding (Morrison & Preston, 2016).

The fermentation of starch substrates by the microbial community in the large intestine is characterised by high acetate production, followed by propionate and relatively less butyrate, though ratios are variable (Herrmann et al., 2017; Warren et al., 2018). The RNA-stable isotope probing studies show that lactate is a precursor to both acetate and propionate, and that acetate is a precursor for butyrate via the Co-A transferase pathway and the butyrate kinase pathway (Boets et al., 2017). For example, *Bifidobacterium adolescentis* can degrade resistant starch leading to the by-products lactate and acetate. Acetate is, in turn, used by *Eubacterium* spp., *Roseburia* spp., and *Coprococcus catus*, resulting in the production of butyrate. *Faecalibacterium prausnitzii*, an abundant butyrate producer in adults, has not been detected in infants younger than approximately 2 years of age (Koga et al., 2016). Figure 2.4 demonstrates a simplified ecological network in which multiple species of bacteria commonly identified in infants perform parts of the metabolic pathways leading to the biosynthesis of SCFAs.

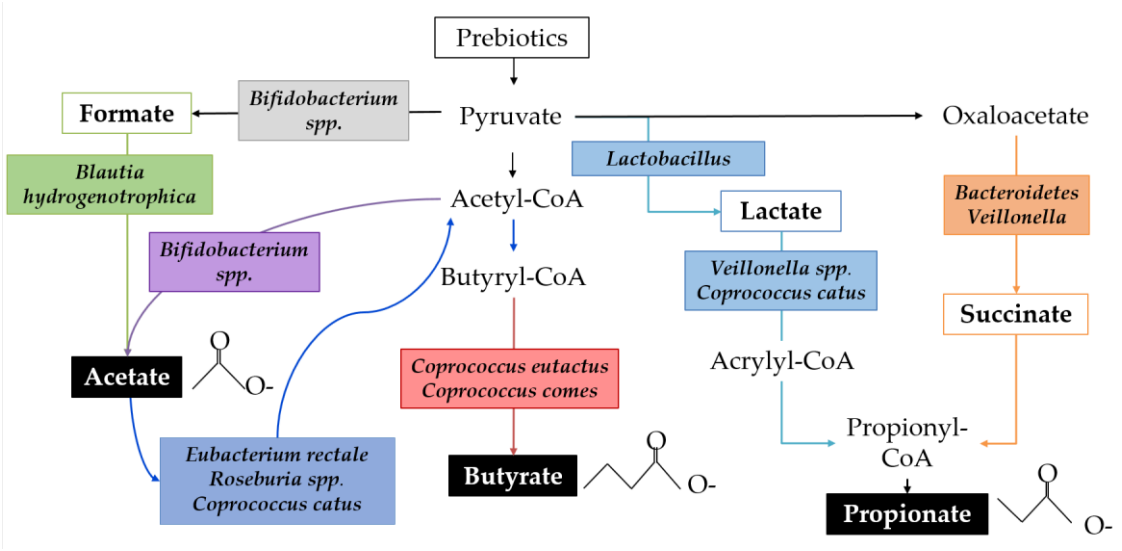


Figure 2. 4. A simplified schematic of the biosynthesis of short-chain fatty acids (SCFAs) by microbial species identified in human infants. Organic acid metabolites are outlined, and SCFAs are highlighted in black boxes. Species of bacteria found in the infant GIT microbiome that are implicated in the corresponding pathway are italicised. *Figure generated by author in PowerPoint.*

SCFAs begin influencing the enteric environment with the introduction of breastmilk and formula. Exclusive breastfeeding is associated with lower absolute concentrations of all SCFAs, except lactate (Nagpal, Kurakawa, et al., 2017). Ratios of SCFAs within total concentrations have been found to be variable. Exclusively breastfed infants are more likely to have higher proportions of acetate, while partially breastfed and formula-fed infants are more likely to have relatively higher proportions of propionate. Exclusively formula-fed infants are likely to have relatively higher proportions of butyrate (Bridgman et al., 2017). However, measuring SCFAs in faecal samples only indicates the remainder of SCFAs that have not been utilised by epithelial cells or absorbed. Absorption is likely to vary with epithelial barrier integrity and maturity, which is known to be influenced by other factors in breastmilk (Cummins & Thompson, 2002; Macfarlane & Macfarlane, 2003).

SCFAs modulate immune factors through multiple mechanisms. They increase the expression of antimicrobial peptides excreted by epithelial cells; modulate the production of cytokines and chemokines; regulate the differentiation, recruitment, and activation of immune cells; and modulate the differentiation of T lymphocytes (Vieira & Vinolo, 2019). Commonly cited anti-inflammatory properties of SCFAs can be attributed to their ability to reduce the production and activity of pro-inflammatory cytokines such as TNF-α and IL-12, often by modulating the activity of neutrophils, dendritic cells, and macrophages (Vinolo, Rodrigues, Nachbar, & Curi, 2011). Alternatively, SCFAs increase the production of other cytokines, such as IL-18, which has been implicated in the repair and maintenance of epithelial integrity (Kalina et al., 2002).

Acetate is a minor energy source for epithelial cells, a major energy source for muscles and brain tissue, has anti-inflammatory properties, decreases the pH of the colon, and is used by cross-feeding species as a co-substrate to produce butyrate (Louis & Flint, 2017; Macfarlane & Macfarlane, 2003). Numerous species from the *Bifidobacterium* genus readily produce

938 acetate from starchy substrates. Anti-inflammatory properties of acetate have been linked to
939 the SCFA-dependent modulation of NF-κB in the COLO320DM adenocarcinoma cell lines, to
940 decreased IL-6 protein release from organ culture, and decreased LPS-stimulated TNFα from
941 neutrophils. However, these dose-dependent effects are less pronounced for acetate than
942 propionate and butyrate (Tedelind, Westberg, Kjerrulf, & Vidal, 2007). Acetate has also been
943 identified as an important metabolite by which some subspecies from the *Bifidobacterium*
944 genus protect against infection, possibly by inhibiting the translocation of toxins from the GIT
945 lumen to the bloodstream (Fukuda, Toh, Taylor, Ohno, & Hattori, 2012). Several *in vitro*
946 studies suggest that the benefits of acetate are largely due to the enhanced epithelial integrity,
947 which imparts protection from infection and inflammation. For instance, *B. longum infantis*
948 *157F*, which is found in breastfed infants and metabolises glucose to acetate, was found to
949 protect against harmful protein translocation across a Caco-2 epithelial barrier assay (Fukuda
950 et al., 2012).

951 Propionate has been associated with health benefits, particularly in adults (Hosseini,
952 Grootaert, Verstraete, & Van de Wiele, 2011). Similar to acetate, propionate is a minor energy
953 source for epithelial cells, decreases the pH of the colon, is anti-inflammatory, and has immune-
954 modulating properties that *in vitro* studies in Caco-2 cell lines suggest are linked to beneficial
955 effects on epithelial barrier integrity (Mariadason, Barkla, & Gibson, 1997). Additionally,
956 propionate decreases liver lipogenesis, serum cholesterol levels, and colorectal carcinogenesis
957 in other tissues (Hosseini et al., 2011). Insulin sensitivity improvements and increased satiety
958 in adults has also been correlated with increased propionate levels (Hosseini et al., 2011; Louis
959 & Flint, 2017; Macfarlane & Macfarlane, 2003). These effects have not been investigated in
960 weaning infants.

961 Butyrate is the preferred energy source for colonic epithelial cells, meaning that little
962 butyrate reaches systemic circulation. Butyrate also decreases the pH of the colon, promotes

epithelial proliferation, prevents colorectal cancer cell proliferation, reduces oxidative stress, is anti-inflammatory, and improves barrier function by stimulating the production of mucins, antimicrobial peptides, and TJ proteins (Louis & Flint, 2017; Macfarlane & Macfarlane, 2003). Gantois et al. found that butyrate also downregulates the expression of virulence genes in *Salmonella enterica* and *S. typhimurium* (Gantois et al., 2006).

Higher faecal butyrate concentrations along with higher relative abundances of butyrate-producing bacterial taxa and relative abundance of the gene encoding butyryl–coenzyme A (CoA):acetate–CoA-transferase have also been associated with reduced incidence of childhood asthma in school-aged kids who were raised on farms as infants (Depner et al., 2020). Butyrate-producing bacteria, such as *Eubacterium rectale, Roseburia spp,* and *Faecalibacterium prausnitzii,* frequently utilise acetate as a substrate (Louis & Flint, 2009). The effects of butyrate have been found to be paradoxical where low concentrations of butyrate (2 mM) promote epithelial barrier function, characterised by increased TER and decreased mannitol flux in the Caco2 cell assay. However, high doses (8 mM) may induce cell apoptosis and disrupt the epithelial barrier, as is characteristic of necrotising enteric colitis (Peng, Li, Green, Holzman, & Lin, 2009). One study identified the benefits of butyrate to be characteristic of cellular differentiation because of the increased dome formation and alkaline phosphatase activity (Mariadason et al., 1997), whereas another identified cell migration, as is needed for epithelial repair, as a beneficial mechanism (Peng et al., 2009). Both studies found that the effects were dependent on protein synthesis and gene transcription but not the beta-oxidation or activation of adenosine 3', 5'-cyclic monophosphate (Mariadason et al., 1997; Peng et al., 2009).

Most investigations into SCFAs have focused on adult populations. In infants, SCFAs are considered beneficial. However, faecal measurements are inappropriate to use as an indicator of a healthy microbiome in infants due to paradoxical effects at high concentrations and the

988 importance of considering the balance of SCFA utilisation by epithelial cells and absorption
989 into the bloodstream (Morrison & Preston, 2016).

2.10. Vitamins

Vitamins are an additional class of secondary metabolites produced by the microbiota with effects on immunity. Commensal bacteria can synthesise essential vitamins, particularly from the B and K groups, the capacity of which is distinct in infants compared to adults. The microbiota in neonates demonstrate the enrichment of genes involved in the production of vitamin K2, retinol, folate, pyroxidal (B6), and biotin (B7), which are involved in bone, vision, and tooth development, glucose conversion, andare upregulated in the neonatal microbiome (Bäckhed et al., 2015). Genes involved in the transport of B12, iron, haemin, and haeme are also enriched in neonates but decline markedly with age, corresponding with increased nutritional demands for iron between 4 and 6 months of age (Bäckhed et al., 2015). Throughout the weaning months, genes involved in the biosynthesis of thiamine (B1), pantothenate (B5), cobalamin (B12), and lysine increase (Bäckhed et al., 2015; Yatsunenko et al., 2012). Methionine degradation, leucine biosynthesis, and tryptophan biosynthesis increase to reach levels comparable to mothers by 12 months of age (Bäckhed et al., 2015).

It has been estimated that B vitamins are produced by 40 to 65% of human commensal microbes, with some microbes having the capacity to produce all eight B-vitamins and some studies demonstrating pathways that complemented those of other organisms (Magnúsdóttir, Ravcheev, de Crécy-Lagard, & Thiele, 2015). These estimates were determined by aligning metagenomes from the human faecal microbiome to an annotated genome on the PubSEED platform (Magnúsdóttir et al., 2015). The order *Bifidobacteriales* contained the most conserved pattern of vitamins B1, B3, and B7 in approximately 35% of the genomes, whereas the phylum *Bacteroidetes* demonstrated biosynthetic pathways of all eight vitamins present in 51% of the genomes (Magnúsdóttir et al., 2015). The family *Prevotellaceae* produce vitamins B2, B5, and B7; the order *Lactobacillales* either contain no biosynthetic pathways or were limited to B2; and the order *Clostridiales* produce only B12 (Magnúsdóttir et al., 2015). The full folate

biosynthesis (B9) pathway is present in 43% of genomes, which is distributed in nearly all genomes from the phylum Bacteroidetes, in most taxa from phyla Fusobacteria and Proteobacteria, and in partial pathways occurring in bacteria from phyla Actinobacteria and Firmicutes (Magnúsdóttir et al., 2015). Vitamin K in one of two forms is reported to be produced by the genera *Bacteroides*, *Enterobacter,* and *Veillonella,* and *Eubacterium lentum*, though the bioavailability of bacterially-derived vitamin K has not been established (Biesalski, 2016). How these genes are differentially expressed in the GIT microbiome of the infant and response to specific types of complementary foods has yet to be explored.

The interactions between microbially-derived vitamins and immune cells are varied and poorly characterised. Of the known pathways, vitamin B6 has been found to serve as a cofactor in immunomodulatory pathways (Ueland, McCann, Midttun, & Ulvik, 2017), vitamin B9 has been implicated in the maintenance of regulatory T cells (Kunisawa, Hashimoto, Ishikawa, & Kiyono, 2012), and vitamin B12 has been found to augment CD8+ T lymphocytes and NK cell activity in deficient patients (Brestoff & Artis, 2013; Tamura et al., 1999). Interestingly, the by-products from vitamin synthesis pathways have recently been implicated in immune cell recognition. Mucosa-associated invariant T cells, which produce IL-17 and IFN-γ, are known to be activated in response to microbe-derived products of the riboflavin biosynthetic pathway that are presented by a monomeric major histocompatibility complex class 1 (MHC-1)-like related molecule (MR1) (Kjer-Nielsen et al., 2012). These MHC and MHC-like molecules are imperative to discriminate self from non-self, enabling protective immunity (de Verteuil, Granados, Thibault, & Perreault, 2012).

2.11. Summary

Complementary feeding merges neonatal nutrition with diverse childhood nutrition during a window of high variability and instability in the GIT microbiome. Starchy foods are common complementary foods based on texture and palatability, but the health-promoting benefits of complex starches during this window are less studied. Currently, no harmful effects or negative outcomes of starch consumption during complementary feeding have been reported. However, certain common complementary foods that contain starch, such as rice and wheat, also contain nutrient-binding compounds such as phytic acid, which can be altered during processing (Gibson, Bailey, Gibbs, & Ferguson, 2010).

Investigations into prebiotic complex starches are warranted based on the evidence that prebiotic HMOs and NDCs alter the GIT microbial community structure, its metabolism, promote immunity, and immune development. However, the utilisation and fermentation of starch structures occur in a complex trophic network governed by keystone species, cross-feeding dynamics, host-microbe interactions, and biogeography of the GIT lumen that are dynamic during the rapid growth and colonisation phase of complementary feeding. Identifying the interactions and characteristics of a healthy infant GIT that result in beneficial clinical outcomes (i.e., absence of atopy and health growth trajectories) remains challenging.

The mechanisms by which starch may contribute to immunity and immune development are varied. While some oligosaccharides are known to act as receptor analogues for pathogens, thus preventing adhesion to the epithelium and consequent infection, this effect has not been explored for complex starch structures from foods. Starch also promotes populations of commensal bacteria with direct immunomodulatory activity in the mucosa and at the intestinal epithelium. These commensal bacteria also produce metabolites, including SCFAs, vitamins, and small molecules that may beneficially alter the luminal environment, support structural

immunity at the epithelial barrier, and promote balanced and appropriate immune responses to commensals and pathogens.

Building upon extensive research into HMOs and prebiotic-supplemented formulas by investigating transitions to more complex starch structures may offer functional insights into the mechanisms that underpin balanced microbiome-mediated immune development during the complementary feeding and facilitate the development and application of complementary foods.

2.12. Aims of the thesis

From this review of the literature, the opportunity to leverage complementary foods to benefit the GIT microbiome, metabolome, and immunity is clear. However, achieving the desired effects requires a deeper understanding of complex interactions and high-quality clinical data. Conducting systems nutrition research is different in infants than adults, presenting both opportunities and challenges to study designs, sampling collection, processing, and data interpretation. The opportunity to introduce foods to a naïve ecosystem of microbes is tempered by the need to account for the influence of early life factors into the eligibility criteria and randomisation of participants (Chapter 3). Despite this inter-individual variation that exists as infants begin to consume solids foods, some patterns of microbiome adaptation and diversification are common throughout the complementary feeding window (Bäckhed et al., 2015; Moore & Townsend, 2019; Muegge et al., 2011). This study aimed to understand broad patterns of adaptation occurring in the developing microbiome of the infant and associated metabolome using faecal samples as a proxy of the large intestine during this period before evaluating the roles of milk feeding, nutrients, and specific foods (Chapter 4).

The basis of infant nutrition research has been around breastmilk and infant formula. Both practices have well-established effects on the GIT microbiome (Brink et al., 2020; Chow et al., 2014; Hanson, Korotkova, & Telemo, 2003; Pluymen et al., 2018). However, how solid food diets and resulting nutrient consumption differ between infants who consume milk from different sources is less understood. Considering the high proportion of the diet that remains milk-based throughout the complementary feeding period, another aim was to understand how the mode of milk feeding (breastfeeding, formula feeding, or a mix of breastmilk and formula feeding) influenced diet diversity, reported nutrient consumption and temporal changes in the faecal microbiome and metabolome during complementary feedings (Chapter 5).

Increasing nutrient demand by the growing infant in the early postnatal life that surpasses what milk alone can provide motivates the introduction of solid foods. Inadequate provisions of specific nutrients during the rapid development taking place in the first 1000 days of life are known to have dire consequences throughout the lifespan (Dewey, 2013; Hibberd et al., 2017; Rackerby, Kim, Dallas, & Park, 2020). This historic emphasis on nutrients in nutrition research in infants has thus far occurred without attention to how these nutrients influence the GIT microbiome. Considering the crucial role that the GIT microbiome plays in energy and nutrient capture, to support itself as well as the human host, another aim was to identify patterns between classic nutrient composition provisioned by the diet and the microbiome composition, functional capacity, and metabolome of faecal samples during complementary feeding (Chapter 6).

More recently, it has become clear that the advantage the GIT microbiota has in capturing energy and nutrients can be attributed to the greater diversity of genes and enzymes that can access and utilise nutrients packaged in food matrices that humans lack the enzymes to break down (J. M. Aguilera, 2019; Ward, Benninghoff, & Hintze, 2020). It seems GIT microbes may be more responsive, in abundance and function, to foods and food groups rather than the specific nutrients contained within the complex matrices, particularly as the infant diet transitions from a complex liquid matrix to an increasingly complex solid matrix (Johnson et al., 2019). Another aim was to understand the networks of microbial species, gene abundances that represent functional capacity, and metabolites that emerge in response to the consumption of distinct food groups at different time points during complementary feeding (Chapter 7).

The broad aim of this dissertation was to connect complementary feeding behaviours and specific food products with the microbiome and metabolome development and biomarkers of immune development using faecal samples as a proxy of changes in the microbiome in the large intestine as outlined below in Figure 2.5. Based on prior research into the development

of the infant GIT microbiome and increasing understanding of how diet and microbiome structure and function are related in adults, several hypotheses were formed. First, it was hypothesized that the younger and less established infant GIT is more susceptible to influence by complementary foods than the older and more mature infant GIT microbiome. Second, it was hypothesized that starchy carbohydrates promote bifidogenic bacteria already abundant in the GIT of breastfed infants, with positive effects on their immune development by direct interactions between probiotic microbes and SCFAs that promote the development of the epithelium and adaptive immunity. Third, it was hypothesised that dietary diversity promotes microbiome diversity in infants as it does adults.

	Chapter 1	General introduction
	Chapter 2	Review of the literature
Research chapters	Chapter 3	Study design, randomisation, and sampling • Analysis of baseline faecal samples to inform randomisation in subsequent studies • Description of the clinical and analytical methods used in the Nourish to Flourish study • Assessment of the successes and failures of protocols and methods used to inform subsequent studies
	Chapter 4	Adaptation of the infant gut microbiome and metabolome during complementary feeding • Characterisation of the shifts in relative abundance of microbial taxa, KEGG pathways, and metabolites in the Nourish to Flourish cohort during early and late complementary feeding

	Chapter 5	Mode of milk-feeding during complementary feeding
		• Characterise the differential effect of breastfeeding, formula feeding, or mixed-feeding on
		  • Estimated nutrient consumption
		  • Dietary diversity
		  • Gut microbiome alpha diversity
		  • Gut microbiome taxonomic composition
		  • Gut microbiome functional capacity (KEGG pathways)
		  • Gut metabolome

	Chapter 6	Dietary nutrients and the infant gut microbiome and metabolome • Correlations between dietary nutrient consumption and gut microbiome composition, KEGG pathways, and metabolites at 9 months of age and 12 months of age using Projection to Latent Structure Regression models and linear correlations
	Chapter 7	Complementary food networks and the infant gut environment • Construction of networks visualising the strongest correlations between food groups at 2 levels of specificity and gut microbiome composition, KEGG pathways, and metabolites at 9 months of age and 12 months of age
	Chapter 8	General discussion

Figure 2. 5. Overview of the thesis structure

Chapter 3. Study design, randomisation, and sampling in nutrition and systems biology research in weaning infants

Conducting nutrition research in weaning infants specific to the development of the microbiome and immune system requires stringent study design, randomisation of participants based on environmental factors, and modifications to sample collection and processing protocols. The purpose of this chapter is to assess the feasibility of conducting systems biology research in weaning infants, and to inform study design and protocol development for subsequent intervention studies. The Nourish to Flourish study was a 6-month open label (non-blinded) parallel-arm intervention pilot study that took place in a cohort of 40 infants in Auckland, New Zealand between October 2018 and October 2019. Reported compliance to the interventions was high (≥80%), although this was not reflected in the detection of specific probiotic species in the probiotic intervention group. Collection of faecal samples was more successful than the collection of blood, saliva, or urine, however modifications to faecal sample collections are needed to accommodate the changing consistency of faecal material as the diet changes from a liquid diet to a predominantly solid diet. Variation of the sequences assigned to microbial taxa and to KEGG pathways at baseline according to 6 relevant environmental factors was also assessed. The small size of the Nourish to Flourish cohort limited statistical significance, but the results indicate that mode of delivery and mode of milk feeding have the strongest influence on the microbiome prior to the introduction of solid foods and should be accounted for either in the inclusion criteria, exclusion criteria, or during the randomisation of participants to intervention groups.

3.1. Introduction

The weaning period is a unique window of opportunity to conduct nutrition research that can inform our understanding of food-microbe interactions in infant populations. The ability to account for environmental factors that have thus far shaped the GIT microbiome diminishes with age and increasing exposure. This age group presents a unique opportunity to explore complex effects from exposures such as diet and advance our understanding of biological interactions and promote lifelong health.

The infant GIT microbiome prior to the introduction of solid foods reflects a relatively small number of complex environmental factors, including gestational age at birth, mode of delivery, maternal microbiome, early life milk feeding, antibiotic exposure, household environment, and day-care attendance (Bäckhed et al., 2015; Koenig et al., 2011; Madan, Farzan, Hibberd, & Karagas, 2012). While these signatures may not last in a detectable manner beyond childhood, or even the first two years of life, the influence on immune development during the transition from innate to adaptive immunity may be long term and thus warrant consideration as potential confounding factors in systems biology investigations (Ajslev, Andersen, Gamborg, Sørensen, & Jess, 2011; Arrieta, Stiemsma, Amenyogbe, Brown, & Finlay, 2014; Avershina et al., 2016; Gensollen, Iyer, Kasper, & Blumberg, 2016). The prevalence and distribution of these characteristics of interest to the microbiome are listed in Table 3.1. Potential relationships between these characteristics and the GIT microbiome and metagenome of participants are investigated in Figure 3.2.

Pilot studies are essential to address the unique logistical and temporal challenges of research with weaning infants. They serve the important purpose of assessing the feasibility of

study interventions, informing calculations to achieve statistical power, and refining study tools and protocols to minimise the risk of failure in larger studies.

The ubiquitous use of nappies on infants overcomes the challenges of faecal sample collection that might exist with older populations. However, significant variability in nappy types, the potential for co-contamination with urine, and rapid changes in stool texture during the introduction of solid foods present increased challenges in establishing standardised sample collection procedures that work with parental habits. Additionally, infants do not provide samples at will during clinic visits, or in-home settings, nor are GIT transit times consistent within and between infants. This challenge means that parents must collect and save samples when available. Each of these seemingly minor variations in protocol may alter the results from sensitive analysis of biological signatures, such as metabolites and microbiome (Choo, Leong, & Rogers, 2015; Hill et al., 2016; Karu et al., 2018; Liebisch et al., 2019; Loftfield et al., 2016; Song et al., 2016; Z. Wang et al., 2018). In addition, infant faecal samples are vulnerable to alterations due to storage and transport methods because any change in abundance of certain species has a disproportionately larger effect in relatively lower diversity samples (Hill et al., 2016).

This chapter aimed to assess the feasibility of conducting systems biology research in weaning infants to inform design and protocol development for subsequent intervention studies. It included specific attention on compliance, collection of biological samples, and assessing environmental factors as potential randomisation criteria to be considered in studied in weaning infants.

3.2. Methods

3.2.1. Clinical study

The *Nourish to Flourish* study was a pilot investigation with a randomised open-label (non-blinded) parallel-arm design (Australia New Zealand Clinical Trials Registry number 12618000157279). Forty infants who primarily consumed breastmilk and had not yet begun consuming solid foods were recruited from the Auckland Metropolitan area between January and December 2018. Inclusion criteria were as follows: healthy, 2-3 months of age at enrolment, will have first foods introduced between 4 and 6 months of age, born ≥32 weeks of age, no developmental disability, no significant health impairments, no feeding problems, not currently taking antibiotics, and not currently taking probiotics or prebiotics. Exclusion criteria was not specified.

Infants were randomised into either of two intervention groups, one consuming a daily dose of commercially available probiotics (1 billion CFU of *Bifidobacterium animalis* ssp *lactis*), and the other consuming five grams of lyophilised kumara (a variety of New Zealand sweet potato, including the skin) daily (Figure 3.1). Throughout the study, six infants withdrew, five were excluded from the analysis for incomplete sample sets, two were excluded from analysis for incomplete diet data records, and two were excluded due to oral antibiotic prescription within 21 days of sample collection (Figure 3.1). The remaining 25 infants were included in the subsequent analyses presented in Chapters 4, 5, 6, and 7. This chapter includes data from 30 infants who had complete data and sample sets at baseline to maximise the participant number included in the randomisation analysis.

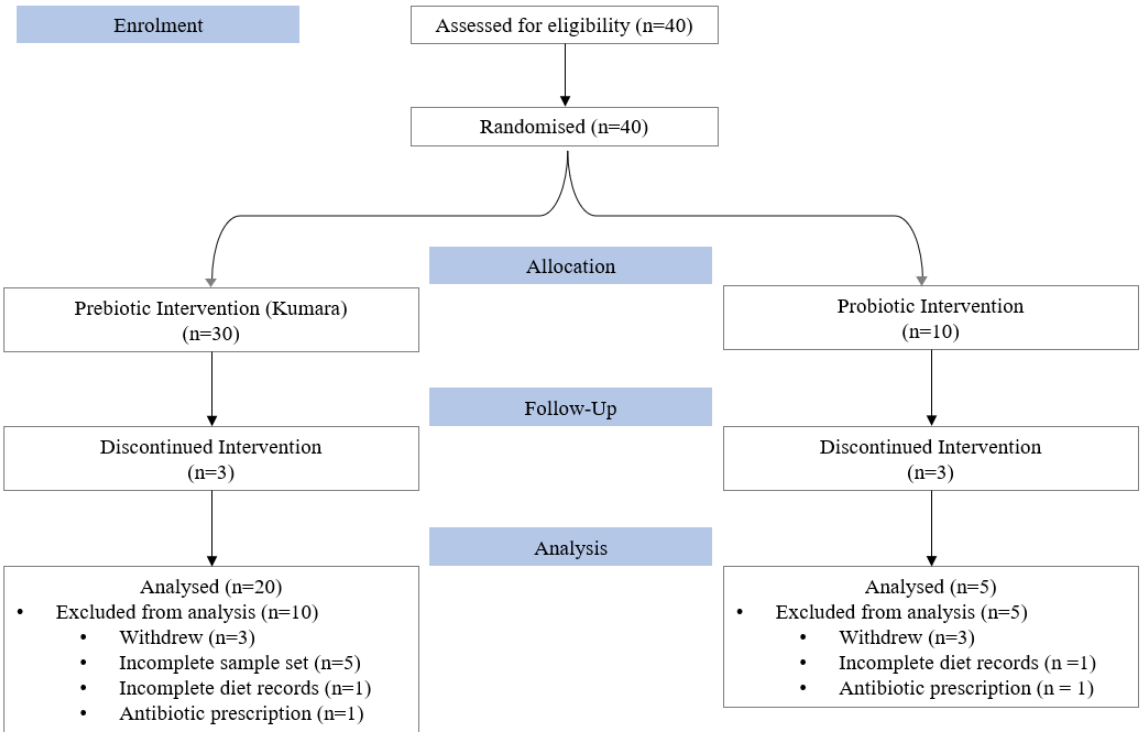


Figure 3. 1. CONSORT flow of Nourish to Flourish pilot study (ANZ Clinical Trials Registry number 12618000157279).

3.2.2. Sample collection

Parents collected stool samples from nappies during the three days before clinic visits. At the 4 months of age clinic visit, a nappy liner was provided to accommodate the liquid consistency of infant stools before introducing solid foods. Samples at later time points were primarily scooped into collection pottles. Parents were asked to store samples in the home freezer (-20 °C). They were provided with Styrofoam boxes and ice packs for transport to the clinic. Stool samples were stored at -80 °C until processed for microbiome and metabolome analyses, which were conducted in two batches: Batch 1 included samples from all infants at approximately 4 months of age, half of the 9-month-old samples, and six of the 12-month-old samples. Batch two had the remaining 9- and 12-month-old samples and a repeated 4-month-old sample from Batch 1 for quality control. Batch one was analysed 10 months prior to Batch 2, each processed upon receipt of samples to the laboratory and analysed for preliminary reporting throughout the study.

Saliva was collected on a sterile cotton stick. Parents placed the cotton stick in the side of the child's mouth and either coerced them or allowed them to chew on the cotton stick for approximately 10 min. Saliva was extracted from the cotton stick into a vial using a syringe and frozen immediately at -80 °C.

3.2.3. Faecal microbiome

A single faecal aliquot was vortexed centrifuged at 3,000 x *g* for 15 min to separate the bacterial pellet from the supernatant, then the supernatant was removed. This process was repeated for low volume samples to achieve approximately 100 mg to conduct DNA isolations. The extraction of DNA from faecal samples was conducted using a Macherey-Nagel DNA isolation kit, quantified using a Nanodrop Spectrophotometer (Thermo Fisher Scientific), and the approximate degree of degradation and fragmentation of DNA was visualised by gel electrophoresis.

The APC Microbiome Institute, Teagasc, part of the Agricultural and Food Development Authority in Cork, Ireland, conducted the shotgun metagenomic sequencing using Illumina NextSeq paired-end 150 bp x 2 sequencing. The subsequent sequence assignments were carried out by the AgResearch Bioinformatics & Statistics team at the Grasslands Research Centre. Libraries were prepared using the Illumina Nextera XT kit. Paired sequences were joined with PEAR version 0.9.6 (Zhang, Kobert, Flouri, & Stamatakis, 2014). Host sequences were detected and removed using the bbduk.sh function from the BBMAP package version 38.22-0 [PM2], with the human genome (Human GRCh38) as reference. Metaxa2 version 2.1.3a (Bengtsson-Palme et al., 2015) was used to identify SSU ribosomal DNA and taxonomy was assigned using the Silva 128 database (Quast et al., 2012). The "blastx" function of DIAMOND version 0.9.22 (Buchfink, Xie, & Huson, 2015) was used to map the reads against the "nr" NCBI database. Megan version 6 ultimate edition (Huson et al., 2016) was used to assign putative functions to the DIAMOND alignment files against the KEGG database (Kanehisa, Furumichi, Sato, Ishiguro-Watanabe, & Tanabe, 2021; Kanehisa & Goto, 2000; Overbeek et al., 2005). Taxonomic assignments were conducted using Metaxa2 against the Silva 1.28 SSU database.

Phylum and species-level taxonomic compositions were used for downstream analyses. Composition tables were filtered to exclude rare species, defined here as species present at <0.001% in fewer than three samples. A total of 116 level 6 Metaxa2 assignments and 184 KEGG pathway assignments (Level 3) were included in subsequent analyses. Total sum scaling within individual samples was used to indicate relative abundances of species and KEGG pathways.

3.2.4. Faecal metabolome

The normalisation of faecal samples by weight was conducted by lyophilising and homogenising samples. Biphasic extraction of 50 mg faecal sample was performed with

methanol:water:methyl tert-butyl ether (6:7:10v/v). An aliquot of the methanol:water phase containing the aqueous metabolites, and an aliquot of the methyl tert-butyl ether phase containing the lipids, were subsequently dried on a heating block (30 °C) under a continuous stream of nitrogen. Lipid extracts were reconstituted with chloroform:methanol (2:1 v/v) with a deuterated phosphatidylethanolamine (PE) internal standard. Aqueous samples were reconstituted with acetonitrile:water (1:1 v/v) with deuterated amino acid internal standards. Instrument methods, including columns, solvents, gradients, and spray voltage, were the same for both liquid chromatography-mass spectrometry (LC-MS) analysis batches. LC-MS metabolomics was performed using a high-resolution Orbitrap (Thermo) system, with polar metabolites separated using HILIC (Hydrophilic Interaction Chromotography) and lipids separated using a CSH-C18 reversed-phase separation (Cabrera, Kruger, Wolber, Roy, & Fraser, 2020; Fraser et al., 2012; Su et al., 2019). Both polar metabolite and lipid LC-MS analyses were performed in positive and negative ionisation modes.

The MSConvert function of ProteoWizard™ was used to convert Thermo RAW data files to mzML files. Peak detection, retention time alignment, grouping and gap-filling were carried out using XCMS (Smith, Want, O'Maille, Abagyan, & Siuzdak, 2006) and in-house scripts in R version 3.6.2 (Ihaka & Gentleman, 1996). The ‘diffreport’ function of XCMS was used to conduct a two-group comparison between (1) pooled quality control (QC) samples and blank samples and (2) QC and participant samples. The QC vs blank comparison was used to identify non-significant ($\alpha = 0.05$) *m/z* features that correspond to background noise contributed by blanks, which were subsequently removed from the QC vs samples list of features. The feature list generated by comparing QC vs samples was further cleaned by browsing extracted ion chromatograms generated by the diffreport function in XCMS and deleting low chromatographic quality *m/z* features that represented background noise. The filtered data

matrices were then subjected to Lowess normalisation using Workflow4Metabolomics (Galaxy) and subsequent filtering based on the coefficient of variance (>30% CV removed).

The resultant data matrix was used for downstream statistical analyses and identification by LipidSearch™, based on comparing the ddMS^2 data obtained from selected samples from *in silico* calculations and databases. The Q-Exactive database with the product search option was selected. Default parameters were used for LipidSearch™ quantitation. For this study, [M-H]$^-$ and [M+HCOO]$^-$ adducts, and [M+H]$^+$, [M+NH$_4$]$^+$, [M+Na]$^+$ adducts were selected for negative and positive ionisation modes, respectively. A retention time tolerance of 0.01 min and mass error tolerance of ± 5 ppm was allowed. Results and identifications from individual MS^2 files (LipidSearch™ results) were concatenated into a single table (one each for positive and negative ionisation modes) and matched with the respective, normalised, peak intensity table (XCMS results). Identification of water-soluble (aqueous) metabolites was conducted by aligning unique *m/z* and retention times with an in-house library of metabolites specific to the instruments and columns used at AgResearch, which has been compiled and tailored from external reference libraries, and searches against the Human Metabolome Database (HMDB) and METLIN. Concentrations of metabolites were log-transformed for subsequent analyses.

3.2.5. Saliva metabolome

Biphasic extraction of 100 µL of saliva was performed with CHCl$_3$:MeOH (50:50 v/v) with internal standards as described above. Aliquots of the aqueous and lipid phases were subsequently dried on a heating block (30 °C) under a continuous stream of nitrogen. Lipid extracts were reconstituted with modified Folch solution (CHCl$_3$:MeOH:H$_2$O, 66:33:1 v/v/v) containing 0.01% (w/v) 16:0 d$_{31}$-18:1-PE internal standard. Aqueous samples were reconstituted with CH$_3$CN:H$_2$O (1:1 v/v). LC-MS metabolomics was performed using a high-resolution Orbitrap (Thermo) system, with polar metabolites separated using HILIC chromatography and lipids separated using a CSH-C18 reversed-phase separation (Cabrera et

al., 2020; Fraser et al., 2012; Su et al., 2019). Both polar metabolite and lipid LC-MS analyses were performed in positive and negative ionisation modes. Raw MS data were processed in the same way as faecal MS data to convert mass spec data to peak tables.

3.2.6. Assessment and analysis of interventions

Due to the uneven distribution of intervention groups in a small cohort (n=30 consuming a food product, and n=10 consuming a commercial probiotic product), powered statistical comparison of the effects of the products compared to each other was impossible. Additionally, the caregivers of the intervention groups were not instructed to avoid infants to consumer the product given to the other group. Preliminary data analysis was done on total counts and relative abundances of sequences assigned to *Bifidobacterium animalis* subs. *lactis*, which was the strain of probiotics in the commercial product, and consumption of kumara was compared from 3-day food records. Compliance was assessed as part of the pilot study aims, including results. Analyses in subsequent chapters treat this cohort as an observational cohort without distinguishing the intervention groups.

3.3. Results

3.3.1. Participant randomisation

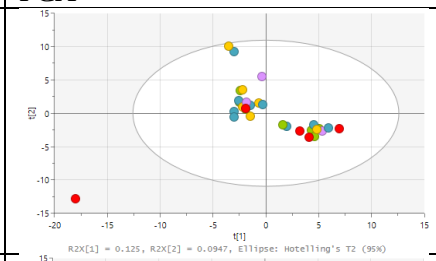
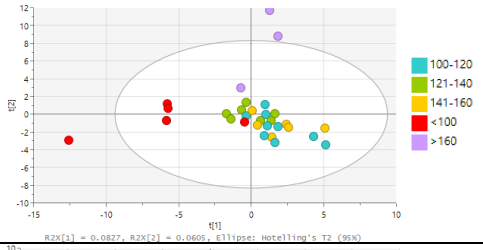
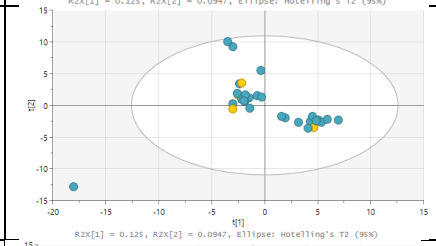
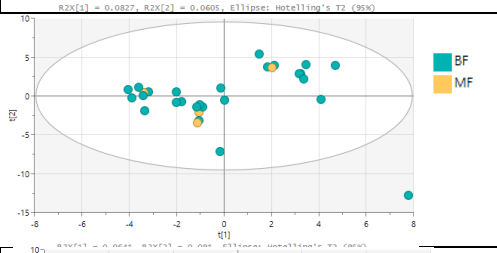
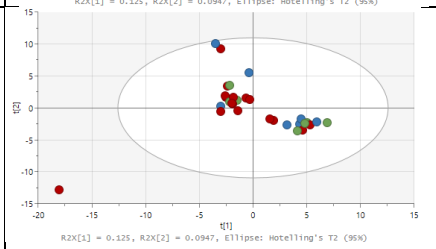
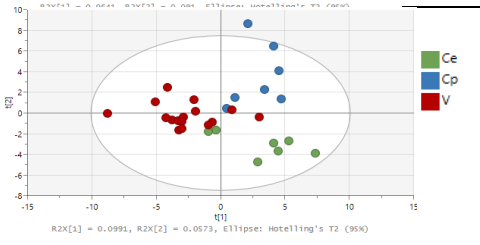
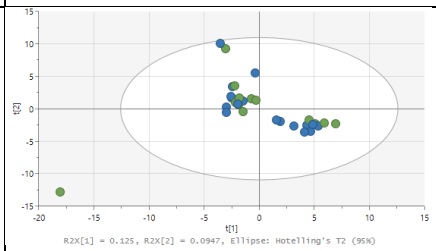
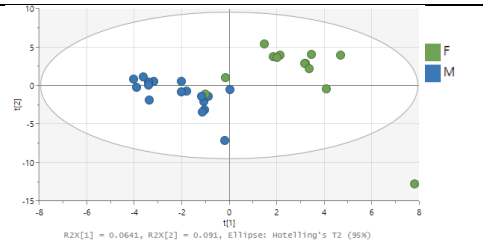
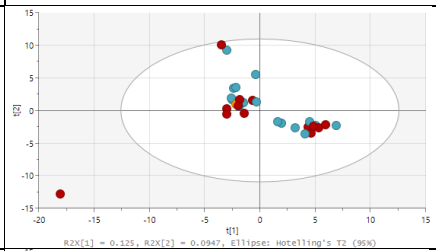
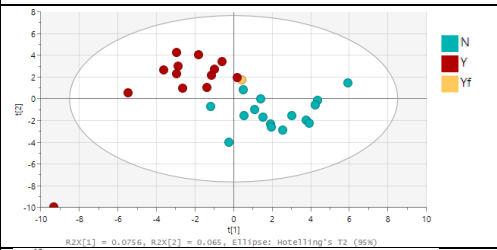
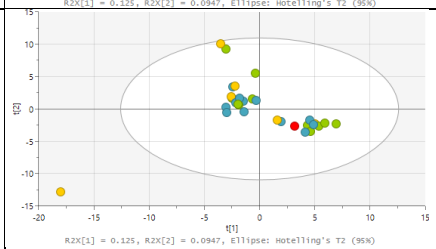
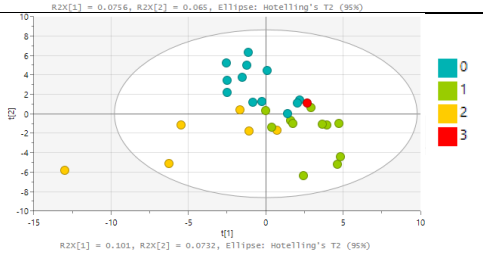
Principal Component Analyses (PCA) of species assigned to taxa and functional KEGG pathways from metagenomic sequencing of baseline samples yielded two clusters and outliers (Figures 3.2A and 3.2B). These clusters were unrelated to the biological and environmental factors that were recorded and analysed in this cohort, indicated by colour in Figure 3.2, and were unrelated to a batch effect, as baseline samples were analysed together.

Partial Least Squared Discriminate Analyses (PLSDA) were conducted to assess the potential predictive power of modelling variance in the baseline microbiome composition and gene abundances based on the environmental factors (Figures 3.2A and 3.2B). Negative Q^2 values indicate that the variable supervising the model is an irrelevant predictor of the variance, despite visual clustering when the model is forced to fit along two components. The Q^2 of the model for mode of delivery (birth) was positive in both composition and gene abundance profiling, indicating predictive relevance of this variable but not statistical significance. R^2 values indicate goodness of fit of the model, with values closer to one indicating better fit, R^2X representing the predictor variables and R^2Y representing the responder variables. These values must still be interpreted in the context of the Q^2 value. Microbiome composition models classified according to breastfeeding status and sex both showed high R^2Y values (≥ 0.80), but the combined R^2X and Q^2 values were unacceptable to indicate the predictive accuracy of the given model. Gene abundance models classified according to siblings, breastfeeding status, and age at sampling showed the highest R^2X values (≥ 0.60) for goodness of fit of the predictor variables, but the R^2Y and Q^2 values showed poor model response and predictive accuracy.

Table 3. 1. Characteristics of the Nourish to Flourish cohort.

	Kumara (n=23)	Probiotic (n=7)	Combined
Sex (n)			
Male	13	4	17
Female	10	3	13
Mode of delivery (n)			
Vaginal	12	4	16
Caesarean Section- unplanned	5	2	7
Caesarean Section- planned	6	1	7
Weight (g) at birth (mean ± SD)	3369.35 ± 422.65	3740.71 ± 378.31	3472 ± 443.07
Length (cm) at birth (mean ± SD)	50.89 ± 2.52	51.71 ± 1.67	51.28 ± 2.40
Siblings (n)			
0	10	2	12
1	9	2	11
2	3	3	6
3	1	0	1
Household pets (n)			
Yes	11	3	14
No	12	4	16
Age (days) at baseline sampling (mean ± SD)	121.91 ± 30.53	138.71 ± 15.90	125.83 ± 29.20
Age (days) at introduction of solids (mean ± SD)	168.83 ± 17.46	172.29 ± 13.18	196.27 ± 16.77
Age (days) at midpoint sampling (mean ± SD)	260.09 ± 17.64	278.57 ± 23.72	264.40 ± 21.11
Age (days) at completion of study (mean ± SD)	354.65 ± 17.45	356.71 ± 24.24	355.13 ± 19.60
Introduction of formula (n)			
Baseline	3	1	4
Midpoint	4	4	8
Endpoint	3	1	4
Compliance (mean ± SD)			
Midpoint	80%	82%	NA
Endpoint	90%	79%	NA
Asthma (n)	0	0	
Eczema (n)	5	3	8
Allergy (n)			
Environmental	1	0	1
Food	1	0	1
Antibiotic consumption (n)			
Baseline	0	0	0
Midpoint	3	3	6
Endpoint	7	2	9
Other prescription medication (n)			
Baseline	0	0	0
Midpoint	4	0	4
Endpoint	2	0	2

A. Sequences assigned to taxa

	PCA	PLS-DA: Forced Fit to 2 components	
Age at sampling			R^2X: 0.14 R^2Y: 0.33 Q^2: -0.10 ANOVA P: NA
Breastfeeding status			R^2X: 0.13 R^2Y: 0.87 Q^2: -0.18 ANOVA P: NA
Mode of delivery*			R^2X: 0.10 R^2Y: 0.33 Q2: 0.06 ANOVA P: 0.57
Sex			R^2X: 0.16 R^2Y: 0.80 Q^2: -0.18 ANOVA P: NA
Pets			R^2X: 0.14 R^2Y: 0.52 Q^2: -0.21 ANOVA P: NA
Siblings			R^2X: 0.17 R^2Y: 0.42 Q^2: -0.11 ANOVA P: NA

1362
1363

B. Sequences assigned to KEGG pathways

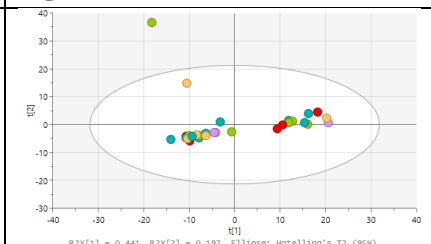
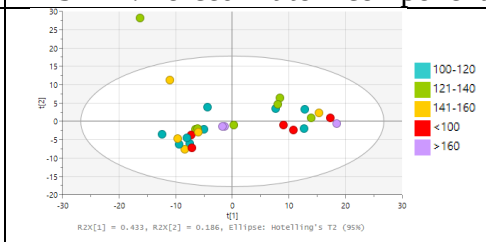
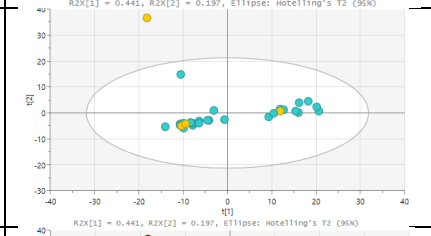
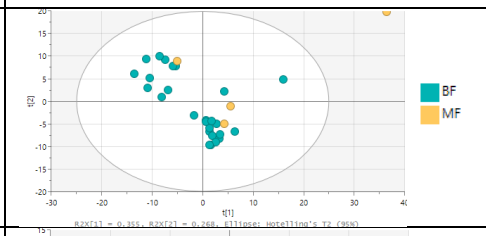
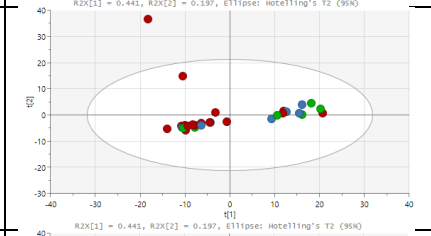
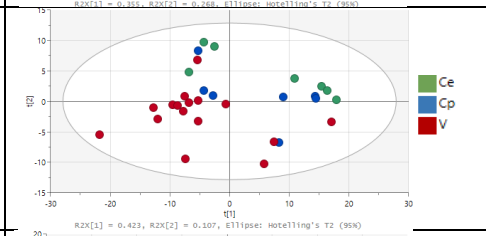
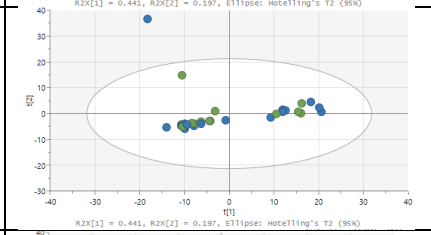
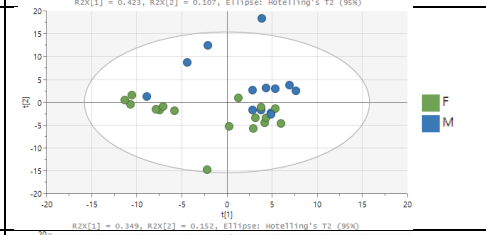
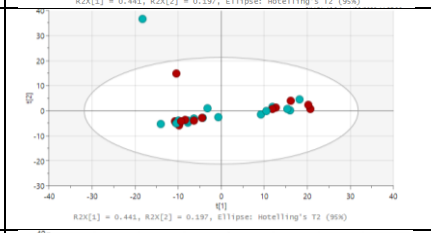
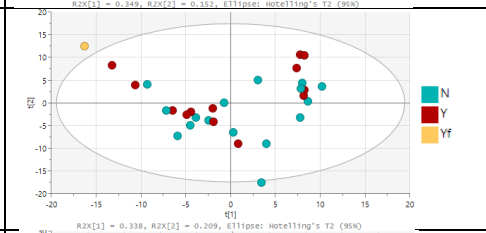
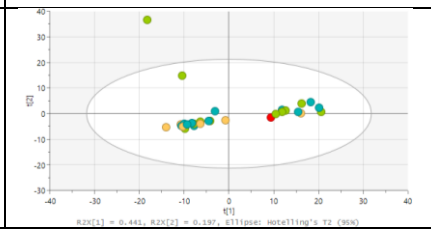
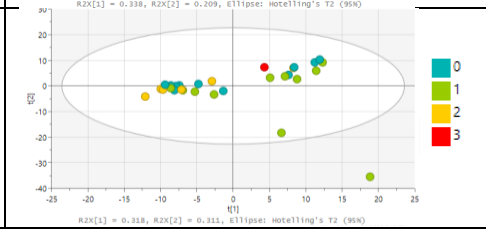
	PCA	PLS-DA: Forced Fit to 2 components	
Age at Sampling			R^2X: 0.62 R^2Y: 0.07 Q^2: -0.16 ANOVA P: NA
Breastfeeding status			R^2X: 0.62 R^2Y: 0.28 Q^2: -0.14 ANOVA P: NA
Mode of delivery*			* R^2X: 0.53 R^2Y: 0.34 Q2: 0.05 ANOVA P: 0.038
Sex			R^2X: 0.50 R^2Y: 0.44 Q^2: -0.02 ANOVA P: NA
Pets			R^2X: 0.55 R^2Y: 0.18 Q^2: -0.17 ANOVA P: NA
Siblings			R^2X: 0.63 R^2Y: 0.11 Q^2: -0.16 ANOVA P: NA

*Figure 3. 2. Variance at baseline in the faecal microbiome composition **A.** by Metaxa2 classifications and predicted KEGG pathways **B.** by KEGG pathways. A principal component analysis (unsupervised) coloured by variables of interest are shown to the left, and Partial Least Squared Discriminate Analysis (PLS-DA) supervised by the variable in question is shown to the right. R2 values indicate predictive accuracy, and Q2 values indicate predictive relevance for the PLS-DA models. *The values for the mode of delivery PLS-DA plots reflect a model built along with a single component rather than force-fit to two components. BF = Breast Fed, MF = Mixed Fed, Ce = Emergency caesarean section birth, Cp = Planned caesarean section birth, V = Vaginal birth, F = Female, M = Male, N = No Pets, Y = Household pets, Yf = Farm animals.*

3.3.2. Sample collection: protocols, contamination, and normalisation

In the Nourish to Flourish cohort, faecal, saliva, and urine samples were collected at each clinic visit. Blood collection was attempted at baseline and endpoints. Breastmilk samples were requested from mothers who were still breastfeeding. Of the sample types collected, faecal and saliva samples were chosen for analysis, which was conducted in two batches ten months apart. Figure 3.3 summarises the successes and challenges of faecal and saliva sample collections from the original 40 infants enrolled in the Nourish to Flourish study.

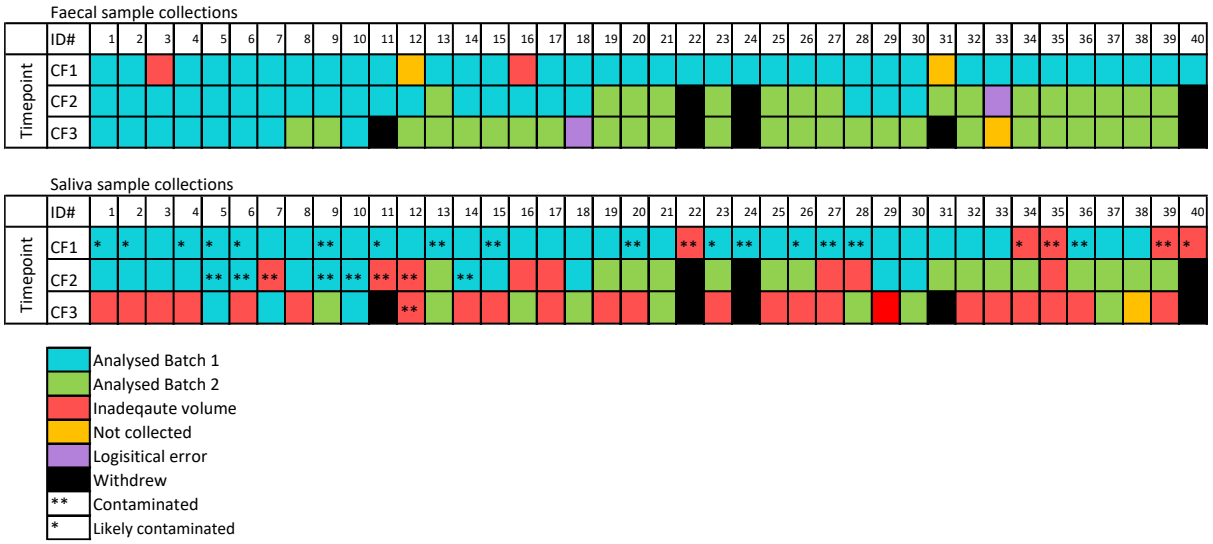


Figure 3. 3. Summary figure of faecal and saliva sample collections and attempted collections for the original cohort. Faecal sample collections are shown above. Saliva Collections are shown below. CF1 refers to the baseline visit at approximately 4 months of age. CF3 refers to the midpoint visit at 9 months of age. CF6 refers to the endpoint visit at 12 months of age. Samples included in different batch analyses are shown in various shades of green: teal in Batch 1, lime in Batch 2. Red squares indicate inadequate volume collected for analysis. Yellow squares indicate samples were not collected. Black squares indicate participant withdrawal. Purple squares indicate a logistical error in sample transport resulting in an extra freeze-thaw cycle. One asterisk in the saliva sample squares indicate likely contamination, and two asterisks indicate known contamination.

3.3.3. Faecal samples

One parent reported personal discomfort with the storage of stool samples in home freezers informed by cultural norms, which led to withdrawal from the study, highlighting the need to provide alternative frozen storage options. Parents were provided lint-free cloths (multiple colours, CHUX brand) to line nappies with, which could be easily removed for immediate storage in a pottle in a home freezer before transferring to the clinic. During baseline visits around 6 months of age when stools are primarily liquid, 38 out of 40 parents chose to utilise nappy liners. Only one parent opted to use the nappy liners during midpoint and endpoint visits.

Protocols for the isolation of microbiota remained consistent between sample collection methods, but extraction of faecal metabolites from nappy liners required minor protocol modifications to account for the absorption of aqueous extraction solvents by the nappy liners. The normalisation of faecal samples by volume was attempted by lyophilising and homogenising samples, as per standard good practices. The volume of sample that was available for LC-MS analysis was too low for successful homogenisation by volume for the majority of samples collected onto CHUX clothes. Normalisation was impossible for samples collected on CHUX clothes due to low volumes, contamination with urine, and variable amounts of CHUX with sample on it from storage in pottles. Due to the absorbent nature of CHUX, metabolite extractions were conducted directly from the clothes, which required a doubling of the aqueous solvent relative to lipid solvent during biphasic metabolite extractions. This protocol was repeated for consistency across all faecal samples.

3.3.4. Saliva contamination

Some infants had recently been fed either food or breastmilk before saliva sampling, typically to increase comfort during clinic visits. Some infants were more amenable to chewing on the saliva collection cotton stick, while others refused, creating extreme variability in sample volumes, which were generally insufficient for multiple biological assays and normalisation by

osmolarity. Younger infants were more amenable to chewing on cotton sticks than older infants, which has been observed in other infant studies focusing on the analysis of saliva samples (Neyraud et al., 2020). Chromatographs of lipid metabolites did not yield sufficient peaks to warrant further lipidomic analysis (data not shown). Normalised peak intensity tables were used to assess contamination in saliva samples. Of the analysed saliva samples, contamination was apparent in aqueous metabolomic datasets (Figure 3.4).

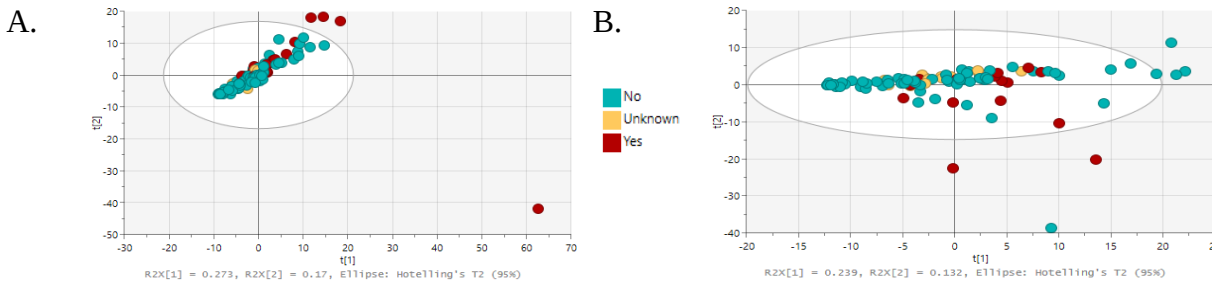


Figure 3. 4. ***Principal component analysis of aqueous metabolites detected by global metabolite profiling in saliva samples, with known and possible contaminations coloured in red and yellow, respectively. A.*** *Analyses included all saliva samples analysed.* ***B.*** *Analysis excluded contaminated outliers present in* ***A.***

3.3.5. Blood, urine, and breastmilk sample collection

The collection of blood and urine samples was unfeasible to analyse in the Nourish to Flourish cohort. Blood collection was attempted in infants whose parents opted to allow blood to be collected at the baseline visit. Collecting venous blood in sufficient quantities for plasma to be extracted and analysed proved unfeasible for this cohort, and subsequent blood collections were omitted. This experience emphasised the importance of ensuring that study teams include human resources such as neonatal phlebotomists.

Urine was collected during clinic visits using cotton pads inserted into nappies. Unfortunately, many baseline urine samples showed contamination with faecal material and were subsequently excluded from the analysis. Urine is an uncommonly used diagnostic fluid in infants for this reason. In some cases, urine samples have been collected using adhesive plastic bags, for example, to assess mannitol:lactulose excretion ratios during the investigation of the leaky gut syndrome and environmental enteric dysfunction (Trehan & Manary, 2015).

Breastmilk collection protocols are well-established and standardised. Mothers who provided breastmilk samples opted for whichever method was preferable before or during the clinic visit. Breastmilk samples were used to quantify immune factors only, which are thought to be consistent in foremilk and hindmilk samples. The breastmilk microbiome and metabolome were not investigated.

3.3.6. Comparison of interventions

Compliance with the interventions was higher in the kumara group (84% overall, 80% between 6 and 9 months of age, 90% between 9 and 12 months of age) than in the probiotic group (80%, 82% between 6 and 9 months of age, 79% between 9 and 12 months of age) (Table 3.1). The proportion of infants who withdrew from the probiotic intervention (30%) was higher than the infants who withdrew from the kumara intervention group (10%) (Figure 3.1). *Bifidobacterium animalis* subs. *lactis* was not identified in any samples collected at baseline prior to the start of the intervention but was identified in four samples at 9 months of age (two

from the kumara intervention group and two from the probiotic intervention group), and seven samples at 12 months of age at the end of the 6-month intervention (six from the kumara intervention group and one from the probiotic intervention group) (data not shown). Statistical comparison was irrelevant for unevenly distributed intervention groups with a high prevalence of zero values. All but two infants consumed kumara (non-intervention) during the three days of food record-keeping around sample collection during the study. Food records of the infants in the kumara group reported consuming more non-intervention kumara than infants in the probiotic intervention group (zero records of consumption) at 12 months of age (data not shown).

3.4. Discussion

The analyses carried out in Chapter 3 aimed to assess the feasibility of conducting systems biology research in infants and collecting various sample types for robust and sensitive downstream microbiome and metabolome analysis. The reported compliance to the interventions was sufficiently high to suggest that long-term probiotic and prebiotic interventions are feasible. Intervention with the prebiotic food product became increasingly well accepted by infants and their caregivers over the course of the study, while compliance to probiotic drops as a supplement decreased over time. Despite high reported compliance to the probiotic intervention, the specific species that was administered (*Bifidobacterium animalis* ssp *lactis*) was not detected in faecal samples of all of the participants in the probiotic intervention group, which suggests a misalignment between reported compliance and actual compliance. This may have been due to biological factors in the GIT, such as proximal colonisation, or heterogeneity of faecal samples collected on liners. This species was detected in several faecal samples from participants in the kumara intervention group, which was not surprising considering it is a popular probiotic strain in infant yogurts and kumara-consuming participants were not instructed to avoid probiotics. Similarly, kumara was reportedly consumed by infants randomised to the probiotic consuming group. The open-label nature of the study may have impacted parental feeding behaviours of the participating infants. Blinding and exclusion criteria are warranted in infant nutrition studies when possible.

Stool sample collection from infants in this age range presents challenges that can be reduced or overcome with minor protocol adjustments. Based on the modification of protocols needed for different methods of stool collection, it is recommended to provide liners only for samples collected prior to the introduction of solid foods and pottles for subsequent collections to accommodate the changing consistency of stool during the first year of life. A single type of non-biodegradable nappy liners without dyes or added chemicals could further reduce variation

and simplify sample processing protocols. A minor alteration in the instructions to parents to fold liners into a square with the centre of the stool sample at the corner and store flat in double Ziploc labelled bags in a home freezer until transport to the -80 °C clinic freezer could maximise the usable sample collected on a nappy liner and make standardisation by volume possible. The use of separation bags with buffer to remove stool samples from liners was suitable for microbiome analysis. Extracting samples from nappy liners for aqueous and lipid metabolite detection is possible with modified protocols, such as increasing the aqueous extraction solvent to account for reabsorption by the paper.

Recommendations can be made from the analysis of variation in faecal microbiome composition and gene content at baseline in the context of six relevant environmental exposures. First, the increased predictive power of classifying PLS-DA models of sequences assigned to taxa support the randomisation of participants to intervention groups according to the mode of delivery. However, this recommendation assumes that the mode of milk feeding (breastfed, formula-fed, or mixed feeding) is consistent across the cohort as stated in the inclusion and exclusion criteria. Surprisingly, there was little clustering of sequences assigned to taxa or KEGG pathways when classified by age. However, reducing the age range for collecting baseline samples, for instance, 100-140 days of age, would likely reduce variance in the microbiome composition.

Saliva is not recommended as a sample matrix for systems biology research in weaning infants, considering the limited range of aqueous metabolites detected, risk of contamination, and variable acceptability of saliva collection methods by weaning infants. Similarly, collecting urine samples is not recommended until later in infancy, when contamination between faecal and urine samples in nappies is less likely. Therefore, rather than collecting less invasive and more variable sample types, focusing resources on high-quality blood collections by neonatal phlebotomists would likely provide more valuable samples with substantially more useful

information and increase parental confidence in clinical protocols and decreased discomfort of the infants.

The primary limitation in this study was unevenly distributed intervention groups that made statistical comparison and interpretation about either intervention irrelevant.

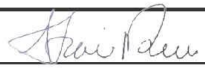
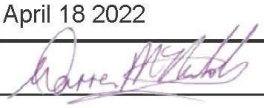
3.5. Conclusions

Systems biology research in weaning infants requires that protocols at homes, in the clinic, and during the sample processing be designed to account for the specificity and high resolution of metagenomic and metabolomic data and accommodate the developing physiology of infants. Accounting for the influence of mode of delivery and mode of milk feeding either through inclusion and exclusion criteria or during randomisation is justified in infant microbiome studies (Figures 3.2 and 3.3). Sample matrices should be chosen based on their physiological relevance to the research question, bearing in mind that fewer sample types collected well and processed consistently are preferable to more sample types of variable quality and usability. Study design principles around even intervention group distribution and blinding to invention products are critical to obtaining relevant and informative results when studying infant populations, even in a pilot study.

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Starin A. McKeen
Name/title of Primary Supervisor:	Prof Warren C. McNabb
Name of Research Output and full reference:	
Adaptation of the infant gut microbiome during the complementary feeding transition	
In which Chapter is the Manuscript /Published work:	Chapter 4
Please indicate:	
• The percentage of the manuscript/Published Work that was contributed by the candidate:	90
and	
• Describe the contribution that the candidate has made to the Manuscript/Published Work:	
The candidate conducted the lab work, wrote the manuscript, carried out the data analysis, created figures, and incorporated revisions from the remaining authors.	
For manuscripts intended for publication please indicate target journal:	
PLOS ONE	
Candidate's Signature:	
Date:	April 18 2022
Primary Supervisor's Signature:	
Date:	May 09 2022

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

1536

1537

Chapter 4. Adaptation of the infant GIT microbiome during complementary feeding

The infant GIT microbiome changes in composition and function during the introduction of solid foods throughout the first year of life. This study aimed to characterise changes in microbiome composition, gene abundances, and associated metabolites in faecal samples of healthy infants over the complementary feeding period. Faecal samples were obtained at three time points from infants participating in the Nourish to Flourish pilot study: before introducing solid foods at approximately 4 months of age, after introducing solid foods at 9 months of age, and after continued diet diversification at 12 months of age. The KEGG and taxonomy assignments were correlated with LC-MS metabolite profiles to identify patterns of co-abundance. The taxonomic composition of the faecal microbiome diversified during the first year of life, while the functional capacity present in the GIT microbiome remained stable. The introduction of solid foods between 4 and 9 months of age corresponded to a larger change in the relative abundance of sequences assigned to KEGG pathways and taxonomic assignments. It also leads to stronger correlations with faecal metabolites at 9 months of age, compared to the magnitude of changes and correlations seen during continued diet diversification between 9 and 12 months of age. Changes in aqueous faecal metabolites were more strongly correlated with KEGG pathway assignments of the faecal microbiome, while changes in lipid metabolites were associated with taxonomic assignments, particularly between 9 and 12 months of age. This study establishes trends in faecal microbiome composition and functional capacity occurring during the complementary feeding period and identifies potential faecal metabolite targets for future investigations.

4.1. Introduction

The infant GIT microbiome undergoes dramatic shifts in composition from the unstable neonatal microbiome towards a more stable microbiome throughout the first year of life (Bäckhed et al., 2015). This progression has been characterised as non-random in response to life events and environmental exposures, such as diet (Koenig et al., 2011). Because the introduction of solid foods is a disruptive event for the infant GIT ecosystem, it provides an opportunity to understand how it adapts in both composition and function in early postnatal life (He et al., 2019). It is accepted that the composition of microbial species diversifies throughout the first year of life. However, it is less clear how the functional capacity of the microbiome changes in response to complementary feeding: the addition of solid foods to the milk-based diet.

At the phylum level, the faecal microbiome of neonatal infants is dominated by approximately 70% of taxa from the Actinobacteria and Proteobacteria phyla and approximately 30% of taxa from the Firmicutes, Bacteroidetes, and Verrucomicrobia phyla (Gerald W Tannock et al., 2013). By adulthood, the faecal microbiome comprises approximately 90% bacteria from the Firmicutes and Bacteroidetes phyla and approximately 10% of bacteria from the Actinobacteria, Proteobacteria, and Verrucomicrobia phyla. The majority of this transition occurs within the first two to three years of life (Faith et al., 2013).

Taxa from the Actinobacteria phylum are efficient utilisers of oligosaccharides, one of the primary components of breastmilk, and thrive in the infant GIT (Ho et al., 2018). Those from the Proteobacteria and Bacteroidetes phyla utilise proteins and saccharides, substantial components of breastmilk and infant formula, which may allow members of these phyla to persist throughout the transition to diverse substrates from solid food (Tramontano et al., 2018). The Firmicutes phylum encompasses a diversity of families and species of microbes that collectively provide the enzymatic capability to utilise carbohydrates, proteins, saccharides,

and fermentative by-products of other microbes such as formate, lactate, succinate, acetate, and various gases (W. Wang, Hu, Zijlstra, Zheng, & Gänzle, 2019). This metabolic diversity allows taxa from the Firmicutes phyla to fill new resource niches that are created during dietary diversification. The GIT microbiome of milk-fed infants adapts to a rapidly changing diet during the complementary feeding period through shifts in populations of microbial species with a vast enzymatic capacity.

This study hypothesised that changes in the composition, rather than gene abundances, of the faecal microbiome of infants are more significant during the period when infants are introduced to solid foods (between 4 and 9 months of age), compared to approximately 3 months later (between 9 and 12 months of age), when the diet has continued to diversify. Therefore, shotgun metagenomics and LC-MS based metabolomics were conducted on faecal samples collected from infants at three different time points to characterise changes in the microbiome and metabolome following a period of dietary diversification. The time points were at 4.1 months of age ± 31 days (before the introduction of solid foods), at 8.7 months of age ± 18 days, and at 11.7 months of age ± 18 days (subsequently referred to as 4, 9, and 12 months of age respectively). In addition, sequences assigned to taxonomy and gene functions (KEGG pathways) and metabolomic data were integrated into a longitudinal analysis to identify key associated features.

4.2. Methods

Methods relating to clinical protocols, microbiome taxonomy assignments, functional group assignments based on gene abundances, and faecal metabolome are explained in Chapter 3.

4.2.1. Statistical and bioinformatic analyses

Statistical analyses were conducted in R version 4.0.2 on total sum scaled taxonomy, metagenome datasets, and log-transformed metabolome data (R Core Team, 2021). These methods maintain positive values, which is critical when assessing the direction of change in relative abundance or concentration. Ordination analyses, including sparse Principal Component Analysis (sPCA) and sparse Partial Least Squared Discriminant Analysis (sPLSDA) and regularised Canonical Correlation, were carried out using the 'MixOmics' package version 6.1.1 in R version 4.0.2 after filtering low-abundance reads and total sum scaling (Rohart, Gautier, Singh, & Lê Cao, 2017). Further identification of variables important to projection (VIPs) was identified using MixOmics, and a variable's scores from each component were combined into a sum VIP score to better understand the contribution of a variable to the given model.

Regularised canonical correlations were based on Pearson's correlations of the delta (change) between time points and were carried out on three components using the clustered image heatmap function in MixOmics before filtering significant correlation coefficients from the resulting matrix. A correlation cut-off of 0.6 was used for the KEGG pathway correlations, whereas a correlation cut-off of 0.5 was used for species correlations based on differences in the significance of correlations between the given datasets. In addition, permutational analysis of variance (PERMANOVA) was conducted in the 'RVAideMemoire' package version 0.9-7.8 at 2000 permutations per time point comparison, and Benjamin-Hochberg false discovery rate (FDR) corrected p-values used to determine significance ($p \leq 0.05$).

4.3. Results

As a pilot investigation, the Nourish to Flourish study demonstrated the safety and feasibility of conducting systems biology research in infants during complementary feeding. Results presented here will focus on the changes in faecal microbial composition and predicted functional capacity over the complementary feeding period and associated changes in faecal metabolite concentrations.

4.3.1. Broad trends in the faecal microbiome composition and function

Before introducing solid foods, taxa from the Actinobacteria phylum, such as the *Bifidobacterium* genus, were the most abundant (Figure 4.1A). Taxa from the Firmicutes phylum gradually increased in relative abundance from approximately 25% of sequences at 4 months of age to approximately 53% assigned to phyla at 12 months of age (Figure 4.1A). Taxa from both Actinobacteria and Proteobacteria phyla decreased in relative abundance (by 10% and 20%, respectively), and members of the Bacteroidetes phylum increased by 8% over time (Figure 4.1A). Species belonging to Verrucomicrobia, Fusobacteria, Cyanobacteria, Saccharibacteria, and Tenericutes phyla were present in much lower proportions at all time points (Figure 4.1A).

Metagenomic sequences were annotated to Signalling and cellular processes KEGG pathways at Level 2 classification at 4 months of age and decreased from approximately 19.4% to 17.1% of total sequences annotated (Figure 4.1B). Genes assigned to Genetic information processing increased from approximately 15.1% at 4 months of age to 17.2% by 12 months of age (Figure 4.1B). Sequences assigned to Metabolism of carbohydrates, proteins, and amino acids remained at approximately 9%, 8%, and 6%, respectively, of total sequences annotated to Level 2 KEGG pathway classifications over time (Figure 4.1B). The variation in the relative abundance of sequences assigned to Level 2 KEGG pathways was smaller than the differences in the relative abundance of sequences assigned to dominant phyla, shown by the shaded areas

around trend lines, and the relative abundances were smaller for KEGG pathways than for phyla (Figure 4.1A-B).

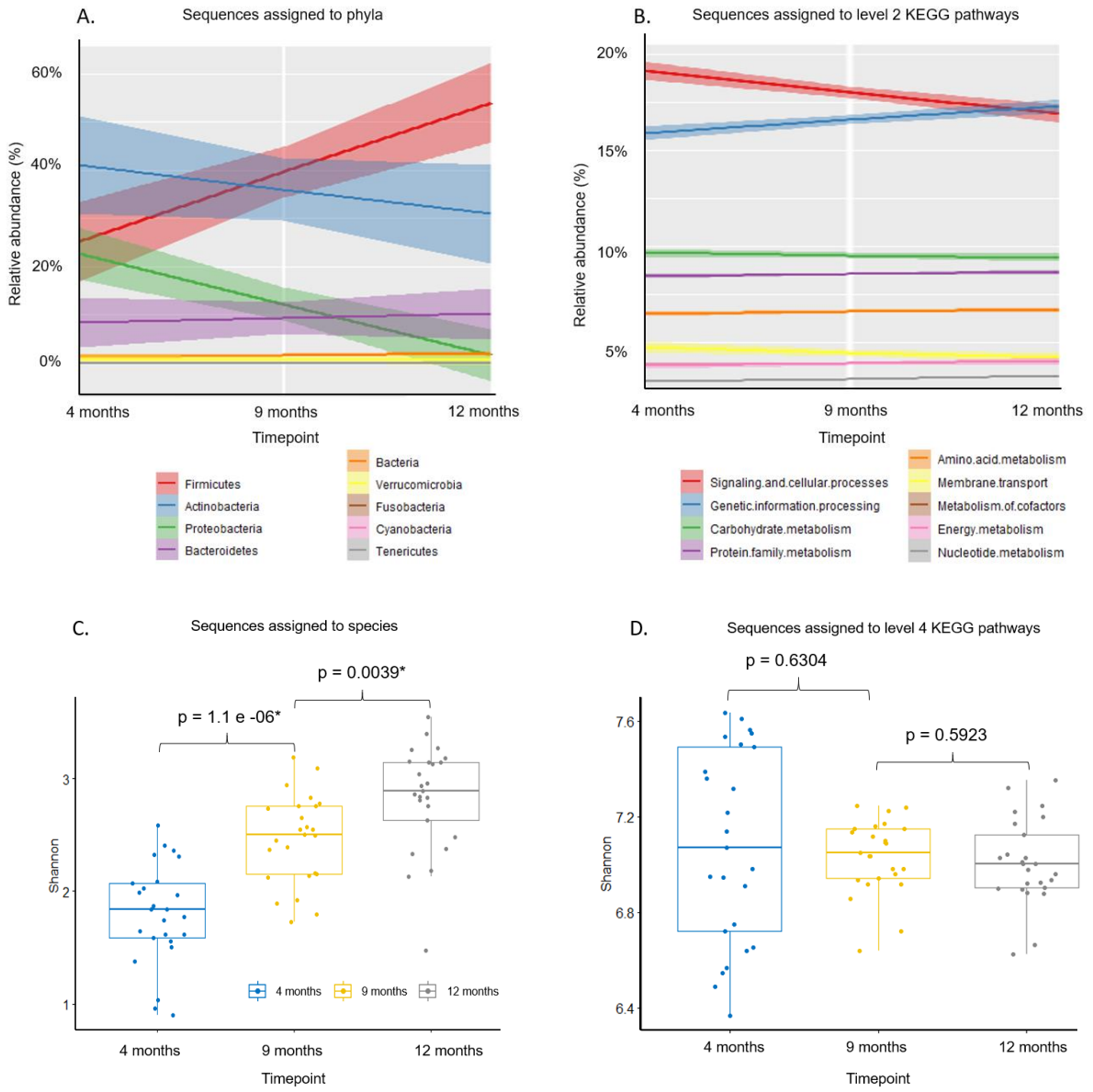


*Figure 4. 1. Panels **A** and **B** show trends in sequences assigned to phyla and Level 2 KEGG pathways at 4 months of age, 9 months of age, and 12 months of age. Panels **C** and **D** show Shannon alpha diversity index at each time point, with significant differences evaluated by t-test between sampling time points for sequences assigned to Level 7 Metaxa2 species and sequences assigned to Level 4 KEGG pathways. The 4-month-old time point is shown in blue, 9 months in orange and 12 months in grey.*

The diversity of species in the infant faecal microbiome increased significantly with the addition of solid foods to the milk-based diet (Figure 4.1C). However, the diversity of functional genes remained, on average, the same, as measured by the Shannon alpha diversity index in the R package "vegan" at Level 7 of Metaxa 2 taxonomic assignments and Level 4 KEGG pathway assignments (Figure 4.1D). The increase in taxonomic alpha diversity was more significant between 4 and 9 months of age than between 9 and 12 months of age (Figure 4.1C). The alpha diversity scores for relative abundances of sequences assigned to KEGG pathways were more variable at 4 months of age, but the difference narrowed at 9 and 12 months of age (Figure 4.1D).

4.3.2. Shifts in the faecal microbiome and metabolome between time points

Both sPCA and sPLS-DA were conducted on the faecal microbiome genus-level taxonomic composition, faecal gene composition aggregated at KEGG Level 3 pathways, and aqueous and lipid metabolites in faecal samples to visualise the changes in the faecal microbiome throughout the complementary feeding period (Figure 4.2). The sPCA did not show any obvious separation between ages. The application of sPLS-DA, a supervised ordination analysis, did not increase the separation between time points. However, the overlap between time point clusters was reduced in the sPLS-DA models of sequences assigned to KEGG pathways, aqueous metabolites, and lipid metabolites of the faecal samples (Figure 4.2D). The percentage of explained variance was lower at 12 months of age than 9 months of age for Metaxa2 assignments, KEGG pathway assignments, and metabolites (Figure 4.2). KEGG pathway assignments had the highest percentage of explained variance at both time points, followed by aqueous metabolites (Figure 4.2).

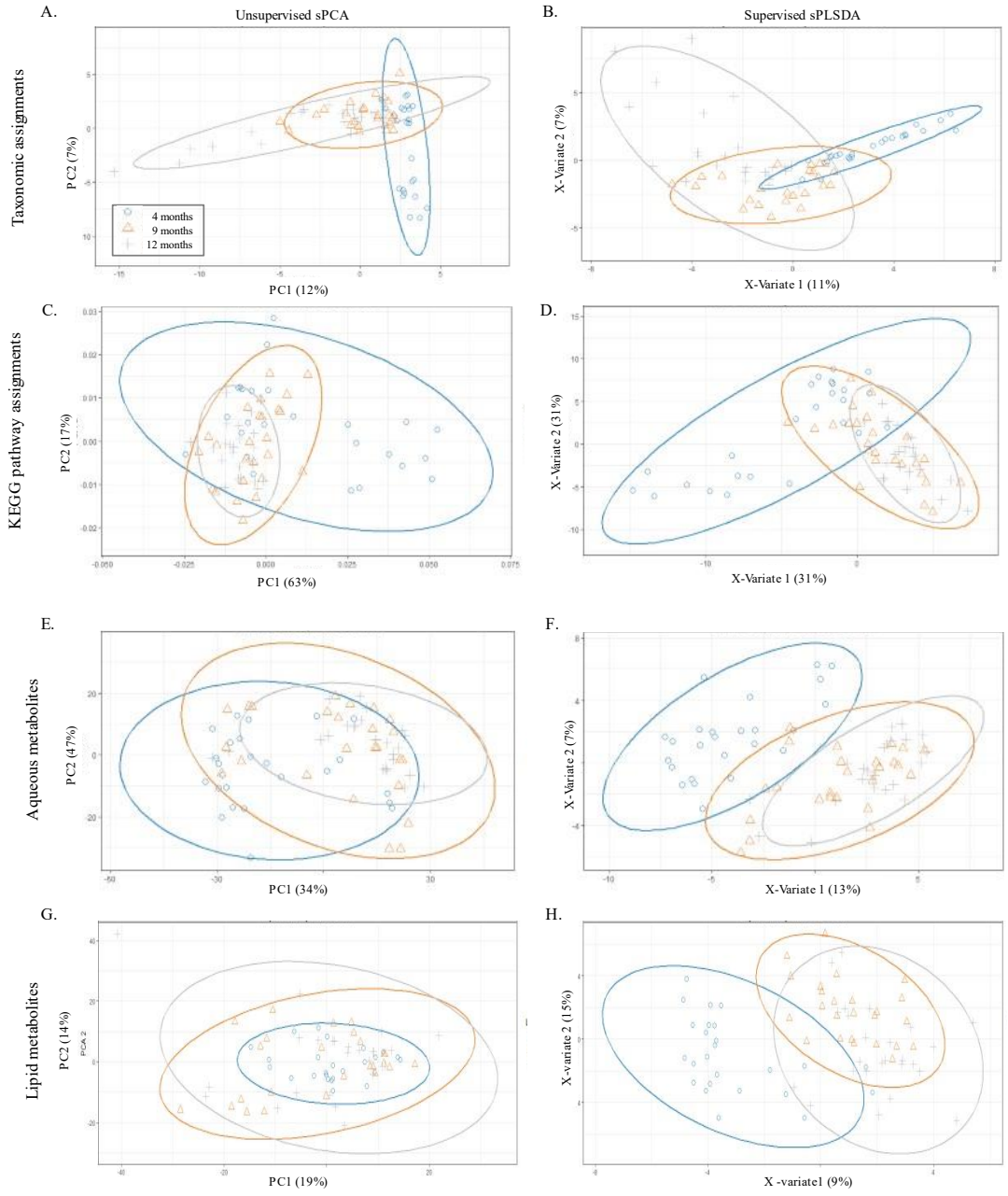


*Figure 4. 2. Ordination analyses of taxonomic assignments are shown in panels **A** and **B**, KEGG pathway assignments in panels **C** and **D**, aqueous metabolites in panels E and F, and lipid metabolites in panels **G** and **H**. Sparse Principal Component Analyses are shown in the left and column (**A, C, E, and G**). Sparse Partial Least Squared Discriminant Analysis plots supervised by time point and cast on three components are shown in the right-hand column (**B, D, F, and H**). Four-month-old samples are depicted in blue, 9-month-old samples in orange, and 12-month-old samples in grey with individuals marked in circles, triangles, and crosses. Ellipses represent 95% confidence of clustered samples. The explained variance (%) is listed in parenthesis on each axis.*

The taxonomic composition of the faecal samples was most similar between individual infants at 9 months of age, compared to at 4 or 12 months of age (Figure 4.2A). Sequences assigned to KEGG pathways became increasingly similar between individuals over time (Figure 4.2C), agreeing with the alpha diversity patterns shown in Figure 3D. Sequences assigned to Metaxa2 classifications and KEGG pathways separated into two distinct clusters at 4 months of age in both sPCA and sPLS-DA models (Figure 4.2C and 4.2D). The differences between these infants diminished at 9 and 12 months of age.

Metabolite profiles in the faecal samples were more variable between infants at all time points than microbiome taxonomy or KEGG pathway assignments, as shown by the scales on which the sPCA ordinations were cast (Figure 4.2). Supervision by time point in sPLS-DA models separated the 4 months of age samples, but the 9 and 12 months of age metabolite profiles were similar in these models (Figure 4.2F and 4.2G). Aqueous metabolite profiles became more similar between infants at the 9-month time point, compared to infants at either 4 or 12 months of age (Figure 4.2E). Lipid metabolite profiles of 4-month-old samples were similar among infants in the sPCA model (Figure 4.2G). However, these similarities diminished at 9 and 12 months (Figure 4.2G).

Separate sPLS-DA models were conducted to compare taxonomy assignments and KEGG pathway assignments of the faecal samples between 4 and 9 months of age and 9 and 12 months of age. VIPs important to sPLS-DA were selected from these two-time point models, and importance was derived by adding variables' VIP scores from each of the three components, thus indicating how that variable contributed to the separation of each time point.

PERMANOVA was used to compare the significance of the difference between the means and spread of relative abundance of individual taxonomic assignments and individual sequences assigned to KEGG pathways of the faecal samples between time points. Variables considered significantly different by PERMANOVA, which constituted a substantial

proportion of the overall abundance, are shown in Tables 4.1 and 4.2 along with cumulative VIP scores, relative abundances, and log2 fold change between time points. Mean relative abundances of zero for any given time point were replaced with a value of 10^{-6} for statistical calculations. Metabolite data were omitted from this analysis considering the small sample size, high variability, and semi-quantitative nature of untargeted metabolite profiling.

*Table 4. 1. Taxonomic compositional changes over time, grouped by phylum and direction of change in relative abundance that were significantly different in either time point comparison by PERMANOVA analysis. Taxa with a false discovery rate (FDR) corrected p-value of >0.05 at either time point were omitted, apart from 1, which was not significant by PERMANOVA but constituted a large proportion of the faecal microbiome indicated by **. Log2 fold change (Log2FC) between mean relative abundances at each time point are coloured in orange for increasing fold change and blue for decreasing fold change. Shading in the variable important to projection (VIP) column indicates the most important variables to the sPLS-DA projections in Figure 4.3, with darker shading indicating higher cumulative VIP scores. Trend arrows indicate the direction of change in abundance between time points; bent arrows show the direction of change between the first and second-time points, followed by the change from the second to third-time points. Additional data for species not included above is available in Appendix 1.*

		Relative abundance (%)			4 vs 9 months			9 vs 12 months			
Phylum	Taxa	4 months	9 months	12 months	Log2 FC	VIP Sum	fdr p-value	Log2FC	VIP Sum	fdr p-value	Trend
Actinobacteria	*Bifidobacteriaceae Bifidobacterium longum*	0.55%	0.33%	0.19%	-0.73	2.71	0.04	-0.80	3.38	0.04	
	Bifidobacteriales	0.02%	0.01%	0.00%	-0.96	3.93	0.01	-1.69	3.13	0.01	
	*Bifidobacteriaceae***	36.82%	30.68%	26.79%	-0.26	2.18	0.50	-0.20	0.31	0.51	
Firmicutes	*Ruminococcaceae Subdoligranulum*	0.00%	0.31%	0.96%	8.19	1.27	0.01	1.63	3.57	0.01	
	Ruminococcaceae Ruminococcus	0.00%	0.10%	0.47%	5.73	1.22	0.01	2.17	3.45	0.02	
	Ruminococcaceae Faecalibacterium	0.01%	1.11%	3.62%	6.66	1.84	0.00	1.70	6.60	0.00	
	Ruminococcaceae	0.18%	0.77%	2.06%	2.13	2.33	0.00	1.42	7.19	0.00	
	Ruminococcaceae Ruminococcaceae	0.01%	0.03%	0.16%	1.86	0.17	0.01	2.67	4.39	0.01	
	Lachnospiraceae Tyzzerella	0.01%	0.02%	0.01%	3.13	1.93	0.01	1.82	2.48	0.01	
	Lachnospiraceae Sellimonas	0.00%	0.02%	0.08%	0.96	0.43	0.03	2.25	3.40	0.03	
	Lachnospiraceae Roseburia	0.01%	0.04%	0.19%	7.42	1.41	0.04	0.78	0.57	0.03	
	Lachnospiraceae Lachnospiraceae	0.00%	0.80%	0.28%	2.34	1.10	0.00	2.15	3.72	0.00	
	Lachnospiraceae Coprococcus	0.00%	0.67%	0.55%	1.69	0.43	0.00	3.12	7.98	0.00	
	Lachnospiraceae Blautia	0.00%	0.03%	0.01%	2.43	3.45	0.01	0.70	1.29	0.00	
	Lachnospiraceae	1.89%	11.00%	13.71%	2.54	7.57	0.00	0.32	0.05	0.00	
	Erysipelotrichaceae Turicibacter	0.00%	0.00%	0.02%	3.86	2.45	0.04	2.56	0.96	0.04	
	Unclassified Clostridia	0.01%	0.06%	0.08%	2.44	8.64	0.00	0.40	0.79	0.00	
	Clostridiales	1.22%	4.08%	6.13%	1.74	6.13	0.00	0.59	2.98	0.00	
	Bacillales	0.00%	0.01%	0.06%	1.16	0.04	0.01	2.49	5.28	0.01	
	Peptostreptococcaceae Intestinibacter	0.00%	0.06%	0.07%	4.78	10.27	0.01	0.15	1.93	0.00	
	Peptostreptococcaceae	0.06%	0.38%	0.37%	2.75	5.04	0.04	-0.07	1.67	0.04	
	Peptostreptococcaceae Peptoclostridium	0.02%	0.16%	0.12%	3.14	5.07	0.03	-0.43	1.89	0.01	
	Veillonellaceae	0.17%	1.36%	1.04%	2.98	7.68	0.01	-0.39	0.55	0.01	
	Firmicutes Unclassified	0.28%	0.91%	0.79%	1.69	3.52	0.04	-0.20	0.93	0.02	
	Clostridiaceae Clostridium sensu stricto	0.03%	0.00%	0.30%	-2.40	2.36	0.01	-0.96	5.91	0.01	
	Clostridiaceae Clostridium neonatale	0.02%	0.00%	0.00%	-6.62	3.23	0.00	-27.75	1.46	0.00	
	Ruminococcaceae Butyricicoccus	0.04%	0.01%	0.31%	-2.52	2.30	0.01	5.62	6.76	0.00	
	Erysipelotrichaceae Holdemanella	0.00%	0.00%	0.06%	-0.36	0.96	0.01	7.38	5.39	0.01	
	Erysipelotrichaceae Erysipelotrichaceae	0.00%	0.00%	0.15%	-0.67	0.35	0.00	6.12	4.60	0.00	
	Bacillaceae	0.00%	0.00%	0.02%	-2.68	0.12	0.00	7.93	5.71	0.01	
	Bacillaceae Bacillus	0.01%	0.00%	0.20%	-0.58	0.03	0.01	5.79	5.87	0.01	
	Christensenellaceae Christensenellaceae	0.00%	0.00%	0.10%	-1.50	1.47	0.00	6.91	6.25	0.01	
Proteobacteria	*Enterobacteriaceae Enterobacteriaceae*	7.21%	2.10%	1.09%	-1.77	2.84	0.00	-1.17	0.88	0.01	
	Enterobacteriaceae Escherichia-Shigella	0.58%	0.06%	0.02%	-1.78	5.65	0.00	-0.94	3.90	0.00	
	Enterobacteriaceae Salmonella enterica PA	0.08%	0.02%	0.01%	-2.07	6.53	0.00	-1.75	6.41	0.00	
	Enterobacteriaceae	16.40%	3.82%	1.87%	-2.10	7.83	0.00	-1.04	6.02	0.00	
	Enterobacteriales	0.05%	0.00%	0.00%	-3.47	7.68	0.00	-0.28	0.55	0.00	
	Gammaproteobacteria Unclassified	0.84%	0.18%	0.13%	-2.19	8.24	0.00	-0.51	1.27	0.00	
	Vibrionaceae Vibrio	0.02%	0.01%	0.01%	-1.26	2.93	0.04	-0.17	0.01	0.03	
	Moraxellaceae Acinetobacter	0.00%	0.00%	0.38%	-0.41	1.93	0.01	9.66	5.38	0.01	
	Pseudomonadaceae Pseudomonas	0.00%	0.00%	0.12%	-1.71	0.20	0.04	8.17	5.22	0.02	
	Proteobacteria Unclassified	0.04%	0.01%	0.03%	-1.53	5.2	0.047	0.84	2.0	0.060	
	Micrococcaceae Rothia	0.05%	0.02%	0.02%	-1.49	3.22	0.04	0.04	0.08	0.03	

Table 4. 2. Changes in KEGG pathway assignments in faecal samples over time, grouped by parent pathways. The table is filtered to include KEGG pathway assignments that were significantly different in either time point comparison by PERMANOVA analysis and had a log2 fold change >|0.2|. Log2 fold change (Log2FC) between mean relative abundances at each time point are coloured in orange for increasing fold change intensity and blue for decreasing fold change intensity. Shading in the variable important to projection (VIP) column indicates the most important variables to the sPLS-DA projections in Figure 4.3, with darker shading indicating higher cumulative VIP scores. Trend arrows indicate the direction of change in abundance between time; bent arrows show the direction of change between the first and second-time points, followed by the change from the second to third-time points. Additional data for species not included above are available in Appendix 2.

KEGG pathways level 1 & 2		KEGG pathways level 3	Relative abundance (%)			4 vs 9 months			9 vs 12 months			Trend
			4 months	9 months	12 months	Log2 FC	VIP Sum	fdr p-value	Log2FC	VIP Sum	fdr p-value	
Brite Hierarchies	Protein families: signaling and cellular processes	ko04812 Cytoskeleton proteins	0.25%	0.32%	0.34%	0.36	5.32	0.00	0.06	0.57	0.00	↗
		ko01504 Antimicrobial resistance genes	0.36%	0.28%	0.26%	-0.34	0.51	0.01	-0.11	2.40	0.01	↘
		ko02044 Secretion system	1.35%	0.89%	0.85%	-0.61	0.71	0.00	-0.06	0.09	0.00	↘
		ko02035 Bacterial motility proteins	0.50%	0.30%	0.29%	-0.73	0.54	0.02	-0.06	0.20	0.01	↘
		ko02000 Transporters	11.19%	9.99%	9.51%	-0.16	4.64	0.00	-0.07	6.09	0.00	↘
	Protein families: metabolism	ko00194 Photosynthesis proteins	0.25%	0.31%	0.34%	0.30	4.01	0.00	0.13	7.05	0.00	↗
		ko01008 Polyketide biosynthesis proteins	0.10%	0.05%	0.03%	-1.09	0.09	0.01	-0.69	1.42	0.01	↘
		ko01005 Lipopolysaccharide bios. proteins	0.43%	0.35%	0.29%	-0.33	0.34	0.03	-0.24	2.54	0.03	↘
		ko01001 Protein kinases	0.65%	0.50%	0.50%	-0.39	7.13	0.00	0.00	0.64	0.00	↘
		ko01009 Protein phosphatases & proteins	0.12%	0.16%	0.16%	0.34	7.28	0.00	-0.01	0.17	0.00	↗
Cellular community - prokaryotes		ko05111 Biofilm formation- *Vibrio cholerae*	0.33%	0.21%	0.17%	-0.61	0.32	0.00	-0.30	5.50	0.00	↘
		ko02026 Biofilm formation- *Escherichia coli*	0.51%	0.35%	0.34%	-0.55	2.32	0.00	-0.02	0.11	0.00	↘
		ko02025 Biofilm formation- *Pseudomonas aeruginosa*	0.25%	0.15%	0.14%	-0.75	8.05	0.00	-0.12	0.77	0.00	↘
Environmental Information Processing		ko04151 PI3K-Akt signaling pathway	0.02%	0.03%	0.04%	0.79	0.83	0.00	0.17	2.15	0.00	↗
		ko02020 Two-component system	1.74%	1.36%	1.34%	-0.35	0.43	0.02	-0.03	1.73	0.03	↘
		ko02060 Phosphotransferase system (PTS)	0.72%	0.55%	0.49%	-0.39	0.15	0.03	-0.17	0.37	0.03	↘
Genetic Information Processing		ko03013 RNA transport	0.02%	0.04%	0.04%	0.55	1.75	0.02	0.24	0.70	0.03	↗
		ko03000 Transcription factors	1.51%	1.13%	1.12%	-0.41	5.52	0.00	-0.02	0.24	0.00	↘
		ko04141 Protein processing in ER	0.04%	0.05%	0.05%	0.43	2.49	0.00	-0.03	0.39	0.01	↗
Metabolism	Xenobiotics	ko00627 Aminobenzoate degradation	0.04%	0.02%	0.02%	-0.67	0.16	0.00	-0.14	0.20	0.01	↘
		ko00362 Benzoate degradation	0.10%	0.06%	0.06%	-0.69	0.09	0.03	-0.01	0.19	0.04	↘
		ko00980 Metabolism by cytochrome P450	0.05%	0.03%	0.03%	-0.72	9.98	0.00	-0.24	2.20	0.00	↘
		ko00982 Drug metabolism - cytochrome P450	0.05%	0.03%	0.03%	-0.72	9.94	0.00	-0.25	2.31	0.00	↘
		ko00930 Caprolactam degradation	0.02%	0.01%	0.01%	-1.38	0.98	0.00	-0.22	0.04	0.00	↘
	Terpenoids and polyketides	ko01053 Biosynthesis of siderophore nr peptides	0.12%	0.05%	0.04%	-1.17	0.02	0.00	-0.56	2.66	0.00	↘
		ko00281 Geraniol degradation	0.04%	0.01%	0.01%	-1.68	2.85	0.00	-0.46	0.68	0.00	↘
		ko00903 Limonene and pinene degradation	0.05%	0.03%	0.03%	-0.52	4.36	0.00	-0.24	2.22	0.00	↘
	Cofactors and vitamins	ko00860 Porphyrin and chlorophyll metabolism	0.36%	0.47%	0.48%	0.38	3.22	0.03	0.03	1.80	0.03	↗
		ko00130 Ubiquinone & terpenoid-quinone bios.	0.19%	0.14%	0.10%	-0.43	0.15	0.00	-0.40	8.56	0.00	↘
		ko00830 Retinol metabolism	0.04%	0.03%	0.02%	-0.45	2.36	0.00	-0.21	1.01	0.00	↘
	Lipid metabolism	ko00121 Secondary bile acid biosynthesis	0.05%	0.06%	0.07%	0.44	5.28	0.00	0.13	0.81	0.00	↗
		ko01040 Biosynthesis of unsaturated fatty acids	0.03%	0.02%	0.02%	-0.57	2.75	0.00	-0.04	0.78	0.00	↘
	Glycan	ko00540 Lipopolysaccharide biosynthesis	0.27%	0.18%	0.14%	-0.54	0.30	0.02	-0.43	2.99	0.03	↘
	Energy	ko00195 Photosynthesis	0.25%	0.31%	0.34%	0.31	4.07	0.00	0.13	7.15	0.00	↗
	Carbohydrate	ko00053 Ascorbate and aldarate metabolism	0.20%	0.16%	0.14%	-0.32	0.19	0.01	-0.18	2.88	0.00	↘
	Secondary metabolite bios.	ko00966 Glucosinolate biosynthesis	0.05%	0.06%	0.06%	0.33	0.14	0.00	0.10	2.73	0.00	↗
		ko00333 Prodigiosin biosynthesis	0.06%	0.07%	0.07%	0.30	0.79	0.01	0.05	0.64	0.00	↗
	Amino acid	ko00310 Lysine degradation	0.24%	0.17%	0.15%	-0.51	0.60	0.00	-0.21	7.33	0.00	↘
		ko00380 Tryptophan metabolism	0.18%	0.12%	0.11%	-0.57	6.11	0.00	-0.22	6.30	0.00	↘
		ko00480 Glutathione metabolism	0.40%	0.32%	0.29%	-0.30	6.33	0.00	-0.15	2.64	0.00	↘
		ko00410 beta-Alanine metabolism	0.14%	0.11%	0.10%	-0.37	0.25	0.02	-0.18	0.91	0.02	↘
		ko00440 Phosphonate & phosphinate metabolism	0.07%	0.05%	0.05%	-0.53	0.55	0.03	-0.04	3.74	0.03	↘

Between 4 and 9 months of age, most species from the Actinobacteria and Proteobacteria phyla decreased in relative abundance, replaced by the increasing relative abundance of species from the Firmicutes phyla. An apparent increase in the relative abundance of taxa from the Bacteroidetes phylum was not significant. Within the Firmicutes phylum, the 4 to 9 months of age comparison was characterised by significant increases in sequences assigned to species belonging to the *Ruminococcaceae, Lachnospiraceae, Peptostreptococcaceae*, *Clostridiales, Veillonellaceae,* and *Erysipelotrichaceae* families (Table 4.1). Some species from *Rumincococcus, Erysipelotoclostridium, Bacillaceae,* and *Christensenellaceae* families decreased in relative abundance at 9 months of age before recovering and surpassing prior abundances at 12 months of age, which strongly influenced the separation of clustering between 9 and 12 months of age in sPLS-DA (Table 4.1). With the exceptions of *Clostridium neonatale*, which was not detected at 12 months of age, and *Clostridium sensu stricto*, species belonging to the Firmicutes phylum were detected at higher relative abundance at 12 months of age compared to at 4 months of age (Table 4.1). Populations of some *Peptostreptococcaceae* subspecies and *Veillonellaceae* families increased significantly during the 4 to 9 months of age comparison before decreasing at 12 months of age (Table 4.1).

Most species from the Actinobacteria and Proteobacteria phyla significantly decreased over time. However, the apparent decrease in the most abundant taxonomy assignment from the *Bifidobacteriaceae* family was not significant by PERMANOVA, and this remained the single most abundant taxonomic assignment at 12 months of age (Table 4.1). In addition, where some taxa from the class *Gammaproteobacteria,* including the *Enterobacteriaceae* family and *Vibrionaceae vibrio,* decreased consistently and significantly over time, other taxa from the class *Gammaproteobacteria* such as *Moraxcellaceae family* and *Pseudomonas genus* increased in relative abundance after a decrease between 4 and 9 months of age (Table 4.1).

More KEGG pathway assignments (57%) were significant by PERMANOVA analysis than Metaxa2 assignments (39%) for at least one-time point comparison. The magnitude of log2 fold-change of sequences assigned to KEGG pathways was smaller than the fold-change of sequences assigned to taxonomy (Tables 4.1-4.2), in line with patterns identified in Figure 4.3. The magnitude of fold-change was greater between 4 and 9 months of age than between 9 and 12 months of age (Table 4.2), in agreement with clustering patterns shown in Figure 4.2. More sequences assigned to KEGG pathways significantly decreased relative abundance between 4 and 9 months of age (Table 4.2), agreeing with diversity scores shown in Figure 4.1 and clustering identified in Figure 4.2.

Sequences assigned to Amino acid metabolism, Xenobiotic degradation and metabolism, and Terpenoid and polyketoid metabolism consistently decreased over time, along with associated Protein families: metabolism pathways (Table 4.2). In particular, reduced Cytochrome P450 degradation pathways strongly influenced the separation between 4 and 9 months of age sPLS-DA clustering. Sequences assigned to Secondary metabolite biosynthesis and Secondary bile acid biosynthesis increased consistently in relative abundance over time, whereas Biosynthesis of lipopolysaccharides and associated Protein family: metabolism pathways decreased in relative abundance (Table 4.2). Of the cellular processes that changed significantly and with a log2 fold change of >|.2|, sequences assigned to the PI3k-Akt signalling pathway and RNA transport increased significantly over time (Table 4.2). Sequences assigned to Biofilm formation from specific species significantly decreased between each time point (Table 4.2).

4.3.3. Correlation and integration of the microbiome with metabolites between time points

Regularised canonical correlations based on Pearson's correlations of the delta (change) between time points of microbiome datasets (taxonomic assignments (Level 6), KEGG pathway assignments (Level 3)), and metabolome datasets were conducted to identify

significantly correlated shifts between time points. Correlations between changes in relative abundances of KEGG pathways and changes in metabolite concentration had stronger coefficients of association and more significant changes in relative abundance than correlations between microbial species and faecal metabolites.

Heatmaps, including only correlations above the cut-off values, are shown (Figures 4.3-4.6). There were more significant positive and negative correlations between 4 and 9 months of age than between 9 and 12 months of age, as seen in Figure 4.1 and Tables 4.1-4.2. Of the taxonomy and KEGG pathway assignments that correlate significantly with metabolites, few overlapped with those significant by PERMANOVA (Tables 4.1-4.2) in the 4- vs 9-month time point comparison. Sequences assigned to KEGG pathways correlated more strongly with aqueous than lipid metabolites, whereas microbial species correlated more strongly with lipids. This pattern was apparent between 9 and 12 months of age (Figures 4.5-4.6).

Between 4 and 9 months of age, unclassified taxa from the Firmicutes phyla, *Gammaproteobacteria* class, and *Enterobacteriales* order were the only taxonomic assignments that correlated strongly with metabolites and changed significantly (PERMANOVA FDR corrected p-value of <0.05; Figure 4.3 and Table 4.1). Of the sequences assigned to KEGG pathways that correlated strongly with changes in metabolite concentrations, the PI3K-Akt signalling pathway and RNA transport pathway assignments increased significantly, and Benzoate degradation pathway assignments decreased significantly (PERMANOVA FDR corrected p-value of <0.05) (Figure 4.4 and Table 4.2). The remaining sequences assigned to taxonomy and KEGG pathways that correlated with metabolites were not significantly different between 4 and 9 months of age by PERMANOVA (Figure 4.4 and Table 4.2).

Significant correlations were found between changes in sequences assigned to each of the four main phyla and changes in lipid and aqueous metabolites between 4 and 9 months of age

(Figure 4.3). In addition, hypoxanthine was strongly correlated with phyla members, positively with Proteobacteria and Bacteroides phyla, and negatively with Actinobacteria (*Actinomycetaceae Varibaculum)* and four species of Firmicutes (Figure 4.3) phyla.

Hypoxanthine was also positively correlated with glycan biosynthesis and metabolism integrating KEGG pathway assignments (Figure 4.4). This positive correlation was part of a larger cluster of aqueous metabolites that positively correlated with sequences assigned to Glycan degradation and biosynthesis, which otherwise did not strongly correlate with any lipid metabolites (Figure 4.4). Microbial Carbohydrate metabolism was negatively correlated with lipid metabolites, and microbial Starch and sucrose metabolism was negatively correlated with five aqueous metabolites (Figure 4.4).

Between 9 and 12 months of age, fewer aqueous metabolites significantly correlated with taxonomic assignments than with KEGG pathways (Figures 4.5-4.6). Changes in several classes of lipid metabolites correlated with changes in sequences assigned to species (Figure 4.5). Taxa from the Firmicutes and Proteobacteria phyla were strongly positively correlated with ceramides, diglycerides, and triglycerides but negatively associated with cytidine, an aqueous metabolite (Figure 4.5). C*lostridium sensu stricto* and species of the *Lachnospiraceae* family were negatively correlated with lipid metabolites (Figure 4.5).

A series of dipeptides and amino acids were associated with sequences assigned to KEGG pathways between 9 and 12 months of age (Figure 4.6). Whereas others were negatively associated with serine in particular: Biosynthesis of unsaturated fatty acids, Selenocompound metabolism, Protein kinases, Transporters, and Quorum sensing were positively associated with dipeptide metabolites, Citrate cycle, Metabolism of cofactors and vitamins, Transcription machinery, and AMPK signalling (Figure 4.6).

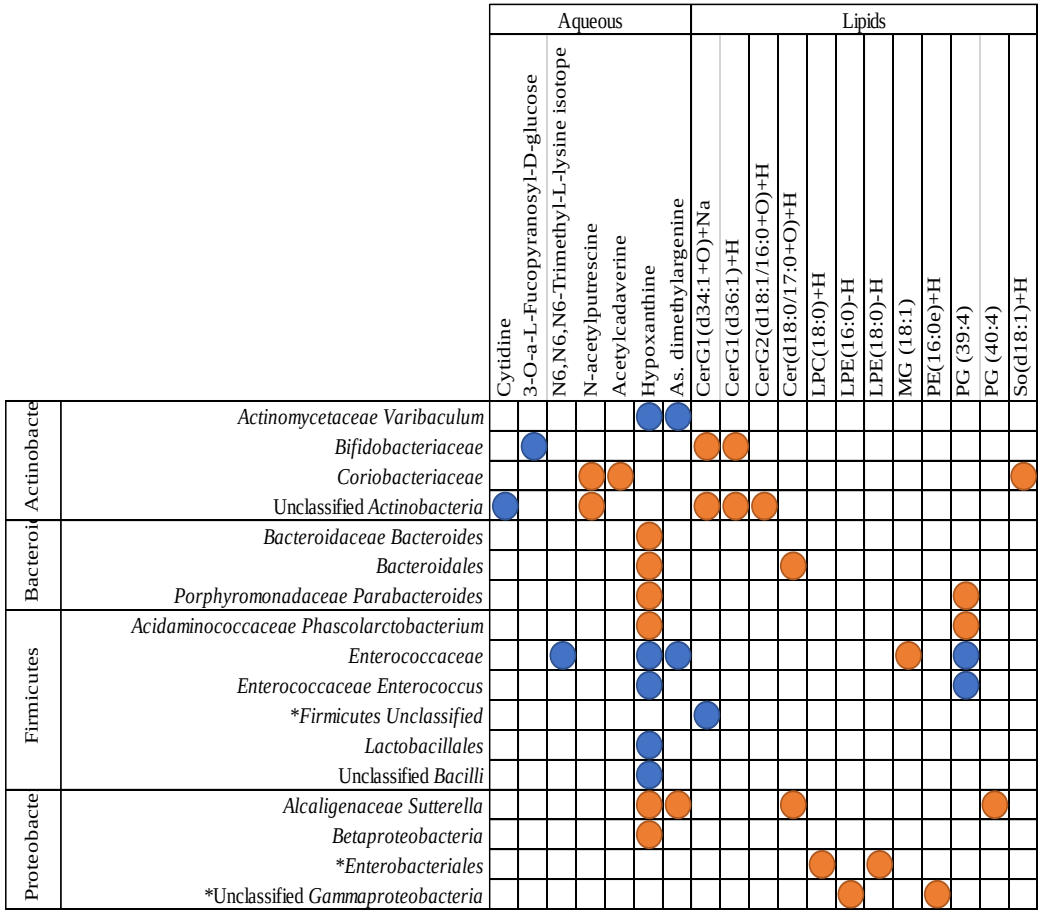


*Figure 4. 3. Filtered regularised canonical correlation heatmap of the delta value of taxonomic assignments and metabolites in faecal samples of infants between 4 and 9 months of age. Orange circles depict significant positive Pearson's correlation coefficients (>0.5), and blue depicts significant negative correlations (<-0.5). Taxonomic assignments that were also significant by PERMANOVA analysis for this time point comparison are indicated by *. CerG1 = Glucosylceramide; CerG2 = Diglycosylceramide; Cer = Ceramide; LPC = Lysophosphatidylcholine; LPE =Lysophosphatidylethanolamine; MG = Monoglyceride; PE = Phosphatidylethanolamine; PG = Phosphatidylglycerol; SO = Sphingosine.*

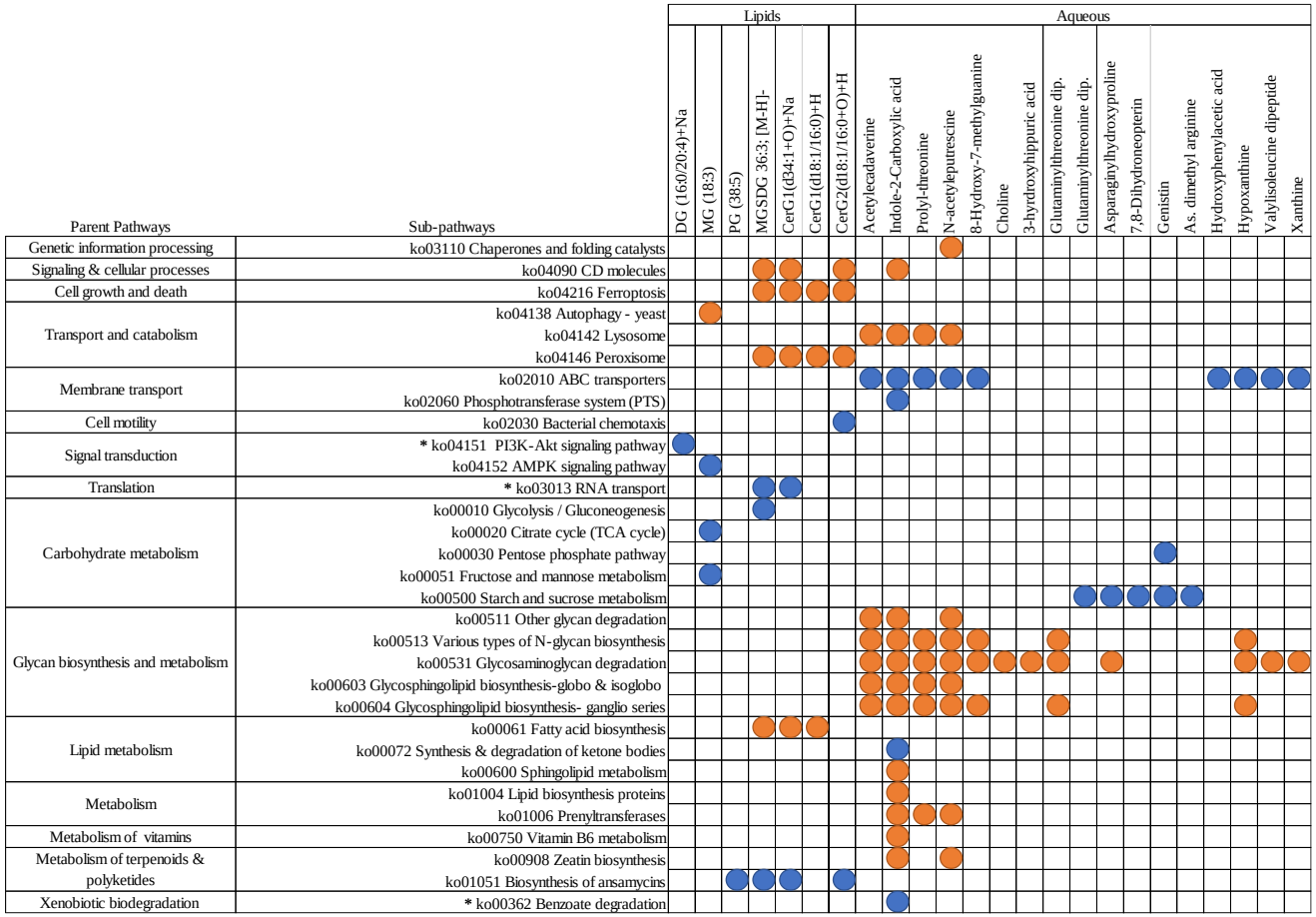


*Figure 4. 4. Filtered regularised canonical correlation heatmap of the delta value of KEGG pathway assignments and metabolites in faecal samples of infants between 4 and 9 months of age. Orange circles depict significant positive Pearson's correlation coefficients (>0.6), and blue depicts significant negative correlations (<-0.6). KEGG pathway assignments that were also significant by PERMANOVA analysis for this time point comparison are indicated by *. DG = Diglyceride; MG = Monoglyceride; PG = Phosphatidylglycerol; MGSDG = Monogalactosyldiacylglycerol; CerG1 = Glucosylceramide; CerG2 = Diglycosylceramide*

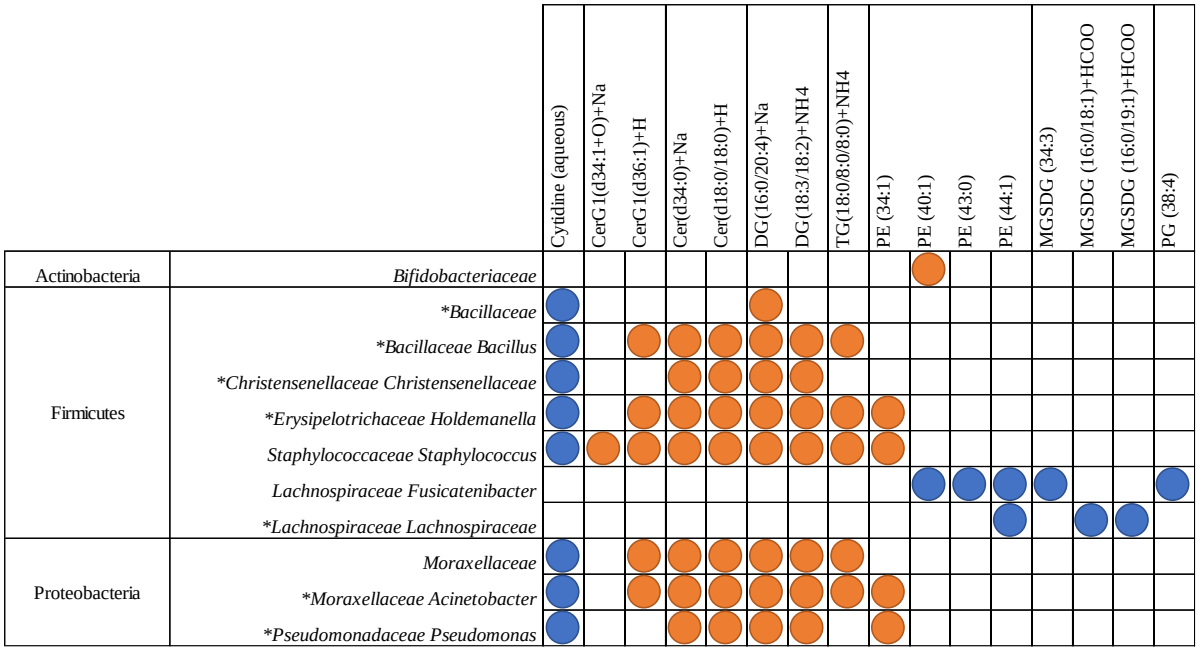


*Figure 4. 5. Filtered regularised canonical correlation heatmap of the delta value of taxonomic assignments and metabolites in faecal samples of infants between 9 and 12 months of age. Orange circles depict significant positive Pearson's correlation coefficients (>0.5), and blue depicts significant negative correlations (<-0.5). Taxonomic assignments that were also significant by PERMANOVA analysis for this time point comparison are indicated by *. Cer = Ceramide; CerG1 = Glucosylceramide; DG = Diglyceride; TG = Triglyceride; PE = Phosphatidylethanolamine; MGSDG = Monogalactosyldiacylglycerol; PG = Phosphatidylglycerol.*

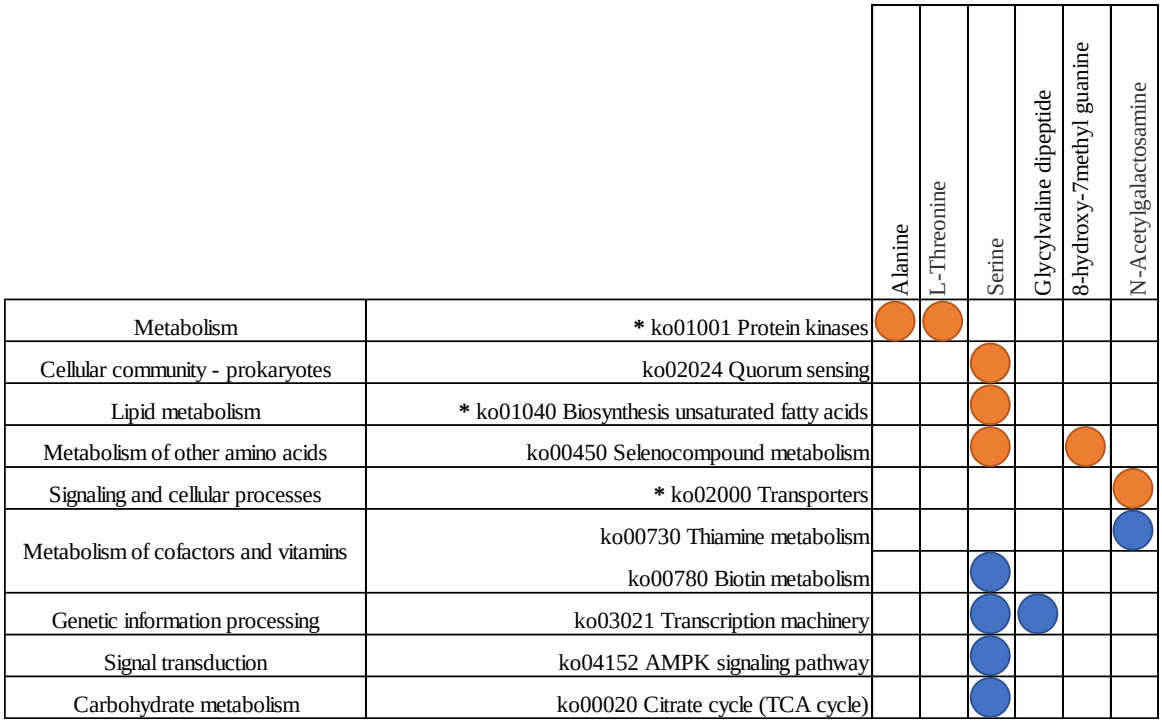


*Figure 4. 6. Filtered regularised canonical correlation heatmap of the delta value of KEGG pathway assignments and metabolites in faecal samples of infants between 9 and 12 months of age. Orange circles depict significant positive Pearson's correlation coefficients (>0.6), and blue represents significant negative correlations (<-0.6). KEGG pathway assignments that were also significant by PERMANOVA analysis for this time point comparison are indicated by *.*

4.4. Discussion

This study investigated how the infant faecal microbiome adapts to the addition and continuation of solid foods during the first year of life by characterising shifts in species composition, relative abundance of gene functions, and identifying significant correlations with faecal metabolites. Faecal samples were utilised to proxy the microbiome and metabolome in the lower GIT. The novel aspect of this study was the identification of metabolites that associate with taxa and predicted gene functions of the faecal microbiota before and after the introduction of solid foods. These highly correlated metabolites may be relevant compounds for further investigations into complementary feeding using targeted metabolite analyses.

The diversity of substrates introduced through complementary foods between 4 and 12 months of age increased the resources available to support a greater diversity of microbes in the faecal samples. In agreement with the hypothesis, the changes in composition occurring between 4 and 9 months of age were more significant than between 9 and 12 months of age. This period corresponds with the introduction of complementary foods between 4 and 9 months of age, rather than continuing complementary feeding between 9 and 12 months of age. A greater proportion of sequences assigned to KEGG pathways was significant by PERMANOVA than sequences assigned to taxonomy, indicating more consistent changes in the relative abundance of gene groups than in species composition. However, the magnitude of changes in the relative abundance of sequences assigned to particular gene functions was smaller than that of changes in microbial composition over the same period. This pattern indicates a degree of functional redundancy across microbial species, particularly genes that are compulsory for survival, and suggests that shifts in functional capacity patterns are similar among infants as the diet transitions from milk to solid foods.

The higher variance (beta diversity) of both taxonomy and KEGG pathway assignments at 4 months of age compared to 9 or 12 months of age suggests that the effects of early-life factors

1909 that influenced the infant GIT microbiome diminished with the addition of solid foods. Effects
1910 of mode of delivery on the GIT microbiome, including increased skin associated species and
1911 functional diversity in Caesarian-born infants, have been reported to persist until approximately
1912 6 months of age (Koenig et al., 2011). However, the magnitude and persistence of these
1913 influences may vary. The mode of milk feeding has been shown to affect the composition of
1914 the faecal microbiome, but as time progresses, differences between breastmilk-fed or formula-
1915 fed infants diminishes (C. J. Stewart et al., 2018). Individual microbial differences that may
1916 have been associated with these early-life factors may become abrogated with the addition of
1917 solid foods, as suggested by the tighter clustering of taxonomic and functional assignments in
1918 ordination analyses at 9 months of age. These findings agree with published studies (de Muinck
1919 & Trosvik, 2018; He et al., 2019). The variance in the faecal microbiome composition at 12
1920 months of age may reflect microbes' preferences for specific structures from the food matrix
1921 that is diversifying and divergent between infants as they age.

1922 The patterns that appeared through correlating the change in microbial sequences assigned
1923 to species or KEGG pathways with the change in faecal metabolite concentrations support
1924 results from PERMANOVA and ordination analyses. More changes occurred with the
1925 introduction of solid foods, rather than the continuation of solid foods. Interestingly, few taxa
1926 and functional pathways that correlate strongly with metabolites were different by
1927 PERMANOVA analysis, particularly from 4 to 9 months of age. This result suggests that the
1928 change in the relative abundance of sequences assigned to taxa or KEGG pathways shown in
1929 Figures 4.5-4.8 (not significantly different by PERMANOVA) were more closely related to
1930 changes in metabolite concentrations than to age.

1931 The propensity for particular taxa to correlate more strongly with lipid metabolites than
1932 aqueous metabolites in faecal samples suggests these taxa may have a direct or indirect role in
1933 the production of these lipids, or that the microbes themselves have a higher content of these

lipids than other members of the community (Sohlenkamp & Geiger, 2015). The increased variation in taxonomic composition and lipid metabolite profiles between infants at 9 and 12 compared to at 4 months of age supports the theory that the lipids identified were from the bacteria themselves. The correlations between the differences in aqueous metabolites and the differences in metabolism and degradation based on KEGG pathway assignments imply shifting dietary patterns and altered utilisation of substrates derived from the digestion of dietary proteins, saccharides, oligosaccharides, and fermentative by-products from other microbes (Tremaroli & Bäckhed, 2012).

Many of the correlated aqueous metabolites detected are products of incomplete protein degradation and putrefaction (Wishart et al., 2018). This series of dipeptides and acetylcadaverine and N-acetylputrescine were negatively correlated with sequences assigned to Starch and sucrose metabolism in the present study. This finding indicates that as protein utilisation decreases, as shown in the reduced amino acid utilisation pathways over time, carbohydrate utilisation increases, which may reflect the composition of proteins and carbohydrates in breastmilk and formula (C. R. Martin, Ling, & Blackburn, 2016). Prior research has identified the faecal microbiome as playing a critical role in the production of essential amino acids that are important for infant growth and development, particularly in the context of breastmilk and formula consumption (Bäckhed et al., 2015). The enrichment of amino acid utilisation and biosynthetic pathways at 4 months of age compared to 9 or 12 months of age may compensate for immature proteolytic enzyme production and secretion in the developing GIT (Bäckhed et al., 2015; Gan, Bornhorst, Henrick, & German, 2018). The incomplete digestion of proteins can have a significant effect on microbial trophic networks, which may influence immune system development through the production of precursors and intermediaries to immune crosstalk throughout the GIT (Dallas et al., 2017; Ma & Ma, 2019).

As trophic networks adapt to shifting resources, interactions between populations of microbial community members also flux in response to metabolite intermediaries. For instance, the strong positive correlation between change in serine concentration, and changes in the percentage of sequences assigned to quorum sensing pathways points to the role of metabolites in the regulation of microbial genes that monitor and respond to population density (Parsek, Val, Hanzelka, Cronan, & Greenberg, 1999; Whitehead, Barnard, Slater, Simpson, & Salmond, 2001).

This study was limited by the small cohort size (n=25), varied age at sampling, analysis of samples in two batches, and alterations in sample processing protocols for samples collected at 4 months of age. Additionally, there were limitations inherent in using shotgun metagenomic sequencing to assign taxonomy and in annotating chemical identities to chromatographic peaks during metabolomic data processing (Chaleckis, Meister, Zhang, & Wheelock, 2019; Hillmann et al., 2018).

The small size of the Nourish to Flourish cohort limits the confidence with which early-life factors such as mode of delivery and mode of milk feeding can be investigated. However, the design of this study allowed for the effects of the introduction of solid foods on the faecal microbiome to be compared with the impact of the continuation of solid foods, using a longitudinal, high-resolution, multi-omic approach. Batch effects were successfully overcome by including samples from all time points in both batches during sample analysis and normalisation approaches specific to each dataset. Furthermore, additional dietary data were collected to accompany the samples used here, enabling further research into specific associations between complementary feeding and the development of the infant GIT microbiome.

4.5. Conclusions

The Nourish to Flourish longitudinal pilot study showed that the infant GIT microbiome (based on faecal samples) responds to the introduction of solid foods with increased microbial taxonomic diversity while the abundance of functional genes remains stable. Ongoing dietary diversification may lead to continued changes in the relative abundances of different species during the transition from liquid to solid diets. The decreased magnitude of changes in the functional pathway assignments of the faecal microbiome from 9 to 12 months of age was indicative of the ongoing stabilisation process. Some changes in relative abundances of microbes and KEGG pathways were more strongly associated with changes in metabolite concentrations than with changes in age.

Overall, the lesser changes in gene functions over time, despite large changes in taxonomic composition, suggest that most of the genes required to use new substrates are already present at 4 months of age. Therefore, it is plausible that the changing environment at that time results in the selection of different microbes that may harbour these genes in different combinations.

Chapter 5. Mode of milk-feeding differentially influences microbiome composition and function during complementary feeding

During complementary infant feeding, dietary patterns are complex, dynamic, and highly individual, but breastmilk and/or formula are a substantial part of the diet that influence the gastrointestinal tract (GIT) microbiome. This study aimed to analyse estimated nutrient consumption, dietary diversification, and associated patterns in the faecal microbiome composition and function in a cohort of 25 infants who were fed by different modes of milk: breastfed, formula-fed, or mixed-fed (breastfed and formula-fed) at three different time points before (4 months of age) and during weaning (9 and 12 months of age). Shotgun metagenomic sequencing and LCMS of the faecal samples were used to profile microbiome composition, microbial gene abundance, and the metabolome. At 4 months of age, the lipid metabolite profile was most influenced by the mode of milk feeding, at 9 months of age, sequences assigned to KEGG pathways differed according to the mode of milk feeding, and at 12 months of age, aqueous metabolites differed among mixed-fed infants who consumed both formula and breastmilk. In addition, the microbial alpha diversity was more strongly correlated with dietary diversity at 9 months of age than 12 months of age. The species that were significantly different by mode of milk feeding demonstrated patterns of diversification and enrichment of species within the Firmicutes, Proteobacteria, and Bacteroides phyla. In summary, the influence of the mode of milk feeding decreased during complementary feeding and changed as the consortia of microbes and their respective genomes shifted throughout the first year of life.

5.1. Introduction

Complementary feeding refers to the transition from breastmilk and/or formula to solid foods from various food groups. During this transition, solid foods complement the milk-based diet up until approximately 9 months of age, and breastmilk and/or formula complement solid foods until energy and nutrient requirements are fulfilled entirely by solid foods (Zealand, 2008). This transition amounts to the most drastic dietary shift individuals experience over the lifespan, affecting all areas of physiological and behavioural development.

Breastmilk and formula constitute a significant part of the diet, but the ratio of milk to solid foods varies by infant and constantly changes throughout complementary feeding. This diversity leads to differences in nutrients and bioactive components obtained from the diet. For example, breastmilk is lower in energy and protein by volume than formula, and breastfed infants consume less milk than formula-fed infants (Haisma et al., 2003; Heinig, Nommsen, Peerson, Lonnerdal, & Dewey, 1993). As a result, the bioactive components of breastmilk are different compared to infant formula. However, with advances in the incorporation of bovine-derived milk fat globule membrane (MFGM), synthetic HMOs, and the addition of probiotic strains, these differences are narrowing (Andreas, Kampmann, & Le-Doare, 2015; Boix-Amorós et al., 2019; Green Corkins & Shurley, 2016; Hanson et al., 2003; Qian et al., 2016).

It has also been proposed that formula-fed infants develop different feeding preferences and behaviours, possibly in response to decreased exposure to flavours from foods that are transferred through breastmilk (Hausner, Nicklaus, Issanchou, Mølgaard, & Møller, 2009). Breastfed infants have higher acceptance of solid foods later in infancy, decreased pickiness, and increased fruit and vegetable consumption in childhood (Hausner et al., 2009; Ventura, 2017). Collectively these differences in nutrient composition and feeding behaviours influence risk for obesity and atopy through mechanisms mediated by the GIT microbiome composition and function (Bokulich et al., 2016; Davis et al., 2017; Fujimura et al., 2016; He et al., 2019;

Iyengar & Walker, 2012; Liu et al., 2016; Morrow, Ruiz-Palacios, Jiang, & Newburg, 2005; Piccolo et al., 2017; Sjögren et al., 2009; Van der Leek, Yanishevsky, & Kozyrskyj, 2017; Ventura, 2017; M. Wang, Monaco, & Donovan, 2016).

The influence of breastmilk compared to or in combination with formula on the GIT microbiome composition and function is best understood prior to introducing solid foods, after which the differences become less apparent (Bäckhed et al., 2015; de Muinck & Trosvik, 2018; He et al., 2019). However, the cessation of breastfeeding is a significant factor in the maturation of the infant GIT microbiome (Bäckhed et al., 2015). The diversity of species in the breastfed infant GIT is lower than in the formula-fed infant, primarily due to the ability of some species of the *Bifidobacterium* genus (i.e., *B longum* subsp. *infantis*) to outcompete other bacteria for utilisation of whole HMOs as an energy source (Azad et al., 2013; Azad et al., 2016; Pannaraj et al., 2017; D. A. Sela et al., 2011). Breastfeeding has also been associated with higher relative abundances of taxa from the *Lactobacillus* genus, whereas mixed-fed infants have been found to have higher abundances of *Veillonella* and *Klebsiella* genera (N. Li et al., 2020). Species associated with formula consumption include *Bacteroides* and *Blautia* genera of the Firmicutes phylum, diverse members of the Proteobacteria phylum, and higher relative abundances of *Bifidobacterium adolescentis* among the Actinobacteria phylum (Davis et al., 2017; Lee et al., 2015; N. Li et al., 2020).

Investigations into the differences in the microbial functional capacity of infants fed different modes of milk prior to the introduction of solid foods have found a small number of enriched pathways for phosphotransferase system (PTS) transporters, methanogenesis, bile acid biosynthesis, and a series of distinct carbohydrate degrading enzymes that resemble an adult metagenome in formula-fed infants (Bäckhed et al., 2015). Whereas in breastfed infants, pathways for oxidative phosphorylation, vitamin biosynthesis, and alpha-amylase enzymes were enriched (Bäckhed et al., 2015). Later in infancy, microbial metagenomes become

increasingly adult-like with the addition of solid foods corresponding to increased carbohydrate utilisation, vitamin biosynthesis, and xenobiotic degradation (Koenig et al., 2011),

The infant GIT metabolome has not been as extensively explored as the microbiome or metagenome. However, key metabolites have been identified in breastmilk and formula feeding. In porcine models, the faecal acetate concentration was higher in animals fed human breastmilk, which was also associated with altered host physiology (Liu et al., 2016). Tryptophan metabolism shifted from serotonin to tryptamine pathways in neonatal formula-fed infants, which may link formula feeding and increased risk for immune dysregulation (Saraf et al., 2017). Further to this point, serotonin has been found to be negatively associated with breastmilk-associated *Bifidobacterium* spp. (Brink et al., 2020). Enrichment of specific sub-pathways of tryptophan metabolism, indoleamine 2,3 dioxygenase subset of the kynurenine pathway, have been reported. The resulting kynurenine metabolite may support immune tolerance via aryl hydrocarbon receptors (Van der Leek et al., 2017).

The analyses reported in Chapter 5 aimed to characterise the differential influences of breastfeeding, formula-feeding, or mixed-feeding (breastfed and formula-fed) on the faecal microbiome and metabolome of infants during the complementary feeding window (6 to 12 months of age). Alpha diversity of the faecal microbiome is expected to be lower in breastfed infants due to the exclusive utilisation of HMOs by *Bifidobacterium infantis*, despite increased dietary diversity arising from higher acceptance of foods predicted in breastfed infants. Metabolite profiles of lipid and aqueous metabolites in faecal samples are expected to be most different between breastfed and formula-fed infants, due to different lipid structures and higher protein content in the formula, with substantial overlap in mixed-fed infants.

Chapter 5 reports the anthropometric data, diet diversification, and sources of nutrients in a cohort of 25 infants who were breastfed, mixed-fed, or formula-fed. These readouts were then integrated into an analysis of faecal microbiome taxonomic and functional gene composition

and faecal metabolites at three time points before (4 months of age) and during the complementary feeding transition (9 and 12 months of age).

5.2. Methods

Clinical protocols and microbiome methods are explained in Chapter 3. Metabolomic and dietary methods are described in Chapter 4. Further dietary methods and statistical methods are outlined below.

5.2.1. Diet

Parents kept 3-day food records on paper hardcopies during the days before sample collection and clinic visits at approximately 9 and 12 months of age. Instructions with examples of records were provided, indicating how to record quantities, brands, recipes, feeding time, and duration of breastfeeding. Records were transferred to the study team during clinic visits, and clinicians clarified ambiguous records with the parents where possible.

Diet records were then input and analysed for food composition using the NZ FOODFiles 2016-Food Composition Database specific to food products available in New Zealand (Sivakumaran et al., 2016). If more than 25% of a food record for a given participant was missing quantities, the record and participant were omitted. Two participants' records were omitted from the analyses as a result. During the preliminary analysis of diet data, records that contained outlying quantities of nutrients were verified for accuracy of transcription and source of anomalous information. Nutrient component values were assumed equal for all breastmilk and amounts of breastmilk consumed were estimated based on the feeding length.

Food group classification at multiple levels was manually assigned using an iterative approach to ensure agreement and consistency. Food groups were collectively defined based on prior literature, tailored to infant diets, specific to New Zealand foods, and assessed against a subset of food records. Researchers A and B independently assigned food groups to the complete dataset, including individual recipe components when provided, based on the agreed classifications. These classifications were cross-checked for variance (3.7% of records) and supplied to Researcher C to designate final classifications (Figure 5.1). Level 1 food groups

referred to broad categories such as fruits, vegetables, or meats. Level 2 food groups included more specific sub-categories such as berries, stone fruits, beef, and fish.

5.2.2. Statistical analyses

Statistical analyses were conducted in R version 4.0.2 (R Core Team, 2021). Alpha diversity metrics were calculated for microbiome composition using the 'vegan' package version 2.5 in R. Sparse PLS-DA models were done in 'MixOmics' package version 6.12.2 in R and tuned to include species that were most important to the projection (Ruiz-Perez, Guan, Madhivanan, Mathee, & Narasimhan, 2020). ANOVA testing was conducted in the 'RVAideMemoire' package version 0.9-7.8 in R.

Table 5. 1. Anthropometric and feeding characteristics of the Nourish to Flourish cohort included in this analysis, including weight for length ratios (WFL). Z-scores were based on the World Health Organisation Child Growth Standards (WHO Multicentre Growth Reference Study Group, 2006).

Sex (n) (%)		
Male		14 (56%)
Female		11 (44%)
Anthropometrics at birth (mean ± SD)		
Weight (kg)		3.48 ± 0.48
Length (cm)		50.94 ± 2.54
WFL (kg/cm)		0.06 ± 0.007
WFL z-score		-0.35 ± 1.10
Age (days) at introduction of solids (mean + SD)		171.76 ± 14.05
4-month time point	Age (days) (mean ± SD)	
	Breastfed (n)	22 (88%)
	Anthropometrics (mean ± SD)	
	Weight	6.58 ± 0.99
	Length	63.05 ± 3.27
	WFL (kg/cm)	0.10 ± 0.01
	*WFL for age z-score**	-0.23 ± 0.66
	Mixed-fed (n)	3 (12%)
	Anthropometrics (mean ± SD)	
	Weight	6.19 ± 0.16
	Length	61.77 ± 1.83
	WFL (kg/cm)	0.10 ± 0.00
	*WFL for age z-score**	-0.38 ± 0.52
9-month time point	Age (days) (mean ± SD)	265.92 ± 18.12
	Breastfed (n)	17 (68%)
	Anthropometrics (mean ± SD)	
	Weight	8.56 ± 0.95
	Length	70.37 ± 2.47
	WFL (kg/cm)	0.12 ± 0.01
	*WFL for age z-score**	0.22 ± 0.95
	Formula-fed (n)	4 (16%)
	Anthropometrics (mean ± SD)	
	Weight	8.61 ± 0.74
	Length	71.41 ± 2.05
	WFL (kg/cm)	0.12 ± 0.01
	*WFL for age z-score**	-0.10 ± 0.33
	Mixed-fed (n)	3 (12%)
	Anthropometrics (mean ± SD)	
	Weight	8.86 ± 0.85
	Length	70.53 ± 2.94
	WFL (kg/cm)	0.13 ± 0.007
	*WFL for age z-score**	0.53 ± 0.12
12-month time point	Age (days) (mean ± SD)	357.84 ± 17.76
	Breastfed (n)	14 (56%)
	Anthropometrics (mean ± SD)	
	Weight	9.28 ± 1.03
	Length	75.92 ± 2.68
	WFL (kg/cm)	0.12 ± 0.01
	*WFL for age z-score**	-0.35 ± 1.22
	Formula-fed (n)	6 (24%)
	Anthropometrics (mean ± SD)	
	Weight	9.49 ± .95
	Length	76.11 ± 1.72
	WFL (kg/cm)	0.12 ± 0.01
	*WFL for age z-score**	-0.17 ± 1.02
	Mixed-fed (n)	5 (20%)
	Anthropometrics (mean ± SD)	
	Weight	9.43 ± 1.01
	Length	76.52 ± 4.80
	WFL (kg/cm)	0.12 ± 0.01
	*WFL for age z-score**	-0.17 ± .22

**Z-scores were calculated using the WHO Child Growth Standards, accounting for days since birth and gender.*

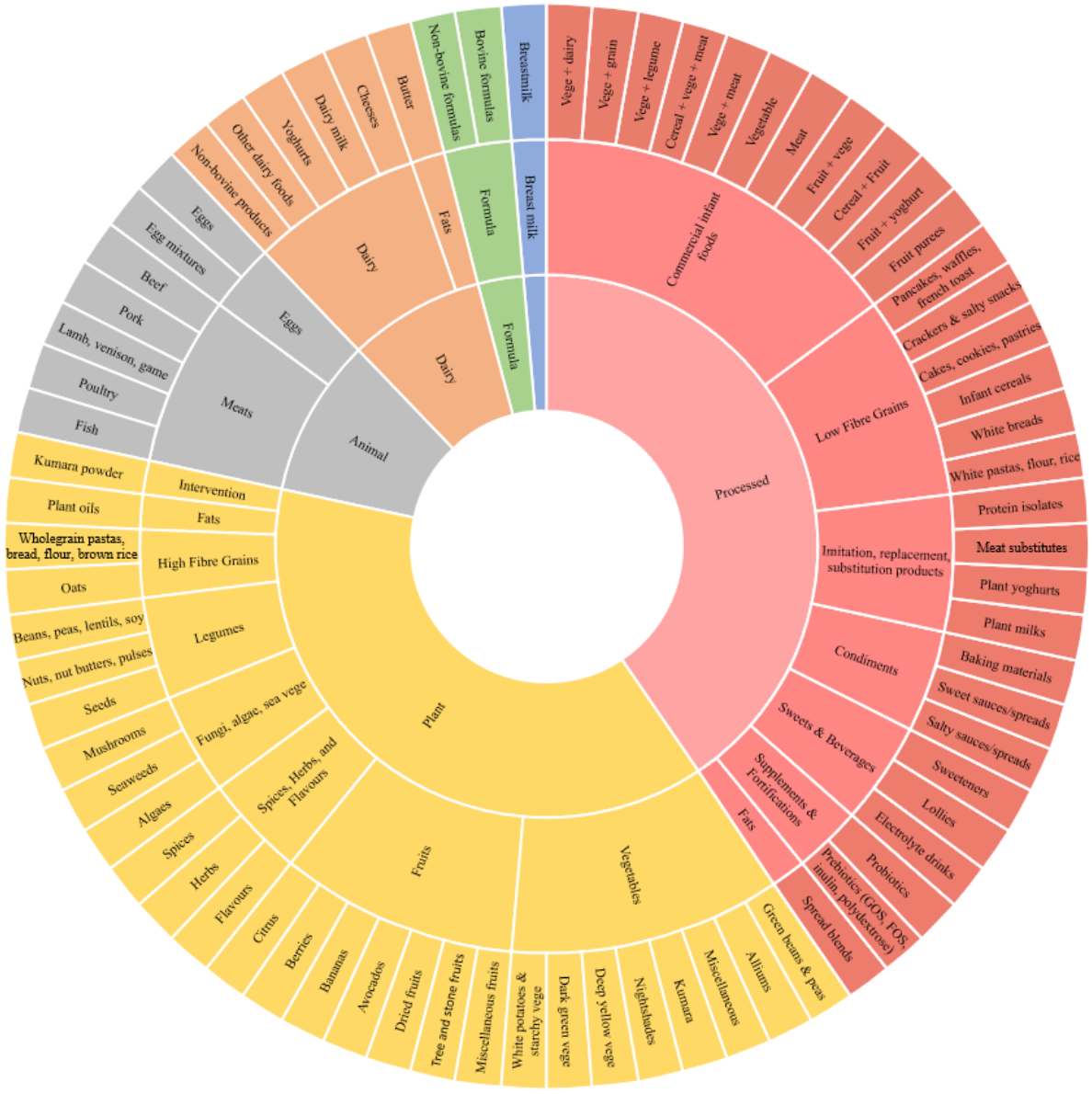


Figure 5. 1. Food group classifications of the Nourish to Flourish cohort. The inner circle depicts the broadest level (Source), followed by Level 1 food groups in the middle ring and Level 2 food groups in the outer ring. The pink colour represents manufactured foods. The yellow colour represents plant-derived products. The orange represents dairy products. The grey represents flesh foods such as meat and fish. Green indicates the formula. Blue represents breastmilk as its food group.

5.3. Results

5.3.1. Milk feeding summary

Infants who consumed formula had higher average daily energy consumption than mixed-fed infants or breastmilk alone at both 9 and 12 months of age (Figure 5.2). Mixed-fed infants in this cohort consumed more energy from non-milk sources than breastfed and formula-fed infants at 9 and 12 months of age (Figure 5.2). At 9 months of age, an average of 6.8 daily breastfeeding sessions (± 3.3) was reported among infants who did not consume any formula (n=17), which amounted to an average of 63% of daily energy consumption (data not shown). At 12 months of age, 14 infants did not consume formula, and the average number of breastfeeding sessions dropped to 4.5 (± 2.7), contributing approximately 38% of daily energy (data not shown). The number of formula-fed infants rose from four at 9 months of age to six at 12 months of age. At 9 months of age, the formula contributed 56% of daily energy from an average of 5 (± 0.86) serves. At 12 months of age, formula contributed 39% of daily energy from an average of 3.3 serves (± 0.68). Three infants were mixed-fed at 9 months of age, which provided 26% and 15%, respectively, of daily energy across an average of 3.5 (± 0.69) breastfeeding sessions and 1.4 (± 0.09) serves of formula (Table 5.1). At 12 months of age, five infants were mixed-fed, and average daily breastfeeding sessions dropped to 2 (± 1.0) and formula serves dropped to 1.4 (± 0.28), contributing 15% and 13% of daily energy, respectively (Table 5.1).

5.3.2. Sources of nutrients

Formula-fed infants consumed higher levels of most macro- and micronutrients than breastfed or mixed-fed infants (Figure 5.2). Breastfed infants consumed similar quantities of fat and more iodine at 9 months of age compared to infants who consumed formula at the same age but with lower fat and similar iodine intakes at 12 months of age. On average, breastfed infants did not meet the adequate intake (AI) requirements for iron, zinc, or vitamin E at either 9 or 12 months of age (Figure 5.2). The statistical significance of these differences could not

be determined due to uneven numbers of breastfed, formula-fed, and mixed-fed infants. Additionally, quantities of breastmilk consumed and the specific nutrients in individual breastmilk were estimated (Figure 5.2).

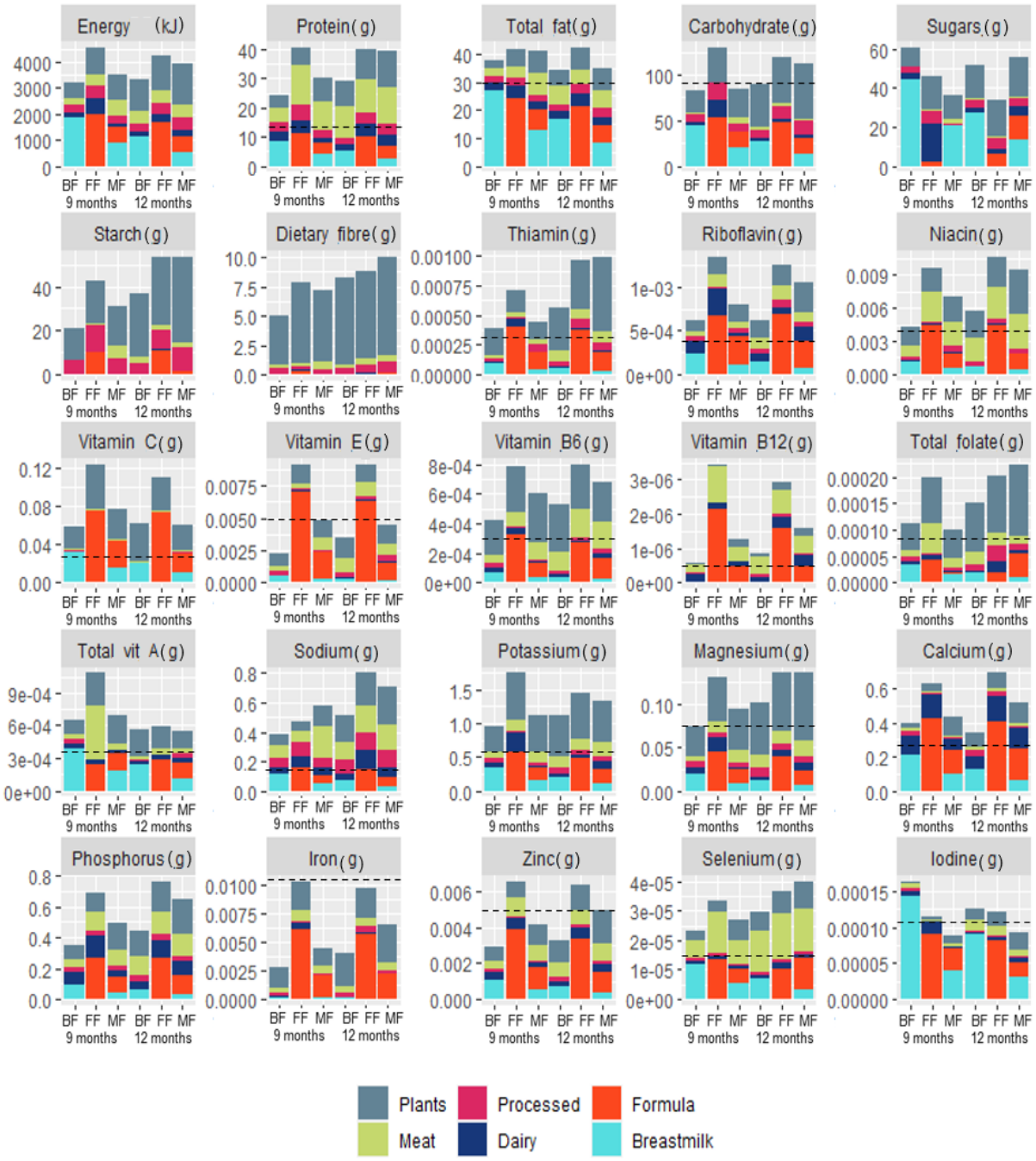


Figure 5. 2. Average nutrients from food sources among breastfed (BF), formula-fed (FF) and mixed-fed (MF) infants at 9 months of age on the left of each graph and 12 months of age on the right, with the adequate intake (AI) of the given nutrient shown in the dashed horizontal black line. If no AI is shown, no AI currently exists under New Zealand child feeding recommendations. Breastmilk is shown in blue, formula in red, dairy in navy, meat in green, plants in grey, and processed foods in pink. Statistical significance was not determined due to uneven group sizes.

5.3.3. Diet diversity and microbiome diversity

Bar charts of individual diets measured by food group contribution to total energy consumption at Level 1 food grouping at 9 and 12 months of age (Figure 5.3a) show that estimated breastmilk and formula consumption was, unsurprisingly, greater at 9 months of age than 12 months of age, and the number of infants consuming formula increased at each time point (Figure 5.3a). However, diet composition by food group was highly variable between individuals (Figure 5.3a). Average diet diversity measured as Shannon Entropy Index of Level 2 food group classifications of formula-fed and mixed-fed infants was higher than breastfed infants at 9 and 12 months of age. However, a greater number of infants who were breastfed compared to infants who were formula-fed or mixed-fed was more likely to contribute to a larger range of dietary diversity scores than the milk feeding behaviour alone (Figure 5.3). In addition, positive correlations between diet diversity and microbiome diversity were stronger at 9 months of age than 12 months of age (Figure 5.3).

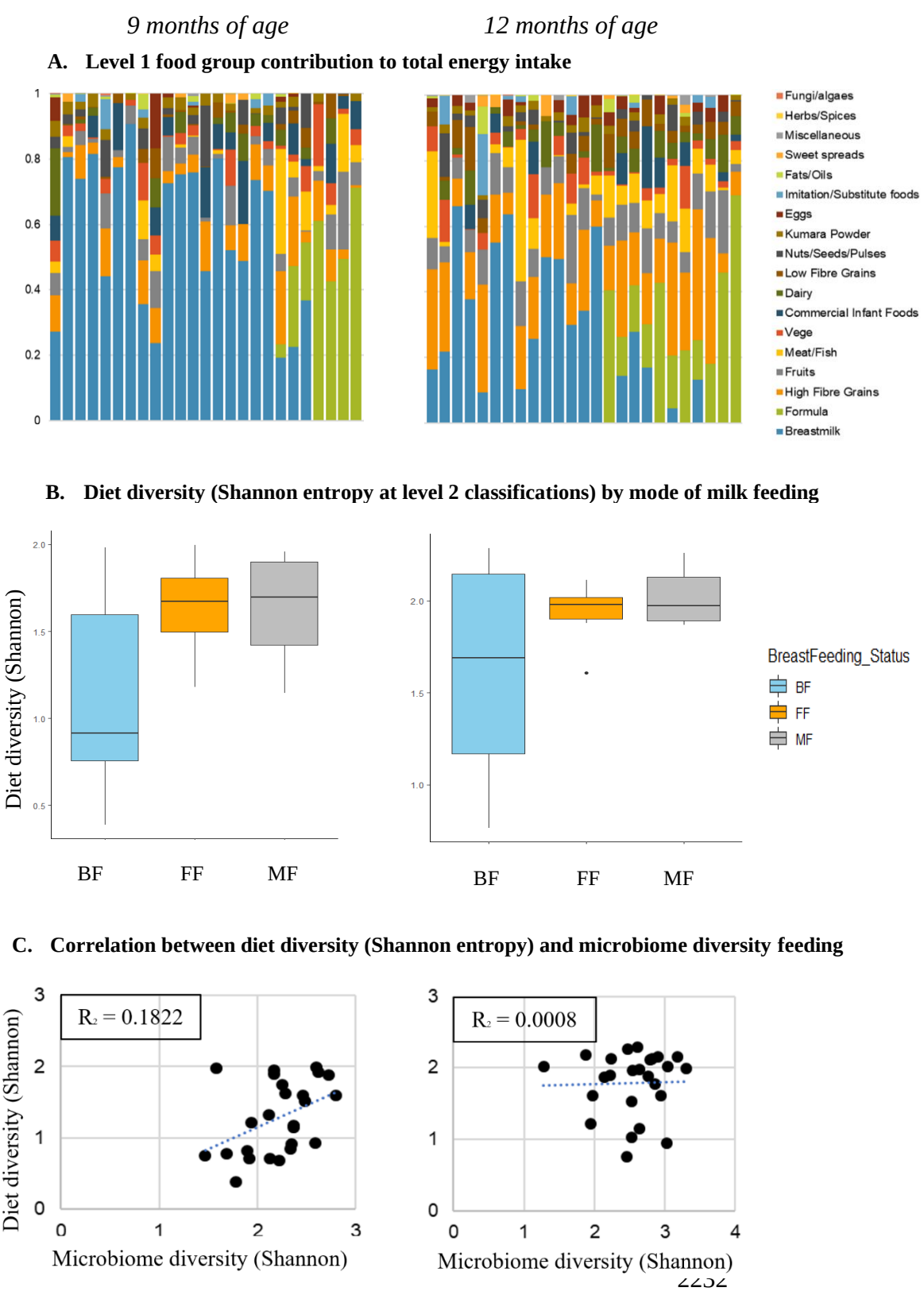


*Figure 5. 3. Results for the 9 months of age time point are shown in the left column, and 12 months of age in the right column for **A** relative abundances of food groups contributing to overall energy intake in infants, **B** diet diversity ratio of breastfed (BF), mixed-fed (MF), and formula-fed (FF) infants, and **C** correlation between Shannon entropy index scores of diet diversification and Shannon entropy index scores of faecal microbiome diversity.*

5.3.4. PLS-DA clustering of microbiome variables

Sequences assigned to taxa in faecal samples of breastfed infants clustered more tightly in sPLS-DA models at 4 months of age than at either 9 or 12 months of age (Figure 5.4). The separation of breastfed and mixed-fed infants was greater at 4 months of age. The explained variance of sequences assigned to taxa increased with age (Figure 5.4). Conversely, the sPLS-DA model of sequences assigned to KEGG pathways had the greatest explained variance at 4 months of age and decreased with age at 9 and 12 months, respectively. Sequences assigned to KEGG pathways of mixed-fed infants were more like breastfed infants than formula-fed infants at 9 months of age than 12 months of age, and these clusters converged at 12 months of age (Figure 5.4).

Similar to sequences assigned to KEGG pathways, sPLS-DA models of lipid metabolites in faecal samples had the highest explained variance at 4 months of age compared to 9 or 12 months of age. Mixed-fed infants clustered with breast-fed infants as opposed to formula-fed infants at 9 months of age, followed by a convergence of clusters at 12 months of age (Figure 5.4). The explained variance of aqueous metabolites was highest at 12 months of age, but mixed-fed infants were again more similar to breastfed infants than formula-fed infants at 9 months of age (Figure 5.4).

Notably, sequences assigned to taxa and aqueous metabolite profiles in faecal samples of mixed-fed infants at 9 and 12 months of age did not overlap neatly between breastfed and formula-fed infants but displayed a unique response to the inclusion of both breastmilk and formula in the diet (Figure 5.4). The distinct profile of aqueous metabolites in mixed-fed infants was explained by ANOVA comparisons shown in Table 3.

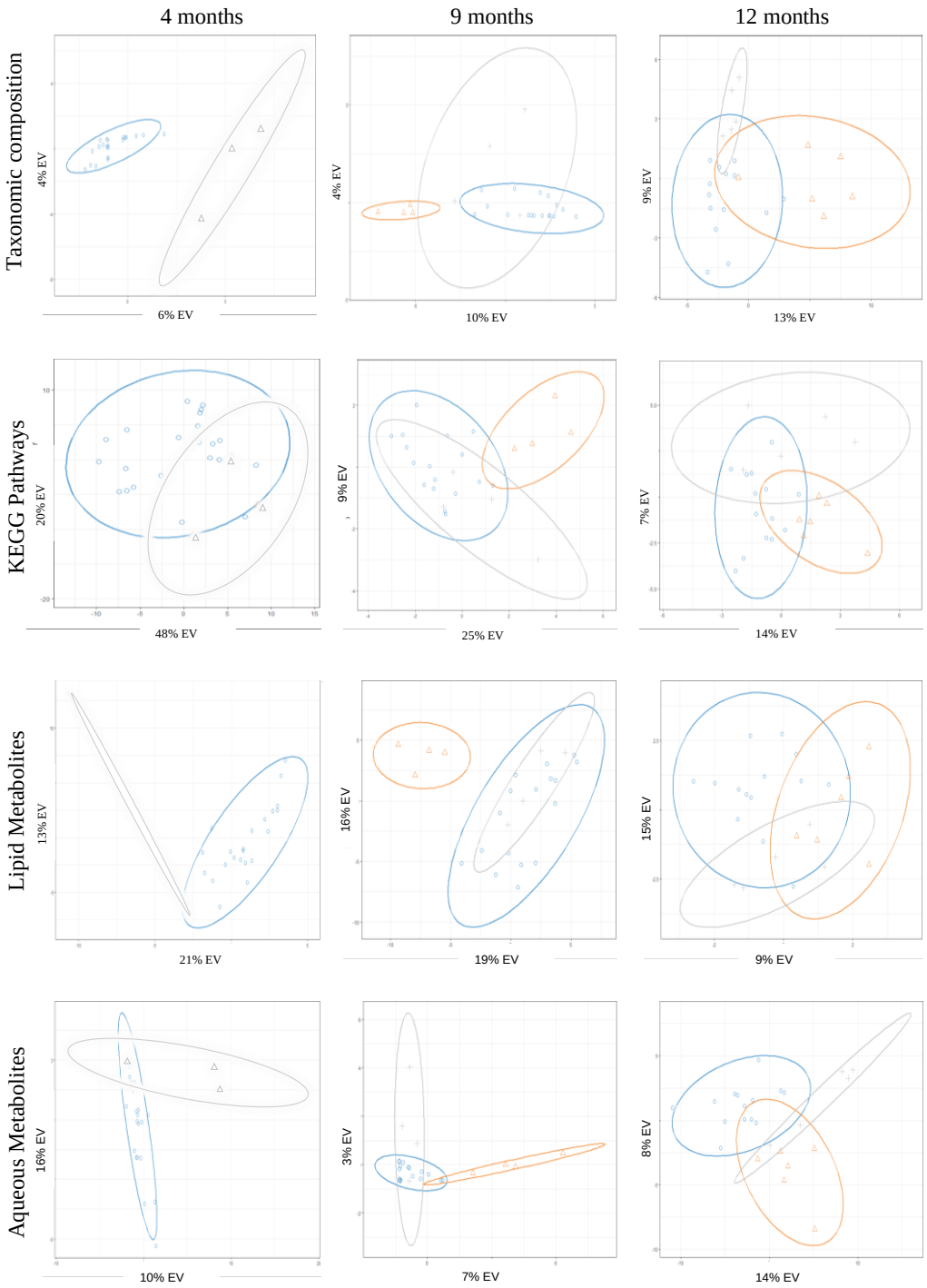


Figure 5. 4. Sparse PLS-DAs at 4, 9, and 12 months of age of composition, function, lipid, and aqueous metabolites of faecal samples. Breastfed infants are shown in blue, mixed-fed in grey, and formula-fed in orange. Explained variance (EV) is indicated on the X and Y axes.

5.3.5. ANOVA comparisons

At 9 months of age, more significantly different sequences were assigned to KEGG pathways between modes of milk feeding by ANOVA than at either 4 or 12 months of age (Table 5.2). Sequences assigned to Energy metabolism, Carbohydrate metabolism, Biosynthesis of secondary metabolites, Beta-alanine metabolism, and Styrene degradation were highest in formula-fed infants and lowest in breastfed infants (Table 5.2). Sequences assigned to Genetic information processing, Proteosome, Metabolic protein families, Cellular processes, Phosphatidylinositol signalling, Nucleotide excision repair, Fatty acid degradation, Biosynthesis of unsaturated fatty acids, Retinol metabolism, and Xenobiotic and Drug metabolism by cytochrome P450 were highest in breastfed infants and lowest in formula-fed infants (Table 5.2).

Mixed-fed infants had significantly higher sequences assigned Prokaryotic defence and Biosynthesis of vancomycin at 4 months of age. At 9 months of age significantly higher sequences were assigned to Nitruolene degradation, and at 12 months of age, they had significantly higher sequences assigned to Penicillin and cephalosporin biosynthesis and Protein processing in the endoplasmic reticulum (Table 5.2). Sequences assigned to Methane metabolism and Carbon fixation in photosynthetic organisms (plants) were highest in formula-fed infants at 9 and 12 months of age compared to 4 months of age (Table 5.2). Thiamine metabolism was significant by ANOVA at both 9 and 12 months of age but was lowest in the mixed-fed infants at 9 months and lowest in breastfed infants at 12 months of age (Table 5.2).

The relative abundance of taxa from the Proteobacteria phylum, *Alcalginaceae parasutterella* and *Desulfovibrionaceae bilophila* were significantly higher in mixed-fed infants than breastfed infants at 4 months of age and the latter at 12 months of age (Table 5.2). At 12 months of age, taxa from the *Moraxellaceae* family (*Moraxellaceae* family and

Acinetobacter genus) and *Psudomonaceae pseudomonas* were significantly highest in formula-fed infants but lowest in mixed-fed infants (Table 5.2).

At 9 months of age, the relative abundance of taxa from the *Bifidobacterium* genus (Actinobacteria phylum) and *Prevotella* genus (Bacteroides phylum) were significantly highest in breastfed infants and lowest in formula-fed infants. However, the relative abundance of taxa from the *Prevotellaceae* family (*Bacteroides* genus) was highest in formula-fed infants and lowest in mixed-fed infants at 9 months of age (Table 5.2).

Similarly, at 12 months of age, the relative abundance of *Rikennellaceae alistipes* and *Porphyromonadaceae barnesiella* from the Bacteroidetes phylum were significantly highest in formula-fed infants and lowest in mixed-fed infants (Table 2). Species belonging to the Firmicutes phylum had significantly lower relative abundance in breastfed infants at 9 and 12 months of age, except for *Eryspelotrichaceae faecalitalea*, which was highest in abundance in mixed-fed infants at 9 months of age (Table 5.2). Species belonging to the *Lachnospiraceae* family and *Erysipelotrichaceae* (Level 3) had the highest relative abundance in formula-fed infants and lowest in breastfed infants at 9 months of age (Table 5.2).

The sequences assigned to *Peptostreptococcaceae intestinibacter* and *Ruminococcaceae ruminococcus* were highest in abundance in formula-fed infants and lowest in breastfed infants at 12 months of age (Table 5.2). The highest number of sequences assigned to *Peptostreptococcaceae Peptostreptoclostridium*, *Staphylococcaceae Staphylococcus*, *Erysipelotrichaceae Holdmanella*, *Christensenellaceae* R-7 group, and *Bacillaceae bacillus* (all of the Firmicutes phylum), were found in formula-fed infants and lowest in mixed-fed infants at 12 months of age (Table 5.2). Sequences assigned to *Erysipelotrichaceae faecalitea* and *Turicibacter* genus were highest in abundance in mixed-fed infants and lowest in formula-fed infants at 12 months of age (Table 5.2). The relative

2316 abundance of *Clostridium sensu stricto* and taxa from the *Erysipelotrichaceae* family was
2317 significantly different between modes of milk feeding at 9 and 12 months of age (Table 5.2).

Table 5. 2. Sequences assigned to KEGG pathways and microbial species in faecal samples that were significantly different (ANOVA $p \leq 0.05$) between breastfed (BF), mixed-fed (MF), and formula-fed (FF) infants at 4, 9, or 12 months (mo) of age. The highest relative abundance is shown in orange, lowest in blue, and mid-value in grey. P values for the time point(s) in which the comparison was significant are shown in the Sig at timepoint column.

KEGG Ortholog Pathways			Sig at Timepoint			Relative abun.		
Level 1	**Level 2**	**Level 3**	**4 mo**	**9 mo**	**12 mo**	**BF**	**MF**	**FF**
BH Protein families	Genetic information processing	03051 Proteasome		0.026				
		04131 Membrane trafficking		0.036				
	Metabolism	01002 Peptidases		0.044				
		01004 Lipid biosynthesis proteins		0.026				
	Signaling cellular processes	02048 Prokaryotic defense system	0.034					
Cellular Processes	Cell growth and death	04214 Apoptosis - fly		0.042				
		04216 Ferroptosis		0.030				
	Prokaryotic cellcommunity	02024 Quorum sensing		0.026				
	Transport & catabolism	04138 Autophagy - yeast		0.004				
		04146 Peroxisome		0.048				
Enviromental Information Processing	Membrane transport	02060 Phosphotransferase system PTS		0.006				
	Signal transduction	04066 HIF 1 signaling pathway		0.008				
		04070 Phosphatidylinositol signaling system		0.028				
		04152 AMPK signaling pathway		0.010				
Genetic Information Processing	Folding sorting and degradation	03050 Proteasome		0.046				
		03060 Protein export		0.034				
		04141 Protein processing in endoplasmic reticulum			0.054			
	Replication and repair	03420 Nucleotide excision repair		0.044				
	Translation	03013 RNA transport		0.006				
Metabolism	Biosynthesis of other secondary metabolites	00311 Penicillin and cephalosporin biosynthesis			0.036			
		00333 Prodigiosin biosynthesis		0.008				
		00405 Phenazine biosynthesis		0.022				
	Carbohydrate metabolism	00010 Glycolysis - Gluconeogenesis		0.002				
		00020 Citrate cycle - TCA cycle		0.014				
		00030 Pentose phosphate pathway		0.018				
		00051 Fructose and mannose metabolism		0.002				
		00620 Pyruvate metabolism		0.002				
		00640 Propanoate metabolism		0.006				
		00650 Butanoate metabolism		0.002				
	Energy metabolism	00190 Oxidative phosphorylation		0.012				
		00680 Methane metabolism		0.044	0.038			
		00710 Carbon fixation in photosynthetic organisms		0.002	0.020			
		00720 Carbon fixation pathways in prokaryotes		0.026				
	Lipid metabolism	00561 Glycerolipid metabolism		0.018				
		00071 Fatty acid degradation		0.046				
		01040 Biosynthesis of unsaturated fatty acids		0.012				
	Cofactors & vitamins	00830 Retinol metabolism			0.022			
		00730 Thiamine metabolism		0.006				
					0.020			
	Amino acids	00410 beta Alanine metabolism		0.006				
	Terpenoids & polyketides	01055 Biosynthesis of vancomycin group antibiotics	0.028					
	Xenobiotics biodegradation & metabolism	00633 Nitrotoluene degradation		0.030				
		00643 Styrene degradation		0.040				
		00980 Xenobiotic metaolism by cytochrome P450			0.008			
		00982 Drug metabolism - cytochrome P450			0.010			
Microbiome Composition								
Phyla	**Family**	**Genus**	**4 mo**	**9 mo**	**12 mo**	**BF**	**MF**	**FF**
Proteobacteria	*Alcaligenaceae*	*Parasutterella*	0.020					
	Desulfovibrionaceae	*Bilophila*	0.002		0.046			
	Moraxellaceae	*Moraxellaceae*			0.044			
		Acinetobacter			0.034			
	Pseudomonadaceae	*Pseudomonas*			0.038			
Actinobacteria	*Bifidobacteriaceae*			0.040				
Bacteroides	*Prevotellaceae*	*Prevotella*		0.032				
		Prevotellaceae		0.036				
	Rikenellaceae	*Alistipes*			0.010			
	Porphyromonadaceae	*Barnesiella*			0.020			
Firmicutes	Unclassified	Unclassified		0.038				
	Lachnospiraceae	*Anaerostipes*		0.002				
		Lachnospiraceae Dorea		0.016				
	Clostridiaceae	*Clostridiales*		0.020				
		Clostridium sensu stricto		0.018	0.004			
	Erysipelotrichaceae	*Erysipelotrichaceae*		0.016	0.002			
		Faecalitalea		0.026				
		Turicibacter			0.044			
		Holdemanella			0.048			
	Bacillaceae	*Bacillaceae*			0.002			
		Bacillus			0.006			
	Christensenellaceae	*R 7 group*			0.012			
	Ruminococcaceae	*Ruminococcus*			0.014			
	Peptostreptococcaceae	*Intestinibacter*			0.026			
		Peptoclostridium			0.030			
	Staphylococcaceae	*Staphylococcus*			0.022			

More and different lipids were significantly different between breastfed and mixed-fed infants at 4 months of age than between infants fed by different modes of milk at 9 or 12 months of age (Table 5.3). At 4 months of age, mixed-fed infants had significantly higher levels of all lipids that emerged as significant by ANOVA (Table 5.3). At 9 and 12 months of age, a series of phosphorylethanolamines was significantly higher in breastfed infants and lowest in formula-fed infants and sphingomyelin (SM(d18:1/16:0)+HCOO) at 9 months of age. Ceramides (Cer(d18:1/18:1)+H, Cer(d34:0)+Na), Digalactosyldiacylglycerol (DGDG(16:0/18:1)+HCOO), and sphingomyelin (SM(d36:1)+H) were highest in mixed-fed infants and lowest in formula-fed infants at 9 months of age (Table 5.3). Monoglycerides (MG(27:1) +H, MG 26:2) +NH_4), diglyceride (DG (21:0) +NH_4), and triglyceride (TG(8:0/8:0/12:0) +NH4) were highest in formula-fed and infants at 9 months of age (Table 5.3). Lysophosphatidylcholine (LPC(18:0)+H) was highest in formula-fed infants and lowest in mixed-fed infants at 12 months of age, whereas phosphatidylglycerol and sphingosine were highest in mixed-fed infants and lowest in formula-fed infants at 12 months of age (Table 5.3).

The aqueous metabolites that were significantly different were detected in the highest concentrations in mixed-fed infants at 4 and 12 months of age, but more aqueous metabolites were significantly highest in formula-fed infants at 9 months of age (Table 5.3). N-acetyl-L-tyrosine, neopterin, and inosine were highest in mixed-fed infants at 4 months of age but highest in formula-fed infants at 9 months of age. Inosine, xanthurenic acid, 3-O-a-L-fucopyranosyl-D-glucose, aspartylglutamine, glutaminylthreonine, and hydroxyprolyl-asparagine were highest in formula-fed infants and lowest in mixed-fed infants at 9 months of age (Table 5.3). Tyrosyl-glutamine and 3-hydroxyhippuric acid were highest in formula-fed infants and lowest in breastfed infants at 9 months of age (Table 5.3). At 12 months of age, the most significantly different aqueous metabolites were highest in mixed-fed infants,

except proline, which was highest in formula-fed infants, and 2-piperidinone, which was highest in breastfed infants (Table 5.3). Four amino acids and peptides (saccharopine, a glutaminylproline fragment, arginine, N6, N6, N6-trimethyl-L-lysine isotope) were highest in mixed-fed infants and lowest in formula-fed infants at 12 months of age (Table 5.3).

Table 5. 3. Faecal metabolites that were significantly different (ANOVA $p \leq 0.05$) between breastfed (BF), mixed-fed (MF), and formula-fed (FF) infants at 4, 9, or 12 months (mo) of age. The highest relative concentration is shown in orange, lowest in blue, and mid-value in grey. White spaces represent the lack of exclusively formula-fed infants at 4 months of age. P values for the time point(s) for the significant comparison are shown in the Sig at timepoint column. Asterisks in the aqueous metabolites indicate the presence of adducts that were detected separately and significant by ANOVA.

Lipid Metabolites		Sig at timepoint			Relative conc.		
Class	**Compound**	**4 mo**	**9 mo**	**12 mo**	**BF**	**MF**	**FF**
Ceramide	Cer(d18:0/18:2)+Na	0.002					
	Cer(d36:1)+Na	0.002					
	Cer(d18:1/18:1+O)+H	0.026					
	Cer(d18:2/16:0)+H	0.010					
	Cer(d18:1/18:2)+H	0.026					
	Cer(d16:0/16:0)+H	0.040		0.030			
	Cer(d18:1/18:1)+H		0.036				
	Cer(d34:0)+Na		0.026				
cPA	cPA(18:0)-H	0.054					
Digalactosyldiacylglycerol	DGDG(16:0/18:1)+HCOO		0.004	0.034			
Fatty Acids	FA(20:4)-H	0.004					
	FA(22:5)-H	0.002					
Lysophosphatidylmethanol	LPMe(18:0)-H	0.004					
Lysophosphatidylcholine	LPC(18:2)+H	0.044					
	LPC(18:0)+Na	0.010					
	LPC(18:1)+H	0.010					
	LPC(18:0)+H	0.030					
				0.048			
Lysophosphatidyl-ethanolamine	LPE(16:0)-H	0.042					
	LPE(18:0)-H	0.014					
Monoglyceride	MG(18:3) +NH4	0.002					
	MG(18:1) +NH4	0.008					
	MG(20:5) +NH4	0.006					
	MG(18:3) +NH4	0.034					
	MG(18:3) +NH4	0.036					
	MG(34:4) +NH4	0.008					
	MG(27:1) +H		0.004				
	MG 26:2) +NH4		0.010				
Diglyceride	DG (21:0) +NH4		0.050				
	DG(18:3/18:2) +NH4	0.030					
Triglyceride	TG(8:0/8:0/12:0) +NH4		0.038				
Phosphorylethanolamine	PE 25:0; [M-H]-		0.050				
	PE 38:1; [M-H]-		0.002	0.002			
	PE 40:1; [M-H]-		0.002	0.004			
	PE 44:1; [M-H]-		0.014	0.040			
Phosphatidylethanolamine	PE (36:0) [M-H]-		0.002				
	PE(18:0e)+H	0.008					
	PE(16:0e)+H	0.048					
	PE(16:0/18:2)+H	0.018					
	PE(16:0e)+H	0.008					
Phosphatidylglycerol	PG (27:0)-			0.004			
SO	So(d18:0)+H			0.052			
Sphingomyelins	SM(d36:1)+H	0.010					
			0.054				
	SM(d18:1/16:0)+HCOO		0.038				

Aqueous Metabolites							
Subclass	**Common name**	**4 mo**	**9 mo**	**12 mo**	**BF**	**MF**	**FF**
Pyrimidines and pyrimidine derivatives	Uracil	0.054		0.048			
Amino acids, peptides, and analogues	Aspartylglutamine	0.034					
	Isoleucyl-Threonine	0.018					
	L-Asparagine	0.054					
Medium-chain hydroxy acids and derivatives	3-Hydroxydodecanoic acid	0.006					
	M266T862	0.016					
	M185T723	0.034					
	M266T858	0.042					
	M434T662 + M435T662*	0.042					
Amino acids, peptides, and analogues	N-Acetyl-L-tyrosine	0.028					
			0.010				
Pterins and derivatives	Neopterin	0.006					
			0.048				
Purine nucleosides	Inosine	0.042					
			0.002				
	M469T944		0.010				
Medium-chain hydroxy acids	3-hydroxyhippuric acid		0.008				
Quinoline carboxylic acids	Xanthurenic acid		0.024				
Fatty acyl glycosides	3-O-a-L-Fucopyranosyl-D-glucose*		0.006				
Amino acids, peptides, and analogues	Aspartylglutamine		0.032				
	Glutaminylthreonine		0.048				
	Hydroxyprolyl-Asparagine		0.012				
	Tyrosyl-Glutamine*		0.002				
	M164T675		0.014				
	M160T896		0.008				
Benzoic acids and derivatives	3-hydroxyhippuric acid			0.016			
Amino acids, peptides, and analogues	Proline			0.028			
	Asymmetric dimethyl arginine****			0.002			
	Citrulline			0.024			
	Glutaminylproline fragment			0.004			
	Saccharopine			0.014			
	Arginine			0.002			
	N6,N6,N6-Trimethyl-L-lysine isotope			0.050			
Carbohydrates and conjugates	Galactitol			0.018			
Purines and purine derivatives	8-Hydroxy-7-methylguanine			0.018			
	7-Methylguanine*			0.042			
Piperidinones	2-piperidinone			0.032			
Indolecarboxylic acids and derivatives	Indole-2-Carboxylic*			0.024			

5.4. Discussion

Formula and breastmilk constituted a large part of the diet at all three time points (4, 9 and 12 months of age). Nutrient composition estimates support prior research findings that formula-fed infants consume more energy, protein, vitamins, and minerals than breastfed infants (Haisma et al., 2003; Heinig et al., 1993). In this cohort, this higher estimated intake of most nutrients analysed can be attributed to infant formula. However, increased dietary diversity of infants fed formula also contributed.

The differences in anthropometry of infants fed different milk types were nominal (Table 5.1), supporting prior evidence indicating that the classic profile of macronutrients and micronutrients is only part of the energy capture, growth, and development story in infants. Complementary feeding studies in low- and middle-income countries have found that microbiome composition, function, and related metabolites of the faecal microbiome are at the core of energy capture, immunomodulation, and epithelial barrier integrity affecting the absorption and utilisation of dietary and microbially produced nutrients (Keusch et al., 2014; Nancy F Krebs et al., 2013).

Enrichment of species belonging to the *Erysipelotrichaceae* family, such as those that were found in infants fed formula in this cohort, may be a good indicator of higher energy consumption and excess energy reaching the lower GIT. This family of bacteria have been found to be positively associated with energy consumption in adults, as well as in infants who had higher energy intake in the New Zealand BLISS study exploring baby-led weaning (Daniels et al., 2015). Relative abundances of taxa from the *Erysipelotrichaceae* family have also been positively associated with dairy consumption in Australian toddlers and have been detected in infants with islet autoimmunity in the Canadian TEDDY study, suggesting a possible connection with immune development (Daniels et al., 2015; Kaakoush, 2015; C. J. Stewart et al., 2018). In adults, the *Erysipelotrichaceae* family is also associated with

elevated inflammatory biomarkers (tumour necrosis factor-alpha) and inflammatory diseases of the GIT (Kaakoush, 2015).

As predicted, breastmilk consumption was negatively associated with microbial diversity, likely due to the bifid-shunt pathway in *Bifidobacterium* subs. *infantis* that preferentially utilises HMOs, effectively outcompeting other microbes for resources (Sela et al., 2008). Species that cannot compete with HMO utilising *Bifidobacterium spp.* for HMOs, have the enzyme capacity to use other substrates, such as from plants, efficiently. However, the reduced dietary diversity reported in breastfed infants also amounts to reduced substrate diversity to support diverse microbial populations in the GIT of infants. This aligns with the distinctive pattern of decreased relative abundance of metabolic pathways, particularly those classified as Carbohydrate metabolism, in breastmilk fed infants compared to mixed-fed and formula-fed infants at 9 months of age.

Mode of milk feeding was more distinct and significant at 9 months of age than 12 months of age, particularly on the microbiome composition, functional gene abundance, and aqueous metabolites of the faecal samples, but not the lipid metabolite profiles. Bäckhed et al. proposed that the cessation of breastmilk has the largest impact on the infant GIT microbiome, which is in part supported by the results at 12 months of age reported here, but the majority of infants in this study were still consuming breastmilk at 12 months of age, which limit the conclusions that can be made agreeing or dissenting that hypothesis (Bäckhed et al., 2015). More apparent in this study was the convergence of the faecal microbiome between individuals, by all metrics assessed, as shown by increased overlap of groups in Figure 5.4, in agreement with de Muink & Trosvik (de Muinck & Trosvik, 2018).

The addition of formula to the breastmilk diet (mixed feeding) had a strong effect on faecal lipid metabolites at 4 months of age prior to the inclusion of solid foods. Mixed-fed infants had similar faecal profiles of lipid metabolites to those of breastfed infants than

formula-fed infants at 9 months of age. The number of lipid metabolites that were significantly different between infants fed different types of milk decreased with age, which may indicate increased lipid digestion prior to the large intestine.

At 9 months of age concentrations of significantly different aqueous metabolites were highest in formula fed infants, which may reflect higher protein consumption from formula and incomplete protein digestion earlier in complementary feeding. However, at 12 months of age concentrations of significantly different aqueous metabolites were highest in mixed-fed infants. One explanation is that the combined effect of breastmilk and formula constituents moving through the GIT at different times results in distinct trophic networks and the increased production of metabolic by-products. Why this would be so apparent at 12 months of age but not 9 months of age is unclear, and existing research on mixed feeding of infants is sparse. The elevated neopterin, a biomarker of GIT inflammation and permeability, in mixed-fed infants at 4 months of age and formula-fed infants at 9 months of age supports previous studies suggesting increased inflammatory responses in formula-fed infants (Castanet et al., 2020; McCormick et al., 2017).

The bioavailability of nutrients from infant formula compared to breastmilk were important contributors to the results of this study but cannot be accounted for from the data and biological samples collected. The structures of vitamins and minerals that are present in the formula may be different to those found in breastmilk, thus affecting their absorption and utilisation (Nancy F. Krebs, 2001). Additionally, breastmilk and formula differentially impact the structural characteristics of the epithelium, such as crypt depth and villi height and density, which may affect the transport of nutrients from the lumen to the bloodstream. For instance, increased inflammation associated with neopterin in formula-fed infants has been associated with epithelial crypt hyperplasia (Cummins & Thompson, 2002).

The minor limitations of the analyses conducted for the Nourish to Flourish study in relation to modes of milk feeding include estimations about quantities of breastmilk consumed, estimations of the nutrient composition of foods included in 3-day diet records, and individual variations in breastmilk composition. The major limitation of this analysis was the small group size with an inequal distribution of infants consuming different types of milk at each time point. An additional covariate not recorded, was if infants fed breastmilk were fed by pump and bottle, which may influence the development of self-regulation of milk and food intake. Factors such as absorption and utilisation of nutrients from milk sources cannot be determined from diet records and faecal samples alone.

5.5. Conclusions

The analysis of the dietary records shows that infants fed formula consumed higher levels of most dietary nutrients during complementary feeding. This was likely due to the nutrient density of formula compared to breastmilk and increased average diet diversity among infants fed formula. This observation corresponded to higher microbial diversity, relative abundances of specific species of the phyla Proteobacteria, Bacteroides, and Firmicutes, and relative abundances of microbial DNA sequences assigned to metabolic pathways in faecal samples as a proxy of the lower GIT, particularly at 9 months of age. Conversely, the addition of formula to the diet had the most significant influence on the faecal lipid metabolite profile prior to introducing solid foods and the most significant influence on the faecal aqueous metabolite profile at 12 months of age.

Chapter 6. Associations between estimated nutrient consumption and the infant GIT microbiome and metabolome during complementary feeding

The food consumed and the nutrients absorbed and metabolised during weaning are crucial for the healthy development of the infant. However, how these nutrients influence the structure and function of the immature and unstable infant GIT microbiome is not well understood. This study aimed to identify associations between dietary nutrients, composition, and gene abundances of the infant faecal microbiome, and metabolome to inform the design of future interventions. Daily nutrient consumption from a cohort of 25 infants was estimated from 3-day diet records collected prior to faecal sample collection at two time points during complementary feeding. Data were integrated using projection to latent structure regressions (PLS-R) and linear correlations. The results show that nutrient intakes from the diet had a stronger correlation with the faecal microbiome composition and gene abundances at 9 than 12 months of age. In contrast, the faecal metabolites were more strongly associated with nutrient intakes at 12 months of age. Of the macronutrients analysed, dietary protein had the strongest influence on the microbiome, whereas dietary fats and carbohydrate showed the weakest associations. Of the dietary minerals analysed, iron, zinc, and phosphorus showed the strongest associations, whereas sodium, selenium, and iodine showed the weakest associations. Vitamin C, E, B6, and B12 were the most influential vitamins, whereas thiamine, folate, and vitamin A showed only weak associations. This study showed that the relationship between dietary components, microbiome composition and functional potential in faeces, as a proxy of the lower GIT microbiome, was stronger at 9 than 12 months of age. This finding indicates a crucial point during weaning when the GIT microbiome may be more susceptible to dietary intervention.

6.1. Introduction

The addition of solid foods to the milk-based infant diet alters the composition of nutrients consumed by the infant and what nutrients are available to the developing infant GIT microbiome. This transition aligns with the increasing needs for energy, changing nutritional requirements to fuel body growth, the developing GIT, and cofactors to facilitate metabolic reactions important for host-microbiota interactions. In addition, achieving adequate intake or recommended daily intake (RDI) of nutrients during the first 1000 days of life supports improved physical and cognitive outcomes throughout the lifespan, as shown by the occurrence of growth failure during weaning in low and middle-income countries (T.-C. Wu & Chen, 2009). However, how these nutrients influence and interact with the immature and unstable GIT microbiome, which has also been implicated in lifelong health and wellbeing, is not well understood.

Dietary carbon and nitrogen sources, from carbohydrates and protein, respectively, are the primary compositional and mechanistic drivers of the GIT microbiome (Holmes et al., 2017; Nicole M Koropatkin, Cameron, & Martens, 2012; Sonnenburg & Sonnenburg, 2014). Lipids are not considered a strong modulator of the microbiome due to high digestibility prior to the large intestine. However, infant formulas supplemented with milk fat globule membrane (MFGM) compared to plant-derived lipids led to differences in the faecal microbiome composition of mice (Bhinder et al., 2017).

The transition from single sources of nutrients (breastmilk and/or formulas) to a more diverse diet, parallels compositional changes in the GIT microbiome. This includes predominant taxa from the Actinobacteria phylum, which harbour enzymes that are specific to certain bond structures in carbohydrates, to species belonging to the Firmicutes phylum, which harbour diverse enzymes with lower specificity for bond structures (Bode, 2012).

Protein degradation has been attributed to species in the Firmicutes and Proteobacteria phyla. However, urease enzymes have recently been discovered in *Bifidobacterium infantis* suggesting more complex maintenance of nitrogen balance in the GIT of the infant (Schimmel, Kleinjans, Bongers, Knol, & Belzer, 2021). The majority of research into protein digestion and utilisation in the infant GIT has investigated the addition of free amino acids or hydrolysed proteins to formula (He, Sotelo-Orozco, Rudolph, Lönnerdal, & Slupsky, 2020). Baby rhesus monkeys fed formula supplemented with free amino acids developed microbiomes that were more similar, by alpha and beta diversity metrics, to baby monkeys fed human milk compared to those fed standard formula (He et al., 2020). The effects of sources of protein from complementary food on the GIT microbiome have not been as well researched *in vivo*, although protein-rich foods (that are also rich in iron and zinc), have been the subject of microbiome specific complementary feeding studies (Qasem et al., 2017).

Micronutrients have been best studied in infants in the context of deficiencies in low and middle-income countries, with particular emphasis on host metabolism and anthropometrics. However, vitamins and minerals are increasingly recognised as important modulators of the GIT microbiome structure and stability due to their activity as co-enzymes in metabolic reactions in species from every branch of life. For instance, B vitamins need to be sourced exogenously from the diet or prototrophic microbes by at least 20% of the microbes (auxotrophs) of the gut microbiota (Sharma et al., 2019). This exchange and sharing of B vitamins from co-abundant species may promote stability in the microbiome (Magnúsdóttir et al., 2015; Rodionov et al., 2019; Sharma et al., 2019).

In infants, vitamin B12 deficiency has not been associated with any differences in relative abundances of species of microbes in faecal samples (Boran et al., 2020). Vitamins A, C, and E have not been extensively studied. However, some associations with the GIT microbiome composition have been suggested in adults. Vitamin B2, in addition to vitamin

C, has been associated with decreases in the relative abundance of taxa from the Proteobacteria phylum, and Vitamin A is associated with increased abundance of members from the Actinobacteria phylum (Pham et al., 2021; Steinert, Lee, & Sybesma, 2020). Genus level associations with vitamin supplementation are more robust than phyla or species level implicating *Alistipes, Bilophila, Clostridium, Collinsella, Sutterella, Faecalibacterium, Coprococcus, Odoribacter* genera, and unidentified genera from the family *Peptostreptococcaceae* (Pham et al., 2021).

Achieving mineral balance is crucial for adequate infant growth. However, higher levels of iron and zinc, in particular from supplemented sources, have been associated with increased oxidative stress and relative abundances of *Clostridia* species compared to infants who consumed iron and zinc through meat-based products (Nancy F Krebs et al., 2013; Podany et al., 2019; Qasem et al., 2017). In addition, the requirement for other minerals, including calcium, magnesium, and potassium, are poorly understood for developing the GIT microbiome of the infant. Several pathways have been suggested connecting probiotic microbial species, prebiotics, SCFAs, and calcium absorption with rapid bone accretion during complementary feeding (Rizzoli, 2019; Yan & Charles, 2017).

Changing quantities of nutrients during the transition from a milk-only diet to a diverse diet of solid foods coincides with a changing consortium of microbial species in the lower GIT over the first two to three years of life. Species that effectively compete for nutrients and are less specific about the substrates they utilise are more likely to become established members of the core GIT microbiome during the complementary feeding period (Koenig et al., 2011; Verster & Borenstein, 2018). However, associations with pathways and metabolic by-products have not been investigated when the diet changes.

Chapter 6 details the patterns of association between dietary nutrient consumption estimated from 3-day food records with multiple aspects of the faecal microbiome

(composition, gene abundance) and the faecal metabolome (metabolite abundances). Faecal samples collected at 9 and 12 months were analysed by shotgun metagenomics sequencing (composition, gene abundance) and LC-MS (metabolites). PLS-R modelling and linear correlations were used to identify patterns of association between estimated nutrient intakes and faecal microbial genera. Additionally, predicted functional pathways (inferred from gene abundances) and lipid and aqueous metabolites present in faeces were also analysed.

6.2. Methods

Microbiome and metabolomic methods are detailed in Chapter 3. Diet records and nutrient composition methods are included in Chapter 5.

6.2.1. Statistical analysis: normalisation and scaling of datasets

KEGG pathway counts at Level 3 and taxonomy at Level 6 (genus) were Total Sum Scaled, followed by unit variance normalisation in SIMCA (Umetrics Version 16). Faecal lipid and aqueous metabolite peak tables were normalised by Lowess normalisation during pre-processing using Galaxy Workflow for Metabolomics. Peak tables were further transformed using log-transformation, which achieved a more Gaussian distribution.

Orthogonal PLS-R (OPLS-R) were conducted in SIMCA (Umetrics Version 16). Faecal datasets (species composition, KEGG pathways, lipid metabolites, aqueous metabolites) were regressed in Type 1 models of three groups of nutrients: macronutrients (protein, lipids, carbohydrate, sugar, starch, and dietary fibre), minerals (phosphorus, potassium, iron, magnesium, selenium, sodium, and zinc), and vitamins (niacin, riboflavin, thiamine, vitamin A, vitamin B6, vitamin B12, vitamin C, vitamin E, and folate).

Variables were selected by comparing models built on the highest and lowest regression coefficients and models, including variables in the top 5-10% of normal probability. Top-performing models were selected based on the smallest p-values for the most nutrients, the highest R^2X (explained variance of predictor values) and R^2Y (responder values), and the steepest slope for the model. The significance of the influence of the estimated nutrient intake on the microbiome variable was interpreted by CV-ANOVA p-values. From each, VIPs were aligned across models and microbiome datasets and included in linear pairwise correlation and filtered heatmaps (Figures 6.4-6.8).

6.3. Results

6.3.1. OPLS-R model performance

Model performance for KEGG pathways and microbial species against nutrient classes (macronutrients, minerals, and vitamins) consumed was stronger at 9 months than 12 months, as shown by the higher R^2X, R^2Y, and R^2 values (Table 6.1). Explained variance of the microbial taxa datasets (R^2X) decreased at 12 months of age, except for aqueous metabolites in the vitamin model at 12 months of age which increased (Table 6.1). However, explained variance (R^2Y) of the nutrients included in models mostly increased at 12 months of age compared to 9 months of age. Predictability (Q^2) is not a strong aspect of OPLS-R models, which was true here, as no consistent patterns emerged (Table 6.1).

Of the macronutrients, the dietary protein was the most significant in all models except for lipid metabolites at 12 months of age (Table 6.1). Zinc was significant in all except the KEGG pathway model at 12 months of age (Table 6.1). Dietary starch was significantly associated with lipid metabolites at 9 and 12 months of age, microbial species, and aqueous metabolites at 9 but not 12 months of age, and KEGG pathways at 12 but not at 9 months of age (Table 6.1). Similarly, dietary carbohydrates and fibre were significantly associated with both microbial species at 9 months and lipid metabolites at 12 months of age (Table 6.1). Carbohydrates were significant in the 9 months lipid metabolite model and the 12 months aqueous metabolite model, whereas dietary fibre was not (Table 6.1). However, dietary fibre was significant in the aqueous metabolite model at 9 months of age, whereas dietary carbohydrates were not. Lipids and sugar were not significant in any model at either time point (Table 6.1).

Minerals were significant in the microbial species and lipid metabolite models at 9 months of age and the aqueous metabolite model at 12 months of age (Table 6.1). This pattern of associations was driven by potassium, magnesium, and calcium (Table 6.1). Of the minerals, zinc was significant to all models except for KEGG pathways at 12 months of

age and was the only mineral significant to KEGG pathways at 9 months of age. Iron was significant to all models except the KEGG pathway models at 9 and 12 months of age (Table 6.1). Phosphorus was not significant to the KEGG pathway model at 9 months of age or the lipid metabolite model at 12 months of age but was the only mineral significant to the KEGG pathway model at 12 months of age (Table 6.1). Sodium was significant to microbial species and both metabolite models at 12 months of age but not at 9 months of age (Table 6.1). Selenium was significant to the lipid metabolite model at 9 months of age and the aqueous metabolite model at 12 months of age. Iodine was not significant to any model at either time point (Table 6.1).

More vitamins were significant in all models at 9 months of age than 12 months of age (Table 6.1), with some exceptions. KEGG pathways and microbial species were similar for vitamin E, while microbial species and lipid metabolites were similar for vitamin B12 at 12 months of age (Table 6.1). Vitamin B6 was significant to all models except for aqueous metabolites at 9 or 12 months of age and lipid metabolites at 9 months of age (Table 6.1). Niacin was significant to KEGG pathway and microbial species models at both 9 and 12 months of age but only significant to the lipid metabolite model at 12 months of age (Table 6.1). Riboflavin was significant to the microbial species model at both 9 and 12 months of age but only to lipid metabolites at 9 months of age and aqueous metabolites at 12 months of age (Table 6.1). Vitamin C was only significant to the microbial species and lipid metabolite model at 9 months of age, and folate was only significant to the microbial species model at 9 months of age (Table 6.1). Vitamin A and thiamine were not significant to any model at either time point (Table 6.1).

								Macronutrients							Minerals									Vitamins								
		Model Metrics	Macro	Mineral	Vitamin		**CV-ANOVA Metrics**	Protein (g)	Total fat (g)	Sugars (g)	Carbohydrate (g)	Starch (g)	Dietary fibre (g)	Potassium (mg)	Magnesium (mg)	Calcium (mg)	Phosphorus (mg)	Iron (mg)	Zinc (mg)	Sodium (mg)	Selenium (µg)	Iodine (µg)	Vitamin A (µg)	Total folate (µg)	Thiamin (mg)	Riboflavin (mg)	Niacin (mg)	Vitamin B6 (mg)	Vitamin B12 (µg)	Vitamin C (mg)	Vitamin E (mg)
9 months of age	KEGG	**R2X**	0.60	0.89	0.71		**SS**	6.28	0.79	0.00	5.50	3.04	4.59	5.11	3.75	2.17	4.94	5.22	7.35	0.00	1.19	0.00	0.00	0.00	0.00	2.03	6.02	5.81	9.43	4.34	9.42
		R2	0.17	0.17	0.23		**F**	3.90	0.37	0.00	3.27	1.59	2.60	2.98	2.04	1.10	2.85	3.06	4.86	0.00	0.57	0.00	0.00	0.00	0.00	1.01	3.68	3.51	7.12	2.43	7.11
		Q2	0.11	0.13	0.14		**p**	**0.04**	0.69	1.00	0.06	0.23	0.10	0.07	0.15	0.35	0.08	0.07	**0.02**	1.00	0.57	1.00	1.00	1.00	1.00	0.38	**0.04**	**0.05**	**0.00**	0.11	**0.00**
		R2Y	0.51	0.67	0.39		**SD**	1.77	0.63		1.66	1.23	1.51	1.60	1.37	1.04	1.57	1.62	1.92		0.77					1.01	1.73	1.70	2.17	1.47	2.17
	Microbes	**R2X**	0.36	0.53	0.40		**SS**	10.87	0.85	0.00	5.96	8.05	6.09	9.94	7.56	5.98	8.29	9.28	9.96	0.51	2.49	0.08	4.21	5.84	5.06	6.99	8.83	10.28	7.89	9.79	12.35
		R2	0.29	0.32	0.46		**F**	9.10	0.40	0.00	3.64	5.55	3.74	7.78	5.06	3.65	5.81	6.93	7.81	0.24	1.27	0.04	2.34	3.54	2.94	4.52	6.40	8.25	5.39	7.58	11.65
		Q2	0.22	0.25	0.33		**p**	**0.00**	0.67	1.00	0.04	**0.01**	**0.04**	**0.00**	0.02	0.04	**0.01**	**0.00**	**0.00**	0.79	0.30	0.96	0.12	**0.05**	0.07	**0.02**	**0.01**	**0.00**	**0.01**	**0.00**	**0.00**
		R2Y	0.51	0.67	0.67		**SD**	2.33	0.65		1.73	2.01	1.75	2.23	1.94	1.73	2.04	2.15	2.23	0.51	1.12	0.20	1.45	1.71	1.59	1.87	2.10	2.27	1.99	2.21	2.48
	Lipids	**R2X**	0.17	0.17	0.62		**SS**	13.55	0.00	0.00	16.70	10.11	4.80	11.86	12.77	9.71	12.30	11.80	15.22	3.09	7.20	0.00	0.00	2.44	4.14	10.50	4.47	8.94	8.84	10.45	12.45
		R2	0.43	0.52	0.45		**F**	6.49	0.00	0.00	11.45	3.64	1.25	10.74	12.51	7.47	11.56	10.65	19.06	1.63	4.71	0.00	0.00	0.56	1.04	3.89	1.15	2.97	2.92	3.86	5.39
		Q2	0.30	0.39	0.28		**p**	**0.00**	1.00	1.00	**0.00**	**0.02**	0.32	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.22	**0.02**	1.00	1.00	0.69	0.41	**0.02**	0.36	**0.04**	**0.05**	**0.02**	**0.00**
		R2Y	0.52	0.67	0.50		**SD**	1.84			2.04	1.59	1.09	2.43	2.53	2.20	2.48	2.43	2.76	1.24	1.90			0.78	1.02	1.62	1.06	1.49	1.49	1.62	1.76
	Aqueous	**R2X**	0.25	0.51	0.32		**SS**	9.37	0.00	0.00	2.74	8.51	7.80	10.77	9.95	10.44	12.63	16.15	15.08	0.00	5.13	3.07	0.00	5.86	6.65	3.72	9.78	6.11	16.75	6.32	15.19
		R2	0.30	0.57	0.60		**F**	7.04	0.00	0.00	1.42	6.04	5.29	2.44	2.13	2.31	3.33	6.18	5.07	0.00	0.82	0.44	0.00	0.65	0.77	0.37	1.38	0.68	4.62	0.72	3.45
		Q2	0.19	0.37	0.31		**p**	**0.00**	1.00	1.00	0.26	**0.01**	0.01	0.07	0.10	0.08	**0.02**	**0.00**	**0.00**	1.00	0.57	0.84	1.00	0.73	0.64	0.92	0.28	0.70	**0.00**	0.68	**0.02**
		R2Y	0.51	0.91	0.80		**SD**	2.16			1.17	2.06	1.97	1.34	1.29	1.32	1.45	1.64	1.59		0.92	0.72		0.86	0.91	0.68	1.11	0.87	1.45	0.89	1.38
12 months of age	KEGG	**R2X**	0.24	0.20	0.30		**SS**	12.85	0.67	0.00	7.60	10.65	7.72	11.03	10.30	3.41	13.62	5.33	9.89	10.32	5.19	0.00	0.00	0.00	0.00	2.67	7.26	6.54	6.37	0.00	5.30
		R2	0.46	0.57	0.26		**F**	5.77	0.14	0.00	2.32	3.99	2.37	2.55	2.26	0.50	3.94	0.86	2.10	2.26	0.83	0.00	0.00	0.00	0.00	1.37	4.77	4.12	3.97	0.00	3.12
		Q2	0.24	0.30	0.10		**p**	**0.00**	0.96	1.00	0.09	**0.02**	0.09	0.06	0.08	0.80	**0.01**	0.54	0.10	0.08	0.56	1.00	1.00	1.00	1.00	0.27	**0.02**	**0.03**	**0.03**	1.00	0.06
		R2Y	0.58	0.67	0.58		**SD**	1.79	0.41		1.38	1.63	1.39	1.36	1.31	0.75	1.51	0.94	1.28	1.31	0.93					1.15	1.90	1.81	1.78		1.63
	Microbes	**R2X**	0.26	0.38	0.29		**SS**	11.34	5.70	0.00	4.89	4.58	3.12	5.29	4.17	7.62	9.11	5.74	9.95	7.34	3.48	0.00	0.75	4.13	5.59	6.59	10.13	4.20	1.52	0.00	0.79
		R2	0.49	0.33	0.22		**F**	4.48	1.56	0.00	1.28	1.18	0.75	3.11	2.31	5.11	6.73	3.46	7.79	4.85	1.86	0.00	0.36	2.29	3.34	4.16	8.03	2.34	0.74	0.00	0.37
		Q2	0.20	0.24	0.14		**p**	**0.01**	0.22	1.00	0.31	0.35	0.57	0.06	0.12	**0.02**	**0.01**	**0.05**	**0.00**	**0.02**	0.18	1.00	0.70	0.13	0.05	0.03	**0.00**	0.12	0.49	1.00	0.69
		R2Y	0.76	0.67	0.56		**SD**	1.68	1.19		1.11	1.07	0.88	1.63	1.44	1.95	2.13	1.69	2.23	1.92	1.32		0.61	1.44	1.67	1.81	2.25	1.45	0.87		0.63
	Lipids	**R2X**	0.26	0.46	0.29		**SS**	4.49	0.54	0.00	6.98	8.83	9.12	4.40	5.80	3.54	5.72	8.30	6.65	7.63	0.95	0.00	0.00	0.17	2.83	5.52	7.05	5.72	3.31	0.00	5.81
		R2	0.30	0.30	0.29		**F**	2.53	0.25	0.00	4.51	6.40	6.74	2.47	3.51	1.91	3.44	5.82	4.22	5.12	0.45	0.00	0.00	0.08	1.47	3.29	4.57	3.44	1.76	0.00	3.52
		Q2	0.20	0.19	0.12		**p**	0.10	0.78	1.00	**0.02**	**0.01**	**0.01**	0.11	**0.05**	0.17	**0.05**	**0.01**	**0.03**	**0.01**	0.64	1.00	1.00	0.92	0.25	0.06	**0.02**	**0.05**	0.20	1.00	**0.05**
		R2Y	0.58	0.67	0.58		**SD**	1.50	0.52		1.87	2.10	2.14	1.48	1.70	1.33	1.69	2.04	1.82	1.95	0.69			0.29	1.19	1.66	1.88	1.69	1.29		1.71
	Aqueous	**R2X**	0.16	0.30	0.66		**SS**	14.27	7.11	0.56	12.07	7.45	7.41	12.61	9.54	13.14	16.46	10.82	17.31	13.19	10.12	0.00	2.58	0.00	0.00	8.51	0.00	4.68	8.74	5.55	6.94
		R2	0.50	0.53	0.25		**F**	7.33	2.11	0.12	5.06	2.25	2.23	12.17	7.25	13.31	24.00	9.03	28.44	13.42	8.02	0.00	1.33	0.00	0.00	6.04	0.00	2.66	6.30	3.31	4.48
		Q2	0.34	0.47	0.09		**p**	**0.00**	0.12	0.97	**0.01**	0.10	0.10	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	1.00	0.29	1.00	1.00	**0.01**	1.00	0.09	**0.01**	0.06	**0.02**
		R2Y	0.58	0.67	0.56		SD	1.89	1.33	0.37	1.74	1.37	1.36	2.51	2.18	2.56	2.87	2.33	2.94	2.57	2.25		1.14			2.06		1.53	2.09	1.67	1.86

Table 6. 1. Orthogonal Projection to Latent Structure Regression model performance at 9 and 12 months of age for KEGG pathways, Microbes, Lipid metabolites, and Aqueous metabolites. Model performance metrics are shaded in greyscale, with higher values shaded darker. R^2X is the cumulative explained variance of the X variables (respective microbiome dataset), R^2Y is the cumulative explained variance of the Y variables (estimated nutrient intakes), R^2 is the explained variation or "goodness of fit," and Q^2 is the predicted variation or "goodness of prediction." SS = sum of squares, F = F value, probability = p-value, SD = standard deviation. Significant CV-ANOVA p values are shown in bold and shaded in grey.

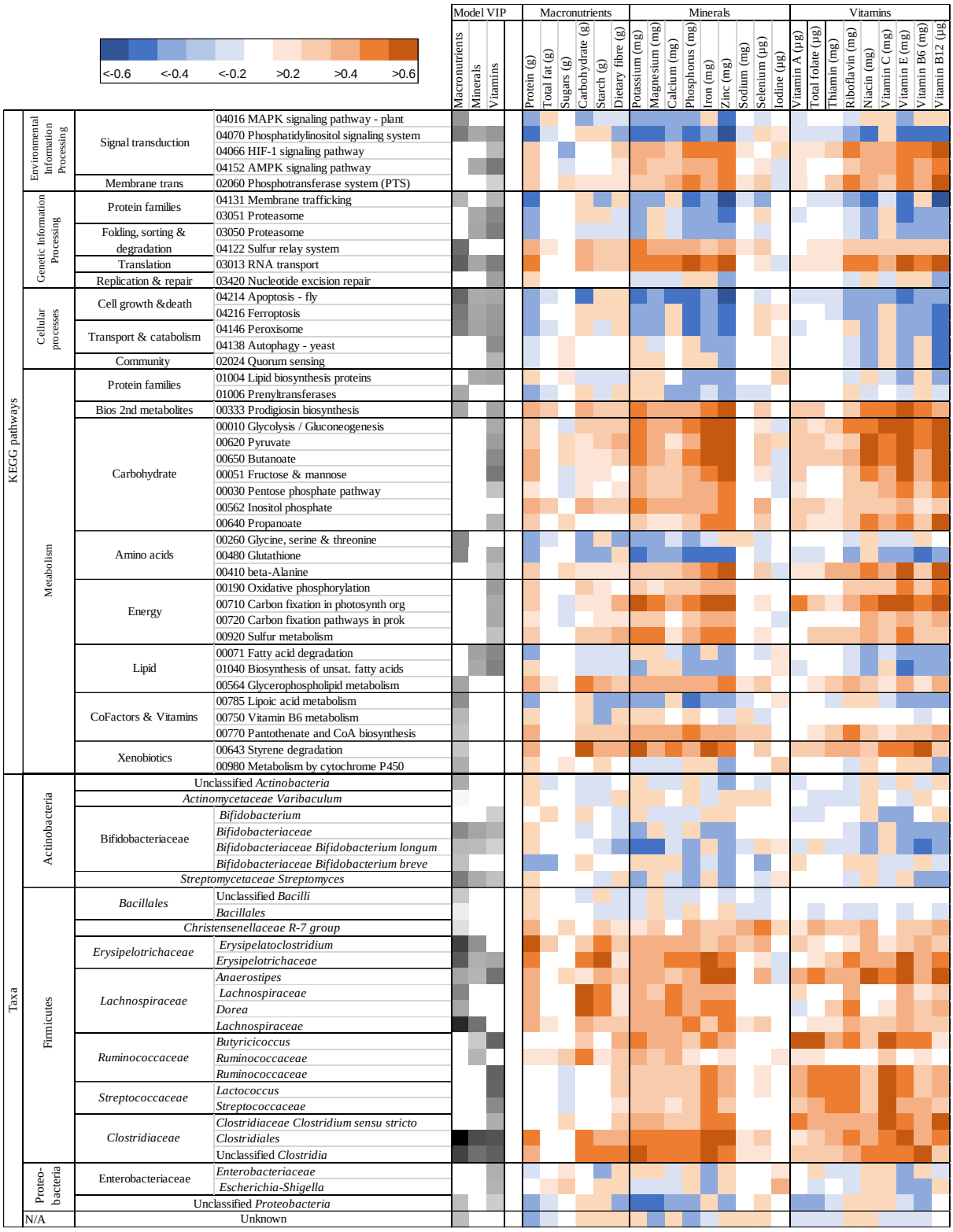


Figure 6. 1. Filtered heatmap indicating the strength and direction of linear correlation between VIPs (KEGG pathways and taxa) of OPLS-Regressions and estimated nutrient intakes included in models at 9 months of age (Table 6.1). The value of the cumulative VIP score is depicted in the greyscale, with black indicating the highest VIP values. Orange indicates positive correlation, and blue indicates negative correlation. KEGG pathway and microbe composition in faecal samples at 9 months (Figure 6.1).

6.3.2. Linear correlations between nutrients and sequences assigned to taxa and KEGG pathways at 9 months of age

In line with the OPLS-R macronutrient model performance, protein levels showed the strongest correlations with sequences assigned to KEGG pathways and microbial species (Figure 6.1). The strongest correlations with protein were negative associations with sequences assigned to the Phosphatidylinositol signalling system and Membrane trafficking. Additionally, protein was positively associated with sequences assigned to RNA transport and relative abundances of members from the *Erysipelotrichaceae* family, *Erysipeloclostridium* genus, and *Clostridiales* order (Figure 6.1).

Carbohydrate consumption was negatively associated with sequences assigned to MAPK signalling pathways, Apoptosis, and Amino acid metabolism and positively associated with Glycerophospholipid metabolism and Styrene degradation (Figure 6.1). In addition, carbohydrates and starch were positively correlated with sequences assigned to *Dorea* genus, *Erysipelotrichaceae*, *Lachnospiraceae*, *Lachnospiraceae*, *Ruminococcaceae*, and *Clostridiaceae* families, , and *Clostridiales* order (Figure 6.1).

Dietary fibre was negatively associated with sequences assigned to the Phosphatidylinositol signalling system and Glycine/serine/threonine degradation, and positively associated with unclassified taxa from the *Clostridia* class (Figure 6.1). Sugar was negatively associated with sequences assigned to the HIF-1 signalling pathway, but otherwise, lipids and sugar were not strongly associated with sequences assigned to any pathways or microbial taxa (Figure 6.1).

Zinc, iron, and phosphorus showed the most widespread patterns of strong positive correlations with sequences assigned to KEGG pathways and microbial species, followed by potassium, magnesium, and calcium (Figure 6.1). These minerals were positively associated with sequences assigned to Energy metabolism pathways, Carbohydrate metabolism pathways, Prodigiosin biosynthesis, Styrene degradation, and Beta-alanine metabolism. However, they

were negatively associated with sequences assigned to Glycine metabolism , Serine metabolism, Threonine metabolism, Glutathione degradation, and Lipoic acid metabolism (Figure 6.1).

Sequences assigned to Cellular processes and Protein families tended towards negative correlations with all nutrients, whereas Environmental and genetic information processing pathways split between positive and negative associations (Figure 6.1). The strongest correlations between minerals and sequences assigned to microbial species belonging to Actinobacteria and Proteobacteria phyla were negative. In particular, sequences assigned to *Bifidobacterium longum* and unclassified taxa from the Proteobacteria phylum were negatively associated with potassium and magnesium (Figure 6.1). Except for species belonging to the *Bacillales* family, members of the Firmicutes phylum were positively associated with minerals (Figure 6.1). Taxa from the families *Erysipelotrichaceae* and *Lachnospiraceae, Anaerostipes* genus, and *Clostridiales* order were strongly associated with iron and zinc (Figure 6.1).

Consistent with OPLS-R models, niacin, vitamin B12, and vitamin E were most strongly associated with sequences assigned to KEGG pathways and microbial species, followed by vitamin C, vitamin B6, and riboflavin (Figure 6.1). The patterns of positive and negative correlations were generally similar to those with minerals. Sequences assigned to Phosphatidylinositol signalling and Membrane trafficking were strongly negatively associated with vitamin E and vitamin B12. Cellular processes were negatively associated with vitamin B12, whereas RNA transport was positively associated with vitamin B12 (Figure 6.1). Vitamin A, folate, and thiamine were not strongly associated with any KEGG pathways, except for a strong positive association between vitamin A and Carbon fixation in photosynthetic organisms, likely genes from plant foods (Figure 6.1). *Ruminococcaceae* family and *Butyricoccus* genus from the Firmicutes phylum showed a strong positive association with vitamin A, vitamin C, and folate, whereas taxa from the Proteobacteria phylum were negatively

associated (Figure 6.1). Vitamin B6 was negatively associated with Glutathione metabolism and *Bifidobacterium longum* and positively associated with Styrene degradation and unclassified taxa from the *Clostridia* class (Figure 6.1).

6.3.3. Linear correlations between nutrients and faecal metabolites at 9 months of age

The most prominent pattern between nutrients consumed and faecal lipid metabolites at 9 months of age was a series of negative associations between faecal lipids (specifically: monogalactosyldiacylglycerol, sphingosine, sphingomyelin, and long-chain unsaturated phosphatidylethanolamines), and nutrients from three classes (macronutrients, vitamins, and minerals) (Figure 6.2). Of the macronutrients, protein and carbohydrates were the only nutrients to show this pattern, and carbohydrates were more strongly negatively associated than proteins (Figure 6.2). All minerals assessed apart from sodium, selenium, and iodine were consistent with this pattern, an important feature of the OPLS-R model with minerals (Figure 6.2). Riboflavin, vitamin B6, vitamin B12, and vitamin E were consistent with this pattern, and these micronutrients were also negatively associated with a ceramide (Figure 2). Vitamin B12 only showed strong negative associations with the phosphatidylethanolamines and ceramide, and vitamin C did not associate strongly with the monogalactosyldiacylglycerol (Figure 6.2).

Sporadic patterns emerged between the rest of the lipids, significant to the OPLS-R models (Figure 6.2). Total dietary fat was moderately negatively associated with sphingosine (Sph d18:1) and phosphatidylethanolamine (PE 36:0), and positively associated with a shorter chain phosphatidylethanolamine (PE 16:0) and a diglyceride (DG 18:2) (Figure 6.2). Sugar and iodine showed a strong negative association with PE (16:0) (Figure 6.2). Starch showed strong negative associations with monogalactosyldiacylglycerol (MGDG), phosphatidylethanolamine (PE (44:1), and phosphatidylcholine (PC 16:0/18:1). These associations were distinct from carbohydrates and dietary fibre (Figure 6.2). Dietary fibre showed weak positive associations with most lipids, apart from ceramide (CerG1 36:1) and PEs (38:1 & 36:0) (Figure 6.2).

The important variables to the OPLS-R macronutrient model were moderate positive associations with lysophosphatidylmethanol (LPMe 18:0), most diglycerides, and lysophosphatidylcholine (LPC 18:0), and no or weak negative associations with phosphatidylcholines (Figure 6.2). Weak positive associations between phosphatidylethanolamines were important to the vitamin OPLS-R model. In particular, zinc and selenium were more strongly associated with phosphatidylethanolamine (PE (16:0)) (Figure 6.2). Additionally, moderate positive associations between minerals, diglyceride (DG (21:0)), and phosphatidylglycerol (27:1), and weak negative associations with diglyceride (DG (34:4)) and phosphatidylcholine (PC (16:0/18:1)) were important to the OPLS-R model (Figure 6.2).

Associations between aqueous faecal metabolites and nutrients consumed at 9 months of age were predominantly weak (Figure 6.2). Some metabolites that were significant to all three nutrient models were positively associated with all nutrients (Figure 6.2). Neopterin was associated with iron, thiamine, and riboflavin. Inosine was associated with iron, zinc, vitamin A, folate, niacin, vitamin B12, vitamin E, and vitamin C (Figure 6.2). Vitamin B12 was also strongly associated with tyrosyl-glutamine, which was part of a series of metabolites that showed weak positive associations with all nutrients apart from sugars (Figure 6.2). Calcium and phosphorus also showed weak negative associations with tryptophanamide, thiamine, valylserine, ornithine, and taurine, which were important to the mineral OPLS-R model (Figure 6.2).

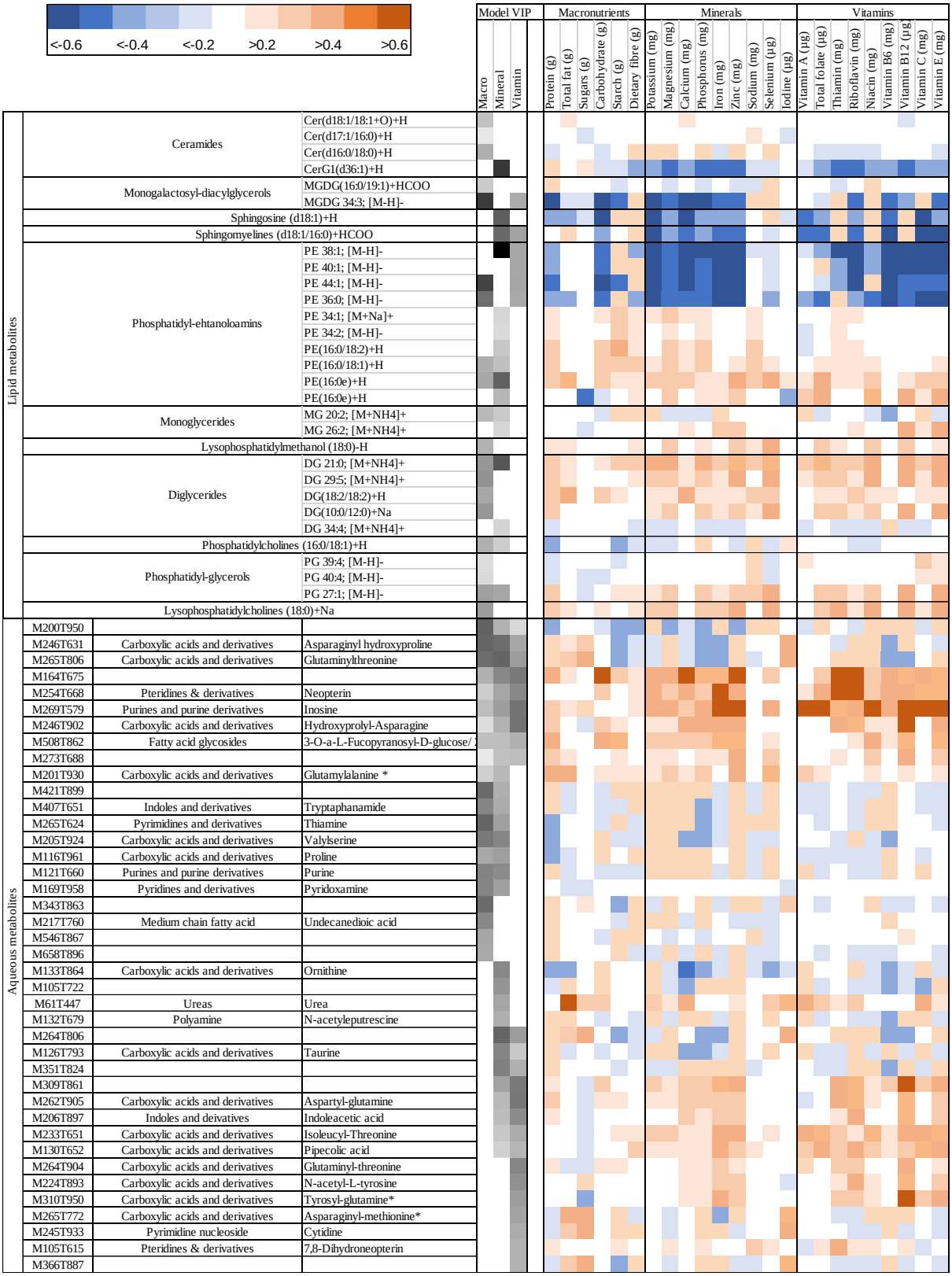


Figure 6. 2. Filtered heatmap indicating the strength and direction of linear correlation between VIPs (metabolites) of OPLS-Regressions and nutrients included in models at 9 months of age (Table 6.1). The value of the cumulative VIP score is depicted in the greyscale, with black indicating the highest VIP values. Orange indicates a positive correlation, and blue indicates a negative correlation, with colour intensity showing the strength of the correlations.

6.3.4. Linear correlations between nutrients and sequences assigned to taxa and KEGG pathways at 12 months of age

The strength of correlations at 12 months of age decreased compared to 9 months of age. In addition, the pathways and microbial species that emerged as significant to the OPLS-R models with strong correlations at 12 months of age were different than at 9 months of age (Figures 6.1-6.4).

At 12 months of age, protein showed strong correlations with a few select sequences assigned to KEGG pathways (ABC transporters, Protein export, Histidine metabolism, and Taurine metabolism) (Figure 6.3). Dietary fat, which was not strongly associated with any parameters at 9 months of age, showed strong negative associations with sequences assigned to Apoptosis, Tryptophan metabolism, and Glycosaminoglycan degradation at 12 months of age (Figure 6.3). Among associations between macronutrients and microbes, dietary fat was strongly associated with sequences assigned to *Veillonellaceae dialister* (Figure 6.3).

Carbohydrates, starch, and dietary fibre (which showed moderate consistency in patterns of association at 9 months of age (Figure 6.1)), showed different patterns of the association at 12 months of age (Figure 6.3). Dietary carbohydrate content was negatively associated with sequences assigned to Base excision repair and positively associated with sequences assigned to Thiamine degradation (Figure 6.3). Starch and dietary fibre content were negatively associated with sequences assigned to Taurine and hypotaurine metabolism, Insect hormone biosynthesis, and Limonene and pinene degradation (Figure 6.3). Only dietary fibre was associated with sequences assigned to Riboflavin metabolism (Figure 6.3). Dietary carbohydrate and starch content were positively associated with sequences assigned to *Peptostreptococcaceae Intestinibacter*. However, only carbohydrate content was strongly associated with sequences assigned to families *Peptostreptococcaceae*, *Peptostreptococcaceae*, *Peptostreptoclostridium*, and *Clostridiaceae Clostridium sensu stricto*. Only starch content was strongly positively associated with sequences assigned to

Lachnospiraceae selimonas and negatively associated with sequences assigned to *Streptomycetaceae streptomyces* (Figure 6.3).

The association between dietary mineral content and sequences assigned to KEGG pathways decreased in strength at 12 months of age compared to 9 months of age (Figures 6.1 and 6.3). Dietary potassium content showed a strong positive correlation with sequences assigned to ABC transporters and a strong negative association with sequences assigned to Protein export and Histidine metabolism (Figure 6.3). Dietary magnesium content showed a similar pattern as starch and dietary fibre with a strong negative correlation with sequences assigned to Insect hormone biosynthesis and Limonene and pinene degradation, and a strong positive association with sequences assigned to Riboflavin metabolism (Figure 6.3). Calcium showed a similar negative pattern of association as dietary fat with Apoptosis and Tryptophan metabolism. Iodine was also similarly negatively associated with sequences assigned to Glycosaminoglycan degradation along with lipids (Figure 6.3). Phosphorus, iron, and zinc were also strongly associated with sequences assigned to Thiamine degradation. Phosphorus and zinc were more strongly associated with ABC transporters than iron. Iron was more strongly associated with sequences assigned to Atrazine degradation and Riboflavin metabolism (Figure 6.3). Zinc and calcium showed stronger negative associations with sequences assigned to Tryptophan metabolism, Retinol metabolism, and Xenobiotic metabolism than other minerals (Figure 6.3). In particular, sequences assigned to Metabolism by cytochrome P450 was significantly negatively associated with zinc and calcium (Figure 6.3).

Sequences assigned to the Actinobacteria phylum were not strongly associated with any minerals, except for *Streptomycetaceae streptomyces*, which showed a strong negative association with magnesium (Figure 6.3). There was a weak positive trend between *Bifidobacterium* genus and some minerals (magnesium and iodine), and a weak positive trend

between geni from the *Acidominococcacea* family and most minerals (Figure 6.3). The taxa from the family *Bacillaceae* were more strongly positively associated with minerals at 12 months than 9 months of age, particularly *Bacillaceae bacillaceae* with zinc and *Bacillaceae bacillus* with calcium (Figures 6.1 and 6.3). *Clostridium sensu stricto* was strongly associated with iron at 12 months of age and, to a lesser extent, phosphorus, as were two species of the family *Peptostreptococcaceae (Intestinibacter* and *Peptoclostridium)*. Taxa assigned to the *Moraxellaceae* and *Pseudomonaceae* families from the Proteobacteria phylum were strongly associated with calcium and zinc at 12 months of age (Figure 6.3).

Vitamins at 12 months of age did not show strong and consistent patterns at 12 months of age compared to 9 months of age (Figures 6.1 and 6.3). All vitamins (except vitamin A) included in this analysis showed no or weak positive associations with sequences assigned to Energy metabolism, Carbohydrate metabolism (apart from Nucleotide and amino sugars), Biosynthesis of phenylpropanoid, Biosynthesis of arginine and Metabolism of cyanoamino acid (Figure 6.3). All vitamins showed no or weak positive associations with sequences assigned to proteins involved in Signalling and cellular repair, DNA replication, and Photosynthesis (Figure 6.3). All vitamins showed neutral or weak correlations with sequences assigned to Glycan biosynthesis and, to a lesser extent, Pyrimidine metabolism(except vitamins A and B12), all of which were important to the OPLS-R model (Figure 6.3).

A strong negative association between sequences assigned to Vitamin E and retinol metabolism, Metabolism by cytochrome p450, and a positive association with Atrazine degradation were also important to the OPLS-R model between KEGG pathways and vitamins (Figure 6.3). Niacin and vitamin B6 showed strong positive associations with sequences assigned to ABC transporters while vitamin B6 and vitamin B12 showed strong negative correlations with sequences assigned to Protein export (Figure 6.3). Niacin also showed strong negative correlations with sequences assigned to Tryptophan, and Taurine and hypotaurine

metabolism, and vitamin B6 showed a strong negative correlation with sequences assigned to Histidine metabolism (Figure 6.3). Vitamin E showed a strong negative correlation with sequences assigned to Fly apoptosis (Figure 6.3). All vitamins showed no positive correlations with Thiamine metabolism and Riboflavin metabolism, but riboflavin, vitamin B6, and vitamin B12 showed strong positive associations (Figure 6.3).

Vitamins showed weak associations with populations of Actinobacteria phylum. Sequences assigned to *Coriobacteriaceae eggerthella*, *Coriobacteriaceae* family, and *Actinomycetaceae actinomyces* were mildly important to the vitamin OPLS-R model, which showed weak negative associations with most vitamins (Figure 6.3). Folate was strongly associated with sequences assigned to *Erysipelotrichaceae erysipeloclostridium,* which was mildly important to the OPLS-R model (Figure 6.3). Sequences assigned to *Rikenellaceae allistipes* from the Bacteroidetes phylum and *Clostridia sensu stricto* and *Clostridiaceae sarcina* from the Firmicutes phylum, were significant to the vitamin OPLS-R model, each of which showed positive associations with thiamine, riboflavin, and niacin. Taxa from the family *Bacillaceae* showed strong positive correlations with vitamin B12 and vitamin E, but these microbes were not significant to the selected OPLS-R model (Figure 6.3). Sequences assigned to families *Peptostreptococcaceae*, *Peptostreptococcaceae*, *Intestinibacter*, and *Peptostreptococcaceae Peptoclostridium* showed weak positive associations with vitamins despite strong models (Figure 6.3). Sequences assigned to *Moraxellaceae acinetobacter* and *Pseudomonaceae pseudomonas* showed strong positive associations with vitamin B12, C, and E, but these were not important to the OPLS-R model (Figure 6.3). Weak negative associations between most vitamins and sequences assigned to *Streptococcus* and *Fusobacterium* genera, and *Veilonellaceae and Gammaproteobacteria* family, were important to the OPLS-R model (Figure 6.3).

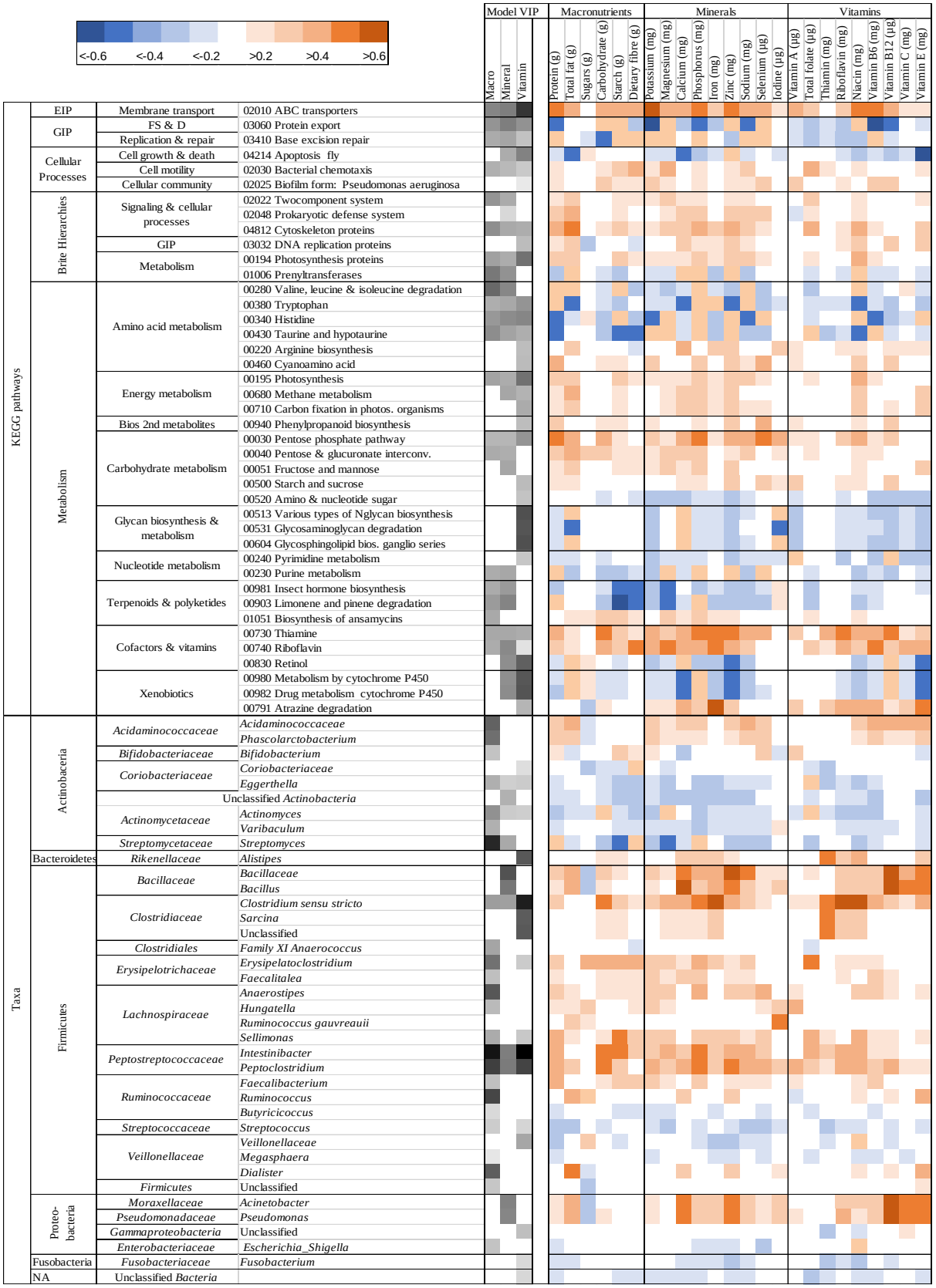


Figure 6. 3. Filtered heatmap indicating the strength and direction of linear correlation between variables import to the projection (VIPs) (KEGG pathways and taxa) of OPLS-Regressions and nutrients included in models at 12 months of age (Table 6.1). The value of the cumulative VIP score is depicted in the greyscale, with black indicating the highest VIP values. Orange indicates positive correlation, and blue indicates negative correlation.

6.3.5. Linear correlations between nutrients and sequences assigned to taxa and KEGG pathways at 12 months of age

At 12 months of age, the number of faecal lipids important to the OPLS-R models decreased, along with the strength of associations (Figures 6.2 and 6.4). The same ceramide (CerG1 36:1) and phosphatidylethanolamines (PE 38:1, 40:1, 36:0) that emerged as strong variables at 9 months of age (Figure 6.2), were also important to all OPLS-R models at 12 months of age (Figure 6.4). However, the pattern of association shifted from 9 to 12 months of age: dietary starch and carbohydrate content showed the strongest negative correlations with these metabolites at 12 months of age, along with calcium, iron, zinc, niacin, vitamin B12, and vitamin E (Figure 6.4).

A monogalactosyl-diacylglycerol (MGDG (16:0/18:2)) and lysophosphatidylcholine (LPC (18:2)) emerged as important to all nutrient OPLS-R models at 12 months of age. The protein showed weaker correlations at 12 months of age compared to 9 months of age, whereas dietary fat and starch increased in relevance (Figures 6.2 and 6.4). Weak positive associations also influenced the OPLS-R macronutrient model with PEs (27:0 and 34:2) and weak negative associations between sphingomyelin (d32:1) and triglyceride (8:0/10:0/12:0) (Figure 6.4). Sodium also showed stronger negative associations (with DG 37:7 and PE 40:1) at 12 months of age than 9 months of age (Figures 6.2 and 6.4). A strong negative association between vitamin A and LPE was important to the OPLS-R vitamin model, but not at 9 months of age (Figure 6.4). Positive and no associations between vitamins and LPC, MGDG, and MGs were important to the OPLS-R vitamin model; in particular, riboflavin with MG (17:2) and vitamin B6 with LPC (18:2) (Figures 6.2 and 6.4). Vitamin B6 and vitamin C, negatively associated with PEs at 9 months of age, showed weak positive associations with PEs at 12 months of age (Figures 6.2 and 6.4). A series of weak positive associations between vitamins and phosphatidylglycerol (PG (40:3)) and weak negative associations with PG (38:4) were also important to the OPLS-R vitamin model (Figure 6.4).

Figure 6. 4. Filtered heatmap indicating the strength and direction of linear correlation between variable important to the projection (VIPs) (Metabolites) of OPLS-Regressions and nutrients included in models at 12 months of age (Table 6.1). The value of the cumulative VIP score is depicted in the greyscale, with black indicating the highest VIP values. Orange indicates a positive correlation, and blue indicates a negative correlation.

Whereas the strength and extensiveness of association between faecal lipids and nutrients decreased at 12 months of age, the associations between faecal aqueous metabolites and nutrients increased in strength and number (Figures 6.3 and 6.4). Additionally, there was substantial overlap between the faecal aqueous metabolites important to multiple OPLS-R models at 9 months of age. There was far less overlap of faecal aqueous metabolites important to multiple models at 12 months of age (Figures 6.2 and 6.4). Only two aqueous metabolites were significant to all three nutrient OPLS-R models: N-acetylputrescine and choline, both positively associated with all nutrients (Figure 6.4). Only five additional aqueous metabolites were significant to the mineral model. The pattern of association that these metabolites showed with nutrients was predominantly weak, apart from a strong negative correlation between cytidine and phosphorus (Figure 6.4). The macronutrient OPLS-R model at 12 months was strongest with more aqueous metabolites included half of which showed a consistent positive association with all nutrients, and half of which showed predominantly negative associations with all nutrients (Figure 6.4).

Similar to the macronutrient model, the vitamin OPLS-R model included 17 additional aqueous metabolites, including purines, carboxylic acids, indoles and derivatives in faecal samples, which showed positive associations with vitamins, particularly riboflavin, vitamins B12, C, and E (Figure 6.4). Conversely, folate was negatively associated with all aqueous metabolites important to the OPLS-R model, and thiamine and niacin did not show any strong positive or negative associations (Figure 6.4).

6.4. Discussion

Estimated nutrient intakes were more strongly and extensively associated with sequences assigned to species, KEGG pathways, and lipid metabolites at 9 than 12 months of age. Faecal aqueous metabolites were associated more strongly and extensively with estimated nutrient intakes at 12 months of age than 9 months of age. Protein showed the strongest patterns of association with sequences assigned to microbes and KEGG pathways at 9 months of age, in line with what has been shown about the strong influence of proteins on infant GIT microbiome composition (Bokulich et al., 2016; Lang et al., 2018; Laursen, Bahl, Michaelsen, & Licht, 2017).

At 12 months of age, the strength of associations was more balanced across macronutrients, indicating that either the nutrients being investigated are no longer as relevant to the composition and function of the infant GIT compared to at 9 months of age, or that there are more confounding factors masking the correlations at 12 months of age. As the taxonomic composition of the GIT matures with age and species become established within trophic networks, the GIT may become more resilient to dietary perturbations, reducing the strength of associations compared to 9 months of age (Figures 6.1 ad 6.3) (Oliphant & Allen-Vercoe, 2019).

Several expected correlations between macronutrients and microbial species were not observed in this cohort. Specifically, members of the *Prevotellaceae* family were not strongly associated with starch and fibre consumption, which may reflect lower abundances of this family among western populations of both adults and infants, related to low starch consumption in Western diets (De Filippo et al., 2010). Furthermore, detection of *Prevotellaceae copri* in mothers was strongly associated with detection of *Prevotellaceae copri* in infants, which presents additional confounding factors, including adult diet, which is not accounted for in this study (Vuillermin et al., 2020).

Increased relative abundance of taxa from the *Erysipelotrichaceae* family was associated with high-fat diets and metabolic disease in adults, whereas in this cohort, a stronger association was observed between members from the *Erysipelotrichaceae* family and protein consumption (Lippert et al., 2017). Interestingly high protein consumption during infancy has been associated with childhood overweight and subsequent metabolic disease. Recent functional analysis of several strains of the *Erysipelotrichaceae* family found that they have the enzymatic capacity to degrade a broad range of substrates, including protein and plant polysaccharides (Kaakoush, N. O., 2015). This functional capacity aligns with the observations in this cohort and supports the theory that the *Erysipelotrichaceae* family may be an efficient scavenger of an excess substrate (which in infants is likely to be protein), and increases energy capture by the microbiome (Hugenholtz et al., 2018; J. Wu et al., 2021).

The representation of KEGG pathways within the metagenome indicates the potential capability, rather than activity in the GIT microbiome. Detection of metabolites provides a snapshot of the metabolic activity that has taken place. The relatively strong associations, both positive and negative, between faecal aqueous metabolites at 12 months of age compared to 9 months of age may be related to the increased diversity of substrates being transformed into secondary bacterial metabolites by microbes in the GIT at 12 months of age (Figures 6.2 and 6.4). This interpretation is supported by the increased number of faecal aqueous metabolites important to macronutrient and vitamin metabolite models at 12 months of age and the lack of overlap of important metabolites to these models (Figure 6.4).

Estimated mineral intakes were more important to the lipid and aqueous metabolite models at 9 months of age than 12. Of these, iron and zinc were important to all models, which aligns with the substantial body of evidence correlating them with the oxidative state of the infant's GIT (Figures 6.2 and 6.4) (Nancy F Krebs et al., 2013; Qasem et al., 2017; Tang et al., 2017). The patterns between sequences assigned to microbes and vitamins aligned with prior evidence

indicating that vitamins B2, C, and E might be influencing the GIT microbiome structure, particularly members from the Firmicutes phyla (Pham et al., 2021).

However, in this cohort, the role of vitamins B6 and B12 were also prominent. Whether these vitamins have an early role in facilitating the stabilisation of the microbiome remains unclear. Sodium, selenium, iodine, vitamin A, folate, and thiamine were not strongly associated with the faecal microbial or metabolic variables included at 9 months of age. These observations align with existing research, both *in vitro* and *in vivo*, on the associations of vitamins and minerals on the GIT microbiome (Steinert et al., 2020). These weak associations could also be attributed to the absorption of these nutrients prior to the large intestine.

One of the most notable differences between 9 and 12 months of age was the appearance of Glycan biosynthesis and metabolism at 12 months of age as an important class of sequences assigned to KEGG pathways of the faecal microbiome. In addition, a pattern of weak negative associations with vitamins and strong negative associations with dietary fat and iodine with Glycan biosynthesis pathways suggests that these pathways were most influenced by breastmilk and decreasing consumption at 12 months of age.

The few species, KEGG pathways, and metabolites in the faecal samples associated at both time points did not show consistent patterns of association, suggesting that the influence of nutrients shifts as species composition changes with age (Figures 6.1-6.4). Alternatively, the lack of consistent patterns of association suggests the possibility of spurious correlations characteristic of a decreasing ratio of signal to noise as the number of variables included in the datasets increases substantially beyond the number of participants included in the study. This limitation is anticipated in systems biology wherein the number of variables is vast, and the interactions and relationships between are often colinear and more complex than can be interpreted with correlations. This study addressed collinearity by employing OPLS-R models to identify the most important variables to the selected models. However, the biological

interpretation of these relationships is limited to the direction and strength of correlation (Palermo, Piraino, & Zucht, 2009).

Additionally, this study was limited by the small sample size and assumptions about the nutrient composition of foods and breastmilk. Estimations of quantities consumed during record keeping and the omission of nutrients that the food composition databases did not sufficiently capture (i.e., cholesterol and resistant starch).

6.5. Conclusions

This study was the first to integrate estimated nutrient composition from the whole diet during complementary feeding with the composition and functional capacity of the faecal microbiota and the faecal metabolome. The results from this study align with research that has investigated specific aspects of the influence of diet on the infant GIT microbiome development during the first year of life.

This study provided information, context, and hypotheses for testing in larger studies with controlled dietary interventions. Subsequent research into the associations between food groups and microbiome composition and function of the lower GIT may reveal links between specific foods as sources of these nutrients and the development of the infant microbiome.

Chapter 7. Complementary food networks in the infant faecal microbiome

The microbial consortium of the infant GIT is presented with rapidly changing substrate from partially digested or digestion resistant foods during the complementary feeding period. This study aims to identify associations between complementary foods (food groups) and composition and gene abundances of the microbiome, and metabolome in faeces of infants at 9 and 12 months of age. The networks associated with the consumption of specific food groups were different at 9 months of age than at 12 months of age, but meat influenced the networks at both time points. At 9 months of age, consumption of meat, vegetables, legumes, and high fibre grains was strongly and positively associated (R-value $\geq |0.6|$) with multiple aspects of the faecal microbiome. Stronger correlations, both positive and negative, between major food groups and faecal microbiome variables at 9 months of age suggest increased plasticity of the GIT microbiome through diet earlier during complementary feeding.

7.1. Introduction

Recent evidence suggests that the GIT microbiome may be more attuned to foods than nutrients (Ercolini & Fogliano, 2018; Johnson et al., 2019). Food group classification better captures biologically relevant food structures and bioactive compounds than standard nutrient composition data.

Proteins and carbohydrates drive broad trophic networks from the proximal to the distal GIT, and vitamins and minerals catalyse reactions as essential cofactors. The food matrix influences the bio-accessibility of these nutrients to host digestive enzymes and microbes in the GIT. In addition, other constituents in plant and animal food products are not included in food composition databases (e.g., plant metabolites and polyamines). These constituents also affect the abundance and function of some microbial species and influence the metabolic activity of microbes and host cells in the GIT.

The food matrix refers to the physical domain that carries the macronutrient and micronutrient constituents of food (J. M. Aguilera, 2019). The liquid matrices of breastmilk and formula are complex but maximise the interactions between nutrients and host digestive enzymes, bile acids, luminal surface, mucus, and pioneering microbes, which increases the bio-accessibility of the nutrients (C. R. Martin et al., 2016). Solid food matrices are complex and a strong determinant of bioavailability and bio-accessibility, which can be altered through processing technologies, storage, fermentation, cooking, and chewing, among other physical and chemical processes (J. M. Aguilera, 2019). The increasingly complex matrices of solid food during the complementary feeding period require and instigate changes in digestion and fermentation in the infant GIT to maximise energy and nutrient capture (Nicholson et al., 2012; Rowland et al., 2018). How the GIT microbiome changes in composition and function to the changing food matrices during complementary feeding may contribute to the physiological, immunological, cognitive, and behavioural development of the infant.

The food matrix is also a strong determinant of trophic networks that form longitudinally from the proximal to the distal colon and influence the biogeography of the microbial consortia and resources niches (T. Wang, Goyal, Dubinkina, & Maslov, 2019). Trophic interactions are formed by the breakdown of complex organic matter (e.g., plant wall hemicellulose) into small metabolite intermediates, which can be used as substrate for other species, forming a cooperative relationship (Gralka, Szabo, Stocker, & Cordero, 2020).

Foods that have been fermented prior to ingestion have an altered food matrix, including bacteria that can transiently complement the microbial communities in the GIT, particularly in the small intestine, whereas alterations in the colon are limited (Derrien & van Hylckama Vlieg, 2015). Plant secondary metabolites are also impactful dietary components but are difficult to capture in nutrient composition databases. Evidence suggests that polyphenols promote the proliferation of beneficial GIT bacteria through utilisation while also inhibiting potential pathogens (Rodríguez-Daza et al., 2021).

Relationships between microbial taxa and gene abundances and faecal metabolites produced in response to food consumption in the infant GIT are poorly understood, attributed to the high functional redundancy in the infant microbiome and the bidirectional relationship between the microbiome and metabolome. This observation has been best demonstrated by Nguyen et al., who found that relative taxonomic abundance from 16SrRNA gene sequences may correlate with metabolite concentrations detected with Nuclear Magnetic Resonance (NMR), but that species relative abundances are not predictive of metabolite concentrations (Nguyen et al., 2021).

The hypothesis behind the analyses reported in this Chapter was that food groups with a high protein content would strongly correlate with an increased relative abundance of taxa from the Bacteroidetes and Firmicutes phyla and a decreased relative abundance of taxa from the Actinobacteria phylum, particularly the *Bifidobacterium* genus. Additionally, based on

evidence in adults, it was hypothesised that food groups with digestion resistant matrices (high fibre grains, legumes, and meat) would be central to more complex networks of microbes, compositionally and functionally, and associated with energy capture and nutrient transformation. Finally, based on the evidence presented in Chapters 4, 5, and 6, it was also hypothesised that stronger correlations between foods and microbial factors would emerge at 9 months of age than at 12 months of age.

7. 2. Methods

The food group classifications at Level 1 and Level 2 and methodologies are described in Chapter 5.

7.2.1. Correlations and network analyses

Spearman's correlations were done in R version 4.0.5 (R Core Team, 2021). Correlations between each dataset separately for food group data and faecal samples collected at 9 and 12 months of age were conducted. Food groups were analysed as a percentage of energy consumed. Faecal samples were analysed for taxonomic relative abundances, KEGG pathway relative abundances, normalised lipid signal intensities, and aqueous metabolite signal intensities.

Correlation tables were imported to Cytoscape Version 3.8.2 (Shannon et al., 2003). Networks were constructed with the MetScape App and edges filtered to include correlations ≥ 0.6 and ≤ -0.6 and merged (Karnovsky et al., 2012). The merged network was then subset by food group so that each network included only one food group that was correlated with at least one taxon, gene abundance assignment, or metabolite. Only the first degree of related nodes and adjacent edges are shown below.

7.3 Results

7.3.1 Level 1 food groups at 9 months of age

At 9 months of age, the broad food groups (Level 1) that strongly correlated with multiple faecal factors were high fibre grains, vegetables, legumes, meats, breastmilk, and formula (Figure 7.1). Food groups that were strongly correlated with a single faecal factor were the low fibre grains, fats, sweets & sweetened beverages, powdered kumara, and herbs & spices groups (Figure 7.1). The fruits, eggs, dairy, imitation & substitute foods, and fungi & sea vegetables groups did not correlate strongly with any of the faecal factors assessed.

Consumption of high fibre grains was positively associated with the lipid phosphatidylethanolamine (16:0/18:2) and the aqueous metabolite methanethiol. The high fibre grain group was also negatively associated with sequences assigned to the family *Prevotellaceae,* the functional pathway for biofilm formation by *Vibrio cholerae,* and the aqueous metabolite 3-O-a-L-Fucopyranosyl D-glucose intensity (Figure 7.1A).

Vegetable consumption showed strong negative associations with aqueous metabolites (acetylcadaverine, phenylpyruvic acid, glutamic acid, and pterin), phosphatidylethanolamines (38:1, 40:1), and sequences assigned to bacterial toxins (Figure 7.1B). However, pterin, phosphatidylethanolamine 40:1, and the bacterial toxin functional group were positively associated with each other, and acetylcadaverine was positively associated with phosphatidylethanolamine (38:1) (Figure 7.1B).

Meat consumption was negatively associated with sequences assigned to Proteasome, Membrane trafficking, Fatty acid degradation, Peroxisome, Ferroptosis, Phosphatidylinositol signalling, and Lipid biosynthesis proteins (Figure 7.1C). It was strongly associated with the relative abundance of sequences assigned to RNA transport, Isoquinolone alkaloid biosynthesis, TAC cycle, and phosphatidylethanolamine (16:0) (Figure 7.1C).

Legume consumption was positively associated with some lipid metabolites including phosphatidylethanolamines (16:0/18:1 and 27:0), lysophosphatidylcholine (18:2), ceramide (d18:1/18:2), and diglyceride (18:3/18:2) (Figure 7.1D).

Formula consumption was at the centre of a network of sequences assigned to microbial taxa, KEGG pathways, and intensity of phosphatidylethanolamine (25:0) (Figure 7.1E). Formula consumption was negatively associated with the relative abundance of the *Bifidobacterium* genus and the KEGG pathways for Yeast autophagy and Lipid biosynthesis proteins. *Bifidobacterium* genus was also positively correlated with these pathways (Figure 7.1E). Formula consumption was positively associated with the relative abundance of sequences assigned to Pyruvate metabolism, Propanoate metabolism, Fructose and mannose metabolism, Pentose phosphate pathway, Butanoate metabolism, Glycolysis and gluconeogenesis (Figure 7.1E). Six of these pathways were also negatively associated with the relative abundance of taxa from the *Bifidobacterium* genus (Figure 7.1 E). The latter three pathways (Pentose phosphate, Butanoate metabolism, Glycolysis and gluconeogenesis) were also positively associated with sequences assigned to unclassified taxa from the Firmicutes phylum and HIF 1 signalling and Phosphotransferase system (Figure 7.1 E). Beta-alanine metabolism and phosphatidylethanolamine (25:0) were positively associated with formula consumption but not strongly associated with any other factor analysed here (Figure 7.1E). Breastmilk consumption was positively associated with histidine and glutaminylthreonine and negatively associated with sequences assigned to the HIF 1 signalling pathway (Figure 7.1 F).

Herbs and spices were positively associated with the relative abundance of the *Coriobacteriaceae* family, while sweets and sweetened beverages were positively associated with that of the *Erysipelotrichaceae* family (Figure 7.1 G). In addition, low fibre grains were positively associated with m-coumaric acid, the powdered kumara intervention product was

positively associated with the relative abundance of the *Bacillales* order, and fats were negatively associated with aqueous N-acetylgalactosamine fragment (Figure 7.1G).

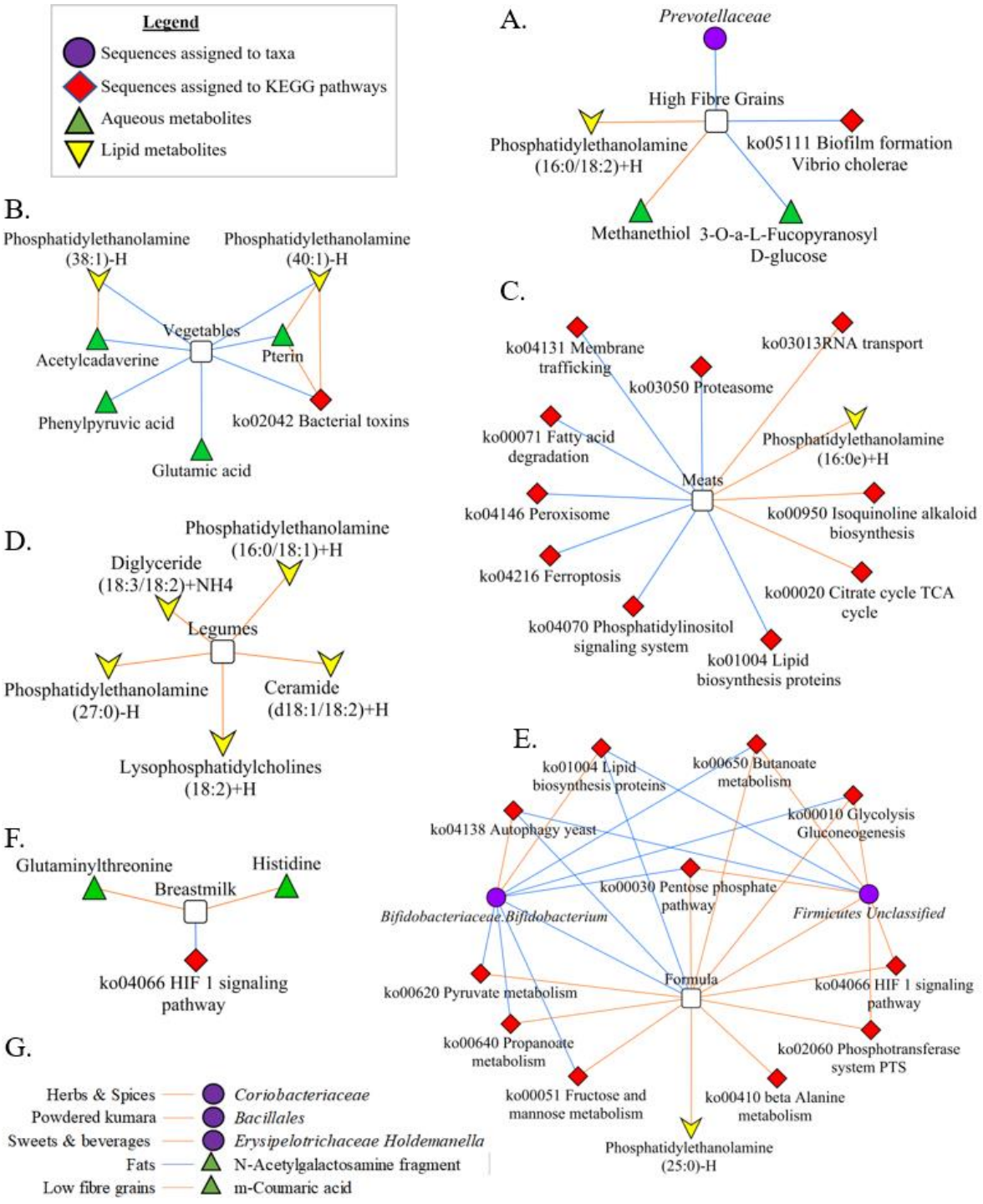


Figure 7. 1. Level 1 food group networks of infants at 9 months of age. ***A.*** *High fibre grains.* ***B.*** *Vegetables.* ***C.*** *Meats.* ***D.*** *Legumes.* ***E.*** *Formula.* ***F.*** *Breastmilk.* ***G.*** *List of foods that show strong correlations with one of the single factors measured in faecal samples. Foods groups are shown in the white square, sequences assigned to taxa are shown in the violet circles, KEGG pathway assignments are shown in red diamonds, faecal aqueous metabolites are shown in green triangles, and faecal lipid metabolites are shown in inverse yellow arrows. Orange lines indicate positive correlations ≥0.60, and blue lines indicate negative correlations ≤-0.60.*

7.3.2. Level 2 food groups at 9 months of age

Several patterns emerged when more specific food group classifications (Level 2) were explored. Beans, peas, lentils, and soy products were strongly correlated with ceramides (d18:1/18:2, d18:0/18:2, d18:1/16:0, d18:1/18:1, d36:1), phosphatidylethanolamines (16:0/18:1, 34:2, 27:0), diglycerides (18:3/18:2, 16:0/20:4), and sphingomyelin (d36:1) (Figure 7.2A). They were also negatively associated with the aqueous metabolite phenylethylamine (Figure 7.2A). Oats and porridge consumption was positively associated with phosphatidylethanolamines (16:0/18:2, 34:2) and lysophosphatidylcholine (18:1) (Figure 7.2B).

White bread consumption was negatively associated with an unidentified aqueous metabolite (M143T603) and sequences assigned to vitamin B6 metabolism (Figure 7.2C). White flour consumption was positively associated with the relative abundance of *Ruminococcus gauvreauii* and taxa of the *Barnesiella* and *Acinetobacter* genera (Figure 7.2D). Infant cereal consumption was positively associated with relative abundances of unclassified *Clostridia* class and *Roseburia* genus but negatively associated with taxa from the *Enterobacteriaceae* family (Figure 7.2E). The consumption of crackers and salty snacks was positively associated with two aqueous metabolites, hypoxanthine and acetylhistamine (Figure 7.2F).

White potatoes and starchy vegetable consumption was negatively associated with phosphatidylethanolamines (38:1, 40:1), ceramide G1(d36:1), and phenylpyruvic acid (Figure 7.2G). On the other hand, consumption of non-starchy alkaloid containing nightshades was negatively associated with the intensity of uracil and positively associated with sequences assigned to the *Akkermansia* genus and RNA degradation, with the latter two parameters positively associated with each other (Figure 7.2H).

Beef consumption was negatively associated with the relative abundance of taxa from the *Bifidobacterium* genus and sequences assigned to pathways Yeast autophagy and Lipid biosynthesis proteins; both pathways were positively associated with the abundance of the *Bifidobacterium* genus (Figure 7.2I).

Miscellaneous fruits (low appearance in food records and not part of higher-level groups) were positively associated with the relative abundance of taxa from the *Butyricoccus* genus and sequences assigned to chromosome and associated proteins (Figure 7.2J). In addition, avocado consumption was positively associated with lipid metabolites: monoglycerides (18:3, 20:2, 20:5), diglyceride (8:0/12:0), and triglyceride (6:0/12:0/18:0).

Figure 7.2L shows a list of correlations between food groups and single faecal factors. Consumption of egg mixtures was positively associated with *Sellimonas* genus. Nut & seed consumption and commercial fruit product consumption were positively associated with the relative abundance of sequences assigned to *Sarcina* genus. Dark green vegetable consumption was negatively associated with the aqueous metabolite pterin. Consumption of the powdered kumara intervention product was negatively associated with the aqueous metabolite aspartyl-glutamine. Dried fruit consumption was positively associated with *Pseudomonas* genus. Consumption of vegetables+meat+grain commercial infant food products was negatively associated with a fatty acid ((22:5)-H). Dairy and legume commercial infant foods, puddings and custards, and butter consumption were all positively associated with the relative abundance of *Holdemanella* genus. Pastries, cakes, and cookies were positively associated with *Lachnospira* genus. Consumption of nut and soy milk and herbs was associated with a relative abundance of *Bilophila* genus. Consumption of spices was positively associated with *Collinsella* genus. Consumption of animal fats other than butter was associated with the relative abundance of *Bacillaceae* family. Consumption of non-dairy yoghurts was positively associated with *Megamonas* genus.

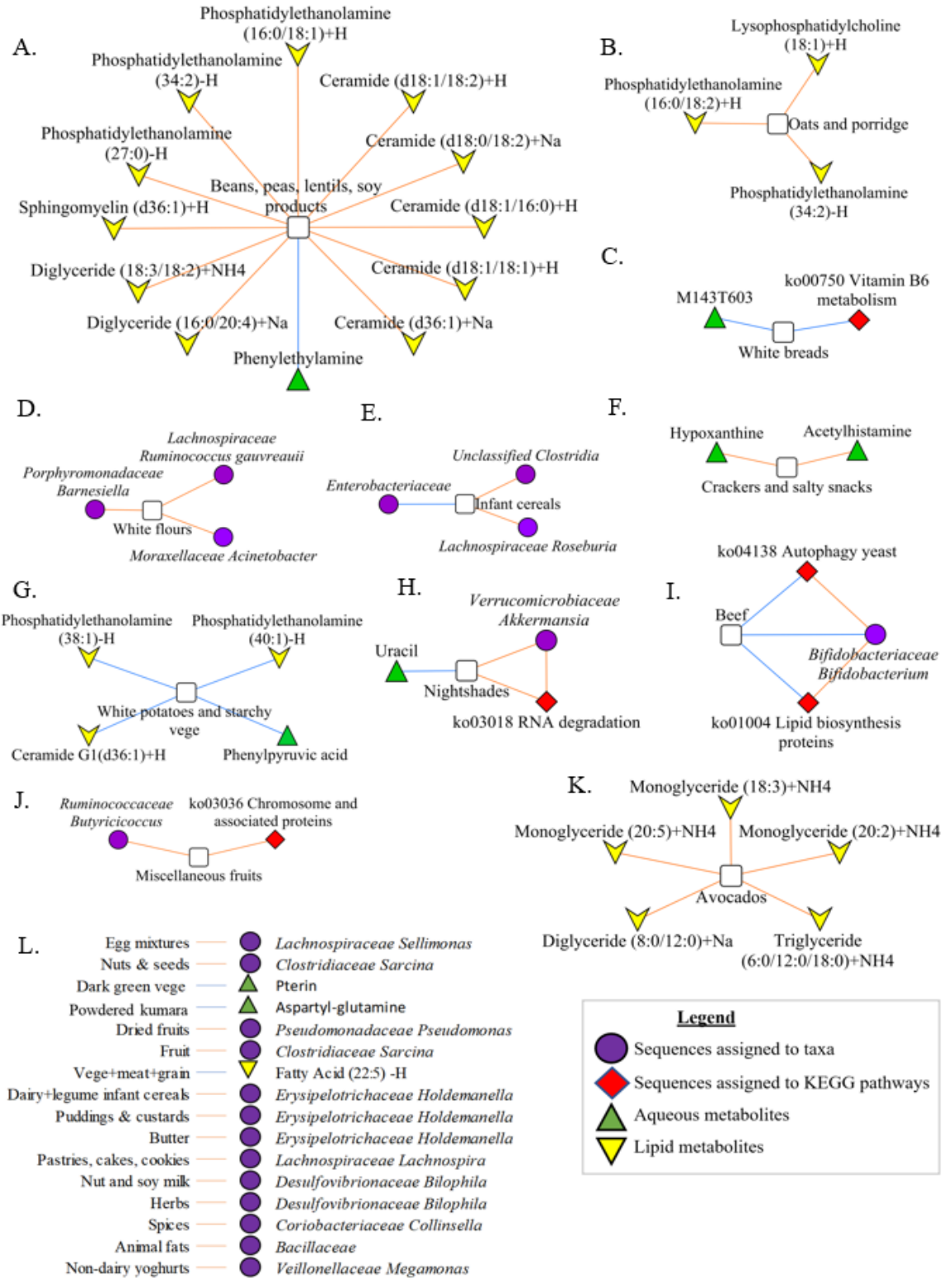


Figure 7. 2. Level 2 food group networks of infants at 9 months of age. ***A.*** *Beans, peas, lentils, and soy products.* ***B.*** *Oats and porridge.* ***C.*** *White bread.* ***D.*** *White flours.* ***E.*** *Infant cereals.* ***F.*** *Crackers and salty snacks.* ***G.*** *White potatoes and starchy vegetables.* ***H.*** *Nightshades.* ***I.*** *Beef.* ***J.*** *Miscellaneous fruits.* ***K.*** *Avocados.* ***L.*** *Foods that show strong correlations with single factors. Foods groups are shown in the white square, sequences assigned to taxa are shown in the violet circles, KEGG pathway assignments are shown in red diamonds, faecal aqueous metabolites are shown in green triangles, and faecal lipid metabolites are shown in inverse yellow arrows. Orange lines indicate positive correlations ≥0.60, and blue lines indicate negative correlations ≤-0.60. Breastmilk and formula were omitted.*

7.3.3. Level 1 food groups at 12 months of age

At 12 months of age, the major food groups (Level 1) strongly correlated with multiple faecal factors were low fibre grains, vegetables, legumes, meats, herbs and spices, breastmilk, and formula (Figure 7.3A-G). Food groups that showed strong correlations with a single faecal factor included eggs, dairy, commercial infant foods, and sweets and beverages (Figure 7.3H). Conversely, high fibre grains, fruits, fats, imitation and substitute foods, fungi and sea vegetables did not correlate strongly with the faecal factors assessed (data not shown).

Low fibre grain consumption was positively associated with glycine (Figure 7.3A). It was negatively associated with unidentified aqueous metabolite intensities and sequences assigned to D glutamine and D glutamate metabolism (Figure 7.3A).

Vegetable consumption was negatively associated with aqueous metabolite 7-methylguanine and ceramides (d18:1/18:1 and d16:1/16:0)(Figure 7.3B). It was positively associated with phosphatidylethanolamine (27:0) (Figure 7.3B). Legume consumption was positively associated with aqueous metabolite trigonelline and relative abundance of *Lachnospira* genus, which were also positively associated with each other (Figure 7.3D).

Consumption of herbs and spices was positively associated with an unclassified bacterium of the *Enterobacteriaceae* family (Figure 7.3E) and was negatively associated with microbial sequences assigned to inositol phosphate metabolism (Figure 7.3E).

Formula consumption was positively associated with relative abundances of taxa from the *Erysipelotrichaceae* family (Figure 7.3F). It was negatively associated with aqueous metabolites (glutaminylthreonine, deoxyuridine, and asparaginyl-methionine) and relative abundances of microbial sequences assigned to the Metabolism of xenobiotics and Metabolism of drugs by cytochrome P450, all of which were positively associated with one another (Figure 7.3F). Breastmilk consumption was positively associated with the intensities of

phosphatidylethanolamine (38:1 and 40:1) and aqueous metabolites (2-phenylbutyric acid, cytidine, deoxyuridine, glutaminylthreonine, and asparaginyl-methionine).

Meat consumption was negatively associated with aqueous metabolites S-methyl benzenecarbothioate and Lysophosphatidylethanolamine (20:0) (Figure 7.3C). Egg consumption was positively associated with the relative abundance of taxa from the *Christensenellaceae* family R7 group (Figure 7.3G). Dairy was positively associated with monoglyceride (20:1), while commercial infant foods were positively associated with monogalactosyldiacylglycerol (16:0/20:1) (Figure 7.3G). Sweets and sweetened beverages were negatively associated with phenylpyruvic acid.

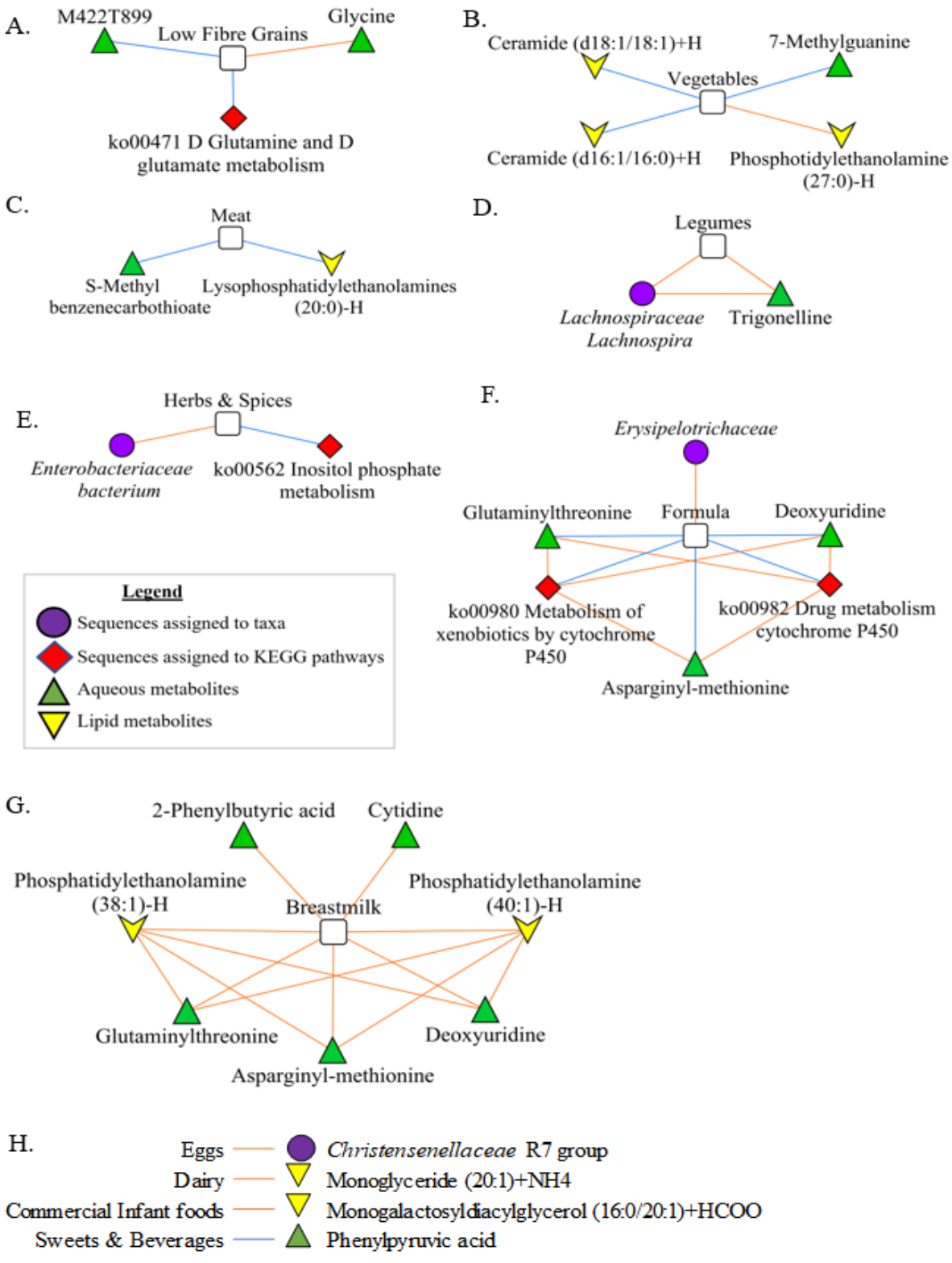


Figure 7. 3. Level 1 food group networks of infants at 12 months of age. ***A.*** *Low fibre grains.* ***B.*** *Vegetables.* ***C.*** *Meats.* ***D.*** *Legumes.* ***E.*** *Herbs & Spices.* ***F.*** *Formula.* ***G.*** *Breastmilk.* ***H.*** *List of foods that show strong correlations with single factors. Foods groups are shown in the white square, sequences assigned to taxa are shown in the violet circles, KEGG pathway assignments are shown in red diamonds, faecal aqueous metabolites are shown in green triangles, faecal lipid metabolites are shown in inverse yellow arrows. Orange lines indicate positive correlations ≥0.60, and blue lines indicate negative correlations ≤-0.60.*

7.3.4. Level 2 food groups at 12 months of age

When more specific food group classifications (Level 2) were explored, fewer networks emerged at 12 months of age than at 9 months of age (Figure 7). Commercial infant food products (vegetables+meat+grain) were positively associated with monogalactosyldiacylglycerol (16:0/19:1) and microbial sequences assigned to ferroptosis and glycosylphosphatidylinositol GPI anchored proteins (Figure 7.4A). Citrus consumption was positively associated with aqueous metabolite pentanoic acid and negatively associated with aqueous metabolite hydroxyadipic acid (Figure 7.4 B).

Consumption of whole-grain pasta, bread, and brown rice was negatively associated with the relative abundance of taxa from the *Coriobacteriaceae* family and *Eggerthella* genus (Figure 7.4E). It was positively associated with histidylvaline (Figure 7.4E).

Nut and soy milk consumption was negatively associated with aqueous metabolites asymmetric dimethyl arginine and serine (Figure 7.4F). Consumption of yoghurts and fermented dairy products was positively associated with monoglyceride (18:3) and riboflavin (Figure 7.4D). Consumption of cheese was positively associated with 7-methylguanine.

Consumption of poultry was negatively associated with aqueous metabolite S-methyl benzenecarbothioate and relative abundance of the taxa from the *Veillonellaceae* family (Figure 7.4C). Of the meat categories, fish consumption was negatively associated with serotonin, pork consumption was negatively associated with microbial sequences assigned to histidine metabolism, and lamb/venison/game was positively associated with microbial sequences assigned to the *Sarcina* genus.

Similarly, consumption of salty spreads and commercial infant foods that contained fruit and grain were also positively associated with the relative abundance of the *Sarcina* genus. Consumption of nuts and pulses was positively associated with sequences assigned with Prokaryotic defence system KEGG pathways. Consumption of beans, peas, lentils, and soy

products was positively associated with the relative abundance of taxa of the *Lachnospira* genus. White potatoes and starchy vegetable consumption was negatively associated with L-lysine. Green beans and peas consumption was positively associated with N-acetylputrescine. Unclassified or miscellaneous vegetable consumption was negatively associated with ceramide G1(d36:1).

Commercial infant foods containing vegetables and meat were positively associated with phosphatidylglycerol (40:3). Consumption of puddings and custards was negatively associated with phenylpyruvic acid. Consumption of dried fruits was negatively associated with the relative abundance of the *Streptococcus* genus. Flavourings and fungi derived substitutes were positively associated with the relative abundance of the *Atopobium* genus. Consumption of plant oils was negatively associated with phosphatidylglycerol (38:4).

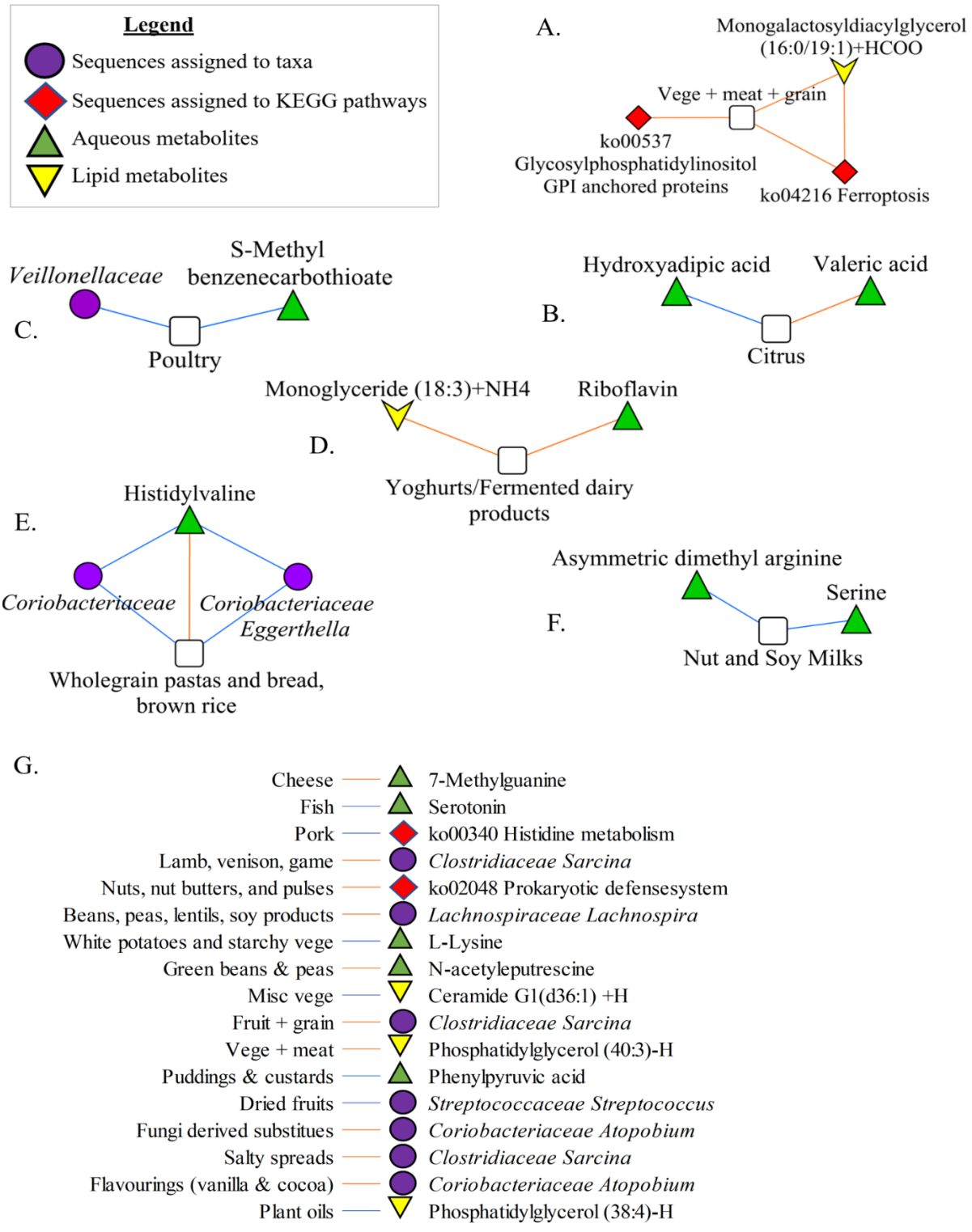


Figure 7. 4. Level 2 food group networks of infants at 12 months of age. ***A.*** *Commercial infant foods: vegetable+meat+grain.* ***B.*** *Citrus.* ***C.*** *Poultry.* ***D.*** *Fermented dairy products.* ***E.*** *Wholegrain pasta, bread, brown rice.* ***F.*** *Nut & soy milk.* ***G.*** *List of foods that show strong correlations with single factors. Foods groups are shown in the white square, sequences assigned to taxa are shown in the violet circles, KEGG pathway assignments are shown in red diamonds, faecal aqueous metabolites are shown in green triangles, and faecal lipid metabolites are shown in inverse yellow arrows. Orange lines indicate positive correlations ≥0.60, and blue lines indicate negative correlations ≤-0.60. Breastmilk & formula were omitted.*

7.4. Discussion

The analysis of the food and microbiome networks of infants aged 9 and 12 months reported in Chapter 7 identified individual and interconnected relationships between food groups at two levels of specificity and relative abundances of species assigned to taxa, KEGG pathways, and lipid and aqueous metabolites. Based on correlations between nutrients and faecal factors described in Chapter 6, there were more strong correlations between foods and faecal factors at 9 months of age than 12 months of age. In line with the stated hypothesis, shifting patterns of nutrient utilisation and energy capture were evident between 9 and 12 months of age and varied by food group.

Protein-rich foods had the greatest number of strong correlations at 9 months of age, but the patterns of correlations were distinct: whereas meat and formula were most strongly associated with gene abundance pathways, legumes were most strongly associated with lipid metabolite intensities. Interestingly, formula consumption was predominantly positively associated with a series of carbohydrate metabolism KEGG pathways, whereas meat was predominantly negatively associated with KEGG pathways from the cellular processes and genetic information processing super-pathways at 9 months of age, all of which emerged as VIPs in Chapter 6 (Table 6.1). Formula and meat consumption shared negative associations with lipid biosynthesis proteins. These findings show that the source of protein differently affects the function of the lower GIT microbiome and identifies the source of the strongest patterns identified between KEGG pathways and nutrients in Chapter 6.

An unexpected finding was the association of the intensity of faecal lipid metabolites with high fibre plant products that are also relatively high in protein such as legumes, oats and porridge, and beans, peas, lentils, and soy. The ceramides in the faecal metabolome that correlated with the consumption of beans, peas, and soy products likely came from the food product itself. These foods are primary sources of ceramides, but due to their incorporation in

complex plant cell matrices may have escaped digestion and absorption at 9 months of age. However, it is unclear how consuming these foods impacts ceramides circulating in the bloodstream (not measured here) or accumulation in organs such as the liver, observed in metabolic disorders, dyslipidaemia, and insulin resistance in adults. Similarly, oats are a rich dietary source of N-acyl-phosphatidylethanolamines, offering the most likely explanation for their emergence in the 9 months of age networks. That these patterns were not apparent at 12 months of age suggests the plant matrices in which these lipids are embedded are degraded more efficiently throughout the GIT of the 12 month old infant than the 9 month old infant, allowing more of these lipids to be utilised and incorporated into cell membranes by either the host or microbiota (Levental et al., 2020).

Avocado consumption was associated with the intensity of mono-, di-, and triglycerides in the faeces at 9 months of age (Figure 7.2 K), which is not surprising considering that these are the primary lipids present in avocado. This pattern was not apparent at 12 months of age, despite consistent avocado consumption at both time points, indicating more efficient lipid digestion in older infants or a lower proportion of avocado in the diet.

In agreement with previous correlations found between fats and sugars in Chapter 6, the relative abundance of taxa from the *Erysipelotrichaceae* family was associated with high energy foods with high fat and sugar content. However, the lack of relationship with the faecal lipids in these food networks suggests that the lipids from these sources were utilised and absorbed more efficiently than lipids from whole grains and legumes.

As hypothesised, taxa from the genus *Bifidobacterium* emerged as a key species in select food group networks. However, the strong primary relationship of this genus was not positive with starchy carbohydrates or breastmilk but rather negative with formula consumption and beef. Most strong associations with taxa appeared in the Level 2 food networks rather than Level 1 food networks, indicative of the enzymatic specificity that microbial species have for

specific substrates. Additionally, most associations between taxa and food groups were positive rather than negative. This is indicative of a statistical bias inherent in total sum scaling of microbial datasets that skews towards capturing populations that relatively increase in abundance rather than the dispersed relative decreases across the remaining microbial community.

Interestingly, high fibre grains at 9 months of age and whole grains at 12 months of age were inversely related to species belonging to families that are considered to have strong carbohydrate degrading enzyme capacity: *Prevotellaceae* and *Coriobacteriaceae* families, respectively. This finding may reflect the specificity of carbohydrate structures or by-products preferred by these bacteria. For instance, taxa from the *Prevotellaceae* family are strong utilisers of starch and found in high abundance in populations that consume starchy diets high in maize (Prasoodanan P. K et al., 2021). In vitro fermentation of a powdered spinach product has also been found enhance Prevotellaceae populations (Parkar, S. G et al., 2021) However, the relative abundance of taxa from the *Prevotellaceae* family was not associated with starchy food groups in this cohort either, which may reflect low abundances of this family of bacteria in populations consuming western diets (Yatsunenko et al., 2012).

Surprisingly, vegetables were inversely correlated with pterin (Figure 7.1 B), specifically dark leafy green vegetables and pterin (Figure 7.2 L). This class of vegetables are a good dietary source of folate, which is a pteridine vitamin (Liang, 2020). Therefore, a possible explanation is that dietary folate downregulated *de novo* folate synthesis by microbes and was well absorbed in the small intestine and colon (Said & Mohammed, 2006).

Similarly, foods rich in bioactive plant secondary metabolites, such as fruits, did not emerge as strongly related to any factors assessed here. Apart from leafy green vegetables, the exceptions were alkaloid-containing nightshades, citrus, and miscellaneous fruit. However, the positive association between non-starchy nightshades, *Akkermansia* genus, and a microbial

RNA degradation pathway was notable. A possible explanation is that the proanthocyanidins commonly found in nightshades promote the growth of taxa from the *Akkermansia* genus (Jayachandran, Chung, & Xu, 2020).

While food group classification captures more microbiome relevant aspects of diet than nutrient composition alone, the effects of processing, cooking, and preparation on the food matrix could not be fully captured through the food records collected here. It can be presumed that the texture of foods fed infants at 9 months of age was smoother than foods at 12 months of age, which may be an additional reason more strong correlations emerged at 9 months than 12 months. Only three networks emerged of the commercial infant foods with a predictable texture. Of those, only one was notable: the vegetables and meat and grain combination were positively associated with the microbial pathways, ferroptosis, and iron-dependent regulated cell death. Both microbial pathways were also positively associated with monogalactosyldiacylglycerol (16:0/19:0), which may accumulate in stressed cells (A. Aguilera et al., 2021). These products may have had supplemental iron in them.

A surprising result was emergence of strong relationships between Level 1 herbs & spices and *Coriobacteriaceae* family at 9 months of age (positive) with *Enterobacteriaceae bacterium* (positive) and Inositol metabolism (negative) at 12 months of age. Similarly, at Level 2 at 12 months of age, fungi, substitutes and flavourings were positively associated with *Coriobacteriaceae Atopobium*. These relationships were particularly unexpected considering the small contribution of these products to the daily diet. However, *Coriobacteriaceae* family are known to activate dietary polyphenols and may be particularly responsive to high potency ingredients during complementary feeding (Clavel, Lepage, & Charrier, 2014).

Fermented or cultured foods did not feature as prominent microbiome modulators, questioning the role of probiotic species obtained through foods on the microbiome, let alone a shift in functional capacity. Yoghurts and fermented dairy products did emerge as positively

associated with monoglyceride and riboflavin, which may reflect monoglyceride based emulsifiers and relatively high riboflavin content in dairy products. The role of lacto-fermented vegetables, such as sauerkraut and kimchi, which have recently been found to have a profound influence on both the structure and function of the adult microbiome, cannot be commented on here, as they were not prominent parts of the diets consumed in this cohort (Taylor et al., 2020; Wastyk et al., 2021).

The statistical relationships between foods and faecal factors presented here offer a starting point for more in-depth studies on complementary feeding and the lower GIT microbiome. However, these correlations are observational and insufficiently powered, and confounding factors and spurious correlations cannot be ruled out. Additionally, relevant associations may not have emerged through this analysis due to the similarity of diets and frequent consumption of common foods in this cohort. Without sufficient variation in diet or faecal factors, correlations cannot emerge.

7.5. Conclusions

The associations presented in Chapter 7 align with previous findings that the faecal microbiome composition and functional capacity are more strongly related to dietary factors at 9 months of age than 12 months of age. However, several distinct patterns emerged that have not previously been identified or explored in relation to complementary food selections. Microbial taxa associate more strongly with more specific food groups as opposed to broader food groups, indicative of substrate specificity. Meat and formula consumption associated with alterations in KEGG pathways at 9 months of age, but not 12 months of age. Additionally, meat was predominantly negatively correlated with KEGG pathways for cellular processes and genetic information processing, whereas formula was predominately positively correlated with KEGG pathways for carbohydrate metabolism. Detection of faecal lipids that associate with plants containing those lipids at 9 months of age, but not 12 months of age, suggest substantial development either in the digestion of lipids or the microbial consortia that allows for plant derived lipids to be accessed prior to excretion between 9 and 12 months of age. Most surprisingly, select associations emerged that likely were not indicative of actual biological relationships between food groups and taxa, but rather the detection of those taxa may be indicative of behaviours such as home cooked and complex meals.

Chapter 8. General discussion

The introduction of solid foods during complementary infant feeding elicits changes in taxonomic composition, relative abundance of genes, metabolites produced, and associated functions of the GIT microbiome. This may affect the maturation of the immune system, cognition, behaviour, and metabolic health of the infant. The GIT microbiome has been extensively researched in the context of breastfeeding and formula feeding, but not as thoroughly in the context of the complementary solid food diet. The infant diet is complex, personalised, and changing rapidly to meet increasing demands for nutrients between 6 and 12 months of age. The research in this dissertation was the first to elucidate broad patterns of association between multiple compositional and functional aspects of the GIT microbiome in the context of dietary nutrients and foods consumed at multiple timepoints during complementary feeding.

Conducting clinical research in infant populations comes with unique challenges, outlined in Chapter 3. Ideally, systems biology triangulates multiple sample types to understand the dynamics of what is entering the system and how it is utilised, absorbed, transformed, and excreted. In this study, the most consistent and informative sample type was faecal samples. Combined with high resolution diet data provided by caregivers and interpreted by dieticians, faecal samples are an acceptable proxy for the processes occurring in more proximal parts of the GIT. Faecal metabolomics holds potential as a non-invasive functional read out of the metabolic activity within the GIT, but challenges remain in identifying and attributing metabolites, aligning them to serum concentrations of relevant biomarkers, and clinical outcomes. For example, many metabolites ingested in foods may be detected in faecal samples relatively intact, while others are transformed and either absorbed or utilised by the host epithelium, or other microbial species residing in the GIT. Additionally, the human host produces metabolites that are excreted frequently through stools that can indicate normal

metabolism, inflammation, epithelial damage, epithelial development, or metabolic abnormalities. The non-targeted faecal metabolomics conducted in this study provide additional faecal metabolites detected in an infant population, along with dietary context from which to understand potential sources or cofactors involved in the production of metabolites.

The research conducted in this PhD dissertation was the first to incorporate multiple aspects of individual infant diets with multiple aspects of the faecal microbiome and metabolome during complementary feeding. The faecal microbiome structure and function as a proxy of the lower GIT were more strongly associated with dietary nutrients and unique food groups of infants at 9 months of age than 12 months of age. However, from these differences, inferences about the developments in nutrient and energy capture and utilisation from various food groups consumed during complementary feeding can be proposed. Additionally, this study and the data generated from it formed the basis for recommendations on study design principles, sampling considerations, and factors to consider during recruitment, enrolment, and randomisation specific to complementary feeding studies.

8.1. Summary of the findings

In line with the hypotheses, the changes occurring in infant faecal microbiome composition, functional potential (KEGG pathways), and metabolome were more numerous and significant earlier in the complementary feeding period (while solid foods were first introduced), than the later complementary feeding period, as solid foods diversified (Chapter 4). The relative abundance of genes and gene functions within the faecal microbiome was more stable than the taxonomic composition. This finding is consistent with previous research showing high functional redundancy across GIT microbiota, suggesting that most of the genes required to make use of changing substrates from a diversifying diet and constitutive genes required for living are present at 4 months of age. Additionally, changes in aqueous metabolites were more strongly associated with changes in KEGG pathways, whereas changes in lipid metabolites were more strongly associated with changes in relative abundance of taxa. This result may be indicative of microbial sources of lipids and supports the potential application of aqueous metabolites as a functional read-out. However, this pattern did not emerge in relation to consumption of food groups (Chapter 7), nor did the individual sequence assignments and metabolites that were associated emerge as significantly associated with mode of milk feeding or dietary nutrients. This suggests that the sequence assignments and associated metabolites identified in Chapter 4 may be more strongly associated with other aspects of development, such as epithelial integrity or immune development, than with the changing diet.

While this study was primarily interested in complementary foods, understanding the impact of the source of milk that comprises a substantial part of the diet, even late in complementary feeding, was crucial. The relative impact of breastmilk and formula on the GIT microbiome could not be distinguished from other confounding factors, such as increased dietary diversity and increased vitamin and mineral consumption, in formula-fed infants compared to breastfed infants from the analyses carried out in Chapter 5. However, by

investigating overlapping patterns with nutrients (Chapter 6) and with food groups (Chapter 7), it became clear that the KEGG pathways that seemed positively associated with formula consumption in Chapter 5 were more nuanced in their associations with protein rich animal products. Of the sequences found to be significantly higher in the formula fed group in Chapter 5 and positively associated with protein consumption in Chapter 6, sequences assigned to Cellular processes, Genetic information processing, and Environmental signalling were actually more strongly associated with meat consumption in Chapter 7, whereas sequences assigned to Carbohydrate metabolism pathways that were found in higher relative abundance in the formula fed group in Chapters 5, were then confirmed to be attributable to formula at 9 months of age in Chapter 7.

This cohort was primarily breastfed, and the microbiome had established with oligosaccharides as a primary substrate; the addition of nitrogen-rich protein to the GIT ecosystem created and expanded ecological niches. However, the lack of strong or predictable modulation by carbohydrates, including the lack of promotion of species from the *Bifidobacterium* genus by starch (Figure 6.1), was not in line with the stated hypotheses. The predominant pattern that emerged in relation to sequences assigned to microbial species was that more associations, and predominantly positive associations, emerged with Level 2 food groups, indicative of enzymatic specificity and preferences for specific substrate structures.

Further analysis into food networks confirmed the importance of accounting for the food matrix and compounds that are not included in standard food composition databases when conducting systems biology research during complementary feeding. This was made clear through the novel finding that lipids from plant foods, such as legumes and avocados, are excreted in higher concentrations at 9 months of age than at 12 months of age. This is presumably due to the less efficient degradation of plant cell walls at 9 months of age than at

12 months of age by the microbiota in the proximal colon, limiting accessibility, absorption, and utilisation of lipids contained within the plant matrix.

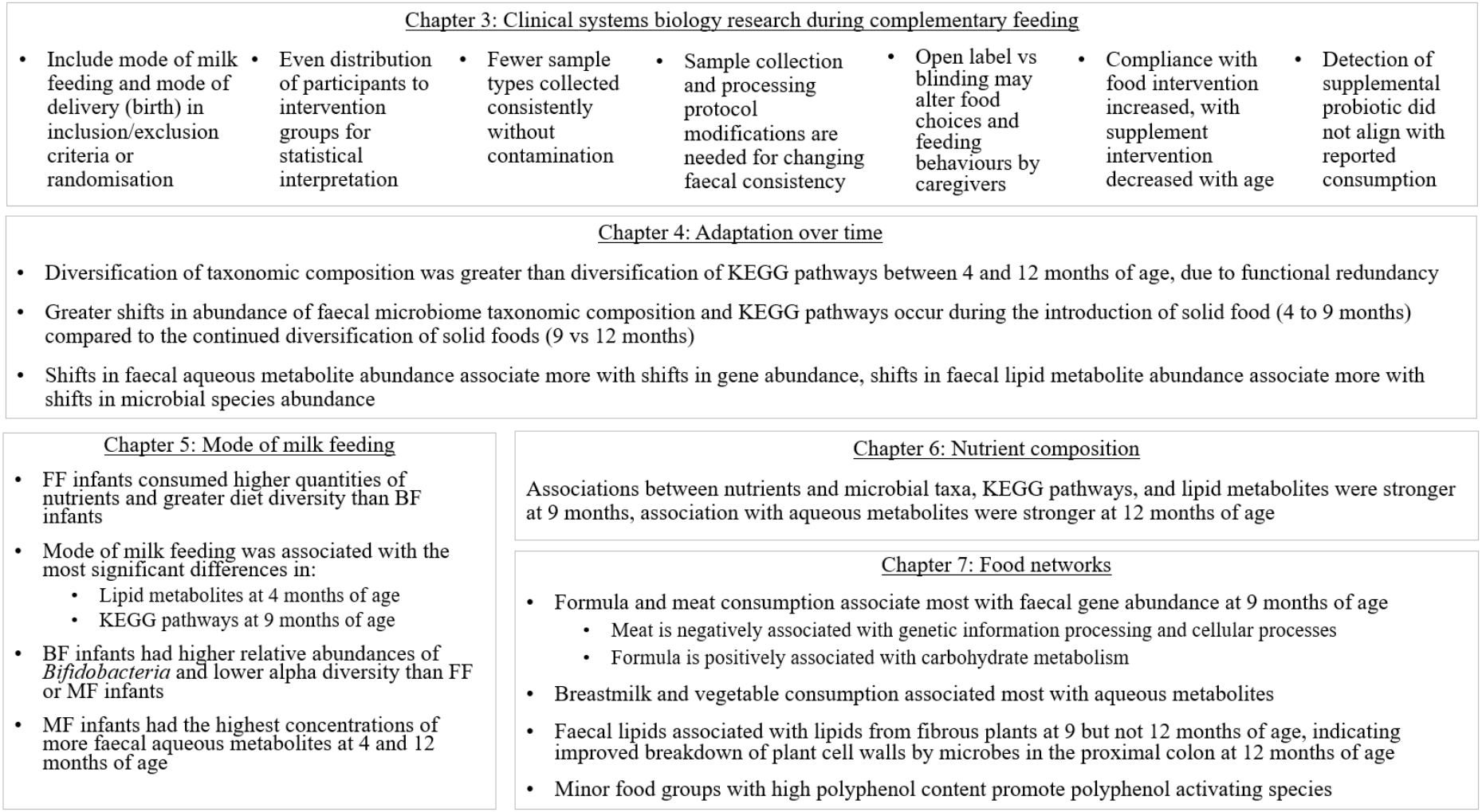


Figure 8. 1. Summary figure of findings. BF = breast fed, FF = formula fed, MF = mixed fed.

8.2. Limitations of the study

The small study size and the homogenous urban population enrolled in this study limit the generalisability of the results. Furthermore, the reliability of this data is impacted by the limitations of diet records, food composition databases, modifications in sample collection at different ages, and the batching of sample analyses. Additionally, confounding factors that may have exerted a significant but unidentified influence on the results may not have been collected or accounted for in this analysis. Finally, due to the lack of immunological data from blood and breastmilk that could not be collected or analysed, the results cannot confirm the immunological relevance of nutrients and food groups.

The study design constrained the statistical methodological choices; uneven intervention groups, dispersed baseline sample collection, the inclusion of both milk-fed and formula-fed infants at baseline, and omission of the stepwise introduction of solid foods and sample collection limit the extent of the findings of this study. Additionally, analysing multi-variate, non-parametric, high dimensional data presents challenges in determining statistical significance.

Participants and their parents were not followed up beyond the 12-month time point to collect information on clinically relevant immunological outcomes such as the onset of asthma, allergies, or atopy. Therefore, the findings of this PhD dissertation do not extend the influence of complementary foods on the long-term health of individuals.

8.3. Implications

Despite these limitations, this study informed recommendations to improve study designs in this unique age group and fill gaps in understanding the relative impact of introducing specific food groups at key ages in infancy. Furthermore, the high-resolution diet data (3-day food records) provided an opportunity to incorporate diet data with faecal

microbiome and metabolome data in a temporal analysis in an urban population not previously explored.

This study highlights the need to conduct complementary diet intervention studies over a period that includes the 9 month of age inflection point, during which the GIT microbiome seems highly susceptible to modulation by dietary factors as complementary feeding patterns are shifting from solid foods complementing the milk-based diet to milk complementing the solid food diet. Additionally, this study reinforces recent evidence collected in adults that the lower GIT microbiome is more attuned to food groups than nutrient composition. This is likely due to better capture of the food matrix and relevant compounds not currently included in nutrient composition databases.

The patterns of microbiome composition dynamics and relatively stable functional capacity partially support prior conclusions drawn by Bäckhed et al., suggesting that the cessation of breastfeeding is the most significant event in the maturation of the GIT microbiome (Bäckhed et al., 2015). This study did not specifically cover cessation of breastfeeding in every infant and thus cannot compare.

8.4. Future perspectives

It is recommended that future studies leverage and examine the inflection point at 9 months of age, considering the patterns identified at that time. Taking note of how and when the shift from solid foods complementing milk to milk complementing solid foods occurs is warranted. Also, feeding behaviours that separate or combine milk and solid foods at mealtimes may change how foods move through the GIT and how the microbiota utilise the substrates presented, and therefore should be accounted for in complementary feeding studies.

Further research is needed to establish the relationships between the metabolites excreted in stools, those present in the lower GIT content, and those absorbed into the bloodstream, along with markers of immune development in the context of the complementary feeding diet.

This study demonstrated the relevance of specific nutrients and food groups to the data collected from faecal samples and how these patterns shift over time. However, a better understanding of the transformation of nutrients in the GIT, absorption from the lumen into the bloodstream, and utilisation by the growing infant are needed to understand how the patterns identified here positively or negatively influence health outcomes.

Future studies in larger cohorts should also leverage the opportunity to understand how individual foods impact the lower GIT microbiome and metabolome. These studies could include the step-wise introduction of food products, collecting more frequent biological samples, and diet data to create a higher resolution temporal map of microbiome development. Additionally, following up with dietary intervention participants annually for several years to collect data about any onset of autoimmune and atopic diseases is needed to understand the clinical relevance of interventions that modulate the lower GIT microbiome for immune development during complementary feeding.

References

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The Placenta Harbors a Unique Microbiome. *Science translational medicine, 6*(237), 237ra265-237ra265.

Abrahamse-Berkeveld, M., Alles, M., Franke-Beckmann, E., Helm, K., Knecht, R., Köllges, R., . . . Bufe, A. (2016). Infant formula containing galacto-and fructo-oligosaccharides and Bifidobacterium breve M-16V supports adequate growth and tolerance in healthy infants in a randomised, controlled, double-blind, prospective, multicentre study. *Journal of nutritional science, 5*.

Aguilera, A., Berdun, F., Bartoli, C., Steelheart, C., Alegre, M., Bayir, H., . . . Pagnussat, G. (2021). C-ferroptosis is an iron-dependent form of regulated cell death in cyanobacteria. *Journal of Cell Biology, 221*(2), e201911005.

Aguilera, J. M. (2019). The food matrix: implications in processing, nutrition and health. *Crit Rev Food Sci Nutr, 59*(22), 3612-3629.

Ajslev, T., Andersen, C., Gamborg, M., Sørensen, T., & Jess, T. (2011). Childhood overweight after establishment of the gut microbiota: the role of delivery mode, pre-pregnancy weight and early administration of antibiotics. *International journal of obesity, 35*(4), 522-529.

Akbari, P., Braber, S., Alizadeh, A., Verheijden, K. A., Schoterman, M. H., Kraneveld, A. D., . . . Fink-Gremmels, J. (2015). Galacto-oligosaccharides protect the intestinal barrier by maintaining the tight junction network and modulating the inflammatory responses after a challenge with the mycotoxin deoxynivalenol in human Caco-2 cell monolayers and B6C3F1 mice. *The Journal of Nutrition, 145*(7), 1604-1613.

Akbari, P., Fink-Gremmels, J., Willems, R. H., Difilippo, E., Schols, H. A., Schoterman, M. H., . . . Braber, S. (2017). Characterizing microbiota-independent effects of oligosaccharides on intestinal epithelial cells: Insight into the role of structure and size. *European Journal of Nutrition, 56*(5), 1919-1930.

Allen, L. H. (2012). B vitamins in breast milk: relative importance of maternal status and intake, and effects on infant status and function. *Advances in Nutrition, 3*(3), 362-369.

Amenyogbe, N., Kollmann, T. R., & Ben-Othman, R. (2017). Early-life host–microbiome interphase: the key frontier for immune development. *Frontiers in pediatrics, 5*, 111.

Anderson, J. M., & Van Itallie, C. M. (2009). Physiology and function of the tight junction. *Cold Spring Harbor perspectives in biology, 1*(2), a002584.

Andreas, N. J., Kampmann, B., & Le-Doare, K. M. (2015). Human breast milk: A review on its composition and bioactivity. *Early human development, 91*(11), 629-635.

Arrieta, M.-C., Stiemsma, L. T., Amenyogbe, N., Brown, E. M., & Finlay, B. (2014). The intestinal microbiome in early life: health and disease. *Frontiers in immunology, 5*, 427.

Asakuma, S., Hatakeyama, E., Urashima, T., Yoshida, E., Katayama, T., Yamamoto, K., . . . Kitaoka, M. (2011). Physiology of Consumption of Human Milk Oligosaccharides by Infant Gut-associated Bifidobacteria *. *Journal of Biological Chemistry, 286*(40), 34583-34592.

Avershina, E., Lundgård, K., Sekelja, M., Dotterud, C., Storrø, O., Øien, T., . . . Rudi, K. (2016). Transition from infant-to adult-like gut microbiota. *Environmental Microbiology, 18*(7), 2226-2236.

Azad, M. B., Konya, T., Maughan, H., Guttman, D. S., Field, C. J., Chari, R. S., . . . Kozyrskyj, A. L. (2013). Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *Cmaj, 185*(5), 385-394.

Azad, M. B., Konya, T., Persaud, R. R., Guttman, D. S., Chari, R. S., Field, C. J., . . . Subbarao, P. (2016). Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: a prospective cohort study. *BJOG: An International Journal of Obstetrics & Gynaecology, 123*(6), 983-993.

Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., . . . Wang, J. (2015). Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell host & microbe, 17*(5), 690-703.

Bager, P., Wohlfahrt, J., & Westergaard, T. (2008). Caesarean delivery and risk of atopy and allergic disesase: meta-analyses. *Clinical & Experimental Allergy, 38*(4), 634-642.

Belkaid, Y., & Segre, J. A. (2014). Dialogue between skin microbiota and immunity. *Science, 346*(6212), 954-959.

Bengtsson-Palme, J., Hartmann, M., Eriksson, K. M., Pal, C., Thorell, K., Larsson, D. G. J., & Nilsson, R. H. (2015). METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Molecular ecology resources, 15*(6), 1403-1414.

Bergström, A., Skov, T. H., Bahl, M. I., Roager, H. M., Christensen, L. B., Ejlerskov, K. T., . . . Licht, T. R. (2014). Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of Danish infants. *Applied and environmental microbiology, 80*(9), 2889-2900.

Bermudez-Brito, M., Rösch, C., Schols, H. A., Faas, M. M., & de Vos, P. (2015). Resistant starches differentially stimulate Toll-like receptors and attenuate proinflammatory cytokines in dendritic cells by modulation of intestinal epithelial cells. *Molecular nutrition & food research, 59*(9), 1814-1826.

Bertelsen, R. J., Jensen, E. T., & Ringel-Kulka, T. (2016). Use of probiotics and prebiotics in infant feeding. *Best practice & research Clinical gastroenterology, 30*(1), 39-48.

Bhinder, G., Allaire, J. M., Garcia, C., Lau, J. T., Chan, J. M., Ryz, N. R., . . . Vallance, B. A. (2017). Milk Fat Globule Membrane Supplementation in Formula Modulates the Neonatal Gut Microbiome and Normalizes Intestinal Development. *Scientific reports, 7*(1), 45274.

Biesalski, H. K. (2016). Nutrition meets the microbiome: micronutrients and the microbiota. *Annals of the New York Academy of Sciences, 1372*(1), 53-64.

Birt, D. F., Boylston, T., Hendrich, S., Jane, J.-L., Hollis, J., Li, L., . . . Whitley, E. M. (2013). Resistant Starch: Promise for Improving Human Health. *Advances in Nutrition, 4*(6), 587-601.

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.

Bode, L. (2012). Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology, 22*(9), 1147-1162.

Boehm, G. n., & Moro, G. (2008). Structural and functional aspects of prebiotics used in infant nutrition. *The Journal of nutrition, 138*(9), 1818S-1828S.

Boets, E., Gomand, S. V., Deroover, L., Preston, T., Vermeulen, K., De Preter, V., . . . 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.

Boix-Amorós, A., Collado, M. C., Van't Land, B., Calvert, A., Le Doare, K., Garssen, J., . . . Munblit, D. (2019). Reviewing the evidence on breast milk composition and immunological outcomes. *Nutrition reviews, 77*(8), 541-556.

Bokulich, N. A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., . . . Blaser, M. J. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Science translational medicine, 8*(343), 343ra382-343ra382.

Boran, P., Baris, H. E., Kepenekli, E., Erzik, C., Soysal, A., & Dinh, D. M. (2020). The impact of vitamin B12 deficiency on infant gut microbiota. *European journal of pediatrics, 179*(3), 385-393.

Bourlieu, C., Ménard, O., Bouzerzour, K., Mandalari, G., Macierzanka, A., Mackie, A. R., & Dupont, D. (2014). Specificity of infant digestive conditions: some clues for developing relevant in vitro models. *Critical reviews in food science and nutrition, 54*(11), 1427-1457.

Brestoff, J. R., & Artis, D. (2013). Commensal bacteria at the interface of host metabolism and the immune system. *Nature immunology, 14*(7), 676-684.

Bridgman, S. L., Azad, M. B., Field, C. J., Haqq, A. M., Becker, A. B., Mandhane, P. J., . . . Turvey, S. E. (2017). Fecal Short-Chain Fatty Acid Variations by Breastfeeding Status in Infants at 4 Months: Differences in Relative versus Absolute Concentrations. *Frontiers in Nutrition, 4*.

Brink, L. R., Mercer, K. E., Piccolo, B. D., Chintapalli, S. V., Elolimy, A., Bowlin, A. K., . . . Shankar, K. (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.

Brockhausen, I. (2003). Sulphotransferases acting on mucin-type oligosaccharides. *Biochemical Society Transactions, 31*(2), 318-325.

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature methods, 12*(1), 59-60.

Caballero-Franco, C., Keller, K., De Simone, C., & Chadee, K. (2007). The VSL# 3 probiotic formula induces mucin gene expression and secretion in colonic epithelial cells. *American Journal of Physiology-Gastrointestinal and Liver Physiology, 292*(1), G315-G322.

Cabrera, D., Kruger, M., Wolber, F. M., Roy, N. C., & Fraser, K. (2020). Effects of short- and long-term glucocorticoid-induced osteoporosis on plasma metabolome and lipidome of ovariectomized sheep. *BMC Musculoskeletal Disorders, 21*(1), 349.

Castanet, M., Costalos, C., Haiden, N., Hascoet, J.-M., Berger, B., Sprenger, N., . . . Picaud, J.-C. (2020). Early Effect of Supplemented Infant Formulae on Intestinal Biomarkers and Microbiota: A Randomized Clinical Trial. *Nutrients, 12*(5), 1481.

Chaleckis, R., Meister, I., Zhang, P., & Wheelock, C. E. (2019). Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics. *Current opinion in biotechnology, 55*, 44-50.

Choo, J. M., Leong, L. E., & Rogers, G. B. (2015). Sample storage conditions significantly influence faecal microbiome profiles. *Scientific reports, 5*(1), 1-10.

Chow, J., Panasevich, M. R., Alexander, D., Vester Boler, B. M., Rossoni Serao, M. C., Faber, T. A., . . . Fahey, G. C. (2014). Fecal metabolomics of healthy breast-fed versus formula-fed infants before and during in vitro batch culture fermentation. *Journal of proteome research, 13*(5), 2534-2542.

Clavel, T., Gomes-Neto, J. C., Lagkouvardos, I., & Ramer-Tait, A. E. (2017). Deciphering interactions between the gut microbiota and the immune system via microbial cultivation and minimal microbiomes. *Immunological reviews, 279*(1), 8-22.

Clavel, T., Lepage, P., & Charrier, C. (2014). The Family Coriobacteriaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Actinobacteria* (pp. 201-238). Berlin, Heidelberg: Springer Berlin Heidelberg.

Cong, X., Xu, W., Janton, S., Henderson, W. A., Matson, A., McGrath, J. M., . . . Graf, J. (2016). Gut microbiome developmental patterns in early life of preterm infants: impacts of feeding and gender. *PloS one, 11*(4), e0152751.

Cornick, S., Tawiah, A., & Chadee, K. (2015). Roles and regulation of the mucus barrier in the gut. *Tissue barriers, 3*(1-2), e982426.

Crittenden, R., Laitila, A., Forssell, P., Mättö, J., Saarela, M., Mattila-Sandholm, T., & Myllärinen, P. (2001). Adhesion of Bifidobacteria to Granular Starch and Its Implications in Probiotic Technologies. *Applied and environmental microbiology, 67*(8), 3469-3475.

Cui, X., Li, Y., Yang, L., You, L., Wang, X., Shi, C., . . . Guo, X. (2016). Peptidome analysis of human milk from women delivering macrosomic fetuses reveals multiple means of protection for infants. *Oncotarget, 7*(39), 63514.

Cummins, A. G., & Thompson, F. M. (2002). Effect of breast milk and weaning on epithelial growth of the small intestine in humans. *Gut, 51*(5), 748-754.

Dallas, D. C., Sanctuary, M. R., Qu, Y., Khajavi, S. H., Van Zandt, A. E., Dyandra, M., . . . German, J. B. (2017). Personalizing protein nourishment. *Critical reviews in food science and nutrition, 57*(15), 3313-3331.

Dallman, P. R. (1988). Nutritional anemia of infancy: iron, folic acid, and vitamin B12. *Infant Nutrition. Philadelphia: Hanley and Belfus Inc, 2*, 16-35.

Daniels, L., Heath, A.-L. M., Williams, S. M., Cameron, S. L., Fleming, E. A., Taylor, B. J., . . . Taylor, R. W. (2015). Baby-Led Introduction to SolidS (BLISS) study: a randomised controlled trial of a baby-led approach to complementary feeding. *BMC Pediatrics, 15*(1), 1-15.

Davis, E. C., Wang, M., & Donovan, S. M. (2017). The role of early life nutrition in the establishment of gastrointestinal microbial composition and function. *Gut Microbes, 8*(2), 143-171.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., . . . Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America, 107*(33), 14691-14696.

De Kivit, S., Tobin, M., Forsyth, C., Keshavarzian, A., & Landay, A. (2014). Regulation of Intestinal Immune Responses through TLR Activation: Implications for Pro- and Prebiotics. *Frontiers in immunology, 5*.

de Muinck, E. J., & Trosvik, P. (2018). Individuality and convergence of the infant gut microbiota during the first year of life. *Nature communications, 9*(1), 1-8.

de Verteuil, D., Granados, D. P., Thibault, P., & Perreault, C. (2012). Origin and plasticity of MHC I-associated self peptides. *Autoimmunity Reviews, 11*(9), 627-635.

Denno, D. M., VanBuskirk, K., Nelson, Z. C., Musser, C. A., Hay Burgess, D. C., & Tarr, P. I. (2014). Use of the lactulose to mannitol ratio to evaluate childhood environmental enteric dysfunction: a systematic review. *Clinical Infectious Diseases, 59*(suppl_4), S213-S219.

Depner, M., Taft, D. H., Kirjavainen, P. V., Kalanetra, K. M., Karvonen, A. M., Peschel, S., . . . Lauener, R. (2020). Maturation of the gut microbiome during the first year of life contributes to the protective farm effect on childhood asthma. *Nature medicine, 26*(11), 1766-1775.

Derrien, M., & van Hylckama Vlieg, J. E. T. (2015). Fate, activity, and impact of ingested bacteria within the human gut microbiota. *Trends in microbiology, 23*(6), 354-366.

Dewey, K. G. (2013). The challenge of meeting nutrient needs of infants and young children during the period of complementary feeding: an evolutionary perspective. *The Journal of nutrition, 143*(12), 2050-2054.

Dewit, O., Dibba, B., & Prentice, A. (1990). Breast-milk amylase activity in English and Gambian mothers: effects of prolonged lactation, maternal parity, and individual variations. *Pediatric research, 28*(5), 502-506.

Dorosko, S. M., MacKenzie, T., & Connor, R. I. (2008). Fecal calprotectin concentrations are higher in exclusively breastfed infants compared to those who are mixed-fed. *Breastfeeding Medicine, 3*(2), 117-119.

Duerkop, B. A., Vaishnava, S., & Hooper, L. V. (2009). Immune Responses to the Microbiota at the Intestinal Mucosal Surface. *Immunity, 31*(3), 368-376.

Duranti, S., Lugli, G. A., Mancabelli, L., Armanini, F., Turroni, F., James, K., . . . Milani, C. (2017). Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. *Microbiome, 5*(1), 1-13.

Engfer, M. B., Stahl, B., Finke, B., Sawatzki, G., & Daniel, H. (2000). Human milk oligosaccharides are resistant to enzymatic hydrolysis in the upper gastrointestinal tract. *The American journal of clinical nutrition, 71*(6), 1589-1596.

Ercolini, D., & Fogliano, V. (2018). Food design to feed the human gut microbiota. *Journal of agricultural and food chemistry, 66*(15), 3754-3758.

Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., . . . Gordon, J. I. (2013). The long-term stability of the human gut microbiota. *Science, 341*(6141), 1237439.

Fallani, M., Amarri, S., Uusijarvi, A., Adam, R., Khanna, S., Aguilera, M., . . . Young, D. (2011). Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology, 157*(5), 1385-1392.

Feehley, T., Plunkett, C. H., Bao, R., Hong, S. M. C., Culleen, E., Belda-Ferre, P., . . . Paparo, L. (2019). Healthy infants harbor intestinal bacteria that protect against food allergy. *Nature medicine, 25*(3), 448-453.

Fergusson, D. M., Horwood, L. J., & Shannon, F. T. (1990). Early Solid Feeding and Recurrent Childhood Eczema: A 10-Year Longitudinal Study. *Pediatrics, 86*(4), 541-546.

Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P., & Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. *Gut Microbes, 3*(4), 289-306.

Fraser, K., Harrison, S. J., Lane, G. A., Otter, D. E., Hemar, Y., Quek, S.-Y., & Rasmussen, S. (2012). Non-targeted analysis of tea by hydrophilic interaction liquid chromatography and high resolution mass spectrometry. *Food chemistry, 134*(3), 1616-1623.

Fujimura, K. E., Sitarik, A. R., Havstad, S., Lin, D. L., Levan, S., Fadrosh, D., . . . Lukacs, N. W. (2016). Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nature medicine, 22*(10), 1187-1191.

Fukuda, S., Toh, H., Taylor, T. D., Ohno, H., & Hattori, M. (2012). Acetate-producing bifidobacteria protect the host from enteropathogenic infection via carbohydrate transporters. *Gut Microbes, 3*(5), 449-454.

Fung, T. C., Artis, D., & Sonnenberg, G. F. (2014). Anatomical localization of commensal bacteria in immune cell homeostasis and disease. *Immunological reviews, 260*(1), 35-49.

Gan, J., Bornhorst, G. M., Henrick, B. M., & German, J. B. (2018). Protein Digestion of Baby Foods: Study Approaches and Implications for Infant Health. *Molecular nutrition & food research, 62*(1), 10.1002/mnfr.201700231.

Gantois, I., Ducatelle, R., Pasmans, F., Haesebrouck, F., Hautefort, I., Thompson, A., . . . Van Immerseel, F. (2006). Butyrate Specifically Down-Regulates Salmonella Pathogenicity Island 1 Gene Expression. *Applied and environmental microbiology, 72*(1), 946-949.

Garrido, D., Kim, J. H., German, J. B., Raybould, H. E., & Mills, D. A. (2011). Oligosaccharide Binding Proteins from Bifidobacterium longum subsp. infantis Reveal a Preference for Host Glycans. *PloS one, 6*(3), e17315.

Garrido, D., Ruiz-Moyano, S., Lemay, D. G., Sela, D. A., German, J. B., & Mills, D. A. (2015). Comparative transcriptomics reveals key differences in the response to milk oligosaccharides of infant gut-associated bifidobacteria. *Scientific reports, 5*(1), 13517.

Geijtenbeek, T. B. H., & Gringhuis, S. I. (2009). Signalling through C-type lectin receptors: shaping immune responses. *Nature Reviews Immunology, 9*(7), 465-479.

Gensollen, T., Iyer, S. S., Kasper, D. L., & Blumberg, R. S. (2016). How colonization by microbiota in early life shapes the immune system. *Science, 352*(6285), 539-544.

Georas, S. N., & Rezaee, F. (2014). Epithelial barrier function: at the front line of asthma immunology and allergic airway inflammation. *Journal of Allergy and Clinical Immunology, 134*(3), 509-520.

Georgi, G., Bartke, N., Wiens, F., & Stahl, B. (2013). Functional glycans and glycoconjugates in human milk. *The American journal of clinical nutrition, 98*(2), 578S-585S.

Gibson, R. S., Bailey, K. B., Gibbs, M., & Ferguson, E. L. (2010). A Review of Phytate, Iron, Zinc, and Calcium Concentrations in Plant-Based Complementary Foods Used in Low-Income Countries and Implications for Bioavailability. *Food and Nutrition Bulletin, 31*(2_suppl2), S134-S146.

Goenka, A., & Kollmann, T. R. (2015). Development of immunity in early life. *Journal of Infection, 71*, S112-S120.

Gonzalez-Perez, G., Hicks, A. L., Tekieli, T. M., Radens, C. M., Williams, B. L., & Lamousé-Smith, E. S. (2016). Maternal antibiotic treatment impacts development of the neonatal intestinal microbiome and antiviral immunity. *The Journal of Immunology, 196*(9), 3768-3779.

Gonzalez-Perez, G., & Lamousé-Smith, E. S. N. (2017). Gastrointestinal Microbiome Dysbiosis in Infant Mice Alters Peripheral CD8+ T Cell Receptor Signaling. *Frontiers in immunology, 8*.

González, R., Klaassens, E. S., Malinen, E., Vos, W. M. d., & Vaughan, E. E. (2008). Differential Transcriptional Response of Bifidobacterium longum to Human Milk, Formula Milk, and Galactooligosaccharide. *Applied and environmental microbiology, 74*(15), 4686-4694.

Goulet, O. (2015). Potential role of the intestinal microbiota in programming health and disease. *Nutrition reviews, 73*(suppl_1), 32-40.

Gralka, M., Szabo, R., Stocker, R., & Cordero, O. X. (2020). Trophic Interactions and the Drivers of Microbial Community Assembly. *Current Biology, 30*(19), R1176-R1188.

Green Corkins, K., & Shurley, T. (2016). What's in the Bottle? A Review of Infant Formulas. *Nutrition in Clinical Practice, 31*(6), 723-729.

Haange, S.-B., Oberbach, A., Schlichting, N., Hugenholtz, F., Smidt, H., von Bergen, M., . . . Seifert, J. (2012). Metaproteome analysis and molecular genetics of rat intestinal microbiota reveals section and localization resolved species distribution and enzymatic functionalities. *Journal of proteome research, 11*(11), 5406-5417.

Haisma, H., Coward, W., Albernaz, E., Visser, G., Wells, J. C. K., Wright, A., & Victora, C. G. (2003). Breast milk and energy intake in exclusively, predominantly, and partially breast-fed infants. *European journal of clinical nutrition, 57*(12), 1633-1642.

Hanson, L. Å., Korotkova, M., & Telemo, E. (2003). Breast-feeding, infant formulas, and the immune system. *Annals of Allergy, Asthma & Immunology, 90*(6), 59-63.

Hartwig, I., Diemert, A., Tolosa, E., Hecher, K., & Arck, P. (2015). Babies galore; or recent findings and future perspectives of pregnancy cohorts with a focus on immunity. *Journal of reproductive immunology, 108*, 6-11.

Hausner, H., Nicklaus, S., Issanchou, S., Mølgaard, C., & Møller, P. (2009). Breastfeeding facilitates acceptance of a novel dietary flavour compound. *e-SPEN, the European e-Journal of Clinical Nutrition and Metabolism, 4*(5), e231-e238.

He, X., Parenti, M., Grip, T., Domellöf, M., Lönnerdal, B., Hernell, O., . . . Slupsky, C. M. (2019). Metabolic phenotype of breast-fed infants, and infants fed standard formula or bovine MFGM supplemented formula: a randomized controlled trial. *Scientific reports, 9*(1), 339.

He, X., Sotelo-Orozco, J., Rudolph, C., Lönnerdal, B., & Slupsky, C. M. (2020). The Role of Protein and Free Amino Acids on Intake, Metabolism, and Gut Microbiome: A Comparison Between Breast-Fed and Formula-Fed Rhesus Monkey Infants. *Frontiers in pediatrics, 7*.

Heinig, M. J., Nommsen, L. A., Peerson, J. M., Lonnerdal, B., & Dewey, K. G. (1993). Energy and protein intakes of breast-fed and formula-fed infants during the first year of life and their association with growth velocity: the DARLING Study. *The American journal of clinical nutrition, 58*(2), 152-161.

Herrmann, E., Young, W., Rosendale, D., Conrad, R., Riedel, C. U., & Egert, M. (2017). Determination of Resistant Starch Assimilating Bacteria in Fecal Samples of Mice by In vitro RNA-Based Stable Isotope Probing. *Frontiers in microbiology, 8*.

Hibberd, M. C., Wu, M., Rodionov, D. A., Li, X., Cheng, J., Griffin, N. W., . . . Osterman, A. L. (2017). The effects of micronutrient deficiencies on bacterial species from the human gut microbiota. *Science translational medicine, 9*(390).

Hill, C. J., Brown, J. R., Lynch, D. B., Jeffery, I. B., Ryan, C. A., Ross, R. P., . . . O'Toole, P. W. (2016). Effect of room temperature transport vials on DNA quality and phylogenetic composition of faecal microbiota of elderly adults and infants. *Microbiome, 4*(1), 1-10.

Hill, C. J., Lynch, D. B., Murphy, K., Ulaszewska, M., Jeffery, I. B., O'Shea, C. A., . . . Tuohy, K. (2017). Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort. *Microbiome, 5*(1), 1-18.

Hillmann, B., Al-Ghalith, G. A., Shields-Cutler, R. R., Zhu, Q., Gohl, D. M., Beckman, K. B., . . . Knights, D. (2018). Evaluating the Information Content of Shallow Shotgun Metagenomics. *Msystems, 3*(6), e00069-00018.

Ho, N. T., Li, F., Lee-Sarwar, K. A., Tun, H. M., Brown, B. P., Pannaraj, P. S., . . . Kuhn, L. (2018). Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. *Nature communications, 9*(1), 4169-4169.

Holmes, A. J., Chew, Y. V., Colakoglu, F., Cliff, J. B., Klaassens, E., Read, M. N., . . . Simpson, S. J. (2017). Diet-Microbiome Interactions in Health Are Controlled by Intestinal Nitrogen Source Constraints. *Cell Metabolism, 25*(1), 140-151.

Hosseini, E., Grootaert, C., Verstraete, W., & Van de Wiele, T. (2011). Propionate as a health-promoting microbial metabolite in the human gut. *Nutrition reviews, 69*(5), 245-258.

Hugenholtz, F., Davids, M., Schwarz, J., Müller, M., Tomé, D., Schaap, P., . . . Kleerebezem, M. (2018). Metatranscriptome analysis of the microbial fermentation of dietary milk proteins in the murine gut. *PloS one, 13*(4), e0194066.

Hugenholtz, F., Ritari, J., Nylund, L., Davids, M., Satokari, R., & de Vos, W. M. (2017). Feasibility of Metatranscriptome Analysis from Infant Gut Microbiota: Adaptation to Solid Foods Results in Increased Activity of Firmicutes at Six Months. *International Journal of Microbiology, 2017*, 9547063.

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., . . . Tappu, R. (2016). MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS computational biology, 12*(6), e1004957.

Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of computational and graphical statistics, 5*(3), 299-314.

Iyengar, S. R., & Walker, W. (2012). Immune factors in breast milk and the development of atopic disease. *Journal of pediatric gastroenterology and nutrition, 55*(6), 641-647.

Jayachandran, M., Chung, S. S. M., & Xu, B. (2020). A critical review of the relationship between dietary components, the gut microbe Akkermansia muciniphila, and human health. *Crit Rev Food Sci Nutr, 60*(13), 2265-2276.

Jeurink, P. V., van Esch, B. C., Rijnierse, A., Garssen, J., & Knippels, L. M. (2013). Mechanisms underlying immune effects of dietary oligosaccharides. *The American journal of clinical nutrition, 98*(2), 572S-577S.

Jiménez, E., Marín, M. L., Martín, R., Odriozola, J. M., Olivares, M., Xaus, J., . . . Rodríguez, J. M. (2008). Is meconium from healthy newborns actually sterile? *Research in Microbiology, 159*(3), 187-193.

Johnson, A. J., Vangay, P., Al-Ghalith, G. A., Hillmann, B. M., Ward, T. L., Shields-Cutler, R. R., . . . Students, P. M. C. (2019). Daily sampling reveals personalized diet-microbiome associations in humans. *Cell host & microbe, 25*(6), 789-802. e785.

Kaakoush, N. O. (2015). Insights into the Role of Erysipelotrichaceae in the Human Host. *Frontiers in cellular and infection microbiology, 5*, 84-84.

Kabat, A. M., Srinivasan, N., & Maloy, K. J. (2014). Modulation of immune development and function by intestinal microbiota. *Trends in immunology, 35*(11), 507-517.

Kalina, U., Koyama, N., Hosoda, T., Nuernberger, H., Sato, K., Hoelzer, D., . . . Böcker, U. (2002). Enhanced production of IL-18 in butyrate-treated intestinal epithelium by stimulation of the proximal promoter region. *European Journal of Immunology, 32*(9), 2635-2643.

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic acids research, 49*(D1), D545-D551.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research, 28*(1), 27-30.

Kaplan, J. L., Shi, H. N., & Walker, W. A. (2011). The role of microbes in developmental immunologic programming. *Pediatric research, 69*(6), 465-472.

Karnovsky, A., Weymouth, T., Hull, T., Tarcea, V. G., Scardoni, G., Laudanna, C., . . . Omenn, G. S. (2012). Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. *Bioinformatics, 28*(3), 373-380.

Karu, N., Deng, L., Slae, M., Guo, A. C., Sajed, T., Huynh, H., . . . Wishart, D. S. (2018). A review on human fecal metabolomics: Methods, applications and the human fecal metabolome database. *Analytica chimica acta, 1030*, 1-24.

Keusch, G. T., Denno, D. M., Black, R. E., Duggan, C., Guerrant, R. L., Lavery, J. V., . . . Tarr, P. I. (2014). Environmental enteric dysfunction: pathogenesis, diagnosis, and clinical consequences. *Clinical Infectious Diseases, 59*(suppl_4), S207-S212.

Kjer-Nielsen, L., Patel, O., Corbett, A. J., Le Nours, J., Meehan, B., Liu, L., . . . McCluskey, J. (2012). MR1 presents microbial vitamin B metabolites to MAIT cells. *Nature, 491*(7426), 717-723.

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., . . . Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences, 108*(Supplement 1), 4578-4585.

Koga, Y., Tokunaga, S., Nagano, J., Sato, F., Konishi, K., Tochio, T., . . . Shibata, R. (2016). Age-associated effect of kestose on Faecalibacterium prausnitzii and symptoms in the atopic dermatitis infants. *Pediatric research, 80*(6), 844-851.

Kollmann, Tobias R., Levy, O., Montgomery, Ruth R., & Goriely, S. (2012). Innate Immune Function by Toll-like Receptors: Distinct Responses in Newborns and the Elderly. *Immunity, 37*(5), 771-783.

Koropatkin, N. M., Cameron, E. A., & Martens, E. C. (2012). How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology, 10*(5), 323-335.

Koropatkin, N. M., & Smith, T. J. (2010). SusG: A Unique Cell-Membrane-Associated α-Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. *Structure, 18*(2), 200-215.

Kovatcheva-Datchary, P., Egert, M., Maathuis, A., Rajilić-Stojanović, M., De Graaf, A. A., Smidt, H., . . . Venema, K. (2009). Linking phylogenetic identities of bacteria to starch fermentation in an in vitro model of the large intestine by RNA-based stable isotope probing. *Environmental Microbiology, 11*(4), 914-926.

Krebs, N., & Hambidge, K. (1986). Zinc requirements and zinc intakes of breast-fed infants. *The American journal of clinical nutrition, 43*(2), 288-292.

Krebs, N. F. (2001). Bioavailability of Dietary Supplements and Impact of Physiologic State: Infants, Children and Adolescents. *The Journal of Nutrition, 131*(4), 1351S-1354S.

Krebs, N. F., Sherlock, L. G., Westcott, J., Culbertson, D., Hambidge, K. M., Feazel, L. M., . . . Frank, D. N. (2013). Effects of different complementary feeding regimens on iron status and enteric microbiota in breastfed infants. *The Journal of pediatrics, 163*(2), 416-423. e414.

Kunisawa, J., Hashimoto, E., Ishikawa, I., & Kiyono, H. (2012). A Pivotal Role of Vitamin B9 in the Maintenance of Regulatory T Cells In Vitro and In Vivo. *PloS one, 7*(2), e32094.

Kwak, M.-J., Kwon, S.-K., Yoon, J.-K., Song, J. Y., Seo, J.-G., Chung, M. J., & Kim, J. F. (2016). Evolutionary architecture of the infant-adapted group of Bifidobacterium species associated with the probiotic function. *Systematic and applied microbiology, 39*(7), 429-439.

Lang, J. M., Pan, C., Cantor, R. M., Tang, W. H. W., Garcia-Garcia, J. C., Kurtz, I., . . . Lusis, A. J. (2018). Impact of Individual Traits, Saturated Fat, and Protein Source on the Gut Microbiome. *mBio, 9*(6), e01604-01618.

Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., . . . Knight, R. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology, 31*(9), 814-821.

Laursen, M. F., Bahl, M. I., Michaelsen, K. F., & Licht, T. R. (2017). First foods and gut microbes. *Frontiers in microbiology*, *8*, 356.

Lawson, M. A., O'Neill, I. J., Kujawska, M., Gowrinadh Javvadi, S., Wijeyesekera, A., Flegg, Z., . . . Hall, L. J. (2020). Breast milk-derived human milk oligosaccharides promote Bifidobacterium interactions within a single ecosystem. *The ISME journal*, *14*(2), 635-648.

Lebenthal, E., Lee, P., & Heitlinger, L. A. (1983). Impact of development of the gastrointestinal tract on infant feeding. *The Journal of pediatrics*, *102*(1), 1-9.

Leder, S., Hartmeier, W., & Marx, S. P. (1999). α-Galactosidase of Bifidobacterium adolescentis DSM 20083. *Current Microbiology*, *38*(2), 101-106.

Lee, S. A., Lim, J. Y., Kim, B.-S., Cho, S. J., Kim, N. Y., Kim, O. B., & Kim, Y. (2015). Comparison of the gut microbiota profile in breast-fed and formula-fed Korean infants using pyrosequencing. *Nutrition research and practice*, *9*(3), 242.

Levental, K. R., Malmberg, E., Symons, J. L., Fan, Y.-Y., Chapkin, R. S., Ernst, R., & Levental, I. (2020). Lipidomic and biophysical homeostasis of mammalian membranes counteracts dietary lipid perturbations to maintain cellular fitness. *Nature communications*, *11*(1), 1339.

Li, M., Wang, M., & Donovan, S. M. (2014). *Early development of the gut microbiome and immune-mediated childhood disorders.* Paper presented at the Seminars in reproductive medicine.

Li, N., Yan, F., Wang, N., Song, Y., Yue, Y., Guan, J., . . . Huo, G. (2020). Distinct gut microbiota and metabolite profiles induced by different feeding methods in healthy chinese infants. *Frontiers in microbiology*, *11*, 714.

Liang, L. (2020). Folates: stability and interaction with biological molecules. *Journal of Agriculture and Food Research*, *2*, 100039.

Liebisch, G., Ahrends, R., Makoto, A., Masanori, A., Bowden, J. A., Ejsing, C. S., . . . Mitchell, T. W. (2019). Lipidomics needs more standardization. *Nature Metabolism, volume 1, issue: November*.

Lippert, K., Kedenko, L., Antonielli, L., Kedenko, I., Gemeier, C., Leitner, M., ... & Hackl, E. (2017). Gut microbiota dysbiosis associated with glucose metabolism disorders and the metabolic syndrome in older adults. Beneficial microbes, 8(4), 545-556.

Lin, A. H. M., & Nichols, B. L. (2017). The digestion of complementary feeding starches in the young child. *Starch-Stärke, 69*(7-8), 1700012.

Liu, Z., Roy, N. C., Guo, Y., Jia, H., Ryan, L., Samuelsson, L., . . . Day, L. (2016). Human breast milk and infant formulas differentially modify the intestinal microbiota in human infants and host physiology in rats. *The Journal of nutrition*, *146*(2), 191-199.

Loftfield, E., Vogtmann, E., Sampson, J. N., Moore, S. C., Nelson, H., Knight, R., . . . Sinha, R. (2016). Comparison of collection methods for fecal samples for discovery metabolomics in epidemiologic studies. *Cancer Epidemiology and Prevention Biomarkers*, *25*(11), 1483-1490.

Lönnerdal, B., & Hernell, O. (2016). An opinion on "staging" of infant formula: a developmental perspective on infant feeding. *Journal of pediatric gastroenterology and nutrition*, *62*(1), 9-21.

Louis, P., & Flint, H. J. (2009). Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiology Letters*, *294*(1), 1-8.

Louis, P., & Flint, H. J. (2017). Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology*, *19*(1), 29-41.

Ma, N., & Ma, X. (2019). Dietary Amino Acids and the Gut-Microbiome-Immune Axis: Physiological Metabolism and Therapeutic Prospects. *Comprehensive Reviews in Food Science and Food Safety*, *18*(1), 221-242.

Macfarlane, S., & Macfarlane, G. T. (2003). Regulation of short-chain fatty acid production. *Proceedings of the Nutrition Society*, *62*(1), 67-72.

Macpherson, A. J., & Uhr, T. (2004). Induction of Protective IgA by Intestinal Dendritic Cells Carrying Commensal Bacteria. *Science*, *303*(5664), 1662-1665.

Madan, J. C., Farzan, S. F., Hibberd, P. L., & Karagas, M. R. (2012). Normal neonatal microbiome variation in relation to environmental factors, infection and allergy. *Current opinion in pediatrics*, *24*(6), 753.

Magnúsdóttir, S., Ravcheev, D., de Crécy-Lagard, V., & Thiele, I. (2015). Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Frontiers in genetics*, *6*, 148.

Maier, T. V., Lucio, M., Lee, L. H., VerBerkmoes, N. C., Brislawn, C. J., Bernhardt, J., . . . Heinzmann, S. S. (2017). Impact of dietary resistant starch on the human gut microbiome, metaproteome, and metabolome. *mBio*, *8*(5), e01343-01317.

Marcobal, A., & Sonnenburg, J. L. (2012). Human milk oligosaccharide consumption by intestinal microbiota. *Clinical Microbiology and Infection*, *18*, 12-15.

Mariadason, J. M., Barkla, D. H., & Gibson, P. R. (1997). Effect of short-chain fatty acids on paracellular permeability in Caco-2 intestinal epithelium model. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *272*(4), G705-G712.

Martin, C. R., Ling, P.-R., & Blackburn, G. L. (2016). Review of Infant Feeding: Key Features of Breast Milk and Infant Formula. *Nutrients, 8*(5), 279.
Martin, R., Nauta, A., Ben Amor, K., Knippels, L., Knol, J., & Garssen, J. (2010). Early life: gut microbiota and immune development in infancy. *Beneficial microbes, 1*(4), 367-382.
McCormick, B. J. J., Lee, G. O., Seidman, J. C., Haque, R., Mondal, D., Quetz, J., . . . Kosek, M. N. (2017). Dynamics and Trends in Fecal Biomarkers of Gut Function in Children from 1-24 Months in the MAL-ED Study. *The American journal of tropical medicine and hygiene, 96*(2), 465-472.
Miller, T. L., & Wolin, M. J. (1996). Pathways of acetate, propionate, and butyrate formation by the human fecal microbial flora. *Appl Environ Microbiol, 62*(5), 1589-1592.
Moore, R. E., & Townsend, S. D. (2019). Temporal development of the infant gut microbiome. *Open biology, 9*(9), 190128.
Morrison, D. J., & Preston, T. (2016). Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes, 7*(3), 189-200.
Morrow, A. L., Ruiz-Palacios, G. M., Jiang, X., & Newburg, D. S. (2005). Human-milk glycans that inhibit pathogen binding protect breast-feeding infants against infectious diarrhea. *The Journal of nutrition, 135*(5), 1304-1307.
Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., González, A., Fontana, L., . . . Gordon, J. I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science, 332*(6032), 970-974.
Nagpal, R., Kurakawa, T., Tsuji, H., Takahashi, T., Kawashima, K., Nagata, S., . . . Yamashiro, Y. (2017). Evolution of gut Bifidobacterium population in healthy Japanese infants over the first three years of life: a quantitative assessment. *Scientific reports, 7*(1), 10097.
Nagpal, R., Tsuji, H., Takahashi, T., Nomoto, K., Kawashima, K., Nagata, S., & Yamashiro, Y. (2017). Ontogenesis of the Gut Microbiota Composition in Healthy, Full-Term, Vaginally Born and Breast-Fed Infants over the First 3 Years of Life: A Quantitative Bird's-Eye View. *Frontiers in microbiology, 8*.
Negele, K., Heinrich, J., Borte, M., von Berg, A., Schaaf, B., Lehmann, I., . . . Group, f. t. L. S. (2004). Mode of delivery and development of atopic disease during the first 2 years of life. *Pediatric Allergy and Immunology, 15*(1), 48-54.
Neyraud, E., Schwartz, C., Brignot, H., Jouanin, I., Tremblay-Franco, M., Canlet, C., & Tournier, C. (2020). Longitudinal analysis of the salivary metabolome of breast-fed and formula-fed infants over the first year of life. *Metabolomics, 16*(3), 1-10.
Nguyen, Q. P., Karagas, M. R., Madan, J. C., Dade, E., Palys, T. J., Morrison, H. G., . . . Frost, H. R. (2021). Associations between the gut microbiome and metabolome in early life. *BMC Microbiology, 21*(1), 1-19.
Nicholson, J. K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., & Pettersson, S. (2012). Host-gut microbiota metabolic interactions. *Science, 336*(6086), 1262-1267.
Nofrarías, M., Martínez-Puig, D., Pujols, J., Majó, N., & Pérez, J. F. (2007). Long-term intake of resistant starch improves colonic mucosal integrity and reduces gut apoptosis and blood immune cells. *Nutrition, 23*(11-12), 861-870.
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), 91-91.
Ong, K. K., Kennedy, K., Castañeda-Gutiérrez, E., Forsyth, S., Godfrey, K. M., Koletzko, B., . . . Schoemaker, M. H. (2015). Postnatal growth in preterm infants and later health outcomes: a systematic review. *Acta paediatrica, 104*(10), 974-986.
Oozeer, R., Van Limpt, K., Ludwig, T., Ben Amor, K., Martin, R., Wind, R. D., . . . Knol, J. (2013). Intestinal microbiology in early life: specific prebiotics can have similar functionalities as human-milk oligosaccharides. *The American journal of clinical nutrition, 98*(2), 561S-571S.
Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., . . . Edwards, R. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic acids research, 33*(17), 5691-5702.
Palermo, G., Piraino, P., & Zucht, H.-D. (2009). Performance of PLS regression coefficients in selecting variables for each response of a multivariate PLS for omics-type data. *Advances and applications in bioinformatics and chemistry : AABC, 2*, 57-70.
Pannaraj, P. S., Li, F., Cerini, C., Bender, J. M., Yang, S., Rollie, A., . . . Bittinger, K. (2017). Association between breast milk bacterial communities and establishment and development of the infant gut microbiome. *JAMA Pediatrics, 171*(7), 647-654.
Parkar, S. G., Rosendale, D. I., Stoklosinski, H. M., Jobsis, C., Hedderley, D. I., & Gopal, P. (2021). Complementary Food Ingredients Alter Infant Gut Microbiome Composition and Metabolism In Vitro. Microorganisms, 9(10), 2089.
Parracho, H., McCartney, A. L., & Gibson, G. R. (2007). Probiotics and prebiotics in infant nutrition. *Proceedings of the Nutrition Society, 66*(3), 405-411.
Parsek, M. R., Val, D. L., Hanzelka, B. L., Cronan, J. E., Jr., & Greenberg, E. P. (1999). Acyl homoserine-lactone quorum-sensing signal generation. *Proceedings of the National Academy of Sciences of the United States of America, 96*(8), 4360-4365.

Paun, A., & Danska, J. S. (2015). Immuno-ecology: how the microbiome regulates tolerance and autoimmunity. *Current opinion in immunology, 37*, 34-39.

Penders, J., Thijs, C., van den Brandt, P. A., Kummeling, I., Snijders, B., Stelma, F., . . . Stobberingh, E. E. (2007). Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. *Gut, 56*(5), 661.

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., . . . Stobberingh, E. E. (2006). Factors Influencing the Composition of the Intestinal Microbiota in Early Infancy. *Pediatrics, 118*(2), 511-521.

Peng, L., Li, Z.-R., Green, R. S., Holzman, I. R., & Lin, J. (2009). Butyrate Enhances the Intestinal Barrier by Facilitating Tight Junction Assembly via Activation of AMP-Activated Protein Kinase in Caco-2 Cell Monolayers. *The Journal of Nutrition, 139*(9), 1619-1625.

Pham, V. T., Fehlbaum, S., Seifert, N., Richard, N., Bruins, M. J., Sybesma, W., . . . Steinert, R. E. (2021). Effects of colon-targeted vitamins on the composition and metabolic activity of the human gut microbiome–a pilot study. *Gut Microbes, 13*(1), 1-20.

Piccolo, B. D., Mercer, K. E., Bhattacharyya, S., Bowlin, A. K., Saraf, M. K., Pack, L., . . . Badger, T. M. (2017). Early postnatal diets affect the bioregional small intestine microbiome and ileal metabolome in neonatal pigs. *The Journal of nutrition, 147*(8), 1499-1509.

Pluymen, L. P., Wijga, A. H., Gehring, U., Koppelman, G. H., Smit, H. A., & Van Rossem, L. (2018). Early introduction of complementary foods and childhood overweight in breastfed and formula-fed infants in the Netherlands: the PIAMA birth cohort study. *European Journal of Nutrition, 57*(5), 1985-1993.

Podany, A., Rauchut, J., Wu, T., Kawasawa, Y. I., Wright, J., Lamendella, R., . . . Kelleher, S. L. (2019). Excess Dietary Zinc Intake in Neonatal Mice Causes Oxidative Stress and Alters Intestinal Host–Microbe Interactions. *Molecular nutrition & food research, 63*(3), 1800947.

Prasoodanan P. K, V., Sharma, A. K., Mahajan, S., Dhakan, D. B., Maji, A., Scaria, J., & Sharma, V. K. (2021). Western and non-western gut microbiomes reveal new roles of Prevotella in carbohydrate metabolism and mouth–gut axis. *npj Biofilms and Microbiomes, 7*(1), 77.

Praveen, P., Jordan, F., Priami, C., & Morine, M. J. (2015). The role of breast-feeding in infant immune system: a systems perspective on the intestinal microbiome. *Microbiome, 3*(1), 1-12.

Prell, C., & Koletzko, B. (2016). Breastfeeding and complementary feeding: recommendations on infant nutrition. *Deutsches Ärzteblatt International, 113*(25), 435.

Qasem, W., Azad, M. B., Hossain, Z., Azad, E., Jorgensen, S., Castillo San Juan, S., . . . Friel, J. (2017). Assessment of complementary feeding of Canadian infants: effects on microbiome & oxidative stress, a randomized controlled trial. *BMC Pediatrics, 17*(1), 54.

Qian, L., Zhao, A., Zhang, Y., Chen, T., Zeisel, S. H., Jia, W., & Cai, W. (2016). Metabolomic Approaches to Explore Chemical Diversity of Human Breast-Milk, Formula Milk and Bovine Milk. *International Journal of Molecular Sciences, 17*(12).

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research, 41*(D1), D590-D596.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rackerby, B., Kim, H. J., Dallas, D. C., & Park, S. H. (2020). Understanding the effects of dietary components on the gut microbiome and human health. *Food Science and Biotechnology, 29*(11), 1463-1474.

Rinne, M., Kalliomaki, M., Arvilommi, H., Salminen, S., & Isolauri, E. (2005). Effect of Probiotics and Breastfeeding on the Bifidobacterium and Lactobacillus/Enterococcus Microbiota and Humoral Immune Responses. *The Journal of pediatrics, 147*(2), 186-191.

Rizzoli, R. (2019). Nutritional influence on bone: role of gut microbiota. *Aging clinical and experimental research*, 1-9.

Rodionov, D. A., Arzamasov, A. A., Khoroshkin, M. S., Iablokov, S. N., Leyn, S. A., Peterson, S. N., . . . Osterman, A. L. (2019). Micronutrient requirements and sharing capabilities of the human gut microbiome. *Frontiers in microbiology, 10*, 1316.

Rodríguez-Daza, M. C., Pulido-Mateos, E. C., Lupien-Meilleur, J., Guyonnet, D., Desjardins, Y., & Roy, D. (2021). Polyphenol-Mediated Gut Microbiota Modulation: Toward Prebiotics and Further. *Frontiers in Nutrition, 8*, 689456-689456.

Rogier, E. W., Frantz, A. L., Bruno, M. E., Wedlund, L., Cohen, D. A., Stromberg, A. J., & Kaetzel, C. S. (2014a). Lessons from mother: long-term impact of antibodies in breast milk on the gut microbiota and intestinal immune system of breastfed offspring. *Gut Microbes, 5*(5), 663-668.

Rogier, E. W., Frantz, A. L., Bruno, M. E., Wedlund, L., Cohen, D. A., Stromberg, A. J., & Kaetzel, C. S. (2014b). Secretory antibodies in breast milk promote long-term intestinal homeostasis by regulating the gut microbiota and host gene expression. *Proceedings of the National Academy of Sciences, 111*(8), 3074-3079.

Rohart, F., Gautier, B., Singh, A., & Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS computational biology, 13*(11), e1005752.

Romano-Keeler, J., & Weitkamp, J.-H. (2015). Maternal influences on fetal microbial colonization and immune development. *Pediatric research, 77*(1), 189-195.

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.

Ruff, W. E., & Kriegel, M. A. (2015). Autoimmune host–microbiota interactions at barrier sites and beyond. *Trends in molecular medicine, 21*(4), 233-244.

Rugtveit, J., & Fagerhol, M. K. (2002). Age-dependent variations in fecal calprotectin concentrations in children. *Journal of pediatric gastroenterology and nutrition, 34*(3), 323.

Ruiz-Moyano, S., Totten, S. M., Garrido, D. A., Smilowitz, J. T., German, J. B., Lebrilla, C. B., & Mills, D. A. (2013). Variation in Consumption of Human Milk Oligosaccharides by Infant Gut-Associated Strains of Bifidobacterium breve. *Applied and environmental microbiology, 79*(19), 6040-6049.

Ruiz-Perez, D., Guan, H., Madhivanan, P., Mathee, K., & Narasimhan, G. (2020). So you think you can PLS-DA? *BMC Bioinformatics, 21*(1), 2.

Ruiz, L., Delgado, S., Ruas-Madiedo, P., Sánchez, B., & Margolles, A. (2017). Bifidobacteria and their molecular communication with the immune system. *Frontiers in microbiology, 8*, 2345.

Said, H. M., & Mohammed, Z. M. (2006). Intestinal absorption of water-soluble vitamins: an update. *Curr Opin Gastroenterol, 22*(2), 140-146.

Saraf, M. K., Piccolo, B. D., Bowlin, A. K., Mercer, K. E., LeRoith, T., Chintapalli, S. V., . . . Yeruva, L. (2017). Formula diet driven microbiota shifts tryptophan metabolism from serotonin to tryptamine in neonatal porcine colon. *Microbiome, 5*(1), 77.

Schack-Nielsen, L., Sørensen, T. I., Mortensen, E. L., & Michaelsen, K. F. (2010). Late introduction of complementary feeding, rather than duration of breastfeeding, may protect against adult overweight. *The American journal of clinical nutrition, 91*(3), 619-627.

Scheiwiller, J., Arrigoni, E., Brouns, F., & Amadò, R. (2006). Human Faecal Microbiota Develops the Ability to Degrade Type 3 Resistant Starch During Weaning. *Journal of pediatric gastroenterology and nutrition, 43*(5), 584-591.

Schimmel, P., Kleinjans, L., Bongers, R. S., Knol, J., & Belzer, C. (2021). Breast milk urea as a nitrogen source for urease positive Bifidobacterium infantis. *FEMS Microbiology Ecology, 97*(3).

Schlinzig, T., Johansson, S., Stephansson, O., Hammarström, L., Zetterström, R. H., von Döbeln, U., . . . Norman, M. (2017). Surge of immune cell formation at birth differs by mode of delivery and infant characteristics—A population-based cohort study. *PloS one, 12*(9), e0184748.

Schroeder, B. O., Birchenough, G. M., Ståhlman, M., Arike, L., Johansson, M. E., Hansson, G. C., & Bäckhed, F. (2018). Bifidobacteria or fiber protects against diet-induced microbiota-mediated colonic mucus deterioration. *Cell host & microbe, 23*(1), 27-40. e27.

Schwartz, C., Chabanet, C., Szleper, E., Feyen, V., Issanchou, S., & Nicklaus, S. (2017). Infant acceptance of primary tastes and fat emulsion: developmental changes and links with maternal and infant characteristics. *Chemical senses, 42*(7), 593-603.

Sela, D., Chapman, J., Adeuya, A., Kim, J., Chen, F., Whitehead, T., . . . German, J. (2008). The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. *Proceedings of the National Academy of Sciences, 105*(48), 18964-18969.

Sela, D. A., Li, Y., Lerno, L., Wu, S., Marcobal, A. M., German, J. B., . . . Mills, D. A. (2011). An infant-associated bacterial commensal utilizes breast milk sialyloligosaccharides. *Journal of Biological Chemistry, 286*(14), 11909-11918.

Sen, P., Mardinogulu, A., & Nielsen, J. (2017). Selection of complementary foods based on optimal nutritional values. *Scientific reports, 7*(1), 1-9.

Sevelsted, A., Stokholm, J., Bønnelykke, K., & Bisgaard, H. (2015). Cesarean Section and Chronic Immune Disorders. *Pediatrics, 135*(1), e92-e98.

Shan, M., Gentile, M., Yeiser, J. R., Walland, A. C., Bornstein, V. U., Chen, K., . . . Cols, M. (2013). Mucus enhances gut homeostasis and oral tolerance by delivering immunoregulatory signals. *Science, 342*(6157), 447-453.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res, 13*(11), 2498-2504.

Sharma, V., Rodionov, D. A., Leyn, S. A., Tran, D., Iablokov, S. N., Ding, H., . . . Peterson, S. N. (2019). B-Vitamin Sharing Promotes Stability of Gut Microbial Communities. *Frontiers in microbiology, 10*(1485).

Shi, N., Li, N., Duan, X., & Niu, H. (2017). Interaction between the gut microbiome and mucosal immune system. *Military Medical Research, 4*(1), 1-7.

Shipman Joseph, A., Berleman James, E., & Salyers Abigail, A. (2000). Characterization of Four Outer Membrane Proteins Involved in Binding Starch to the Cell Surface ofBacteroides thetaiotaomicron. *Journal of Bacteriology, 182*(19), 5365-5372.

Shipman Joseph, A., Cho Kyu, H., Siegel Hilary, A., & Salyers Abigail, A. (1999). Physiological Characterization of SusG, an Outer Membrane Protein Essential for Starch Utilization byBacteroides thetaiotaomicron. *Journal of Bacteriology, 181*(23), 7206-7211.

Shoaf, K., Mulvey, G. L., Armstrong, G. D., & Hutkins, R. W. (2006). Prebiotic galactooligosaccharides reduce adherence of enteropathogenic Escherichia coli to tissue culture cells. *Infection and immunity, 74*(12), 6920-6928.

Sibley, E. (2004). Carbohydrate intolerance. *Current opinion in gastroenterology, 20*(2), 162-167.

Sierra, C., Bernal, M.-J., Blasco, J., Martínez, R., Dalmau, J., Ortuno, I., . . . Vidal, M.-L. (2015). Prebiotic effect during the first year of life in healthy infants fed formula containing GOS as the only prebiotic: a multicentre, randomised, double-blind and placebo-controlled trial. *European Journal of Nutrition, 54*(1), 89-99.

Sivakumaran, S., Huffman, L., Gilmore, Z., & Sivakumaran, S. (2017). New Zealand FOODfiles 2016 manual. The New Zealand Institute for Plant & Food Research Limited and Ministry of Health.

Sjögren, Y. M., Tomicic, S., Lundberg, A., Böttcher, M. F., Björkstén, B., Sverremark-Ekström, E., & Jenmalm, M. C. (2009). Influence of early gut microbiota on the maturation of childhood mucosal and systemic immune responses: gut microbiota and immune responses. *Clinical & Experimental Allergy, 39*(12), 1842-1851.

Smilowitz, J. T., Lebrilla, C. B., Mills, D. A., German, J. B., & Freeman, S. L. (2014). Breast Milk Oligosaccharides: Structure-Function Relationships in the Neonate. *Annual review of nutrition, 34*(1), 143-169.

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical chemistry, 78*(3), 779-787.

Sohlenkamp, C., & Geiger, O. (2015). Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiology Reviews, 40*(1), 133-159.

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), e00021-00016.

Sonnenburg, Erica D., & Sonnenburg, Justin L. (2014). Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates. *Cell Metabolism, 20*(5), 779-786.

Steinert, R. E., Lee, Y.-K., & Sybesma, W. (2020). Vitamins for the gut microbiome. *Trends in molecular medicine, 26*(2), 137-140.

Stewart, C. J., Ajami, N. J., O'Brien, J. L., Hutchinson, D. S., Smith, D. P., Wong, M. C., . . . Metcalf, G. A. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature, 562*(7728), 583-588.

Stewart, M. L., Timm, D. A., & Slavin, J. L. (2008). Fructooligosaccharides exhibit more rapid fermentation than long-chain inulin in an in vitro fermentation system. *Nutrition Research, 28*(5), 329-334.

Su, M., Subbaraj, A. K., Fraser, K., Qi, X., Jia, H., Chen, W., . . . Young, W. (2019). Lipidomics of Brain Tissues in Rats Fed Human Milk from Chinese Mothers or Commercial Infant Formula. *Metabolites, 9*(11), 253.

Swennen, K., Courtin, C. M., & Delcour, J. A. (2006). Non-digestible oligosaccharides with prebiotic properties. *Critical reviews in food science and nutrition, 46*(6), 459-471.

Tamura, J., Kubota, K., Murakami, H., Sawamura, M., Matsushima, T., Tamura, T., . . . Naruse, T. (1999). Immunomodulation by vitamin B12: augmentation of CD8+ T lymphocytes and natural killer (NK) cell activity in vitamin B12-deficient patients by methyl-B12 treatment. *Clinical & Experimental Immunology, 116*(1), 28-32.

Tanaka, M., & Nakayama, J. (2017). Development of the gut microbiota in infancy and its impact on health in later life. *Allergology International, 66*(4), 515-522.

Tang, M., Frank, D. N., Hendricks, A. E., Ir, D., Esamai, F., Liechty, E., . . . Krebs, N. F. (2017). Iron in micronutrient powder promotes an unfavorable gut microbiota in Kenyan infants. *Nutrients, 9*(7), 776.

Tannock, G. W., & Ercolini, D. (2021). Building Robust Assemblages of Bacteria in the Human Gut in Early Life. *Applied and environmental microbiology, 87*(22), e01449-01421.

Tannock, G. W., Lawley, B., Munro, K., Pathmanathan, S. G., Zhou, S. J., Makrides, M., . . . Lowry, D. (2013). Comparison of the compositions of the stool microbiotas of infants fed goat milk formula, cow milk-based formula, or breast milk. *Applied and environmental microbiology, 79*(9), 3040-3048.

Taylor, B. C., Lejzerowicz, F., Poirel, M., Shaffer, J. P., Jiang, L., Aksenov, A., . . . Miller-Montgomery, S. (2020). Consumption of fermented foods is associated with systematic differences in the gut microbiome and metabolome. *Msystems, 5*(2), e00901-00919.

Tedelind, S., Westberg, F., Kjerrulf, M., & Vidal, A. (2007). Anti-inflammatory properties of the short-chain fatty acids acetate and propionate: a study with relevance to inflammatory bowel disease. *World journal of gastroenterology, 13*(20), 2826-2832.

Thompson, A. L., Monteagudo-Mera, A., Cadenas, M. B., Lampl, M. L., & Azcarate-Peril, M. A. (2015). Milk-and solid-feeding practices and daycare attendance are associated with differences in bacterial diversity, predominant communities, and metabolic and immune function of the infant gut microbiome. *Frontiers in cellular and infection microbiology, 5*, 3.

Thomson, P., Medina, D. A., & Garrido, D. (2018). Human milk oligosaccharides and infant gut bifidobacteria: Molecular strategies for their utilization. *Food Microbiology, 75*, 37-46.

Tramontano, M., Andrejev, S., Pruteanu, M., Klünemann, M., Kuhn, M., Galardini, M., . . . Patil, K. R. (2018). Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. *Nature Microbiology, 3*(4), 514-522.

Trehan, I., & Manary, M. J. (2015). Management of severe acute malnutrition in low-income and middle-income countries. *Archives of disease in childhood, 100*(3), 283-287.

Tremaroli, V., & Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature, 489*(7415), 242-249.

Triantis, V., Bode, L., & van Neerven, R. J. (2018). Immunological effects of human milk oligosaccharides. *Frontiers in pediatrics, 6*, 190.

Turroni, F., Milani, C., Duranti, S., Mancabelli, L., Mangifesta, M., Viappiani, A., . . . Ferrarini, A. (2016). Deciphering bifidobacterial-mediated metabolic interactions and their impact on gut microbiota by a multi-omics approach. *The ISME journal, 10*(7), 1656-1668.

Turroni, F., Serafini, F., Foroni, E., Duranti, S., Motherway, M. O. C., Taverniti, V., . . . Roversi, T. (2013). Role of sortase-dependent pili of Bifidobacterium bifidum PRL2010 in modulating bacterium–host interactions. *Proceedings of the National Academy of Sciences, 110*(27), 11151-11156.

Ueland, P. M., McCann, A., Midttun, Ø., & Ulvik, A. (2017). Inflammation, vitamin B6 and related pathways. *Molecular Aspects of Medicine, 53*, 10-27.

Van De Walle, J., Sergent, T., Piront, N., Toussaint, O., Schneider, Y.-J., & Larondelle, Y. (2010). Deoxynivalenol affects in vitro intestinal epithelial cell barrier integrity through inhibition of protein synthesis. *Toxicology and applied pharmacology, 245*(3), 291-298.

Van der Leek, A. P., Yanishevsky, Y., & Kozyrskyj, A. L. (2017). The kynurenine pathway as a novel link between allergy and the gut microbiome. *Frontiers in immunology, 8*, 1374.

Ventura, A. K. (2017). Does breastfeeding shape food preferences links to obesity. *Annals of Nutrition and Metabolism, 70*(Suppl. 3), 8-15.

Verster, A. J., & Borenstein, E. (2018). Competitive lottery-based assembly of selected clades in the human gut microbiome. *Microbiome, 6*(1), 1-17.

Victora, C. G., Bahl, R., Barros, A. J. D., França, G. V. A., Horton, S., Krasevec, J., . . . Rollins, N. C. (2016). Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *The Lancet, 387*(10017), 475-490.

Vieira, A. T., & Vinolo, M. A. R. (2019). Regulation of immune cell function by short chain fatty acids and their impact on arthritis. In *Bioactive Food as Dietary Interventions for Arthritis and Related Inflammatory Diseases* (pp. 175-188): Elsevier.

Vinolo, M. A. R., Rodrigues, H. G., Nachbar, R. T., & Curi, R. (2011). Regulation of Inflammation by Short Chain Fatty Acids. *Nutrients, 3*(10), 858-876.

Vogt, L., Ramasamy, U., Meyer, D., Pullens, G., Venema, K., Faas, M. M., . . . de Vos, P. (2013). Immune Modulation by Different Types of β2→1-Fructans Is Toll-Like Receptor Dependent. *PloS one, 8*(7), e68367.

Vuillermin, P. J., O'Hely, M., Collier, F., Allen, K. J., Tang, M. L. K., Harrison, L. C., . . . the, B. I. S. I. G. (2020). Maternal carriage of Prevotella during pregnancy associates with protection against food allergy in the offspring. *Nature communications, 11*(1), 1452.

Wada, J., Ando, T., Kiyohara, M., Ashida, H., Kitaoka, M., Yamaguchi, M., . . . Yamamoto, K. (2008). Bifidobacterium bifidum Lacto-N-Biosidase, a Critical Enzyme for the Degradation of Human Milk Oligosaccharides with a Type 1 Structure. *Applied and environmental microbiology, 74*(13), 3996-4004.

Walker, A. (2009). Probiotics stick it to the man. *Nature Reviews Microbiology, 7*(12), 843-843.

Walker, W. A. (2013). Initial intestinal colonization in the human infant and immune homeostasis. *Annals of Nutrition and Metabolism, 63*(Suppl. 2), 8-15.

Wang, M., Monaco, M. H., & Donovan, S. M. (2016). *Impact of early gut microbiota on immune and metabolic development and function.* Paper presented at the Seminars in Fetal and Neonatal Medicine.

Wang, T., Goyal, A., Dubinkina, V., & Maslov, S. (2019). Evidence for a multi-level trophic organization of the human gut microbiome. *PLoS computational biology, 15*(12), e1007524.

Wang, W., Hu, H., Zijlstra, R. T., Zheng, J., & Gänzle, M. G. (2019). Metagenomic reconstructions of gut microbial metabolism in weanling pigs. *Microbiome, 7*(1), 48.

Wang, Y., Mortimer, E. K., Katundu, K. G. H., Kalanga, N., Leong, L. E. X., Gopalsamy, G. L., . . . Rogers, G. B. (2019). The Capacity of the Fecal Microbiota From Malawian Infants to Ferment Resistant Starch. *Frontiers in microbiology, 10*.

Wang, Z., Zolnik, C. P., Qiu, Y., Usyk, M., Wang, T., Strickler, H. D., . . . Qi, Q. (2018). Comparison of fecal collection methods for microbiome and metabolomics studies. *Frontiers in cellular and infection microbiology*, 301.
Ward, R. E., Benninghoff, A. D., & Hintze, K. J. (2020). Food matrix and the microbiome: Considerations for pre-clinical chronic disease studies. *Nutrition Research*.
Warren, F. J., Fukuma, N. M., Mikkelsen, D., Flanagan, B. M., Williams, B. A., Lisle, A. T., . . . Gidley, M. J. (2018). Food starch structure impacts gut microbiome composition. *Msphere, 3*(3).
Wastyk, H. C., Fragiadakis, G. K., Perelman, D., Dahan, D., Merrill, B. D., Feiqiao, B. Y., . . . Han, S. (2021). Gut-microbiota-targeted diets modulate human immune status. *Cell, 184*(16), 4137-4153. e4114.
Whitehead, N. A., Barnard, A. M., Slater, H., Simpson, N. J., & Salmond, G. P. (2001). Quorum-sensing in Gram-negative bacteria. *FEMS Microbiol Rev, 25*(4), 365-404.
WHO Multicentre Growth Reference Study Group. (2006). WHO Child Growth Standards based on length/height, weight and age. Acta Paediatrica (Oslo, Norway: 1992). Supplement, 450, 76-85.
Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., . . . Scalbert, A. (2018). HMDB 4.0: the human metabolome database for 2018. *Nucleic acids research, 46*(D1), D608-D617.
Wong, J. M. W., de Souza, R., Kendall, C. W. C., Emam, A., & Jenkins, D. J. A. (2006). Colonic Health: Fermentation and Short Chain Fatty Acids. *Journal of Clinical Gastroenterology, 40*(3).
Wormser, G. P., & Frank, M. (2005). Medication Adherence in HIV/AIDS Edited by Jeffrey Laurence Larchmont, NY: Mary Ann Liebert, 2004. 396 pp., illustrated. $99.00 (cloth). *Clinical Infectious Diseases, 41*(1), 132-132.
Wu, J., Liu, M., Zhou, M., Wu, L., Yang, H., Huang, L., & Chen, C. (2021). Isolation and genomic characterization of five novel strains of Erysipelotrichaceae from commercial pigs. *BMC Microbiology, 21*(1), 1-12.
Wu, T.-C., & Chen, P.-H. (2009). Health consequences of nutrition in childhood and early infancy. *Pediatrics & neonatology, 50*(4), 135-142.
Yan, J., & Charles, J. F. (2017). Gut microbiome and bone: to build, destroy, or both? *Current osteoporosis reports, 15*(4), 376-384.
Yassour, M., Vatanen, T., Siljander, H., Hämäläinen, A.-M., Härkönen, T., Ryhänen, S. J., . . . Xavier, R. J. (2016). Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Science translational medicine, 8*(343), 343ra381-343ra381.
Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., . . . Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature, 486*(7402), 222-227.
Young, B. E. (2017). Chapter 2 - Breastfeeding and Human Milk: Short and Long-Term Health Benefits to the Recipient Infant. In J. M. Saavedra & A. M. Dattilo (Eds.), *Early Nutrition and Long-Term Health* (pp. 25-53): Woodhead Publishing.
Young, B. E., & Krebs, N. F. (2013). Complementary feeding: critical considerations to optimize growth, nutrition, and feeding behavior. *Current pediatrics reports, 1*(4), 247-256.
Yuan, C., Gaskins, A. J., Blaine, A. I., Zhang, C., Gillman, M. W., Missmer, S. A., . . . Chavarro, J. E. (2016). Association Between Cesarean Birth and Risk of Obesity in Offspring in Childhood, Adolescence, and Early Adulthood. *JAMA Pediatrics, 170*(11), e162385-e162385.
Zealand, N. (2008). *Food and Nutrition Guidelines for Healthy Infants and Toddlers (Aged 0-2): Background Paper*. (0478317360). Ministry of Health
Zealand, N. (2012). Food and nutrition guidelines for healthy infants and toddlers (aged 0-2): a background paper.
Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics, 30*(5), 614-620.
Zmora, N., Zilberman-Schapira, G., Suez, J., Mor, U., Dori-Bachash, M., Bashiardes, S., . . . Brik, R. B.-Z. (2018). Personalized gut mucosal colonization resistance to empiric probiotics is associated with unique host and microbiome features. *Cell, 174*(6), 1388-1405. e1321.

Appendices

Appendix 1 *Taxa included in analyses with mean relative abundance at 4, 9 and 12 months of age, as well as Log2 fold change (Log2FC) between time points, permanova p values with FDR correction and associated significance (FDR p <0.05) denoted with *.*

Taxa	4 month mean %	9 month mean %	12 month mean %	4 to 9 month Log2FC	4 to 9 month Permanova fdr p value	Sig	9 to 12 month Log2FC	4 to 9 month permanova fdr p value	Sig
Acidaminococcaceae	0.0038%	0.0105%	0.0083%	1.4615	0.3471		-0.3349	0.3288	
Acidaminococcaceae Phascolarctobacterium	0.0321%	0.0659%	0.0471%	1.0389	0.8020		-0.4845	0.7944	
Actinomycetaceae Actinomyces	0.0672%	0.0763%	0.0266%	0.1841	0.4141		-1.5208	0.3974	
Actinomycetaceae Varibaculum	0.0215%	0.0100%	0.0035%	-1.1001	0.3538		-1.5019	0.3573	
Alcaligenaceae	0.0010%	0.0071%	0.0151%	2.7708	0.1063		1.0919	0.1278	
Alcaligenaceae Parasutterella	0.0229%	0.0520%	0.3051%	1.1819	0.0499	*	2.5521	0.0580	
Alcaligenaceae Sutterella	0.0005%	0.1260%	0.0509%	8.0999	0.2356		-1.3088	0.2512	
Archaea	0.0056%	0.0023%	0.0114%	-1.2917	0.5192		2.3185	0.5236	
Bacillaceae	0.0005%	0.0001%	0.0189%	-2.6849	0.0034	*	7.9290	0.0053	*
Bacillaceae Bacillus	0.0055%	0.0036%	0.2023%	-0.5849	0.0076	*	5.7946	0.0053	*
Bacillales	0.0047%	0.0105%	0.0588%	1.1585	0.0076	*	2.4865	0.0053	*
Bacteroidaceae	0.0313%	0.0237%	0.0119%	-0.4019	0.4804		-0.9987	0.4638	
Bacteroidaceae Bacteroides	5.4570%	8.7384%	6.1525%	0.6793	0.6405		-0.5062	0.6336	
Bacteroidales	1.1279%	2.3327%	1.6021%	1.0483	0.6087		-0.5420	0.6057	
Betaproteobacteria	0.0035%	0.0080%	0.0115%	1.2051	0.1207		0.5312	0.1075	
Bifidobacteriaceae	36.8184%	30.6825%	26.7852%	-0.2630	0.5026		-0.1960	0.5114	
Bifidobacteriaceae Bifidobacterium	1.5812%	1.0421%	0.9381%	-0.6016	0.3140		-0.1516	0.3237	
Bifidobacteriaceae Bifidobacterium breve	0.1246%	0.0409%	0.0254%	-1.6078	0.1796		-0.6880	0.1671	
Bifidobacteriaceae Bifidobacterium longum	0.5491%	0.3305%	0.1896%	-0.7323	0.0424	*	-0.8017	0.0362	*
Bifidobacteriales	0.0171%	0.0088%	0.0027%	-0.9550	0.0120	*	-1.6850	0.0070	*
Burkholderiales	0.0044%	0.0066%	0.0054%	0.5977	0.7263		-0.3122	0.7210	
Christensenellaceae Christensenellaceae	0.0022%	0.0008%	0.0953%	-1.4994	0.0034	*	6.9138	0.0053	*
Clostridiaceae Clostridium sensu stricto	0.0267%	0.0004%	0.3038%	-2.3990	0.0093	*	-0.9623	0.0112	*
Clostridiaceae Sarcina	8.2987%	1.5734%	0.8076%	-6.2229	0.5063		9.7336	0.5236	
Clostridiaceae Clostridium neonatale	0.0222%	0.0002%	0.0000%	-6.6231	0.0034	*	-27.7470	0.0034	*
Clostridiaceae Unclassified	0.6958%	0.0871%	0.4025%	-2.9978	0.3602		2.2082	0.3573	
Clostridiales	1.2236%	4.0759%	6.1331%	1.7360	0.0034	*	0.5895	0.0034	*
Clostridiales Family XI Anaerococcus	0.0320%	0.0017%	0.0020%	-4.2405	0.5840		0.2157	0.5631	
Coriobacteriaceae	0.1520%	0.2887%	0.3235%	0.9260	0.6312		0.1639	0.6336	
Coriobacteriaceae Atopobium	0.0176%	0.0010%	0.0008%	-4.1246	0.2230		-0.2524	0.2159	
Coriobacteriaceae Collinsella	1.2882%	0.9326%	2.4260%	-0.4660	0.4804		1.3793	0.4924	
Coriobacteriaceae Eggerthella	0.0331%	0.1599%	0.1201%	2.2735	0.1796		-0.4133	0.1671	
Desulfovibrionaceae Bilophila	0.0032%	0.0162%	0.0103%	2.3525	0.5442		-0.6454	0.5367	
Enterobacteriaceae	16.4042%	3.8237%	1.8650%	-2.1010	0.0034	*	-1.0357	0.0034	*
Enterobacteriaceae Citrobacter	0.0217%	0.0139%	0.0064%	-0.6422	0.1950		-1.1092	0.1900	
Enterobacteriaceae Enterobacter	0.0155%	0.0034%	0.0146%	-2.2022	0.6405		2.1176	0.6337	
Enterobacteriaceae bacterium	7.2147%	2.1006%	1.0923%	-1.7728	0.0034	*	-1.1715	0.0083	*
Enterobacteriaceae Escherichia-Shigella	0.5785%	0.0587%	0.0221%	-1.7801	0.0034	*	-0.9435	0.0034	*
Enterobacteriaceae Klebsiella	0.0129%	0.0038%	0.0017%	-3.2998	0.2271		-1.4113	0.2375	
Enterobacteriaceae Salmonella enterica Paratyphi A	0.0850%	0.0202%	0.0060%	-2.0745	0.0034	*	-1.7520	0.0034	*
Enterobacteriales	0.0461%	0.0041%	0.0034%	-3.4750	0.0034	*	-0.2763	0.0034	*
Enterococcaceae	0.2747%	0.2852%	0.1806%	0.0541	0.9860		-0.6589	0.9799	
Enterococcaceae Enterococcus	0.4118%	0.3675%	0.2397%	-0.1642	0.9017		-0.6167	0.9065	
Erysipelotrichaceae	0.4773%	0.5348%	0.5040%	0.1639	1.0000		-0.0856	1.0000	
Erysipelotrichaceae Coprobacillus	0.0000%	0.0109%	0.0149%	33.3396	0.2697		0.4517	0.2630	
Erysipelotrichaceae Erysipelatoclostridium	2.6876%	0.4544%	0.4396%	-2.5643	0.3818		-0.0477	0.3846	
Erysipelotrichaceae Erysipelotrichaceae	0.0035%	0.0022%	0.1512%	-0.6710	0.0034	*	6.1158	0.0034	*
Erysipelotrichaceae Faecalitalea	0.0224%	0.0034%	0.0102%	-2.7086	1.0000		1.5679	1.0000	
Erysipelotrichaceae Holdemanella	0.0005%	0.0004%	0.0626%	-0.3585	0.0107	*	7.3847	0.0053	*
Erysipelotrichaceae Turicibacter	0.0002%	0.0026%	0.0156%	3.8577	0.0370	*	2.5619	0.0438	*
Firmicutes Unclassified	0.2814%	0.9073%	0.7906%	1.6892	0.0397	*	-0.1987	0.0211	*
Fusobacteriaceae Fusobacterium	0.0130%	0.0445%	0.0714%	1.7787	0.5063		0.6830	0.5114	
Gammaproteobacteria Unclassified	0.8406%	0.1836%	0.1291%	-2.1944	0.0034	*	-0.5084	0.0034	*
Lachnospiraceae	1.8861%	11.0013%	13.7104%	2.5442	0.0034	*	0.3176	0.0034	*
Lachnospiraceae Anaerostipes	0.0039%	0.0204%	0.0164%	7.3603	0.0580		-0.3030	0.0632	
Lachnospiraceae Blautia	0.0000%	0.0256%	0.0110%	2.4293	0.0093	*	0.7027	0.0034	*
Lachnospiraceae Coprococcus	0.0041%	0.6748%	0.5470%	1.6946	0.0034	*	3.1244	0.0034	*
Lachnospiraceae Dorea	0.5486%	2.9549%	4.8092%	7.9527	0.3918		-0.5741	0.4079	
Lachnospiraceae Eubacterium rectale	0.0022%	0.0072%	0.0627%	2.4002	0.2272		-0.3121	0.2222	
Lachnospiraceae Fusicatenibacter	0.0008%	0.2035%	0.1367%	4.8055	0.2126		0.6406	0.2211	
Lachnospiraceae Hungatella	0.0034%	0.0952%	0.1484%	1.8439	0.1947		-0.9112	0.1828	
Lachnospiraceae Lachnoclostridium	0.0213%	0.0763%	0.0406%	0.4701	0.6120		0.4872	0.6057	
Lachnospiraceae Lachnospira	0.2689%	0.3725%	0.5221%	8.2041	0.8020		-1.4999	0.8116	
Lachnospiraceae Lachnospiraceae	0.0027%	0.7969%	0.2818%	2.3385	0.0034	*	2.1493	0.0034	*
Lachnospiraceae Roseburia	0.0084%	0.0426%	0.1891%	7.4194	0.0424	*	0.7754	0.0329	*
Lachnospiraceae Ruminococcus gauvreauii	0.0023%	0.3945%	0.6753%	34.5779	0.6259		-1.2172	0.6057	
Lachnospiraceae Ruminococcus gnavus	0.0015%	0.0029%	0.0139%	0.3362	0.9809		-0.6246	0.9799	
Lachnospiraceae Sellimonas	0.0026%	0.0227%	0.0800%	0.9639	0.0334	*	2.2453	0.0265	*
Lachnospiraceae Tyzzerella	0.0144%	0.0182%	0.0118%	3.1253	0.0135	*	1.8172	0.0083	*
Lactobacillaceae	0.0016%	0.0319%	0.0059%	4.3464	0.2697		-2.4252	0.2630	

Taxa	4 month mean %	9 month mean %	12 month mean %	4 to 9 month Log2FC	4 to 9 month Permanova fdr p value	Sig	9 to 12 month Log2FC	4 to 9 month permanova fdr p value	Sig
Lactobacillaceae Lactobacillus	0.3914%	1.9382%	1.0520%	2.3078	0.3538		-0.8815	0.3448	
Lactobacillaceae Lactobacillus rhamnosus	0.0045%	0.0154%	0.0013%	1.7712	0.0819		-3.5513	0.0819	
Lactobacillales	0.1751%	0.2601%	0.1748%	0.5708	0.6745		-0.5732	0.6650	
Micrococcaceae Rothia	0.0469%	0.0167%	0.0171%	-1.4930	0.0370	*	0.0396	0.0336	*
Moraxellaceae	0.0079%	0.0017%	0.0189%	-2.1841	0.0974		3.4513	0.0974	
Moraxellaceae Acinetobacter	0.0006%	0.0005%	0.3753%	-0.4081	0.0076	*	9.6612	0.0100	*
Negativicutes	0.0025%	0.0090%	0.0090%	1.8366	0.1947		0.0051	0.1908	
Pasteurellaceae	0.0114%	0.0911%	0.0490%	2.9989	0.1950		-0.8928	0.1957	
Pasteurellaceae Haemophilus	0.0242%	0.2433%	0.1323%	3.3310	0.2199		-0.8784	0.1900	
Peptostreptococcaceae	0.0571%	0.3842%	0.3655%	2.7514	0.0416	*	-0.0718	0.0357	*
Peptostreptococcaceae Intestinibacter	0.0023%	0.0636%	0.0707%	4.7760	0.0076	*	0.1532	0.0034	*
Peptostreptococcaceae Peptoclostridium	0.0179%	0.1579%	0.1176%	3.1388	0.0254	*	-0.4256	0.0112	*
Porphyromonadaceae Barnesiella	0.0005%	0.0052%	0.0152%	3.4959	0.2126		1.5567	0.2159	
Porphyromonadaceae Parabacteroides	0.2242%	0.5905%	0.2460%	1.3969	0.7753		-1.2635	0.7944	
Prevotellaceae	0.0029%	0.0524%	0.0318%	4.1959	0.6394		-0.7212	0.6530	
Prevotellaceae Prevotella	0.0130%	0.5417%	0.4308%	5.3793	0.6745		-0.3305	0.6563	
Pseudomonadaceae Pseudomonas	0.0014%	0.0004%	0.1248%	-1.7064	0.0376	*	8.1666	0.0239	*
Rikenellaceae Alistipes	0.0040%	0.1023%	0.0856%	4.6916	0.2733		-0.2570	0.2630	
Ruminococcaceae	0.1751%	0.7676%	2.0576%	2.1323	0.0034	*	1.4226	0.0034	*
Ruminococcaceae Butyricicoccus	0.0355%	0.0062%	0.3053%	-2.5212	0.0064	*	5.6235	0.0034	*
Ruminococcaceae Faecalibacterium	0.0110%	1.1116%	3.6189%	6.6572	0.0034	*	1.7028	0.0034	*
Ruminococcaceae Flavonifractor	0.0477%	0.0710%	0.1081%	0.5746	0.3454		0.6057	0.3448	
Ruminococcaceae Ruminiclostridium	0.0102%	0.0501%	0.0990%	2.2917	0.2272		0.9813	0.2375	
Ruminococcaceae Ruminococcaceae	0.0069%	0.0252%	0.1599%	1.8634	0.0150	*	2.6658	0.0070	*
Ruminococcaceae Ruminococcus	0.0019%	0.1035%	0.4667%	5.7312	0.0120	*	2.1728	0.0163	*
Ruminococcaceae Subdoligranulum	0.0011%	0.3104%	0.9611%	8.1873	0.0107	*	1.6305	0.0070	*
Selenomonadales	0.0138%	0.0642%	0.0250%	2.2210	0.2667		-1.3602	0.2630	
Staphylococcaceae Staphylococcus	0.0353%	0.0199%	0.1333%	-0.8276	0.0819		2.7459	0.0760	
Streptococcaceae	0.0056%	0.0139%	0.0150%	1.3124	0.5840		0.1060	0.5876	
Streptococcaceae Lactococcus	0.0007%	0.3313%	0.0399%	8.8431	0.9550		-3.0553	0.9601	
Streptococcaceae Streptococcus	0.7379%	1.3085%	1.8564%	0.8264	0.4804		0.5046	0.4235	
Streptomycetaceae Streptomyces	0.1083%	0.0975%	0.0857%	-0.1517	0.7260		-0.1851	0.7405	
Unclassified Actinobacteria	0.9834%	0.7191%	0.6182%	-0.4517	0.2697		-0.2181	0.2667	
Unclassified Bacilli	0.0284%	0.0479%	0.0533%	0.7579	0.3602		0.1514	0.3714	
Unclassified Bacteria	1.4164%	1.8471%	2.0130%	0.3830	0.2939		0.1241	0.2866	
Unclassified Bacteroidetes	0.0180%	0.0501%	0.0379%	1.4753	0.5154		-0.4004	0.5066	
Unclassified Clostridia	0.0107%	0.0582%	0.0768%	2.4373	0.0034	*	0.3999	0.0034	*
Unclassified Proteobacteria	0.0418%	0.0144%	0.0258%	-1.5324	0.0469	*	0.8351	0.0552	
Unknown	0.1270%	0.0676%	0.1747%	-0.9107	0.3650		1.3698	0.3573	
Veillonellaceae	0.1729%	1.3633%	1.0428%	2.9788	0.0076	*	-0.3867	0.0083	*
Veillonellaceae Dialister	0.0019%	0.1673%	0.3750%	6.4474	0.2697		1.1644	0.2630	
Veillonellaceae Megamonas	0.0466%	3.0303%	0.4491%	6.0236	0.3575		-2.7542	0.3573	
Veillonellaceae Megasphaera	0.0196%	0.4455%	1.0922%	4.5031	0.2491		1.2937	0.2630	
Veillonellaceae Veillonella	3.5527%	6.0813%	3.9140%	0.7755	0.3471		-0.6357	0.3448	
Verrucomicrobiaceae Akkermansia	1.0222%	0.4354%	0.9406%	-1.2312	0.9550		1.1111	0.9601	
Vibrionaceae Vibrio	0.0208%	0.0087%	0.0077%	-1.2573	0.0370	*	-0.1691	0.0322	*

4324

Appendix 2 *KEGG pathways included in analyses with mean relative abundance at 4, 9 and 12 months of age, as well as Log2 fold change (Log2FC) between time points, permanova p values with FDR correction and associated significance (FDR p <0.05) denoted with *.*

KEGG pathway	4 month mean %	9 month mean %	12 month mean %	4 to 9 month Log2FC	4 to 9 month Permanova fdr p value	Sig	9 to 12 month Log2FC	4 to 9 month permanova fdr p value	Sig
ko03000.Transcription.factors.	1.5062%	1.1316%	1.1150%	-0.4125	0.0017	*	-0.0213	0.0018	*
ko03009.Ribosome.biogenesis.	1.2119%	1.2624%	1.3314%	0.0589	0.0017	*	0.0768	0.0018	*
ko03011.Ribosome.	0.8666%	0.9639%	0.9816%	0.1535	0.0244	*	0.0263	0.0330	*
ko03012.Translation.factors.	0.5309%	0.6043%	0.6222%	0.1866	0.0017	*	0.0423	0.0018	*
ko03016.Transfer.RNA.biogenesis.	2.3256%	2.5929%	2.6611%	0.1570	0.0017	*	0.0375	0.0018	*
ko03019.Messenger.RNA.biogenesis.	0.7728%	0.8344%	0.8673%	0.1107	0.0017	*	0.0558	0.0018	*
ko03021.Transcription.machinery.	0.6830%	0.7984%	0.8172%	0.2252	0.0017	*	0.0337	0.0018	*
ko03029.Mitochondrial.biogenesis.	1.1648%	1.3270%	1.3724%	0.1881	0.0017	*	0.0485	0.0018	*
ko03032.DNA.replication.proteins.	1.3743%	1.5738%	1.6622%	0.1955	0.0017	*	0.0788	0.0018	*
ko03036.Chromosome.and.associated.proteins.	1.2516%	1.3303%	1.3509%	0.0880	0.0017	*	0.0222	0.0018	*
ko03051.Proteasome.	0.1056%	0.1108%	0.1136%	0.0699	0.8874		0.0360	0.8766	
ko03110.Chaperones.and.folding.catalysts.	0.9103%	0.9819%	0.9986%	0.1092	0.0172	*	0.0244	0.0113	*
ko03400.DNA.repair.and.recombination.proteins.	3.5172%	3.9165%	4.0279%	0.1551	0.0044	*	0.0405	0.0031	*
ko04131.Membrane.trafficking.	0.3421%	0.3661%	0.3472%	0.0977	0.7528		-0.0766	0.7594	
ko00194.Photosynthesis.proteins.	0.2507%	0.3088%	0.3377%	0.3009	0.0017	*	0.1291	0.0018	*
ko01001.Protein.kinases.	0.6539%	0.5001%	0.4998%	-0.3868	0.0017	*	-0.0007	0.0018	*
ko01002.Peptidases.	1.9755%	2.0491%	2.0921%	0.0528	0.0105	*	0.0300	0.0066	*
ko01003.Glycosyltransferases.	0.3685%	0.3028%	0.2809%	-0.2832	0.0164	*	-0.1082	0.0184	*
ko01004.Lipid.biosynthesis.proteins.	0.6646%	0.6475%	0.6514%	-0.0376	0.9053		0.0087	0.9117	
ko01005.Lipopolysaccharide.biosynthesis.prote	0.4336%	0.3455%	0.2934%	-0.3277	0.0282	*	-0.2355	0.0277	*
ko01006.Prenyltransferases.	0.1415%	0.1426%	0.1392%	0.0104	0.8632		-0.0347	0.8588	
ko01007.Amino.acid.related.enzymes.	1.7511%	1.9556%	1.9740%	0.1594	0.0017	*	0.0135	0.0018	*
ko01008.Polyketide.biosynthesis.proteins.	0.0957%	0.0450%	0.0279%	-1.0870	0.0055	*	-0.6911	0.0124	*
ko01009.Protein.phosphatases.&.associated.proteins.	0.1239%	0.1569%	0.1560%	0.3412	0.0017	*	-0.0084	0.0018	*
ko01011.Peptidoglycan.biosynthesis.&.degradation.proteins.	1.1544%	1.2086%	1.2632%	0.0662	0.0017	*	0.0637	0.0018	*
ko00536.Glycosaminoglycan.binding.proteins.	0.0447%	0.0374%	0.0342%	-0.2581	0.8097		-0.1295	0.8021	
ko00537.Glycosylphosphatidylinositol..GPI	0.0742%	0.0688%	0.0695%	-0.1088	0.9359		0.0133	0.9323	
ko01504.Antimicrobial.resistance.genes.	0.3563%	0.2812%	0.2605%	-0.3415	0.0105	*	-0.1100	0.0102	*
ko02000.Transporters.	11.1944%	9.9927%	9.5069%	-0.1638	0.0017	*	-0.0719	0.0018	*
ko02022.Two.component.system.	0.6974%	0.5363%	0.5626%	-0.3790	0.0390	*	0.0690	0.0387	*
ko02035.Bacterial.motility.proteins.	0.5028%	0.3026%	0.2898%	-0.7324	0.0153	*	-0.0624	0.0066	*
ko02042.Bacterial.toxins.	0.1848%	0.1828%	0.1593%	-0.0161	0.7518		-0.1987	0.7594	
ko02044.Secretion.system.	1.3521%	0.8866%	0.8500%	-0.6088	0.0017	*	-0.0608	0.0018	*
ko02048.Prokaryotic.defense.system.	0.8460%	0.8667%	0.9287%	0.0350	0.1375		0.0995	0.1215	
ko04090.CD.molecules.	0.0697%	0.0746%	0.0707%	0.0983	0.9226		-0.0779	0.9236	
ko04147.Exosome.	1.7135%	1.8888%	1.8852%	0.1405	0.0379	*	-0.0027	0.0422	*
ko04812.Cytoskeleton.proteins.	0.2512%	0.3221%	0.3363%	0.3584	0.0017	*	0.0622	0.0018	*
ko04112.Cell.cycle...Caulobacter.	0.3987%	0.4769%	0.4986%	0.2583	0.0017	*	0.0640	0.0018	*
ko04214.Apoptosis...fly.	0.0237%	0.0224%	0.0216%	-0.0762	0.6831		-0.0554	0.6656	
ko04216.Ferroptosis.	0.1704%	0.1732%	0.1694%	0.0228	0.9885		-0.0321	0.9865	
ko04217.Necroptosis.	0.2549%	0.2984%	0.3192%	0.2274	0.0017	*	0.0972	0.0018	*
ko02030.Bacterial.chemotaxis.	0.1625%	0.1188%	0.1214%	-0.4520	0.3529		0.0319	0.3627	
ko02040.Flagellar.assembly.	0.1867%	0.1059%	0.0929%	-0.8176	0.0616		-0.1888	0.0592	*
ko02024.Quorum.sensing.	1.8498%	1.8060%	1.7838%	-0.0346	0.9133		-0.0179	0.9117	
ko02025.Biofilm.formation...Pseudomonas.	0.2547%	0.1518%	0.1396%	-0.7464	0.0017	*	-0.1211	0.0018	*
ko02026.Biofilm.formation...Escherichia.coli.	0.5061%	0.3459%	0.3416%	-0.5489	0.0017	*	-0.0182	0.0031	*
ko05111.Biofilm.formation...Vibrio.cholerae.	0.3273%	0.2140%	0.1738%	-0.6131	0.0017	*	-0.3003	0.0018	*
ko04138.Autophagy...yeast.	0.0305%	0.0300%	0.0326%	-0.0212	0.9133		0.1161	0.9117	
ko04142.Lysosome.	0.1718%	0.2136%	0.1900%	0.3140	0.6016		-0.1685	0.6283	
ko04146.Peroxisome.	0.2546%	0.2552%	0.2483%	0.0033	0.9638		-0.0397	0.9658	
ko02010.ABC.transporters.	2.7671%	2.5299%	2.5805%	-0.1293	0.2796		0.0286	0.2457	
ko02060.Phosphotransferase.system..PTS..	0.7202%	0.5493%	0.4870%	-0.3909	0.0264	*	-0.1737	0.0278	*
ko03070.Bacterial.secretion.system.	0.6128%	0.5389%	0.5446%	-0.1855	0.1158		0.0154	0.1117	
ko02020.Two.component.system.	1.7401%	1.3607%	1.3354%	-0.3549	0.0207	*	-0.0270	0.0264	*
ko04016.MAPK.signaling.pathway...plant.	0.1084%	0.0899%	0.0867%	-0.2700	0.0368	*	-0.0521	0.0417	*
ko04066.HIF.1.signaling.pathway.	0.1886%	0.2280%	0.2231%	0.2742	0.0017	*	-0.0319	0.0018	*
ko04068.FoxO.signaling.pathway.	0.0246%	0.0181%	0.0153%	-0.4433	0.0701		-0.2437	0.0724	
ko04070.Phosphatidylinositol.signaling.system.	0.0593%	0.0569%	0.0561%	-0.0581	0.7518		-0.0214	0.7594	
ko04151.PI3K.Akt.signaling.pathway.	0.0194%	0.0336%	0.0377%	0.7881	0.0017	*	0.1659	0.0031	*
ko04152.AMPK.signaling.pathway.	0.0290%	0.0328%	0.0314%	0.1774	0.8542		-0.0658	0.8575	
ko03018.RNA.degradation.	0.7165%	0.7993%	0.8315%	0.1578	0.0017	*	0.0569	0.0018	*
ko03050.Proteasome.	0.0303%	0.0276%	0.0272%	-0.1384	0.9053		-0.0179	0.8898	
ko03060.Protein.export.	0.4146%	0.4707%	0.4769%	0.1830	0.0017	*	0.0191	0.0018	*
ko04122.Sulfur.relay.system.	0.1249%	0.1380%	0.1584%	0.1438	0.0523		0.1988	0.0568	*
ko04141.Protein.processing.in.ER	0.0374%	0.0504%	0.0494%	0.4282	0.0032	*	-0.0279	0.0066	*
ko03030.DNA.replication.	0.7552%	0.8334%	0.8610%	0.1422	0.0257	*	0.0470	0.0278	*
ko03410.Base.excision.repair.	0.3547%	0.3705%	0.3779%	0.0629	0.0792		0.0284	0.0528	*
ko03420.Nucleotide.excision.repair.	0.7288%	0.8035%	0.8298%	0.1408	0.1355		0.0464	0.1188	
ko03430.Mismatch.repair.	0.8607%	0.9817%	1.0198%	0.1898	0.0017	*	0.0549	0.0018	*

KEGG pathway	4 month mean %	9 month mean %	12 month mean %	4 to 9 month Log2FC	4 to 9 month Permanova fdr p value	Sig	9 to 12 month Log2FC	4 to 9 month permanova fdr p value	Sig
ko03440.Homologous.recombination.	1.0163%	1.1278%	1.1747%	0.1501	0.0044	*	0.0588	0.0031	*
ko03020.RNA.polymerase.	0.3370%	0.3902%	0.3919%	0.2114	0.0017	*	0.0063	0.0018	*
ko00970.Aminoacyl.tRNA.biosynthesis.	1.3881%	1.5944%	1.6343%	0.1999	0.0017	*	0.0357	0.0018	*
ko03008.Ribosome.biogenesis.in.eukaryotes.	0.0527%	0.0499%	0.0516%	-0.0813	0.9185		0.0508	0.9190	
ko03010.Ribosome.	0.8666%	0.9639%	0.9816%	0.1536	0.0249	*	0.0263	0.0278	*
ko03013.RNA.transport.	0.0246%	0.0359%	0.0424%	0.5462	0.0189	*	0.2397	0.0277	*
ko00220.Arginine.biosynthesis.	0.5794%	0.6291%	0.6528%	0.1186	0.0017	*	0.0535	0.0018	*
ko00250.Alanine..aspartate.&.glutamate.metabolism.	1.1533%	1.2817%	1.2924%	0.1523	0.0017	*	0.0121	0.0018	*
ko00260.Glycine..serine.&.threonine.metabolism	0.9838%	0.9752%	0.9561%	-0.0127	0.4790		-0.0285	0.5119	
ko00270.Cysteine.and.methionine.metabolism.	1.1260%	1.1864%	1.1933%	0.0754	0.0844		0.0084	0.0911	
ko00280.Valine..leucine.&.isoleucine.degradation.	0.2825%	0.2708%	0.2385%	-0.0612	0.0032	*	-0.1828	0.0018	*
ko00290.Valine..leucine.&.isoleucine.biosynthesis.	0.5832%	0.6150%	0.6309%	0.0766	0.2542		0.0369	0.2531	
ko00300.Lysine.biosynthesis.	0.5797%	0.6521%	0.6819%	0.1698	0.0017	*	0.0644	0.0044	*
ko00310.Lysine.degradation.	0.2402%	0.1689%	0.1465%	-0.5079	0.0017	*	-0.2061	0.0018	*
ko00330.Arginine.and.proline.metabolism.	0.3982%	0.3571%	0.3511%	-0.1573	0.0390	*	-0.0243	0.0285	*
ko00340.Histidine.metabolism.	0.2647%	0.3116%	0.3057%	0.2355	0.0044	*	-0.0275	0.0056	*
ko00350.Tyrosine.metabolism.	0.2456%	0.2180%	0.2156%	-0.1719	0.0368	*	-0.0160	0.0293	*
ko00360.Phenylalanine.metabolism.	0.1885%	0.1689%	0.1611%	-0.1589	0.3527		-0.0681	0.3928	
ko00380.Tryptophan.metabolism.	0.1841%	0.1243%	0.1070%	-0.5659	0.0017	*	-0.2165	0.0018	*
ko00400.Phenylalanine..tyrosine.&.tryptophan.biosynthesis.	0.6095%	0.6760%	0.6806%	0.1494	0.0095	*	0.0096	0.0066	*
ko00261.Monobactam.biosynthesis.	0.2549%	0.2513%	0.2505%	-0.0202	0.8410		-0.0049	0.8426	
ko00311.Penicillin.and.cephalosporin.biosynthesis.	0.0362%	0.0346%	0.0317%	-0.0648	0.8270		-0.1269	0.8242	
ko00332.Carbapenem.biosynthesis.	0.0783%	0.0858%	0.0861%	0.1310	0.2057		0.0055	0.1949	
ko00333.Prodigiosin.biosynthesis.	0.0583%	0.0718%	0.0745%	0.3005	0.0055	*	0.0538	0.0031	*
ko00401.Novobiocin.biosynthesis.	0.1118%	0.1261%	0.1289%	0.1739	0.0017	*	0.0324	0.0018	*
ko00405.Phenazine.biosynthesis.	0.0508%	0.0549%	0.0506%	0.1140	0.3519		-0.1173	0.3480	
ko00521.Streptomycin.biosynthesis.	0.2931%	0.3160%	0.3285%	0.1083	0.2676		0.0560	0.2878	
ko00524.Neomycin..kanamycin.&.gentamicin.biosynthesis.	0.0500%	0.0553%	0.0512%	0.1459	0.8017		-0.1125	0.7971	
ko00525.Acarbose.and.validamycin.biosynthesis	0.0975%	0.0965%	0.0989%	-0.0140	0.9595		0.0355	0.9658	
ko00940.Phenylpropanoid.biosynthesis.	0.2486%	0.2804%	0.3114%	0.1740	0.2057		0.1512	0.2096	
ko00950.Isoquinoline.alkaloid.biosynthesis.	0.0559%	0.0589%	0.0579%	0.0737	0.7518		-0.0235	0.7589	
ko00960.Tropane..piperidine.and.pyridine.alkaloid.biosynthesis.	0.1169%	0.1072%	0.1026%	-0.1251	0.0249	*	-0.0638	0.0327	*
ko00966.Glucosinolate.biosynthesis.	0.0475%	0.0599%	0.0642%	0.3349	0.0032	*	0.0988	0.0018	*
ko00010.Glycolysis...Gluconeogenesis.	1.3685%	1.3559%	1.3429%	-0.0134	0.9006		-0.0138	0.9029	
ko00020.Citrate.cycle..TCA.cycle..	0.6374%	0.6633%	0.6194%	0.0573	0.6039		-0.0986	0.5742	
ko00030.Pentose.phosphate.pathway.	0.8751%	0.9129%	0.9080%	0.0610	0.0903		-0.0077	0.0827	
ko00040.Pentose.&.glucuronate.interconversions.	0.4507%	0.4659%	0.4761%	0.0480	0.8333		0.0312	0.8360	
ko00051.Fructose.and.mannose.metabolism.	0.6741%	0.7091%	0.6586%	0.0729	0.8400		-0.1065	0.8242	
ko00052.Galactose.metabolism.	1.3713%	1.4104%	1.4438%	0.0405	0.6714		0.0338	0.6656	
ko00053.Ascorbate.and.aldarate.metabolism.	0.2008%	0.1606%	0.1417%	-0.3225	0.0055	*	-0.1809	0.0018	*
ko00500.Starch.and.sucrose.metabolism.	1.8831%	1.7369%	1.8598%	-0.1166	0.2676		0.0987	0.2759	
ko00520.Amino.sugar.&.nucleotide.sugar.metabolism.	1.4274%	1.5109%	1.5059%	0.0821	0.1237		-0.0048	0.1215	
ko00562.Inositol.phosphate.metabolism.	0.0927%	0.0964%	0.0929%	0.0571	0.9414		-0.0538	0.9460	
ko00620.Pyruvate.metabolism.	1.1956%	1.2027%	1.1813%	0.0085	0.9133		-0.0259	0.9117	
ko00630.Glyoxylate.&.dicarboxylate.metabolism.	0.8051%	0.7903%	0.7518%	-0.0269	0.3883		-0.0721	0.3947	
ko00640.Propanoate.metabolism.	0.7616%	0.7048%	0.6660%	-0.1119	0.0172	*	-0.0817	0.0366	*
ko00650.Butanoate.metabolism.	0.6755%	0.6790%	0.6623%	0.0075	0.9053		-0.0360	0.9117	
ko00660.C5.Branched.dibasic.acid.metabolism.	0.3002%	0.3091%	0.3104%	0.0422	0.7518		0.0060	0.7594	
ko00190.Oxidative.phosphorylation.	0.6758%	0.7777%	0.7946%	0.2027	0.0032	*	0.0309	0.0044	*
ko00195.Photosynthesis.	0.2479%	0.3078%	0.3369%	0.3123	0.0017	*	0.1304	0.0018	*
ko00680.Methane.metabolism.	0.7654%	0.7906%	0.7986%	0.0467	0.0746		0.0146	0.0703	
ko00710.Carbon.fixation.in.photosynthetic.	0.5582%	0.5915%	0.5875%	0.0836	0.0310	*	-0.0099	0.0221	*
ko00720.Carbon.fixation.pathways.prokkaryotes.	0.8798%	0.9867%	0.9634%	0.1655	0.0117	*	-0.0344	0.0056	*
ko00910.Nitrogen.metabolism.	0.5499%	0.5312%	0.5179%	-0.0500	0.4417		-0.0366	0.4577	
ko00920.Sulfur.metabolism.	0.2802%	0.2369%	0.2392%	-0.2419	0.3382		0.0135	0.3547	
ko00511.Other.glycan.degradation.	0.6352%	0.7521%	0.6812%	0.2436	0.5678		-0.1428	0.5591	
ko00513.Various.types.of.N.glycan.biosynthesis.	0.0626%	0.0912%	0.0783%	0.5425	0.6016		-0.2206	0.5793	
ko00531.Glycosaminoglycan.degradation.	0.1366%	0.1785%	0.1637%	0.3857	0.5678		-0.1245	0.5591	
ko00540.Lipopolysaccharide.biosynthesis.	0.2663%	0.1835%	0.1367%	-0.5372	0.0216	*	-0.4251	0.0330	*
ko00550.Peptidoglycan.biosynthesis.	0.7277%	0.8375%	0.8960%	0.2029	0.0017	*	0.0973	0.0018	*
ko00572.Arabinogalactan.biosynthesis...Mycobacterium.	0.0405%	0.0426%	0.0352%	0.0706	0.6494		-0.2725	0.6392	
ko00603.Glycosphingolipid.biosynthesis..globo.&.isoglobo.series	0.1730%	0.2173%	0.2123%	0.3288	0.2676		-0.0330	0.2531	
ko00604.Glycosphingolipid.biosynthesis...ganglio.series.	0.0626%	0.0912%	0.0782%	0.5419	0.5857		-0.2210	0.5742	
ko00061.Fatty.acid.biosynthesis.	0.5617%	0.6109%	0.6386%	0.1212	0.3423		0.0640	0.3204	
ko00071.Fatty.acid.degradation.	0.4147%	0.3434%	0.3355%	-0.2724	0.0207	*	-0.0334	0.0189	*
ko00072.Synthesis.&.degradation.of.ketone.bodies.	0.0243%	0.0187%	0.0188%	-0.3762	0.4417		0.0074	0.4577	
ko00120.Primary.bile.acid.biosynthesis.	0.0296%	0.0389%	0.0367%	0.3947	0.0861		-0.0845	0.0828	
ko00121.Secondary.bile.acid.biosynthesis.	0.0451%	0.0612%	0.0670%	0.4430	0.0017	*	0.1299	0.0018	*
ko00561.Glycerolipid.metabolism.	0.3269%	0.3346%	0.3452%	0.0339	0.7449		0.0449	0.7517	
ko00564.Glycerophospholipid.metabolism.	0.3457%	0.3173%	0.3299%	-0.1238	0.5306		0.0563	0.5204	
ko00600.Sphingolipid.metabolism.	0.4365%	0.5085%	0.5060%	0.2203	0.3423		-0.0071	0.3671	
ko01040.Biosynthesis.of.unsaturated.fatty.acids.	0.0311%	0.0209%	0.0204%	-0.5722	0.0044	*	-0.0361	0.0044	*
ko00130.Ubiquinone.&.other.terpenoid.quinone.biosynthesis.	0.1853%	0.1373%	0.1041%	-0.4324	0.0017	*	-0.4003	0.0018	*
ko00670.One.carbon.pool.by.folate.	0.3954%	0.4410%	0.4408%	0.1574	0.0017	*	-0.0006	0.0018	*

KEGG pathway	4 month mean %	9 month mean %	12 month mean %	4 to 9 month Log2FC	4 to 9 month Permanova fdr p value	Sig	9 to 12 month Log2FC	4 to 9 month permanova fdr p value	Sig
ko00730.Thiamine.metabolism.	0.4217%	0.4692%	0.4917%	0.1540	0.0017	*	0.0676	0.0018	*
ko00740.Riboflavin.metabolism.	0.1298%	0.1289%	0.1193%	-0.0102	0.6152		-0.1115	0.6176	
ko00750.Vitamin.B6.metabolism.	0.1848%	0.1876%	0.1798%	0.0221	0.6494		-0.0619	0.6283	
ko00760.Nicotinate.&.nicotinamide.metabolism.	0.4937%	0.4886%	0.4749%	-0.0151	0.8410		-0.0410	0.8242	
ko00770.Pantothenate.and.CoA.biosynthesis.	0.4835%	0.5153%	0.5281%	0.0918	0.0017	*	0.0355	0.0018	*
ko00780.Biotin.metabolism.	0.2026%	0.2197%	0.2165%	0.1169	0.8042		-0.0211	0.7840	
ko00785.Lipoic.acid.metabolism.	0.0464%	0.0389%	0.0340%	-0.2530	0.0055	*	-0.1966	0.0031	*
ko00790.Folate.biosynthesis.	0.3449%	0.3550%	0.3434%	0.0414	0.7518		-0.0479	0.7594	
ko00830.Retinol.metabolism.	0.0380%	0.0278%	0.0241%	-0.4503	0.0017	*	-0.2066	0.0018	*
ko00860.Porphyrin.and.chlorophyll.metabolism.	0.3608%	0.4686%	0.4772%	0.3773	0.0350	*	0.0261	0.0272	*
ko00410.beta.Alanine.metabolism.	0.1435%	0.1110%	0.0981%	-0.3713	0.0244	*	-0.1776	0.0189	*
ko00430.Taurine.and.hypotaurine.metabolism.	0.1377%	0.1276%	0.1256%	-0.1100	0.0179	*	-0.0232	0.0189	*
ko00440.Phosphonate.&.phosphinate.metabolism.	0.0735%	0.0510%	0.0497%	-0.5266	0.0313	*	-0.0374	0.0310	*
ko00450.Selenocompound.metabolism.	0.5215%	0.4947%	0.4888%	-0.0761	0.0455	*	-0.0171	0.0571	*
ko00460.Cyanoamino.acid.metabolism.	0.3234%	0.3647%	0.4027%	0.1735	0.0748		0.1428	0.0666	
ko00471.D.Glutamine.&.D.glutamate.metabolism	0.1277%	0.1505%	0.1519%	0.2376	0.0017	*	0.0132	0.0018	*
ko00473.D.Alanine.metabolism.	0.0979%	0.1084%	0.1113%	0.1470	0.0044	*	0.0390	0.0031	*
ko00480.Glutathione.metabolism.	0.3967%	0.3222%	0.2898%	-0.3002	0.0017	*	-0.1532	0.0018	*
ko00281.Geraniol.degradation.	0.0384%	0.0120%	0.0087%	-1.6757	0.0017	*	-0.4597	0.0018	*
ko00523.Polyketide.sugar.unit.biosynthesis.	0.1459%	0.1547%	0.1613%	0.0846	0.6526		0.0606	0.6656	
ko00900.Terpenoid.backbone.biosynthesis.	0.3429%	0.3828%	0.3967%	0.1588	0.0017	*	0.0514	0.0018	*
ko00903.Limonene.and.pinene.degradation.	0.0492%	0.0344%	0.0291%	-0.5157	0.0017	*	-0.2400	0.0018	*
ko00908.Zeatin.biosynthesis.	0.0284%	0.0322%	0.0324%	0.1831	0.0451	*	0.0092	0.0483	*
ko00981.Insect.hormone.biosynthesis.	0.0265%	0.0272%	0.0239%	0.0347	0.8333		-0.1870	0.8242	
ko01051.Biosynthesis.of.ansamycins.	0.1148%	0.0991%	0.0944%	-0.2122	0.0356	*	-0.0698	0.0387	*
ko01053.Biosynthesis.of.siderophore.group.nr.peptides	0.1214%	0.0541%	0.0368%	-1.1651	0.0017	*	-0.5557	0.0044	*
ko01055.Biosynthesis.of.vancomycin.group.antibiotics.	0.0551%	0.0513%	0.0536%	-0.1026	0.8632		0.0614	0.8588	
ko09113.Global.maps.only.	0.0451%	0.0411%	0.0453%	-0.1360	0.8090		0.1428	0.7985	
ko00230.Purine.metabolism.	1.7736%	1.8561%	1.8568%	0.0656	0.0179	*	0.0005	0.0078	*
ko00240.Pyrimidine.metabolism.	1.0619%	1.1472%	1.1708%	0.1115	0.0017	*	0.0293	0.0018	*
ko00362.Benzoate.degradation.	0.1000%	0.0619%	0.0617%	-0.6907	0.0313	*	-0.0064	0.0417	*
ko00625.Chloroaklane.&.chloroalkene.degradation.	0.1701%	0.1406%	0.1451%	-0.2742	0.0663		0.0447	0.0713	
ko00626.Naphthalene.degradation.	0.1315%	0.1067%	0.1096%	-0.3014	0.0379	*	0.0396	0.0528	*
ko00627.Aminobenzoate.degradation.	0.0366%	0.0230%	0.0208%	-0.6700	0.0017	*	-0.1430	0.0056	*
ko00633.Nitrotoluene.degradation.	0.0561%	0.0494%	0.0365%	-0.1833	0.1998		-0.4376	0.1949	
ko00643.Styrene.degradation.	0.0153%	0.0115%	0.0128%	-0.4129	0.7083		0.1584	0.7381	
ko00791.Atrazine.degradation.	0.0156%	0.0165%	0.0132%	0.0842	0.9053		-0.3194	0.9117	
ko00930.Caprolactam.degradation.	0.0235%	0.0090%	0.0078%	-1.3750	0.0017	*	-0.2208	0.0031	*
ko00980.Metabolism.of.xenobiotics.by.cytochrome.P450.	0.0507%	0.0308%	0.0261%	-0.7179	0.0017	*	-0.2415	0.0018	*
ko00982.Drug.metabolism...cytochrome.P450.	0.0507%	0.0308%	0.0260%	-0.7191	0.0017	*	-0.2458	0.0018	*
ko00983.Drug.metabolism...other.enzymes.	0.2675%	0.2754%	0.2785%	0.0420	0.4790		0.0165	0.4577	

4328

Appendix 3. *Lipid metabolites included in analyses with annotations, median mass (m/z), median retention time in seconds (rt), and valence.*

Lipid	Median m/z	Median rt (seconds)	Valence
So(d18:1)+H	300.2893	90.6687	ve+
So(d18:0)+H	302.3047	104.1785	ve+
MG(18:2)+H	355.2839	65.4578	ve+
MG 17:2; [M+NH4]+	358.2948	40.9098	ve+
MG 17:2; [M+NH4]+	358.2950	56.4968	ve+
MG 18:3; [M+NH4]+	370.2949	199.5786	ve+
MG 18:3; [M+NH4]+	370.2944	42.2456	ve+
MG 18:3; [M+NH4]+	370.2945	59.0423	ve+
MG 18:1; [M+NH4]+	374.3262	208.2640	ve+
So(d24:0)+H	386.3994	262.3042	ve+
MG 20:5; [M+NH4]+	394.2944	179.8314	ve+
MG 20:5; [M+NH4]+	394.3010	225.9909	ve+
MG 20:2; [M+NH4]+	400.3411	126.5315	ve+
MG 20:2; [M+NH4]+	400.3422	213.6014	ve+
MG 20:1; [M+NH4]+	402.3561	127.4914	ve+
DG(8:0/12:0)+Na	423.3073	231.1439	ve+
MG 22:4; [M+NH4]+	424.3415	101.1744	ve+
MG 22:2; [M+NH4]+	428.3726	189.1778	ve+
DG 21:0; [M+NH4]+	432.3684	215.6697	ve+
DG(10:0/12:0)+Na	451.3393	271.6578	ve+
PE(16:0e)+H	454.2918	110.4404	ve+
PE(16:0e)+H	454.2918	404.7121	ve+
DG(12:0/12:0)+Na	479.3704	315.0194	ve+
PE(18:0e)+H	482.3238	176.8222	ve+
MG(27:1)+H	483.4402	398.2040	ve+
MG 26:2; [M+NH4]+	484.4437	398.2227	ve+
LPC(16:0)+H	496.3393	108.7744	ve+
LPC(16:0)+H	496.3392	77.7938	ve+
LPC(18:0e)+H	510.3917	200.4979	ve+
Cer(d16:1/16:0)+H	510.4872	375.1542	ve+
Cer(d16:0/16:0)+H	512.5031	393.0604	ve+
LPC(16:0)+Na	518.3202	110.4280	ve+
LPC(18:2)+H	520.3391	86.8546	ve+
lysoPC 18:1; [M+H]+	522.3543	121.7223	ve+
LPC(18:0)+H	524.3706	174.9517	ve+
LPC(18:0)+H	524.3713	201.8440	ve+
Cer(d17:1/16:0)+H	524.5035	401.5223	ve+
DG 29:5; [M+NH4]+	534.4193	367.5207	ve+
Cer(d18:2/16:0)+H	536.5032	382.6523	ve+
Cer(d16:0/18:2)+H	536.5025	405.4259	ve+
Cer(d18:1/16:0)+H	538.5180	426.6776	ve+
Cer(d16:0/18:0)+H	540.5342	384.1472	ve+
TG(8:0/8:0/12:0)+NH4	544.4560	378.0697	ve+
LPC(18:0)+Na	546.3522	176.9351	ve+
Cer(d18:1/16:0+O)+H	554.5133	405.3322	ve+
Cer(d18:1/18:2)+H	562.5186	393.2536	ve+
Cer(d34:0)+Na	562.5159	444.3866	ve+
Cer(d18:1/18:1)+H	564.5333	433.0468	ve+
Cer(d18:0/18:0)+H	568.5651	495.1868	ve+
Cer(d18:0/17:0+O)+H	570.5444	406.7762	ve+
TG(8:0/10:0/12:0)+NH4	572.4876	425.0792	ve+
DG 31:0; [M+NH4]+	572.5238	381.3417	ve+
TG(8:0/10:0/12:0)+Na	577.4418	424.7958	ve+
Cer(d18:1/18:1+O)+H	580.5290	368.1170	ve+
TG(33:0p)+H	581.5132	343.5467	ve+
DG 32:1; [M+NH4]+	584.5314	372.7194	ve+
Cer(d18:0/18:2)+Na	586.5187	409.1560	ve+
Cer(d18:1/20:3)+H	588.5338	406.5859	ve+
Cer(d36:1)+Na	588.5344	447.8219	ve+
MG(34:4)+NH4	592.5293	471.9896	ve+
TG(8:0/12:0/12:0)+NH4	600.5193	472.5558	ve+
DG 34:4; [M+NH4]+	606.5139	372.7392	ve+
Cer(d18:1/20:0+O)+H	610.5762	497.3957	ve+
DG(18:2/18:2)+H	617.5128	241.2975	ve+
TG(18:0/8:0/8:0)+NH4	628.5499	472.4217	ve+
DG(18:3/18:2)+NH4	632.5228	405.5564	ve+
DG(16:0/20:4)+Na	639.4936	439.3621	ve+
DG 37:7; [M+NH4]+	642.5120	359.5073	ve+

Lipid	Median m/z	Median rt (seconds)	Valence
TG(6:0/12:0/18:0)+NH4	656.5815	518.3783	ve+
DG 38:6; [M+NH4]+	658.5351	359.5893	ve+
SM(d32:1)+H	675.5427	320.6272	ve+
SM 33:0; [M]+	691.5678	368.6894	ve+
CerG1(d18:1/16:0)+H	700.5714	384.4641	ve+
CerG1(d18:0/16:0)+H	702.5857	403.8713	ve+
SM(d34:1)+H	703.5725	369.8539	ve+
PE(16:0/18:2)+H	716.5224	321.4063	ve+
PE(16:0/18:1)+H	718.5380	363.7540	ve+
CerG1(d18:0/16:0+O)+H	718.5818	390.0669	ve+
CerG1(d36:1)+H	728.6025	434.4186	ve+
SM(d36:1)+H	731.6033	426.9389	ve+
CerG1(d34:1+O)+Na	738.5473	368.9983	ve+
PE 34:1; [M+Na]+	740.5218	297.1750	ve+
PC(16:0/18:1)+H	760.5833	432.6865	ve+
CerG2(d18:1/16:0+O)+H	878.6188	356.9982	ve+
FA(20:4)-H	303.2332	194.6611	ve-
FA(22:5)-H	329.2490	201.2894	ve-
cPA(18:0)-H	419.2575	190.5491	ve-
LPMe(18:0)-H	451.2830	199.6275	ve-
LPE(16:0)-H	452.2783	110.5347	ve-
LPE(18:0p)-H	464.3154	195.6803	ve-
LPE(18:0)-H	480.3101	176.7309	ve-
LPE(20:0p)-H	492.3473	240.9356	ve-
MGMG(16:0)+HCOO	537.3275	126.9163	ve-
OAHFA(18:2/18:1)-H	559.4727	422.4298	ve-
OAHFA(18:2/18:0)-H	561.4895	459.6248	ve-
OAHFA(18:1/18:0)-H	563.5046	494.5790	ve-
LPC(18:0)+HCOO	568.3628	174.5876	ve-
PE 25:0; [M-H]-	592.3958	254.5163	ve-
PE 27:0; [M-H]-	620.4257	263.8740	ve-
PE 27:0; [M-H]-	620.4255	283.6201	ve-
PG 27:1; [M-H]-	649.4128	231.5340	ve-
PG 27:0; [M-H]-	651.4233	224.6613	ve-
PE 34:2; [M-H]-	714.5089	323.3438	ve-
OAHFA(18:1/31:0)-H	745.7051	430.4391	ve-
PE 36:0; [M-H]-	746.5794	401.1478	ve-
PG(16:0/18:1)-H	747.5179	366.3311	ve-
SM(d18:1/16:0)+HCOO	747.5659	370.7628	ve-
MGDG 34:3; [M-H]-	751.5315	368.1728	ve-
PE 38:1; [M-H]-	772.5931	432.5271	ve-
MGDG 36:3; [M-H]-	779.5731	349.4803	ve-
PG 38:5; [M-H]-	795.5131	313.1455	ve-
PG 38:4; [M-H]-	797.5301	328.7574	ve-
MGDG(16:0/18:2)+HCOO	799.5589	416.4379	ve-
PE 40:1; [M-H]-	800.6264	480.9434	ve-
MGDG(16:0/18:1)+HCOO	801.5747	454.2862	ve-
PG 39:4; [M-H]-	811.5530	314.3345	ve-
MGDG(16:0/19:1)+HCOO	815.5902	486.7226	ve-
PG 40:4; [M-H]-	825.5707	329.8479	ve-
PG 40:3; [M-H]-	827.5889	490.2816	ve-
MGDG(16:0/20:1)+HCOO	829.6065	498.7475	ve-
PE 43:0; [M-H]-	844.6852	540.5393	ve-
PE 44:1; [M-H]-	856.6889	567.7303	ve-
DGDG(16:0/18:1)+HCOO	963.6276	419.0568	ve-

4331

4332

Appendix 4. *HILIC (aqueous) metabolites included in analyses with annotations, median mass (m/z), median retention time in seconds (rt), and valence.*

Annotation	HILIC (aqueous)	Median m/z	Median rt (seconds)	Valence
	M61T446	60.0401	445.6472	ve+
	M61T447	61.0401	446.6472	ve+
	M172T447	172.0606	447.1242	ve+
	M191T447	191.0744	447.1286	ve+
	M190T447	190.0707	447.1352	ve+
	M130T447	130.0503	447.1436	ve+
Glycine/Asparagine	M133T453	133.0611	453.1684	ve+
Cytidine	M244T466	243.9900	466.4442	ve+
	M247T490	247.1436	490.0481	ve+
Hypoxanthine	M159T499	159.0278	499.4561	ve+
	M137T500	137.0459	499.9109	ve+
	M138T500	138.0489	500.4332	ve+
Methylguanine	M166T502	166.0724	501.9333	ve+
Glutamic acid	M189T516	189.0868	515.8054	ve+
Adenine	M136T520	136.0619	520.4282	ve+
Xanthine	M153T521	153.0409	521.0790	ve+
7-Methylguanine	M166T525	166.0723	524.8556	ve+
Valeric acid	M144T527	144.1018	526.9302	ve+
8-Hydroxy-7-methylguanine	M182T548	182.0673	548.2498	ve+
	M118T557	118.0866	557.1974	ve+
Inosine	M269T579	269.1219	578.5630	ve+
L-Lysine	M146T579	146.1176	579.0733	ve+
	M292T589	292.1742	589.2366	ve+
Asparaginylhydroxyproline	M246T593	246.1088	592.5101	ve+
Trigonelline	M138T599	138.0551	599.3409	ve+
	M166T600	166.0865	599.6542	ve+
	M167T600	167.0897	599.8028	ve+
	M120T600	120.0811	599.8054	ve+
	M143T603	143.1682	603.2322	ve+
	M142T603	142.1646	603.3100	ve+
	M96T603	96.1596	603.3100	ve+
	M217T614	217.1549	613.9880	ve+
7,8-Dihydroneopterin	M256T615	256.1035	615.0550	ve+
	M105T615	105.1107	615.4220	ve+
	M222T624	222.0967	623.9120	ve+
Thiamine	M265T624	265.1510	624.0750	ve+
N-Acetylgalactosamine	M244T624	244.0793	624.2500	ve+
Hydroxyadipic acid	M204T624	204.0865	624.3700	ve+
	M207T626	207.1493	625.5320	ve+
N-Acetylgalactosamine fragment	M173T627	173.1650	626.9420	ve+
	M465T627	465.1674	627.0760	ve+
Genistin	M433T629	433.2352	629.2420	ve+
Deoxyuridine	M246T631	246.1087	630.5780	ve+
	M162T637	162.0764	637.4160	ve+
Indoleacrylic acid	M188T639	188.0707	639.2780	ve+
	M210T640	210.1283	639.6720	ve+
Choline	M187T640	187.1553	640.2030	ve+
	M192T640	192.0954	640.4160	ve+
	M209T641	209.1226	640.5400	ve+
	M162T641	162.1127	640.7910	ve+
	M168T641	168.1129	640.9920	ve+
	M183T641	182.6275	640.9990	ve+
	M163T641	163.1156	641.0040	ve+

Annotation	HILIC (aqueous)	Median m/z	Median rt (seconds)	Valence
	M252T641	252.1426	641.4430	ve+
Allantoin	M159T644	159.1491	643.7080	ve+
Indole-2-Carboxylic	M160T645	160.1526	645.1580	ve+
	M133T645	133.0319	645.2240	ve+
	M72T646	72.0813	645.6960	ve+
Isoleucyl-Valine	M231T648	231.1899	647.9760	ve+
	M230T648	230.1866	648.2120	ve+
	M233T651	233.1496	650.6620	ve+
Tryptophanamide	M407T651	407.2196	650.7150	ve+
	M130T652	130.0867	652.1720	ve+
	M123T655	123.0557	654.9960	ve+
	M141T656	141.0658	655.9780	ve+
	M116T659	116.0711	658.9960	ve+
	M117T659	117.0743	659.0380	ve+
	M139T659	139.0945	659.0770	ve+
Acetylhystidine	M154T659	154.0976	659.4440	ve+
Phenylethylamine	M70T659	70.0656	659.4480	ve+
	M136T659	136.0868	659.4780	ve+
	M121T660	121.0650	659.5400	ve+
	M146T660	146.0812	660.2020	ve+
	M155T661	155.1009	661.2800	ve+
	M434T662	434.1901	661.9440	ve+
	M435T662	435.1935	661.9860	ve+
	M173T662	173.1395	662.3900	ve+
Acetylcadaverine	M128T663	128.1072	662.7570	ve+
	M146T663	146.1368	663.4260	ve+
	M145T663	145.1337	663.4820	ve+
	M146T664	146.1300	663.9570	ve+
	M174T664	174.1428	664.0740	ve+
	M72T667	72.0813	666.6670	ve+
Histidylvaline/neopterin	M254T668	254.1606	668.1050	ve+
Prolyl-threonine	M277T669	277.1761	669.3140	ve+
	M100T674	100.0761	674.0260	ve+
	M164T675	164.1284	675.0280	ve+
	M137T677	137.0707	677.1310	ve+
	M133T678	133.1215	678.2640	ve+
	M114T679	114.0917	678.8000	ve+
N-acetyleputrescine	M132T679	132.1216	679.2880	ve+
	M150T682	150.1127	682.3760	ve+
	M265T685	265.1109	685.3420	ve+
	M273T688	273.2538	687.7350	ve+
	M76T689	76.0762	688.8040	ve+
	M204T692	204.0867	691.6350	ve+
m-Coumaric acid	M165T704	165.0548	703.5780	ve+
	M182T704	182.0810	703.6640	ve+
	M105T722	105.0743	721.5170	ve+
	M86T723	86.0606	722.5840	ve+
	M279T728	279.1014	728.2640	ve+
	M407T728	407.1363	728.3520	ve+
Glycylvaline dipeptides	M175T741	175.1081	740.7540	ve+
	M252T758	252.1444	758.2380	ve+
Undecanedoic acid	M217T760	217.1295	759.9180	ve+
Alanine	M90T766	90.0554	765.8310	ve+
	M91T766	91.0587	765.9480	ve+
Asparginyl-methionine	M265T772	265.1223	771.7480	ve+
	M264T772	264.1180	772.2400	ve+
Glycerophosphocholine	M258T778	258.1104	777.9050	ve+
L-Threonine	M120T780	120.0659	779.9040	ve+

Annotation	HILIC (aqueous)	Median m/z	Median rt (seconds)	Valence
	M126T793	126.0223	792.8580	ve+
	M149T805	149.0638	804.8040	ve+
	M148T805	148.0606	804.8050	ve+
	M84T806	84.0448	805.5880	ve+
Glutaminylthreonine	M265T806	265.1222	805.8160	ve+
	M130T806	130.0503	806.0100	ve+
	M264T806	264.1186	806.2720	ve+
	M147T808	147.0763	807.7950	ve+
Glycine	M76T809	76.0398	809.1060	ve+
	M161T810	161.0920	809.5500	ve+
	M292T815	292.1032	815.1460	ve+
	M311T815	311.1173	815.2720	ve+
	M310T815	310.1120	815.2870	ve+
	M351T824	351.1582	823.8970	ve+
	M106T832	106.0503	832.2120	ve+
	M134T847	134.0450	846.6520	ve+
Serine	M88T847	88.0398	846.6560	ve+
	M116T847	116.0345	847.1500	ve+
	M266T858	266.1589	858.3200	ve+
	M454T860	454.1559	859.9170	ve+
	M309T861	309.1182	860.5800	ve+
3-Oa-L-Fucopranosy-D-Glutamate	M507T861	507.2105	861.4530	ve+
	M266T862	266.1237	861.6060	ve+
	M508T862	508.2151	861.7170	ve+
	M506T862	506.2084	861.8680	ve+
	M238T862	238.0383	862.2390	ve+
	M85T863	85.0289	862.5940	ve+
D-mallose	M188T863	188.0915	863.2840	ve+
	M343T863	343.1218	863.3540	ve+
Ornithine	M132T864	132.1135	864.1000	ve+
	M115T864	115.0870	864.3340	ve+
	M133T864	133.1169	864.4960	ve+
	M546T867	546.2024	867.3990	ve+
	M482T868	482.1981	868.4860	ve+
	M634T869	634.2185	868.5800	ve+
	M657T876	657.2241	875.5100	ve+
Reduced Riboflavin	M714T879	714.2434	878.9660	ve+
	M366T887	366.1401	887.0720	ve+
	M855T888	855.3174	887.7140	ve+
	M854T888	854.3117	887.7600	ve+
	M876T888	876.2959	887.9420	ve+
	M871T888	871.3418	887.9520	ve+
	M514T888	514.2000	888.5000	ve+
	M92T889	92.0711	888.6110	ve+
	M513T889	513.2002	888.7020	ve+
	M512T889	512.1958	888.9830	ve+
N-acetyl L-tyrosine	M224T893	224.1125	893.1550	ve+
	M303T894	303.1548	893.7980	ve+
	M658T896	658.2531	895.6840	ve+
	M160T896	160.0970	896.0260	ve+
Indoleacetic acid	M206T897	206.1019	897.1180	ve+
Hydroxyprolyl-asparagine	M246T902	246.0944	901.8680	ve+
	M527T902	527.1379	901.9450	ve+
Aspartyl-glutamine/glutaminylthreonine	M264T904	264.0659	904.4070	ve+
	M262T905	262.0686	904.7520	ve+
Asymmetric dimethyl arginine	M241T910	241.1064	909.9400	ve+
	M204T911	204.1532	910.5750	ve+
	M203T911	203.1498	910.9280	ve+
Methanethiol	M125T915	124.9229	915.3360	ve+

Annotation	HILIC (aqueous)	Median m/z	Median rt (seconds)	Valence
N6,N6,N6-Trimethyl-L-lysine isotope	M190T921	190.1631	920.9670	ve+
	M205T924	205.1545	924.1840	ve+
	M175T925	175.1442	924.8720	ve+
Serotonin	M199T930	199.0845	929.7260	ve+
Glutamyl-alanine	M201T930	201.0825	929.8670	ve+
	M259T930	258.8995	930.3460	ve+
	M107T930	106.9510	930.4230	ve+
	M245T933	245.0536	932.8560	ve+
3-hydroxyhippuric acid	M218T938	218.0423	937.8590	ve+
	M471T938	471.2191	937.8880	ve+
	M202T940	202.0683	940.1000	ve+
Tyrosyl-glutamine	M309T949	309.1654	949.2660	ve+
	M310T950	310.1693	949.6310	ve+
Histidine	M200T950	200.0401	950.4500	ve+
	M331T951	331.1470	951.2350	ve+
	M156T952	156.0769	951.6480	ve+
	M130T958	130.0864	957.7200	ve+
	M191T958	191.0764	957.7260	ve+
	M148T958	148.1162	957.7300	ve+
	M147T958	147.1129	957.7320	ve+
	M169T958	169.0943	957.7420	ve+
	M84T958	84.0812	957.7780	ve+
Proline	M116T961	116.0709	960.8580	ve+
	M175T980	175.1191	980.1960	ve+

4335

4336

4337 Appendix 5. *Metagenome sequence quality data for individual samples collected.*

	Raw		Trimmed and host removed		Kingdom and domain assignment counts			
Sample	Number of read pairs	Base-pairs	Number of read pairs	Base-pairs	Bacteria	Fungi	Eukaryota - non fungi	Viruses
ICF01-Base_S1	5,266,750	1,481,693,717	5,256,149	1,236,736,517	4,713,022	0	0	0
ICF01-End_S59	6,239,157	1,831,189,568	6,235,719	1,621,370,331	5,831,458	0	0	0
ICF01-Mid_S39	6,141,886	1,702,285,078	6,127,699	1,378,217,586	5,373,798	0	0	0
ICF02-Base_S2	5,920,357	1,689,719,523	5,912,320	1,446,323,477	5,490,265	0	3963	0
ICF02-End_S60	4,677,583	1,362,951,219	4,673,537	1,230,498,080	4,363,679	0	0	0
ICF02-Mid_S40	6,026,763	1,694,972,468	6,017,824	1,377,741,881	5,453,745	0	0	0
ICF03-Base_S3	3,814,233	1,106,147,691	3,810,459	985,880,092	3,406,279	0	20191	0
ICF03-End_S61	5,033,251	1,489,332,395	5,031,241	1,355,642,198	4,722,581	0	0	0
ICF03-Mid_S41	5,249,218	1,464,143,030	5,238,583	1,220,498,547	4,749,190	2777	3790	0
ICF04-Base_S4	4,143,675	1,205,423,738	4,139,839	1,081,429,772	3,771,741	0	9116	0
ICF04-End_S62	4,492,531	1,322,866,442	4,490,535	1,163,601,130	4,218,264	0	0	0
ICF04-Mid_S42	4,365,700	1,238,395,860	4,358,581	1,061,459,903	3,956,098	0	0	0
ICF05-Base_S5	6,297,680	1,844,912,003	6,292,843	1,655,712,418	5,941,602	0	0	0
ICF05-End_S63	5,736,358	1,683,380,950	5,733,304	1,495,467,830	5,399,426	0	0	0
ICF05-Mid_S43	5,731,766	1,683,594,756	5,728,052	1,516,055,206	5,426,741	0	0	0
ICF06-Base_S6	5,549,164	1,531,265,673	5,534,838	1,236,643,279	4,983,147	0	3868	0
ICF06-End_S64	6,348,923	1,826,606,531	6,342,371	1,572,123,119	5,789,314	0	0	0
ICF06-Mid_S44	5,117,046	1,464,412,781	5,109,069	1,266,737,009	4,709,949	0	0	0
ICF07-Base_S7	9,035,187	2,412,118,441	9,008,071	1,852,678,344	7,649,998	0	6856	0
ICF07-End_S65	5,911,564	1,752,789,745	5,909,525	1,584,566,214	5,469,518	0	0	0
ICF07-Mid_S45	4,839,193	1,435,486,568	4,837,212	1,318,681,754	4,478,354	0	0	0
ICF08-Base_S8	4,117,990	1,206,929,758	4,114,670	1,111,436,008	3,876,678	0	0	0
ICF08-Mid_S46	4,577,084	1,362,500,372	4,575,250	1,285,264,517	4,354,549	0	0	0
ICF09-Base_S9	4,244,325	1,265,190,092	4,242,848	1,204,131,765	4,045,512	0	0	0
ICF09-Mid_S47	5,000,269	1,473,585,196	4,997,317	1,346,540,391	4,713,226	0	0	0
ICF10-Base_S10	4,555,500	1,343,263,136	4,553,315	1,216,401,817	4,297,243	0	0	0
ICF10-End_S66	5,081,389	1,501,227,175	5,079,228	1,348,948,398	4,770,336	0	0	0
ICF10-Mid_S48	4,815,013	1,415,544,962	4,811,906	1,292,753,138	4,523,271	0	0	0
ICF11-Base_S11	3,438,572	1,001,910,501	3,435,935	884,034,371	3,232,315	0	0	0
ICF11-Mid_S49	4,775,538	1,403,584,139	4,772,796	1,265,614,440	4,540,483	0	0	0
ICF12-Mid_S50	5,912,769	1,715,185,741	5,906,908	1,501,267,969	5,493,447	0	0	0
ICF13-Base_S12	5,642,974	1,669,335,906	5,640,066	1,557,900,597	5,363,950	0	0	0
ICF14-Base_S13	3,145,262	937,080,016	3,143,956	900,452,412	3,031,569	0	0	0
ICF14-Mid_S51	5,964,464	1,737,971,691	5,960,408	1,508,695,971	5,536,239	0	0	0
ICF15-Base_S14	5,483,404	1,612,662,242	5,479,855	1,471,356,296	5,191,222	0	0	0
ICF15-Mid_S52	4,734,561	1,270,400,804	4,721,549	972,330,658	3,990,030	0	0	0
ICF16-Base_S15	5,658,223	1,678,156,170	5,656,189	1,549,516,476	5,371,658	0	0	0
ICF16-Mid_S53	5,168,358	1,533,948,495	5,166,409	1,413,043,211	4,917,236	0	0	0
ICF17-Base_S16	4,677,976	1,361,117,244	4,674,201	1,224,294,016	4,418,167	3962	3824	0
ICF17-Mid_S54	6,437,497	1,885,512,601	6,433,370	1,686,962,527	5,967,838	0	0	0
ICF18-Base_S17	4,702,884	1,354,477,942	4,696,733	1,193,804,496	4,350,504	0	2263	0
ICF18-Mid_S55	5,842,547	1,723,251,743	5,839,649	1,589,385,532	5,557,147	0	3073	0
ICF19-Base_S18	3,987,917	1,170,170,067	3,984,970	1,074,442,063	3,771,733	0	0	0
ICF20-Base_S19	4,763,938	1,411,633,926	4,762,008	1,313,357,429	4,546,216	0	0	0
ICF21-Base_S20	3,816,236	1,082,190,748	3,810,572	918,856,974	3,517,015	0	3173	0
ICF22-Base_S21	4,649,944	1,340,543,264	4,644,179	1,180,586,226	4,319,197	0	0	0
ICF23-Base_S22	4,496,249	1,295,707,312	4,490,981	1,140,378,911	4,179,643	0	2877	0
ICF24-Base_S23	4,395,954	1,276,211,495	4,392,203	1,137,426,569	4,114,771	0	3202	0
ICF25-Base_S24	5,283,072	1,557,409,454	5,279,868	1,435,195,717	4,945,501	3172	3897	0
ICF26-Base_S25	592,490	170,105,744	591,433	148,827,769	544,276	0	0	0
ICF27-Base_S26	4,714,999	1,343,402,936	4,707,803	1,145,648,794	4,327,581	0	3285	0
ICF28-Base_S27	4,128,494	1,219,736,530	4,125,944	1,133,805,108	3,883,306	0	0	0
ICF28-Mid_S56	6,189,246	1,800,531,283	6,184,246	1,588,973,434	5,709,650	0	0	0
ICF29-Base_S28	5,309,196	1,532,125,180	5,304,396	1,316,291,855	4,930,649	0	0	5398
ICF29-Mid_S57	7,341,657	2,164,957,817	7,338,055	1,930,945,640	6,952,500	0	5323	0
ICF30-Base_S29	4,916,671	1,456,333,818	4,914,682	1,346,133,162	4,673,921	0	0	0
ICF30-Mid_S58	4,457,031	1,307,709,634	4,454,555	1,175,540,478	4,202,388	0	0	0
ICF32-Base_S30	4,035,622	1,195,577,167	4,033,645	1,114,366,137	3,819,083	2826	6512	0
ICF33-Base_S31	4,157,389	1,178,033,063	4,151,019	1,003,888,152	3,656,133	0	3076	0
ICF34-Base_S32	5,270,199	1,502,442,946	5,263,730	1,272,540,252	4,835,791	0	6605	0
ICF35-Base_S33	4,098,407	1,213,760,666	4,096,501	1,128,477,366	3,884,835	0	0	0
ICF36-Base_S34	6,207,673	1,692,052,626	6,193,617	1,321,134,513	5,233,703	0	2854	0
ICF37-Base_S35	4,289,382	1,224,842,498	4,284,219	1,049,812,030	3,963,586	2336	4272	0
ICF38-Base_S36	3,944,005	1,155,453,021	3,941,373	1,038,199,981	3,724,045	0	0	0
ICF39-Base_S37	4,157,044	1,224,795,009	4,154,323	1,131,616,943	3,887,521	0	0	0
ICF40-Base_S38	4,969,346	1,430,261,816	4,963,131	1,245,098,354	4,569,600	0	4595	0

4338