Exploring the faecal microbiome associated with preterm birth

Thesis submitted by:

Jacob A F Westaway

College of Public Health, Medical and Veterinary Sciences

Centre for Tropical Bioinformatics and Molecular Biology

James Cook University

This dissertation is submitted for the degree of Doctor of Philosophy

09/03/21

Supervisors:

Associate Professor Donna Rudd

Dr Yogavijayan Kandasamy

Dr Roger Huerlimann

Dr Robert Norton

Dr Catherine Miller

Dr David Watson

Statement of contributions of others

Assistance	Contribution	Name	Affiliation
Intellectual support	Project plan and development	Ass/Prof Donna Rudd	James Cook University
		Dr Roger Huerlimann	James Cook University
		Dr Yogavijayan Kandasamy	Queensland Health
		Dr Catherine Miller	James Cook University
		Dr Robert Norton	Queensland Health
		Dr David Watson	Queensland Health
	Editorial support	Ass/Prof Donna Rudd	James Cook University
		Dr Roger Huerlimann	James Cook University
		Dr Yogavijayan Kandasamy	Queensland Health
		Dr Catherine Miller	James Cook University
		Dr Robert Norton	Queensland Health
		Dr David Watson	Queensland Health
		Dr Kyran Staunton	James Cook University
Financial support	Research	Education and Research Trust Account (SERTA) research grant	Townsville University Hospital
		FNQHF Grant	Far North Queensland Hospital Foundation
		AIMs PhD Seeding Grant	Australian Institute of Medical Scientists
	Stipend	Australian Government Research Training Program Scholarship	Australian Government
Technical support	Shotgun metagenomic sequencing	Microba Life Sciences	Microba Life Sciences
	16S rRNA amplicon sequencing	Sandra Infante-Villamil	James Cook University
		Dr Roger Huerlimann	
		Dr Tiffany Kosch	

	Patient recruitment and sample collection	Helena McInnes	Queensland Health
		Nicole Dionysius	Queensland Health
		Dr Neil Archer	Queensland Health
		Donna Martin	Queensland Health
Editorial assistance	Formatting of Chapter 2	Ms Katherine Fowler	Professional editor

Acknowledgements

First and foremost, I would like to express my most sincere thank you to my primary supervisor, Associate Professor Donna Rudd. I feel fortunate to have had such a strong willed and determined, yet kind and caring supervisor, one that truly embodies the student-first mentality. I am happy to have had a supervisor that put my personal and professional development before her own agenda. Donna's support, encouragement, guidance, and trust have been invaluable over the course of my PhD, and I hope this will continue into my research career.

I would also like to express my gratitude to the rest of my advisory team. Dr Roger Huerlimann, Dr Yogavijayan Kandasamy, Dr Robert Norton, Dr Catherine Miller, Dr David Watson, and Dr Tiffany Kosch. Each of my supervisors brought their own expertise and experience to a project that required a truly multidisciplinary team. Additionally, I would like to thank all others involved across the different studies. This includes Dr Kyran Staunton, Sandra Infante-Villamil, Helena McInnes and Nicole Dionysius, who all contributed their relative expertise.

I would also like to give a special thanks to one of my secondary supervisors' Dr Roger Huerlimann, for his technical support and guidance. Without Roger's help with all things molecular and bioinformatics, and the many emails back and forth, my PhD would never have got off the ground and I would not be on the path to becoming a bioinformatician.

I would also like to thank all members of the Cohort Doctoral Studies Program, both those who lead the program and my peers, and to the close circle of JCU post-graduate students and post-doctoral researchers that I have surrounded myself with over the course of my PhD. I am forever grateful for the support and guidance provided, and the friendships fostered.

Finally, this PhD would never have been possible without my family. Thank you to my parents for both putting a rent-free roof over my head and supporting me. Additionally, and most importantly, a massive thank you to my partner Juliette (JuJu), who has supported me financially, emotionally, and mentally, and made this experience far easier. Thank you for understanding the countless long days, work filled weekends and evenings, and trips away to the laboratory. I am so very grateful for your support and sacrifice, and cannot imagine how immensely difficult this PhD would've been without you.

Publications

Arising from thesis and published

Westaway, J.A.F., Huerlimann, R., Miller, C.M. *et al.* Methods for exploring the faecal microbiome of premature infants: a review. *Maternal Health, Neonatology and Perinatology* 7, 11 (2021). *https://doi.org/10.1186/s40748-021-00131-9*

Westaway, J.A.F., Huerlimann, R., Kandasamy, Y. *et al.* The bacterial gut microbiome of probiotic-treated very-preterm infants: changes from admission to discharge. *Pediatr Res* (2021). *https://doi.org/10.1038/s41390-021-01738-6*

Westaway, J.A.F., Huerlimann, R., Kandasamy, Y. *et al.* To Probiotic or Not to Probiotic: A Metagenomic Comparison of the Discharge Gut Microbiome of Infants Supplemented With Probiotics in NICU and Those Who Are Not. *Front. Pediatr* 10:838559 (2022). *https://doi.org/10.3389/fped.2022.838559*

Westaway, J.A.F., Huerlimann, R., Kandasamy, Y. *et al.* Exploring the long-term colonisation and persistence of probiotic-prophylaxis species on the gut microbiome of preterm infants: a pilot study. *Eur J Pediatr* (2022). *https://doi.org/10.1007/s00431-022-04548-y*

Arising from thesis and accepted for publication

Submitted to *Australian Journal of Medical Science*: A validation of at-home infant stool sample collection devices for determining the faecal microbiome.

Other publications during candidature

Tsang, H. H., Domingos, J. A., **Westaway, J. A. F.,** Kam, M. H. Y., Huerlimann, R., & Bastos Gomes, G. (2021). Digital droplet PCR-based environmental DNA tool for monitoring *Cryptocaryon irritans* in a marine fish farm from Hong Kong. *Diversity*, *13*(8), 1-15. [350]. *https://doi.org/10.3390/d13080350*

Awards and Presentations During Candidature

Awards

Research training program stipend: Australian Government, 2018-2021.

Runner up, Early candidature presentations: James Cook University (JCU) My Research Rules, 2019.

Finalist: James Cook University (JCU) Three Minute Thesis, 2019.

Best research oral presentation: Australian Institute of Medical Scientist Nation Meeting, 2021.

Presentations

Australian Institute of Medical Scientists Conference, 2018. *The maternal microbiome over pregnancy.* Oral presentation.

Townsville Hospital Research Showcase, 2019. *Exploring the Microbiome of Preterm Infants at the Townsville Hospital*. Oral presentation.

Queensland Perinatal Consortium Conference, 2019. *Discovering the faecal microbiome associated with preterm birth*. Oral presentation.

Townsville Hospital Research Showcase, 2020. Characterising the bacterial gut microbiome of very preterm infants from the Townsville Hospital and exploring the influence of clinical variables. Poster presentation.

Cairns Symposium, 2020. To probiotic or not to probiotic: comparing the microbiome of probiotic-supplemented and non-supplemented preterm infants. Oral presentation.

International Human Microbiome Consortium, 2021. Exploring the bacterial microbiome of probiotic-treated preterm infants. Poster presentation.

Australian Institute of Medical Scientists National Meeting, 2021. Exploring the faecal microbiome associated with preterm birth. Oral presentation.

Centre for Tropical Bioinformatics and Molecular Biology (CTBMB) Conference, 2021. Exploring the faecal microbiome of preterm infants in North QLD.

I have also given oral presentations in My Research Rules (2019), JCU and CPHMVS 3MT competitions (2019), at three COHORT Program internal conferences (2018-2020), and three AITHM laboratory meetings (2018-2021).

Abstract

Introduction

Preterm birth disrupts gut microbiome development. The resulting preterm microbiome is characterised by low diversity and commensal microbe abundance, in combination with a greater number of potential pathogens. This irregular microbiome may contribute to disease burden, with diseases like necrotising enterocolitis (NEC) almost exclusively affecting the most premature. Probiotic prophylaxis during admission appears to mitigate disease risk through modulation of the gut microbiome. As a result, neonatal intensive care units (NICU) across Australia, including the Townsville University Hospital's (TUH) NICU, include probiotic prophylaxis for those infants at high risk of NEC. This includes infants born < 32 weeks and/or < 1500g. The relationship between the gut microbiome and infant health, especially in the context of those born preterm, makes the gut microbiome an important and modifiable factor in neonatal care. Thus, this thesis set out to dig further into the impact of clinical variables on a unique cohort of entirely probiotic-supplemented infants and explored the expansion of the criteria for probiotic supplementation.

Methods

This thesis used a combination 16S rRNA metabarcoding and shotgun metagenomics to characterise the gut microbiome of preterm infants born in North Queensland Australia. Infants born in this region at < 32 weeks and/or < 1500g are supplemented with the probiotic Infloran®. Mixed effects modelling was used to explore and account for the impact of other potential microbiome covariates. The thesis combines results from four studies across three results chapters. Chapter 3 assessed the effect of several covariates at admission and discharge in a unique cohort of entirely probiotic-supplemented preterm infants, and Chapter 4 explored differences in the microbiome of these same infants at discharge to other preterm infants who fall outside the criteria for probiotic prophylaxis. Chapter 5 explores the impact of probiotic prophylaxis post-discharge and begins with a validation study of the sampling method used.

Results

In Chapter 3, the gut microbiome of probiotic-supplemented very preterm infants changed significantly over time, with mixed effects modelling demonstrating significant associations with several covariates. Lower alpha diversity was associated with infants experiencing

unfavourable outcomes, specifically sepsis and retinopathy of prematurity (ROP). Additionally, chorioamnionitis, preeclampsia, sepsis, necrotising enterocolitis and ROP were also all associated with differential abundance of several taxa. Along with these associations, widespread colonisation of probiotic taxa was observed. In Chapter 4, supplementation with the probiotic Infloran® was demonstrated to have a significant association with favourable microbiome metrics, specifically greater diversity and abundance of probiotic taxa, suggesting those preterm infants who fall outside the criteria for supplementation may be at a relative disadvantage. However, in Chapter 5, I observed that this difference between infants who do and do not receive Infloran® during admission does not persist.

Conclusion

The outcomes of this thesis improve our understanding of the relationship between unfavourable outcomes, both maternal and infant, and the gut microbiome of preterm infants, and add credence to the argument for expanding the criteria for probiotic supplementation. The impact of maternal health on the preterm infant microbiome suggests considerations should made with maternal interventions, and the associations with ROP suggest the gut microbiome is a potential target for future interventions, however, more work is needed. Lastly, due to the developmental benefits of key taxa, the differences in microbial populations associated with probiotic prophylaxis suggest expansion of the inclusion criteria could be beneficial. Supporting the development of the preterm infant gut microbiome has the potential to help reduce the disease and developmental burden of these infants.

Table of Contents

List of Figures

List of Tables

1. Introduction

With between ten- and one hundred-trillion microbial cells, and only one- to ten-trillion human cells, we are more microbe than we are human. The impact these mutualistic microbes have on human health has led some to refer to the human microbiome as the "hidden" metabolic organ [1]. The significance of this relationship begins early in life, with several studies demonstrating early patterns of gut microbial colonisation playing an important role in healthy development and future health [2-6]. Disruption in these normal patterns of colonisation, as seen in preterm birth, can lead to the development of acute and chronic disease. Furthermore, similar irregular microbial colonisation patterns in preterm infants have also been linked to key clinical variables, such as mode of delivery and diet [7], maternal health [8], and diseases like necrotising enterocolitis (NEC) and sepsis [9]. However, probiotic supplementation may correct these microbial imbalances and mitigate disease risk [10-13]. As a result, probiotic supplementation has become a common supplementation modality in NICU's, particularly for the most premature of infants. Thus, this thesis will explore the changes in gut bacterial microbiome of probiotic-supplemented, very preterm infants (< 32 weeks), seeking to characterise microbial population dynamics at admission and discharge, in association with several clinical variables and relative to non-treated (probiotic), moderate to late preterm (32 to < 37 weeks) infants.

The human gut microbiome

The human gut microbiome is an integral component of human physiology that significantly impacts health [14], and the disruption of which has been implicated in several diseases [15,16]. Despite limited understanding of the complex mechanisms that define this relationship between humans and their microbiome, evidence suggests a level of dependence for humans on these microbes [17]. A dependency that stems from our coevolution with these microbes over millennia, the result of which has been an assemblage of different microbial species and humans into a single ecological unit. As a result, the human microbiome is intertwined in host physiology, contributing to everything from immunity and metabolism [18], to infant development [19].

The infant gut microbiome

The gut microbiome plays a critical role in healthy infant development, immunity and metabolism [14,18,19], with microbial development occurring in parallel with infant development. Early colonisers play a crucial role in immunity and tolerance, with early colonisers interacting with epithelial and lymphoid tissue in a state of controlled inflammation via apoptotic stimuli, reactive oxygen species synthesis and Toll-like receptor signalling [20]. This crosstalk between microbes with the epithelium and gut-associated lymphoid tissue aids in the development of innate immune defences and promotion of pathogen recognition, which also leads to gene expression, promotion of epithelial turnover, increased mucous synthesis, peristalsis and antimicrobial secretion in mucous [21,22]. All these changes contribute to the development of tolerance to commensal microbes and food whilst providing a barrier against the entry of potentially pathogenic microbes. However, as colonisation occurs through both vertical (mother) and horizontal (environment) routes, largely influenced by extraneous variables, the colonisation process is sensitive to a host of factors (*Figure 1.1*). These factors can contribute to microbial imbalances, or dysbiosis, perturbing normal development and contributing to disease [14].

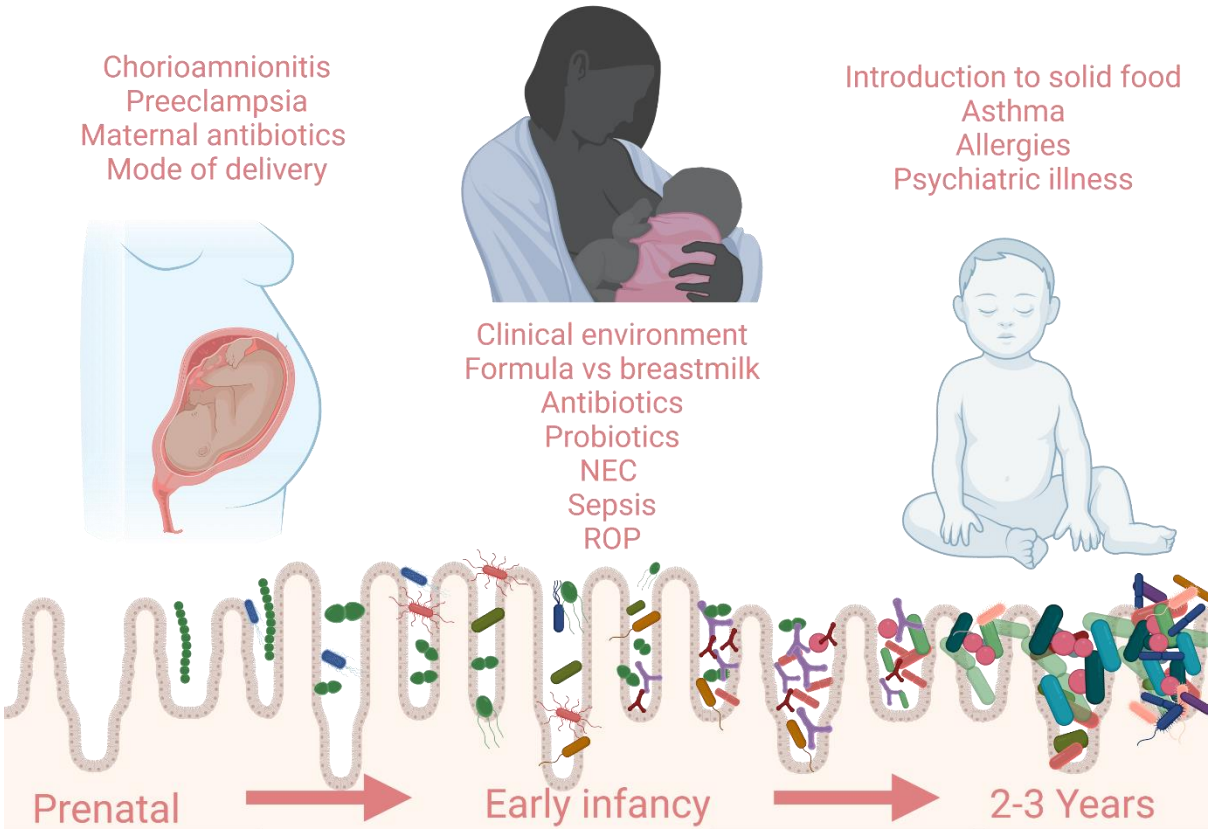


Figure 1.1 Summary of the choreographed progression of microbial colonisation in infants, *and previously described covariates of the infant microbiome, including both potential causes and consequences of microbial dysbiosis. Annotation for necrotising enterocolitis; NEC: retinopathy of prematurity; ROP. Figure 1.1 created using www.biorender.com.*

Vertical colonisation

Vertical inoculation means the maternal microbiome and health status play an important role in shaping the infant gut microbiome, and in turn their health [23]. This process of microbial transmission occurs during key colonising events such as birth and breastfeeding, with the mother passing on key commensal microbes [24]. An unfortunate consequence of this vertical transmission is the potential to pass on undesirable microbes, which could contribute to both disease and the loss of microbial diversity over generations [25]. The impact of this loss in microbial diversity is demonstrated by both the shared pathogens of NEC and urinary tract infections [26], and altered infant microbiomes associated with maternal antibiotic treatment [27,28]. Although the current dogma suggests the transmission of these microbes occurs initially during delivery [29], some evidence suggests maternal-infant translocation of microbes during pregnancy.

Pre-natal development

The potential prenatal, or in utero, stage of microbiome development is the least understood. There is a growing body of contested research that suggests vertical translocation from mother to foetus during pregnancy [30]. This includes studies identifying microbes in the meconium, the amniotic fluid, and the placenta [31,32], as well animal models identifying genetically tagged microbes in the offspring of mothers fed that tagged microbe [33]. How this would occur is unclear, but one proposed mechanism is that projecting dendritic cells embedded in tight-junctions between epithelial cells of the intestine can sample microbes and translocate them from the gut-lumen and into the maternal lymphatic and circulatory systems and subsequently the infant [34,35]. However, whilst compelling, many studies suggesting in utero translocation lack rigour [19], and many of the early gut microbiome colonisers of full-term vaginally delivered infants are of vaginal and faecal origin [36]. This suggests that birth is the first major colonising event for the gut microbiome.

Birth

Mode of delivery has a significant impact on microbial composition [8,37]. The microbes that infants are exposed to during birth play a critical role in the development of immune and metabolic systems [38]. Vaginally delivered infants have higher abundances of vaginally derived microbes, including *Bacteroides* and *Lactobacillus* [37,39]. *Lactobacillus* spp. are dominant in the vaginal canal and play a critical role in the colonisation of other commensal microbes [39]. *Lactobacillus* spp. also produce anti-microbial compounds, lactic acid and hydrogen peroxide that alter the environmental conditions and become deterrents for colonisation by potentially pathogenic microbes [40,41]. In contrast, caesarean born infants acquire greater abundances of skin dwelling microbes like *Staphylococcus* [37,42], in combination with delayed colonisation of key commensal microbes – Lactobacillus [37], Bifidobacterium [37,43] and Bacteroides [37,44-47]. The caesarean-derived microbiome, that also includes greater diversity [8], may contribute to associations between this mode of delivery and diseases like Irritable Bowel Syndrome (IBS), adiposity, Celiac Disease, Asthma and NEC [48]. Caesarean birth, the incidence of which is increasing [39], and set to reach 28.5% of deliveries by 2030 [49], is one of the greatest contributors to disruption of the infant gut microbiome, along with antibiotic use and formula feeding [28].

Diet

Breast feeding, much like vaginal birth, is an evolutionary adaptation for mammals that is integral to the development and health of both the infant and their gut microbiome. The dynamic composition of breastmilk is 'designed' to satisfy the needs of infants at different life stages. For example, what colostrum (the first milk produced after birth) lacks in nutritional value, it makes up in immunologic and growth factors [50]. However, over time, the key components of breastmilk shift from immunoglobulins and growth factors to lipids, proteins, human milk oligosaccharides (HMOs) and probiotics. The specific makeup of breastmilk supports everything from epithelial cell proliferation and maturation to development of the immune system and the fight against pathogens [50,51]. Although formula is designed to promote both growth and development, it is difficult to replicate the complexity of breastmilk, which contains greater than 200 HMOs and 400 unique proteins.

When considering the gut microbiome, HMOs and probiotics are of most interest. HMOs are human-indigestible prebiotic glycans that selectively nourish commensal microbes. Digestion of HMOs produces short chain fatty acids that serve as nourishment for both commensal microbes and human cells in the gastrointestinal tract (GI). Additionally, along with lactic acid, HMOs help to maintain the acidic environment of the GI tract and protect against potentially pathogenic bacteria, through anti-adhesion effects, glycome-modifying effects and functioning as selectin-ligand analogues [52]. In addition to HMOs, breastmilk contains microbes commonly found in the infant gut [53], providing around 10^4-10^6 bacterial cells a day [54]. The presence of HMOs and live microbes in breastmilk may be why breastmilk is protective against disease [48,55]..

Due to the differing components of breastmilk and formula, diet has a significant impact on the gut microbiome of infants. The gut microbiome of breastfed infants is lower in diversity [56], but higher in abundance of commensal microbes [37,57], such as *Lactobacillus* and *Bifidobacterium* [37]. Although diversity may be an important factor for human health [15], the presence of specific microbes in early infancy may be more important for microbiome and GI development [6,58,59]. Microbes such as *B. breve*, *B. longum*, *B. dentium*, *B. infantis* and *B. pseudocatenulatum* [60,61], aid in proper establishment of gut flora and are commonly observed in breastfed infants. In contrast, the gut microbiome composition in formula-fed infants, including those on mixed feeds, lacks these key microbes and shifts towards adult-like composition sooner. The result is a microbiome of higher diversity and one dominated by

Staphylococcus, *Streptococcus*, *Enterococcus* and *Clostridium* [62,63]. Additionally, formula fed infants are more likely to harbour potential pathogens, such as *Clostridioides difficile* [64].

Maternal health

As much of the infant's microbial inoculation occurs through maternal-infant exchange, maternal health and medical interventions can also influence the infant microbiome. This includes interventions like antibiotics [27] and diseases like chorioamnionitis [8], and potentially other maternal microbiome-altering diseases, like type 2 diabetes [65] and preeclampsia [66]. Distortions in the maternal microbiome are passed on to the infant through previously described vertical transmission, demonstrated by increases in Proteobacteria [67] and other microbial alterations [46,56] observed in infants with antibiotic-treated mothers. The maternal influence on the infant is so great, that a study in 2016, conducted by Arboleya et. al. found that maternal antibiotic treatment had a greater effect than infant antibiotic treatment [67]. Vertical transmission of microbes from mother to infant means that maternal health can have a significant impact on infant health.

Characteristic microbiome of preterm infants

Preterm birth disrupts gut microbiome development [68]. This disruption results in a microbiome that is low in diversity [69,70], with fewer commensal microbes [71-73], high inter-individual variation [37,71], and a greater number of potential pathogens [72,74]. This includes potential pathogens *Klebsiella pneumoniae* [74] and *Clostridioides difficile* [72], and common commensals *Bifidobacterium* [8,71-73,75], *Lactobacillus* [69,71,76] and *Bacteroides* [37,72,77]. Unfortunately, reduced levels of *Bifidobacterium* have also been shown to persist throughout the first month of life [43], especially for the most preterm (<33 weeks gestation) [78,79]. This delayed colonisation could result from a breakdown in the relationship between the commensal microbes and intestinal cells, as sufficient maturation of the gut, specifically during the perinatal period, is essential for bifidobacterial colonisation [78]. However, these commensal microbes will eventually increase in abundance as the gut microbiome develops.

Over time the preterm infant microbiome experiences abrupt changes in composition [80,81] and increases in diversity [56,74,81,82], eventually becoming more similar to that of full-term infants [68]. The gut of vaginally delivered full-term infants immediately after birth provides a highly

aerobic environment for microbial colonisation. Therefore, colonisation of full-term infants begins with facultative anaerobes, such as *Escherichia coli* and different *Staphylococcus* spp. These microbes aid in a shift from an aerobic to anaerobic environment, allowing the colonisation of obligate-anaerobic commensal microbes, like *Bifidobacterium* and *Lactobacillus*. In contrast, the pre-term microbiome has a greater abundance of facultative anaerobes [37,73], fewer aerobes [83] and delayed colonisation of obligate-anaerobes [84], that do not become well represented until 12 weeks of age [84]. Although the pre- and full-term gut microbial communities are almost indistinguishable at two years of age, some evidence suggests the delay in colonisation of obligate-anaerobes may be long lasting with lower abundance of *Bifidobacterium* and *Lactobacillus* being observed at two-years of age [85]. So, despite significant changes over time, slight differences may persist, which could have long-term impacts on health.

Disease

The microbial disruption caused by preterm birth contributes to a higher disease burden in preterm infants [86]. This includes a higher risk of NEC, sepsis [9,87,88], asthma [89] and type 1 diabetes [90]. Such diseases are likely the result of some combination of microbial intolerance, disrupted immune system development and microbial imbalances. As the immune system is developmentally regulated, disruption already occurs because of prematurity. However, as microbial colonisation also supports immune system development, the irregularities of the preterm microbiome relative to that of those born full-term may contribute to further perturbation. Additive to this is an imbalanced microbiome that is unable to aid in the fight against pathogens, but rather promotes pathogen growth. Without a full-functioning immune system or commensal microbes to fight off and outcompete pathogens, preterm infants are at a greater risk of disease than their full-term counterparts.

Necrotising Enterocolitis

Necrotising Enterocolitis, characterised by intestinal inflammation and subsequent necrosis of the bowel, affects 4-13% of very low birth weight (< 1500 g) preterm infants, with 20-30% overall mortality [91]. However, even with its high prevalence and highly consequential outcomes its aetiology is not well understood. Although the exact aetiology is unknown, NEC is thought to result from a combination of microbial dysbiosis and an immature

gastrointestinal tract, with diet playing an integral role [92,93]. NEC onset has also been linked to antibiotic exposure [26,94], but human milk [55] and probiotics [95] are protective. The disruption of the microbiome in combination with immaturity of the intestinal mucosa causes a hyper-inflammatory response, and subsequently, NEC lesions [93]. However, NEC does not have a uniform microbial aetiology [96] and associations have been made with several pathogens, including many taxa belonging to the Proteobacteria phylum [97,98], including *Clostridium* [99,100] and *Klebsiella* [101], as well as low microbial diversity [69,94] and fewer commensal microbes [98,102]. Although various microbes have been implicated in the disease, the inconsistencies in the causative pathogen identified, coupled with the presence of these same microbes in healthy infants [26,99] suggests a polymicrobial aetiology.

Despite a complex pathology that is not well understood, the involvement of the gut microbiome in NEC has been well demonstrated by a consistent reduction in the incidence of NEC through probiotic supplementation. Several systematic reviews and meta-analyses, the largest of which contains 23 studies and > 7,000 infants, have shown that despite some level of heterogeneity in the literature, probiotic supplementation can reduce both incidence and associated death of NEC in very preterm infants [11,103], with the most effective probiotic formulations being those that contain *Lactobacillus* and *Bifidobacterium* species [10]. However, as stated by Sawh et al., heterogeneity in both probiotic organisms and treatment regimens has made it difficult to deduce best-practices [10]. In addition, the fact that infants developing NEC have very different gut microbial community structures between NICUs suggests that a successful probiotic protocol in one unit may not work in another [104]. Lastly, a point of concern is that in many countries, probiotics are considered supplements, and thus do not fall under the same stringent pharmaceutical regulations as other drugs. However, despite these lingering questions and concerns, it is clear that probiotic supplementation can aid in the prevention of NEC through modulation of the gut microbiome.

Sepsis

Like NEC, the gut microbiome plays an aetiological role in late-onset sepsis (LoS) [105], a blood-stream infection that affects 20% of extremely preterm infants [106]. Also, similarly to NEC, there is no single causative pathogen, and so it is not surprising that LoS has been positively associated with several potential pathogens in the gut. Positive associations include, but are not limited to, *Staphylococcus* [87,107-109] and Enterobacteriaceae [87]. In addition,

low diversity [81]and a lower abundance of *Bifidobacterium* [110] have also been associated with the disease. These irregular microbial profiles may increase the risk of LoS via disruption of the mucosal barrier, with resultant translocation of luminal contents [111,112]. With increased gut permeability, microbes or endotoxins can make their way into the lymphatic and circulatory systems, causing the systemic immune response known as sepsis.

NICU Environment

Due to higher disease and developmental burden, many preterm infants spend the early weeks of life in a neonatal intensive care unit (NICU). These units specialise in looking after preterm and sick newborn infants, and have strict protocols for hygiene, sanitation, feeding, visitation, and treatment. These NICU-specific protocols may contribute to NICU-specific microbiomes, as demonstrated by clustering of infants microbial populations by NICU, or within the same NICU but at different time points [113]. This may result in-part from different microbes present within different NICU environments, such as the dense biofilms observed in feeding tubes [82]. These feeding tubes are required for ongoing nutritional support in the early weeks/months of life but may also become a site that supports the growth of nosocomial bacteria. However, as treatment regimens can also modulate or perturb microbial populations, as demonstrated by reduced diversity and altered bacterial profiles in antibiotic-treated infants [54], and the promotion of microbes and increased diversity [13,114] in probiotic supplemented infants, differing intervention protocols may also be selecting for specific microbial populations.

Antibiotics

The disproportionate disease burden placed on preterm infants leads to microbiome-altering treatment with antibiotics, a staple in preterm neonatal care. Despite being life-saving in many contexts, antibiotics can disrupt microbial acquisition, resulting in reduced diversity [8] and altered bacterial profiles [54,115,116]. This includes reductions in commensal microbes, like *Bifidobacterium* spp. [46,117], and an increased abundance of potential pathogens, like *Staphylococcus* spp. [109]. Treatment with antibiotics selects for microbes harbouring antibiotic resistant (ABR) genes. Unfortunately, many of these ABR-harbouring taxa are pathogenic [81,118], which is why antibiotic-treatment has also been associated with acute diseases, like NEC [94,119] and Sepsis [108,109], and chronic diseases like Asthma [120]. However, despite

associations with chronic diseases later in life, evidence suggests that antibiotics have a limited effect on the microbiome post-discharge [42]. The mechanism by which antibiotics exert long lasting effects on infants may be through disruption of immunological programming via microbial disruption. So, although antibiotics have an important role to play in neonatal care, unnecessary, excessive, or untargeted use may be counterproductive in promoting better health outcomes.

Probiotics

Probiotic supplementation is now common practice for very preterm infants (<32 weeks gestational age) in Australia. A 2015 study identified that sixteen of the twenty NICUs in Australia use *L. acidophilus* and *B. bifidum* (Infloran®) in some variation of a protocol similar to 2×10^9 colony-forming units per day from the first feed and through to 34 weeks' corrected age [121]. Despite some heterogeneity [68,122], cumulative evidence suggests such probiotic supplementation protocols contributes to infant health through positive modulation of the gut microbiome. Potential causes of the heterogeneity includes the use of different probiotic species [11], methodological variability [123], probiotic impurities/irregularities [114,124], and confounding variables [10,125]. However, despite this heterogeneity, it is now well accepted that probiotic supplementation increases the abundance of commensal microbes (including the probiotic strains), as well as a reduction in potential pathogens in infants [12,13]. These positive changes may also persist, with several studies showing Bifidobacterium remains viable after probiotic supplementation [13,126], and that the differences in the microbiome of supplemented and non-supplemented infants remain post-discharge [42]. These probiotic-induced microbial changes may support metabolic and immune-system development, helping to prevent both acute and chronic pathologies, which would explain the reduced NEC-incidence in probiotic-treated infants across several clinical trials [10,11]. However, if the benefits of probiotics extend beyond acute mitigation of disease, then could they also benefit older preterm infants? In addition, if probiotics induce positive microbial modulation, then what clinical associations previously identified still exist in an entirely probiotic-supplemented cohort?

Aims

The overall aim of this thesis was to describe the faecal microbiome of preterm infants using a combination of metagenomic sequencing technologies in a cohort of preterm infants born in North QLD, Australia, and to explore its relationship to a host of variables.

Aim 1: Explore changes in the bacterial microbiome of very preterm probiotic-supplemented infants from admission to discharge. This aim is addressed in Chapter 3.

Aim 2: Explore the effect of clinical variables (both maternal and infant) on the developing very preterm probiotic-supplemented infant bacterial microbiome. This aim is addressed in Chapter 3.

Aim 3: Investigate differences in the bacterial microbiomes of probiotic-supplemented and non-supplemented preterm infants during hospital admission. This aim is addressed in Chapter 4.

Aim 4: Determine if probiotic microbes colonise the infant gut. This aim is addressed in Chapters 4 and 5.

Aim 5: Investigate differences in the bacterial microbiomes of probiotic-supplemented and non-supplemented preterm infants post-discharge from the hospital. This aim is addressed in Chapter 5.

Final Notes

This thesis contains published papers, and thus the papers are written for both different audiences and specific journals. Minor adjustments were made to the formatting to suit the style of an academic thesis, such as changing the language to first person singular. However, the styles may differ slightly depending on the journal requirements.

The next chapter (Chapter 2) is a methodological review that assesses the most common practices used for characterising the preterm infant microbiome and will elucidate the reasoning behind my selection of methodologies in the subsequent chapters.

2. Methods for Exploring the Faecal Microbiome of Premature Infants: A review

In the introduction, I highlight the variability in results from different studies, and suggest that this, at least in part, stems from variability in the methodologies used. Research into the microbiome is still a relatively new frontier of science. The rapid progression of this area of research has been due, in part, to rapidly evolving genomics technologies and methods. Although this progress will undoubtedly benefit microbiome research, it also makes it difficult to navigate study design and both the interpretation and synthesis of literature. Thus, to better understand the varying methodologies involved in exploring the preterm infant gut microbiome, my second chapter is a review titled "Methods for exploring the faecal microbiome of premature infants: a review", which was published in Maternal Health, Neonatology and Perinatology (https://doi.org/10.1186/s40748-021-00131-9).

Abstract

The preterm infant gut microbiome plays an important part in infant health and development, and recognition of the implications of microbial dysbiosis in preterm infants has prompted significant research into these issues. The approaches to designing investigations into microbial populations are many and varied, each with its own benefits and limitations. The technique used can influence results, contributing to heterogeneity across studies. This review aimed to describe the most common techniques used in researching the preterm infant microbiome, detailing their various limitations. The objective was to provide those entering the field with a broad understanding of available methodologies, so that the likely effects of their use can be factored into literature interpretation and future study design. I found that although many techniques are used for characterising the preterm infant microbiome, 16S rRNA short amplicon sequencing is the most common. 16S rRNA short amplicon sequencing has several benefits, including high accuracy, discoverability and high throughput capacity. However, this technique has limitations. Each stage of the protocol offers opportunities for the injection of bias. Bias can contribute to variability between studies using 16S rRNA high throughout sequencing. Thus, I would recommend that the interpretation of previous results and future study design be given careful consideration.

Introduction

The preterm infant gut microbiome has become an important, modifiable factor in the field of neonatal intensive care. Compared with infants born full-term, the characteristic microbiome of preterm infants (born <37 weeks gestation) is dysbiotic: highly variable [37,71,115], low in diversity [8,26,69], low in common commensals [37,69,75], and harbouring more potential pathogens [72,74]. This dysbiotic microbiome composition puts immune-compromised preterm infants at an increased risk of acute and chronic disease, and developmental abnormalities [38,101,127]. Preterm infants are also more likely to be born via caesarean section, be formula fed, receive antibiotics, and spend much of their early life in a clinical environment, all of which occurrences have the potential to exacerbate the microbial dysbiosis [37,56,113,115].

Unfortunately, the understanding this microbial composition is confounded by the diversity of investigative methods used. Methodologies for examining the microbiome are complex, technically challenging and vary between laboratories. Therefore, it is impossible to rule out

protocol bias as a factor contributing to the variability seen between studies. In fact, a number of studies have demonstrated the role of methodological bias in influencing the outcomes of microbiome analysis [128,129], thus contributing to significant heterogeneity in results between studies.

This review aimed to describe the most common techniques used in researching the preterm infant microbiome. I chose to focus on those studies investigating preterm infants, due to the explosion of interest into this area, although these techniques can also be applied to microbiome study of full-term infants. This magnified interest likely stems from the disproportionate health burden placed on preterm infants and its link to the microbiome. The objective of this review was to provide those entering the field, particularly those in neonatal clinical care, with a broad understanding of the different methods used, so that literature interpretation and future study design can be enhanced. This review identifies and describes the most commonly used methods for examining the preterm infant's microbiome and maps this information against studies comparing efficacy of techniques. This process is designed to illuminate which techniques will be most appropriate for the examination of the microbiome of preterm infants

Methods

Search and Eligibility Criteria

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach (*Figure 2.1*), was taken to search for relevant literature up until August 2020. Studies investigating the preterm microbiome were identified via searches in the SCOPUS and PubMed databases using the search terms; "Microbiota" AND "Infant" AND "Premature" AND "Faeces". Journal articles describing a wide variety of study designs, sample sizes, interventions, comparators and outcomes were included in this review. Articles were excluded if they were not original studies, were case studies, were not in English, were unable to be accessed or did not specifically investigate the preterm infant microbiome specifically. Reviews found in the initial search were also used to locate other papers that addressed the review question.

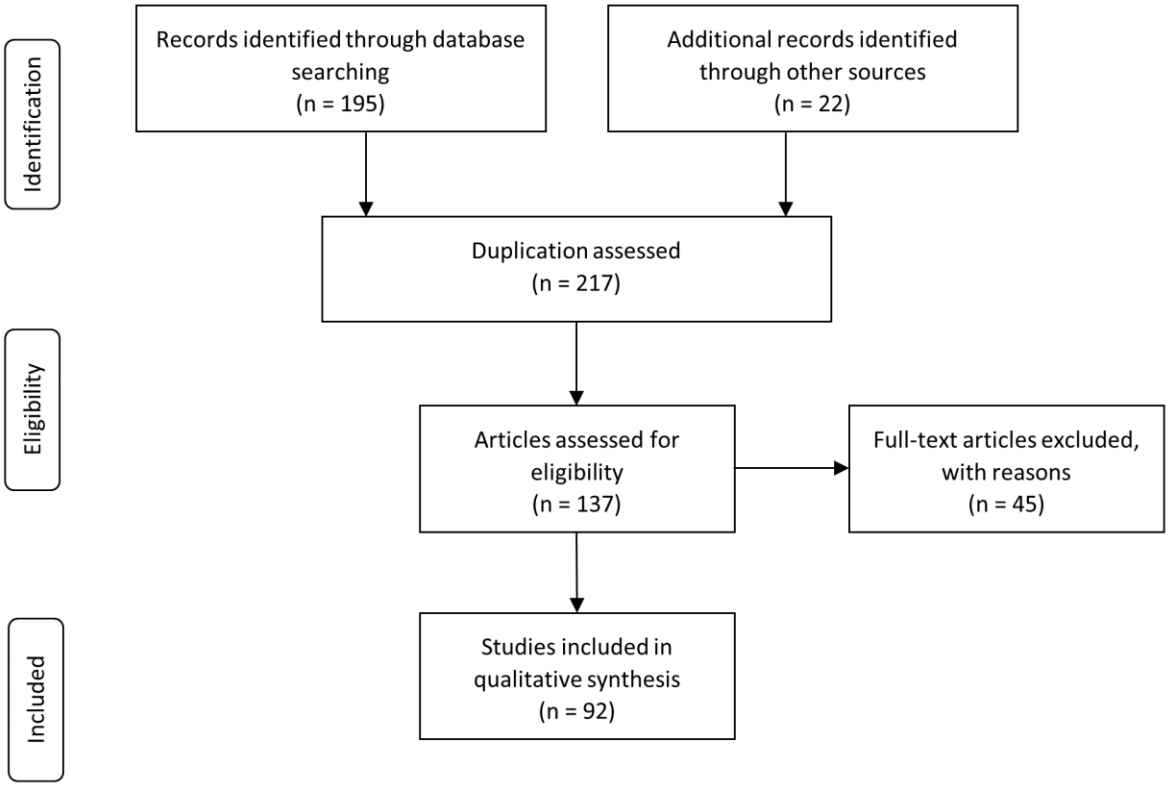


Figure 2.1 PRISMA flow diagram describing the process of study collection and inclusion.

Data Collection Process

A standardised data collection protocol was established to extract all relevant information for qualitative analysis. Author, date of publication, aims/hypotheses, a summary of the methods, a summary of the findings and limitations were recorded. Methodology-specific information was also collected for primary techniques, secondary techniques, storage and DNA extraction. Emphasis was placed on 16S rRNA short amplicon sequencing, as this methodology was the most common primary technique, and further information was collected for target variable regions, platform, pipeline and reference databases.

Results

The review of the literature explored the methodological diversity used in the study of the preterm infant microbiome, and a summary of the major techniques used across the studies is presented in *Figure 2.2*. The outcome of the systematic review is summarised in *Figure 2.1*.

Two hundred and seventeen articles were identified. Of these 137 articles remained after duplicates were removed, with a further 45 articles removed after assessing the full article for eligibility. A total of 92 articles were reviewed. There was a surprising lack of detail displayed in the methods section of many studies, despite there being several technical choices at each stage of the workflow with the potential to contribute to bias. The summary information is based only on the data that was made available.

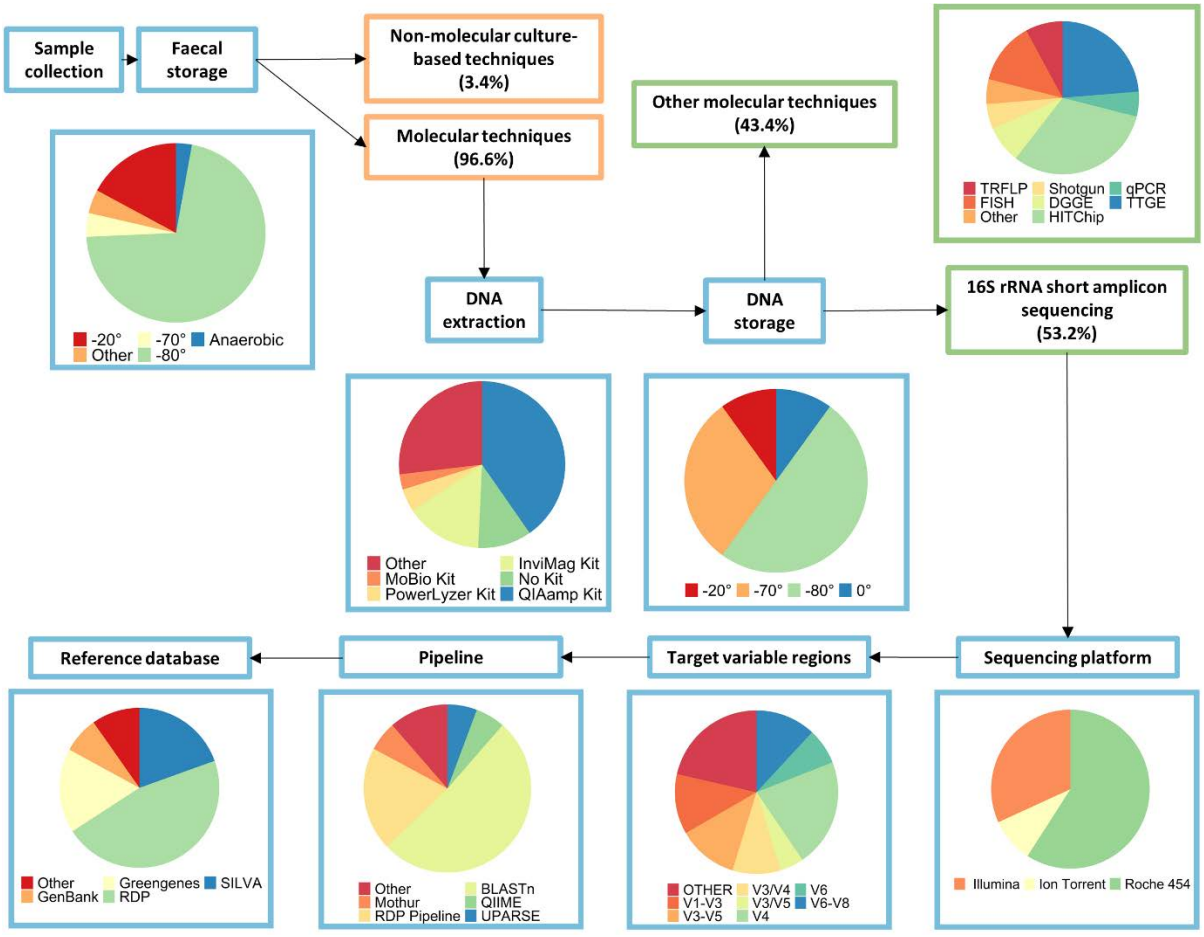


Figure 2.2 Flow diagram describing different workflows and proportions of techniques used for microbiome analysis in preterm infants.

Primary Techniques

Thirteen techniques for characterising the preterm infant gut microbiome were identified (*Figure 2.2*). A wide range of techniques have been used as the primary tool for microbial compositional analysis including:

- Traditional culture-based techniques [130-134],

- 16S rRNA short amplicon sequencing [8,9,13,42,45-47,54,67,69,71,73,75,77,80,84,87,91,96,97,99,102,108-110,113,117,118,122,135-154],
- Shotgun metagenomics [115],
- Temperature gradient gel electrophoresis (TGGE) [70,72,78,79,101,155-158],
- Denaturing gradient gel electrophoresis (DGGE)[74,107,116,159,160],
- HITChip [76,85],
- Fluorescent In-Situ Hybridisation Analysis (FISH) [161,162],
- Terminal restriction fragment length polymorphism (TRFLP) [94,163,164],
- Quantitative PCR (qPCR) [43,47,68,72,83,114,165-170],
- Long-read nanopore sequencing [171], and
- Random amplified polymorphic DNA/pulsed-field gel electrophoresis (RAPD/PFGE)[82].

Molecular techniques dominated, specifically 16S rRNA short amplicon sequencing, which made up 53.2% of primary techniques. qPCR was the second most commonly used technique, used in 13% of studies, with the remaining techniques being used by ≤10% of studies each. Traditional non-molecular techniques using selective and differential agar medium represented a very small fraction of the primary techniques used (5.4%), and a further four studies used culture techniques as a secondary method.

Storage Conditions and Extraction Protocol on DNA for Molecular Techniques

Sample storage protocols were consistent across the molecular techniques. However, the DNA extraction techniques used were highly variable. Freezing at -80°C dominated storage methods for both faeces (71.4%) and DNA (50%), with non-freezing protocols only used in 2.9% of studies. DNA extraction was the area demonstrating the greatest variability, with 15 different methods utilised. The QIAamp DNA Stool Kit, a kit that combines heat, chemical and enzymatic lysis was the most commonly used (40.3%), with using no kit at all (14.9%) the second most common option. The PowerLyzer PowerSoil kit was third (10.4%), with the other twelve kits making up the remaining 34.4%.

16S rRNA Amplicon Sequencing – Specific Techniques

The most common molecular technique was 16S rRNA short amplicon sequencing. However, the methods used were highly variable in sequencing platforms, variable target regions, and pipelines. Roche 454 (57.8%) sequencing platform was the most commonly utilised of the four, with Illumina second (31.1%). The use of this platform has increased in recent years. V4 was the most common variable target region used (22.7%). However, there were sixteen unique combinations used across the 92 studies. Of the eight pipelines used, QIIME/QIIME2 (Quantitative Insights Into Microbial Ecology) [172] made up half of the pipelines used, with Mothur [173] a distant second (20%). The Ribosomal Database Project (46.3%) was the predominant reference database used, with SILVA (19.5%) and Greengenes (17.1%) being used in most of the remaining studies. Techniques specific to 16S rRNA short amplicon sequencing varied greatly, despite it being the most common method.

Trends Over Time

Many outdated techniques and tools are being abandoned for newer, more robust methods. High throughput molecular techniques have become more commonly used over time, especially in 16S rRNA amplicon sequencing, which was first used in 2004. This upward trend in 16S rRNA amplicon sequencing is coupled with a decline in both fingerprinting- and culture-based techniques, with all culture-based studies occurring prior to 2015. There is also a trend towards the use of Illumina platforms and pipelines that use error modelling within 16S rRNA amplicon sequencing.

Discussion

Techniques for examining the microbiome can be categorised into two main groups, molecular and non-molecular. Molecular techniques have become dominant due to their depth of analysis, speed and cost reduction. Nevertheless, there are several techniques to choose from even within molecular methodologies, and within a given molecular technique there is variation possible in protocols. This lack of consistency can contribute to the inconsistencies in results between studies into the preterm microbiome.

Culture Based Approaches

Few studies still rely solely on traditional non-molecular methods for microbiome characterisation, with the most recent study under review occurring in 2014 [132]. Non-molecular techniques are based on traditional microbiological methods that involve growing microbial communities on predetermined growth media under strict laboratory conditions designed to optimise growth. Methods vary depending on the type of micro-organisms present and downstream applications. Techniques include broth culture, enrichment and microbial identification. Examples of growth media include Luria Broth, also known as Lysogeny Broth [174], which is common for the cultivation of *Escherichia coli*, and selective agars such as blood, MacConkey [175] or Xylose Lysine Deoxycholate agar [176], which are specific for other taxa prevalent in the gastrointestinal tract. Microorganisms are placed in growth medium and left to grow under strict conditions, giving them time to grow into individual colonies. Colony morphology can then be used to determine specific taxa and colony counts, and are used for the calculation of concentrations and serial dilutions. These techniques are primarily used to identify specific microorganisms of interest due to their specificity, and employed as diagnostic tools for the detection of pathogenic species.

Non-molecular techniques can be useful despite their limitations in sensitivity and specificity, particularly for anaerobic species, as well as for discovery and scaling. They can improve the robustness of results via identification of specific species of interest or identification of unidentified sequences that may belong to a known organism [177], when used in combination with molecular techniques, such as 16S rRNA high throughputs sequencing. Other major benefits of non-molecular techniques include that the materials are inexpensive and that the protocol requires limited equipment. However, specific culturing conditions that select for specific microbes, of which there must be prior knowledge, mean that many species can go undetected [178,179]. Moreover, they are time consuming and labour intensive when performed at large scale. Thus, traditional culture techniques have largely been displaced by molecular techniques due to these time and labour issues, as well as these older techniques' restricted insight into microbial communities.

Molecular Based Approaches

Molecular techniques, including 16S rRNA high throughput sequencing, fingerprinting, microarrays and quantitative PCRs, are rapid, sensitive and highly specific, particularly for commensal organisms. Molecular techniques have rapidly replaced non-molecular techniques for use in identifying microbiome composition since their advent, due to these benefits (*Figure 2.2*). The utilisation of genetic information to differentiate between taxa has made a more detailed exploration possible, and may provide information on the abundance and composition of these microbial communities beyond those routinely grown in the laboratory. The most described microbiota include bacterial communities, which can be identified through utilisation of the variable regions of the 16S ribosomal RNA (16S rRNA) gene, which is flanked by highly conserved regions. DNA is extracted from faeces in this method, commonly using a commercially derived extraction kit. It is amplified by PCR and then differentiated into groups based on similarity to identify the taxa present, allowing deep community sampling. Samples must first be collected and stored and the DNA extracted for all 16S and other molecular techniques.

Sample Collection

Sample collection protocols will vary depending on study design. However, the timing of collection is an important factor to consider when comparing results across studies or during study design. Most studies provide specific time points, based on the gestational or post-gestational age of the infants. However, there are studies that group samples together more broadly, for example, binning samples together as “early-infancy” or meconium. Meconium is the earliest stool of a mammalian infant, comprised of a thick tar-like substance that lines the intestine of the unborn infant. Typically, meconium is not released until after birth. However, sometimes it will be released into the amniotic fluid prior to birth. It can also be released at different time points post-delivery, typically within the first three to five days. It may be that accurate comparisons cannot be made across studies when using such broad definitions, as the infant microbiome is dynamic, with choregraphed abrupt changes in composition [81,84]. Therefore, the timing of collection is an important factor to take into consideration when interpretating the literature and planning future study design.

Impact of Storage Conditions and Extraction Protocol on DNA for Molecular Techniques

Sample Storage

Storage conditions influence the stability and constitution of faecal microbial communities [180], which could prejudice study conclusions. Inadequate storage can lead to continued growth of specific organisms, altering the proportions of taxa/genera in a sample, and can lead to DNA/RNA fragmentation. Studies show DNA/RNA fragmentation can occur after 24 hours when samples are stored at room temperature [180], and that significant changes in bacterial communities can occur in samples after this time [180-182]. Potentially, changes can occur in as little as 30 minutes, as demonstrated by Gorzalek, et al. [183]. Currently, available storage methods include freezing or refrigeration at different temperatures, and the use of anaerobic incubation systems, aqueous storage/transport mediums and faecal occult blood tests.

Freezing and refrigeration

Optimal sample storage conditions depend upon the duration of storage. If samples are to be processed immediately, storage on ice for up to 48 hours [184], or 4°C for 24 hours [185] appear to be sufficient for sample preservation. However, immediate freezing of faecal samples to inhibit bacterial growth is the optimal procedure for longer term storage. Long term storage of faecal samples at -80°C has been shown to yield microbiota similar to that of fresh samples [181,186,187]. Storage of faecal samples at -20°C also has shown similar efficacy in sample preservation across several studies [180,188,189]. However, this appears to be time limited, with some studies reporting changes in taxa over longer-term storage at -20°C for storage times greater than a week, resulting in significant changes to *Bacteroides* spp. [190] and for up 53 days in the Firmicutes to Bacteroidetes ratio [191]. Storage at -80°C produces the most consistent results and appears to be the most common preservation method (*Figure 2.2*). It was common for samples to be stored at -20°C or 4°C temporarily, or to use non-freeze methods until storage at lower temperatures was possible in situations where immediate freezing was not possible, such as with at home collection.

Other storage techniques

Immediate freezing of faecal samples can be logistically difficult, especially for large-scale population-based studies, and freeze-thawing effects may significantly diminish sample integrity. Therefore, other preservation methods may be better suited for some protocols. These preservation techniques include chemical and drying preservation such as DNA/RNA Shield and anaerobic incubation systems. Preservation buffers, aqueous reagents that stabilise and protect cellular DNA/RNA, like DNA/RNA Shield (Zymo Research) and RNAlater (Thermofisher), may also preserve genetic integrity for weeks without refrigeration or freezing [192-198]. Using these buffers ensures that samples are also protected from the potential stress caused by freeze-thawing effects. These buffers and other non-freeze preservation methods are a good alternative when freezing is not feasible.

Some potential issues have been highlighted with non-freezing preservation methods, despite these methods being a more practical alternative. Preservation buffers may result in lower diversity [199,200] relative to immediate freezing. Moreover, some older preservation buffers may impede downstream DNA extraction and amplification of target genes [192]. Anaerobic incubation systems, like Anaerocult®, are only effective for storage of anaerobic strains and thus have obvious limitations. Most current research still supports the efficacy of preservation buffers, despite several studies highlighting their limitations. Thus, both freezing at -80°C and suspension in a stabilisation buffer are acceptable practices when considering all the available storage options.

DNA Extraction

The first step for molecular analysis is DNA extraction, which can be carried out using commercially available kits. Extraction is an important step in molecular techniques, which involves separating DNA from the other cellular material contained within samples of interest. The process involves cell lysis, or the disruption of cell walls, separation of the DNA from the other cell components and its subsequent isolation. DNA extraction can be laborious and carries a high risk of sample contamination, as a significant amount of handling of the biological material is involved. Fortunately, there are several commercially available extraction kits that make the process less laborious, more streamlined and more reproducible due to the widespread interest in the human microbiome.

The amount of tissue needed for DNA extraction and sequencing is dependent on the extraction methods and downstream application respectively. Generally, 100 to 1000 nanograms of DNA is required for whole genome sequencing, and as little as 1 to 10 nanograms for amplicon sequencing [201]. The amount of stool required will be dependent on the efficacy of the protocol in extracting the DNA, which is dependent on the methods used for the given protocol. The QIAamp DNA Stool Kit, the most commonly used in the preterm infant microbiome field, is optimised for 190-220 milligrams of stool, but as mentioned above, the amount of stool required is kit-dependent.

Unfortunately, different extraction protocols and kits can contribute greatly to variation in microbial community structure [202,203], introducing bias to outcomes [190,204,205]. In one study the extraction method was demonstrated to be the second-greatest contributing factor to variation [202]. This variation arises in large part due to different methods of homogenisation and lysis. These steps are critical, as different stool fractions can contain different microbial compositions, and different microbes are lysed better by different techniques.

Microbial cell wall structure differs between Gram-negative and Gram-positive bacteria and require different lysis methods for DNA extraction. In Gram-negative bacteria, the cell wall is thin and made up of both a peptidoglycan and phospholipid bilayer containing lipopolysaccharides, whereas Gram-positive bacteria have a thick peptidoglycan cell wall. As a result, Gram-positive microbes require more vigorous lysis methods and Gram-negative microbes are more easily lysed [190].

Different forms of homogenisation or lysis can, therefore, contribute to bias by not effectively disrupting the cell wall of all microbes present in a sample or, conversely, by destroying the DNA of easily lysed cells. Mechanical, chemical, and enzymatic lysis methods can also produce different proportions of taxa, with mechanical methods producing higher bacterial numbers and greater diversity [190]. Two comprehensive studies that explored several kits, including the QIAamp DNA Stool Kit, found that the International Human Microbiota Standards (IHMS) Protocol Q, that includes mechanical lysis, performs best across several parameters [203,206]. Despite this, this review found that the most common extraction method utilised in preterm infant microbiome studies was the Qiagen QIAamp DNA Stool Kit. This method uses a combination of heat, chemical and enzymatic lysis, with some studies adapting the protocol to add mechanical lysis through bead beating. Unfortunately, different bead-beating instruments have been shown to produce bias [203,207,208]. Despite this, mechanical

disruption is essential for comprehensive profiling of the human gut microbiome [208,209], and until a standardised protocol is established, researchers must be careful to consider the bias generated through different kits.

Molecular Techniques

Fingerprinting Methods

The increasing usage of 16S rRNA amplicon sequencing has been matched with a reduction in the use of other techniques, like different fingerprinting methods. Fingerprinting methods are more cost effective and faster to perform [210], although high throughput sequencing techniques provide a broad detailed analysis of microbial communities. These techniques are favoured in comparison to traditional culture methods, as they provide greater sensitivity and specificity for individual organisms, and can be used to analyse large numbers of samples [210]. Broadly speaking, fingerprinting methods provide a profile of microbial communities that uses amplification of a target gene (commonly the 16S rRNA gene) and the utilisation of gel electrophoresis to observe physical separation of amplicons, allowing exploration of highly abundant taxa. Fingerprinting methods have been used in studies exploring the preterm microbiome (18.5%), although they are currently less common. These techniques include denaturing/temperature gradient gel electrophoresis (D/TGGE) and terminal restriction fragment length polymorphism (TRFLP), as well as denaturing high performance liquid chromatography (dHPLC) in a single study. dHPLC uses liquid chromatography to identify polymorphisms [211], while the others rely on electrophoresis to differentiate between sequences, although all are considered fingerprinting methods.

Denaturing/temperature gradient gel electrophoresis

Gradient electrophoresis is the size dependent movement and separation of dispersed nucleic acids through an acrylamide gel. As DNA has a negative charge, it moves through the acrylamide gel or molecular mesh towards the positive electrodes at a rate that is inversely proportional to the size of the nucleic acid sequences, thus allowing the differentiation of different sized sequences [212]. More detailed exploration is achieved by applying either a temperature (TGGE) or chemical gradient (DGGE) to denature the samples as they move across acrylamide gel, based on the chemical make-up of the sequences [213,214].

Both DGGE and TGGE differ in the mechanism of DNA denaturation. In DGGE the nucleic acids are exposed to increasingly extreme chemical conditions, leading to the denaturation of the DNA in a stepwise process. This allows for visualisation of the sequence differences by their position on the gel. The method relies on differences in the ability to denature the bases, which is determined by base pair sequences to separate genes by size. In contrast, TGGE uses a temperature gradient in combination with the electrophoresis. Strands separate across the gel depending on base-pair content, with smaller molecules travelling faster [214] as the temperature increases.

Terminal restriction fragment length polymorphism (TRFLP)

TRFLP also uses electrophoresis to differentiate between sequences based on terminal restriction fragment size, like TGGE and DGGE. This allows sequence identification for microbial community profiling [215]. The method involves PCR amplification of a target gene with fluorescently labelled primers and subsequent digestion with restriction enzymes. The sizes of the different terminal fragments are then determined by separating the fluorescently tagged terminal fragments via capillary or polyacrylamide electrophoresis in a sequencing gel, creating unique banding patterns allowing identification of microorganisms [216]. T-RFLP has high throughput capability and can be highly sensitive, but it also has limited accuracy as incomplete or non-specific digestion can lead to overestimation of diversity. Homology of sequences can contribute to an underestimation of taxa present [217]. Furthermore, libraries must be built prior to analysis.

dHPLC

dHPLC uses liquid chromatography to identify DNA polymorphisms [218], unlike TGGE, DGGE and T-RFLP. DNA strands are separated into hetero- and homoduplexes using an ion-pair, reverse-phase liquid chromatography on a poly alkyl column matrix. [218], following partial heat denaturation. The presence of polymorphisms is revealed by the differential retention of these homo- and heteroduplex DNA fragments [219]. Heteroduplexes are double stranded DNA that have formed during PCR amplification that are mismatched at the site of mutation. Mismatched double stranded DNA fragments have reduced retention on the column matrix, and subsequently in a reduced retention time, thus allowing for identification of

polymorphisms. As a result, dHPLC can be useful as a screening test for mutations that may be involved in diseases or associated with antibiotic resistance [211,220]. The method has also been used to differentiate between taxa at species depth by applying the same underlying principle of scanning for mutations to the detection of sequence variations between PCR-amplified bacterial 16S rRNA genes [221], and also as a tool for re-sequencing of genomes [219].

Limitations (Pros and Cons)

Some consider dHPLC to be the optimal fingerprinting method [70], potentially allowing identification of bacteria at the species and/or biotype levels [221]. However, these techniques require extensive downstream processing, can produce PCR bias [178,222] and have limited detection depth, as it is difficult to relate banding patterns created in gels to species or lineages created by fingerprinting methods [223]. Thus, fingerprinting methods are usually limited to identification at the order/family level [224], and to only the most abundant organisms. This methodology also makes it difficult to combine data from multiple studies into a single analysis [223]. Fingerprinting techniques can be useful for exploring dominant members of microbial communities, including clustering of communities based on dominant members [225]. However, their application is limited in describing entire microbial communities.

Phylogenetic Microarrays

Microarrays were originally developed to monitor gene expression, but their application has been expanded to include comparative genomics, DNA sequencing analysis, single-nucleotide polymorphism (SNP) analysis and microbial detection [226], including studies on the preterm infant microbiome [76,85]. Microarrays are microscopic slides printed with probes made of predefined oligonucleotide sequences complementary to the small subunit (SSU) rRNA. The oligonucleotide probes detect gene expression or mRNA transcripts expressed by specific genes and extracted from target organisms. Reverse transcriptase converts mRNA into complementary DNA (cDNA), and this cDNA is fragmented and fluorescently labelled and added to the microarray [227]. cDNA then binds complementary oligonucleotide probes via hybridisation, and measurement of the observed fluorescent intensity at a given probe is an indication of the abundance of predetermined sequences that are chosen prior to analysis and

are of interest [227]. This makes phylogenetic oligonucleotide arrays (phyloarrays), including HITChip, suited to the analysis of microbial communities.

HITChip is an ecosystem-specific phylogenetic microarray developed for microbial detection in the human gastrointestinal tract [228,229], and is the only microarray to be used in studies on the preterm infant gut microbiota. HITChip is an oligonucleotide microarray that uses 4,800 oligonucleotide probes based on two hypervariable regions (V1 and V6) of the 16S rRNA gene, identifying 1,140 phylotypes. Phylotypes were designed following the analysis of 16,000 human gastrointestinal tract 16S rRNA gene sequences [229]. As a result, HITChip is highly specific to the human gastro-intestinal tract microbiome and provides a high level of diversity.

Limitations (Pros and Cons)

Benefits of microarrays include ease of use, speed and cost [230], and potential for investigating microbial gene functionality [231]. These intermediate methodologies allow processing of large sample sizes, while providing more taxonomic depth, like fingerprinting methods. Microarrays target the ribosomal RNA gene, allowing comparisons of diversity and taxonomy, and thus display similar robustness [232,233], like 16S rRNA sequencing. However, when compared to high throughput techniques, phylogenetic arrays are limited when assessing new lineages, as they can only detect predefined taxa [223,234]. Other methodologies are better suited when there is potential for taxonomic or gene discovery, as microarrays are limited to predefined taxa. Microarrays are not commonly used in studies on the preterm infant gut microbiome, as it is a relatively new area of study, and other methodologies may be better suited for characterising this niche.

qPCR and Fluorescent In-situ hybridization

Polymerase chain reaction (PCR) is a highly sensitive molecular technique that was originally developed for detection of DNA/RNA sequences, but has since progressed beyond purely nucleic acid detection. Quantitative polymerase chain reaction (qPCR) builds on standard PCR by providing the quantity of amplified genes. qPCR also differs from standard methods as it monitors the amplification of targeted DNA molecules in real time or during PCR instead of at the end. This process allows not only detection, but also quantification and

characterisation of nucleic acids [235]. Fluorescent dye is added to the PCR reaction in dye-based qPCR, and the fluorescent signal increases proportionately to the quantity of DNA being replicated. This allows quantification of DNA after each cycle. However, qPCR only allows one target to be examined at a time, thus throughput is limited. The more accurate probe-based qPCR provides one way around this drawback, by simultaneously examining multiple targets via recognition of sequence-specific probes. The fluorescent signal from the probe in probe-based qPCR is proportional to the target sequence that is present in the reaction [236], as it is in dye-based qPCR.

Fluorescent in situ hybridisation (FISH) is another probe-based technique. FISH is a molecular technique that uses complimentary binding to identify or quantify cDNA that can be used for microbial identification, like microarrays and qPCR [237]. FISH uses fluorescently labelled DNA probes that match specific DNA sequences that can be observed under a microscope, allowing direct quantification of specific taxa. Fluorescent oligonucleotide probes are created for targets, either 16S or 23S rRNA sequences. The target and probe sequences are denatured with heat or chemicals, and mixed together prior to hybridisation. Hybridisation then occurs between complementary target and probe sequences, with fluorescence microscopy facilitating detection of hybridisation via observations of fluorescently labelled cDNA. This target-specific methodology facilitates high accuracy when targeting specific microbes.

Limitations (Pros and Cons)

Both qPCR and in-situ hybridisation can provide highly accurate quantification [238], can be highly sensitive [239], and can produce similar results to metagenomic methods when considering the main intestinal microbial groups [136]. However, they are limited in their application, as prior knowledge of sequences is required, like fingerprinting and microarrays. Thus, these methods have no discovery power and no capacity for assessing diversity. FISH has been designed to examine the major microbial groups present in preterm infants but is based on groups present in full-term infants [161,162]. However, predefining taxa in this way is a significant limitation, as preterm infants are known to have significantly different microbial populations to infants born full-term [67,136]. Moreover, both qPCR and FISH are not scalable, and therefore are only effective for low target numbers. qPCR or in-situ hybridisation methods may be beneficial when specific populations are being targeted, as they have limited

bias and are cheaper compared to sequencing methods, but they are not suitable for projects mapping entire microbial ecosystems, like that of the preterm infant gut microbiome.

Sequencing Techniques

DNA sequencing is the process of nucleic acid sequence determination, and covers a broad range of techniques across three generations of sequencing. The first generation of sequencing began with a low throughput technique, Sanger sequencing, which only sequenced a single DNA fragment at a time [240]. Sanger uses a labour-intensive cell-based amplification step, involving cloned sequences being placed into plasmids for amplification, prior to extraction and purification [240]. The second generation of sequencing techniques, often referred to as next-generation sequencing (NGS) or high-throughput sequencing (HTS), involved 16S rRNA Metabarcoding and Metagenomics (shotgun sequencing). NGS refers to any sequencing method using the concept of parallel processing. This parallel processing increased the volume of reads per run to millions, vastly improving efficiency, as did the development of a cell-free system. NGS also runs elongation and detection steps in parallel, again improving efficiency [241]. However, NGS technologies are limited in that they use short reads (bp), which create a computational challenge when assembling or mapping to genomes. A third generation of sequencing was developed to overcome this challenge, long read or single molecule direct sequencing. The capacity of all sequencing technologies to produce large volumes of relatively accurate data, coupled with the continual reduction in cost, has led to their adoption across most modern studies investigating microbial populations. 16S rRNA high-throughput amplicon sequencing (metabarcoding) is now the most common method used for studies specifically characterising the preterm infant's gut microbiome (*Figure 2.2*). However, other methods, including shotgun metagenomics, and third generation single molecule direct (long-read sequencing), have also been applied. All techniques described have their strengths and limitations.

Next Generation Sequencing

16S rRNA amplicon sequencing

16S rRNA amplicon sequencing, or metabarcoding, has become the most common technique for characterisation of the preterm infant microbiome since it was first used in 2004. 16S

rRNA metabarcoding uses high throughput sequencing to target variable regions of the 16S rRNA gene, allowing accurate identification of microbial community composition [242-244]. The 16S rRNA gene codes for 16S ribosomal RNA, a component of the 30S small subunit of prokaryotic ribosomes. The 16S rRNA gene is highly conserved across taxa, but also has several variable regions allowing differentiation between taxa, due to a slow rate of evolution. The variable regions are conserved enough that most taxa can be characterised, but variable enough that taxa can be differentiated. There are nine of these hypervariable regions that range in base pair length and are involved in the secondary structure of the small ribosomal subunit. The regions vary in conservation, and thus different regions correlate with different levels of taxonomic resolution. The protocol for 16S rRNA gene amplicon sequencing involves DNA extraction, PCR amplification of the variable target region(s), grouping of sequences into OTUs, ASVs or an equivalent, and then mapping these sequence variants to a reference database for taxonomic identification.

16S rRNA metabarcoding is the most common technique for characterising the preterm infant. However, despite this there is no predominant protocol. There are a myriad of options at all stages of the workflow, all of which can introduce bias that alters outputs, which is supported by the observation that samples cluster by study [128]. Technical differences in how samples are collected and stored, how DNA is extracted, the primers that are selected and variable regions targeted, the sequencing platform, bioinformatic pipelines and reference databases could all produce systemic bias that obscures biological differences [128,129]. As the 16S rRNA protocol is the most common technique for characterising the preterm infant gut microbiome, a more detailed explanation of its varied protocols is discussed below. Caution should be used during both interpretation of the literature and study design until a standardised protocol is agreed upon.

Selection of Variable Regions and Primer Bias

Once DNA is extracted, and prior to sequencing, the target DNA from the variable region of interest must be amplified via PCR. However, there is much debate on which variable sub-region to target and matching primers to use. Indexing primers are complementary base pair sequences that are required to ‘select’ and amplify variable sub-regions. These 16S rRNA variable sub-regions can vary by up to 40% in taxa between samples analysed with the same

pipeline [245], and it has been argued that the most critical step for accurate rDNA amplicon analysis is the choice of primers [246], as primer selection can alter coverage [247].

Samples with the same extraction and storage protocols have been demonstrated to cluster by primer selection [128]. This is because poor primer selection can influence quantitative abundances [248], and contribute to under-representation or over-representation of taxa [98,249] or selection against particular taxa [223,250,251]. For example, most primers may inadequately detect Bifidobacterium [252], possibly over-exaggerating the low levels already observed in preterm infants. For identifying species, targeting of these hypervariable sub-regions is limiting, as different sub-regions show bias in the taxa that they can identify due to limited variability within the sub-region itself. Thus, while V1-V3 may be good for *Escherichia* and *Shigella* species, *Klebsiella* will require V3-V5, and *Clostridium* and *Staphylococcus* require V6-V9 sequencing [253]. As a result, studies that target hypervariable sub-regions must settle for taxonomic resolution at the genus level. Thus, the only way to ensure good taxonomic identification would be to sequence the entire 16S gene, given these limitations and the bias that can be introduced through variable region and primer selection. However, high error rates and cost are still major deterrents.

Arguments have been made for targeting the V4-V6 [254], V4 [248,254,255], and V3-V4 [245,247,256] sub-regions, with V4 being the most common for characterising the microbiome of preterm infants, when targeting hypervariable sub-regions for high throughput sequencing (*Figure 2.2*). The Earth Microbiome Project [257] recommends the V4 sub-region, and it has been demonstrated to have low PCR and sequencing errors due to complete overlap of paired end sequences [255]. Other work has also shown that the phylogenetic relationships based on V4 were closest to that entire 16S rRNA gene [254]. However, some evidence suggests that targeting the V4 region may not be as accurate as previously thought [253]. The debate about which region is best is ongoing, but research conducted by Almeida et al. makes a convincing argument for the use of the V3-V4 region above all else. It compared variable regions across different combinations with pipelines and reference databases for both mock communities and simulations [245], and found the V3-V4 region consistently produced the most reliable taxonomic inferences. Taken together, the frequent use of V4 and the findings across studies, targeting either the V4 or V3-V4 sub regions may be best practice until standardisation occurs. Additionally, consideration should be given when making comparisons across studies that use different variable regions.

NGS Platform

Several sequencing platforms are available for 16S rRNA short read amplicon sequencing, with Illumina MiSeq, Roche 454 (originally 454 Life Sciences) and Thermo Fisher's Ion Torrent Personal Genome Machine (PGM) all being used in the context of the preterm infant microbiome. Roche 454 has historically been the dominant platform, as NGS technologies began with it. However, Illumina now dominates the market, with its consistent growth and the eventual abandonment of the Roche 454 sequencing platform in 2016. It is important to understand the differing methods across platforms, and their limitations and biases for accurate interpretation of the literature, although most modern sequencing technologies will opt for Illumina sequencing.

Illumina sequencing technology facilitates massively parallel sequencing by using optical signals to detect base pairs in real time. DNA libraries, containing fragments that vary between 100-150bp, are loaded onto a flow cell and placed in the sequencer for this process. The sequences bind to the flow cell via complementary adaptors. A process called clonal bridge amplification or cluster generation then amplifies each read, creating a spot (cluster) on the flow cell (slide) with thousands of copies of the same DNA strand. Then, through a process coined sequencing by synthesis, fluorescently tagged nucleotides bind to the complementary bases on the DNA strand via repeated cycles of single-base extension. A fluorescent signal (the colour of which is dependent on the base) is emitted upon incorporation of each nucleotide, and a picture taken, indicating what nucleotide was added. Once the forward DNA strand is read, the reads are washed away, and the process is repeated for the reverse strand. Computers then construct the sequence by detecting the base at each site in each image.

Roche 454 relies on the production of sequence clusters, like Illumina, but through a process called clonal emulsion PCR (emPCR). In emPCR, single stranded DNA fragments (up to 1kb) from a DNA library are attached to the surface of a bead, rather than a slide, with one bead for each DNA fragment. The reads bind to the bead via complementary adaptors. The beads are then compartmentalised into single wells containing emulsified oil, and are subjected to thermal cycling to achieve clonal amplification. This process produces many copies of the original template, as in Illumina's clonal bridge amplification. The slide

containing the wells is then flooded with one of the four nucleoside triphosphates (NTP) that bind to their complements, releasing a light signal upon addition. The original NTP mix is washed away and the next NTP is added and the cycle is repeated. The light intensities are then plotted on a graph for each sequence read, with graphs then used to determine the sequence computationally.

Ion Torrents PGM also uses clonal-emPCR, but differs from Roche 454 both in how it determines the nucleotide sequences and the size of DNA fragments. Ion Torrents PGM or proton sequencing uses DNA fragments of ~200bp, which are again bound to beads via adaptors. These then undergo PCR and are washed with different NTPs. It then exploits the release of hydrogen ions, which occurs through the addition of an NTP to a DNA polymer for nucleotide sequence determination. The release of hydrogen ions causes changes in pH that are used to determine the DNA sequences.

Limitations (Pros and Cons)

No platform is without its limitations: limitations that can contribute to platform-associated biases and study-based clustering [128], despite significant developments in the sequencing field. For example, Roche can have high sequencing error rates associated with A and T bases [258], high error rates in homopolymer regions resulting from accumulated variance in light intensity [259-261], and can have up to 15% of sequences resulting from artificial amplification [262]. Ion Torrent is also subject to high homopolymer error rates [261,263,264], as well as organism-specific read truncation, due to the similar methods of Roche and Ion Torrent, in which multiple nucleotides can be incorporated during a single cycle [265]. Illumina still have their own systematic base-calling biases [266], even though platforms produce comparatively lower error rates [261]. These include production of homopolymer-associated sequencing errors [258], different quality reads across different sequencing tiles [267], increased single-base errors associated with GGC motifs [268] and different sequencing error rates at the different read ends [269]. In the two dominant platforms in preterm infant studies Roche 454 and Illumina, the differences caused by platform are minor [248,258]. However, the lower error rates, higher throughput [261] and higher read quality [265] achieved by Illumina, results in higher quality data. This allows stringent quality control parameters, resulting in more reliable outputs for downstream analyses [248].

Bioinformatics and Reference Databases

Bioinformatics is an interdisciplinary field of science that combines biology, computer science and statistics, in order to process large amounts of biological data, such as that produced by 16S rRNA gene amplicon sequencing. Bioinformatic tools like QIIME [270] and Mothur [173] are required to clean up and make inferences on microbial composition from data that is not human-readable post sequencing and prior to downstream analysis. Bioinformatic tools or pipelines need to be both precise and reliable in order to produce accurate biological conclusions using the vast amounts of genetic data that is being produced with sequencing. These tools convert raw data into interpretable taxonomic abundances by comparing sequencing reads in the form of OTUs, ASVs or an equivalent (sequence variants that represent a true sequence) [271] to a defined reference database, identifying the taxa present in samples by assigning the most likely taxonomic lineages. The accuracy of the taxonomy classifications produced is then reliant on both the diversity and breadth of annotated sequences in the reference databases [245], as well as the accuracy of the ever improving algorithms used by bioinformatic pipelines.

There is no agreement on optimal practices, although the bioinformatic pipelines are rapidly changing and improving, and many researchers are unaware of the biases associated with using different tools. Combinations of different software packages, databases and targeted regions can produce vastly different levels of accuracy when examining mock communities and running simulations [245]. When comparing several bioinformatic tools: QIIME, QIIME 2, Mothur and MAPseq [272], Almeida et al. found that QIIME 2 was the optimal tool in regards to detection sensitivity and composition prediction. QIIME 2 had the largest proportion of classified sequences at the most accurate relative abundances [245]. However, MAPseq was more precise, with fewer genera being miss-assigned. A more recent study which compared the most popular current bioinformatic pipelines for 16S rRNA gene amplicon sequencing, found that DADA2 was the best choice for studies requiring the highest possible biological resolution, but that USEARCH-UNOISE3 [273] had the best overall performance [274]. The common theme running through USEARCH-UNOISE3, DADA2 and QIIME 2 (uses DADA2/deblur plugins) is their denoising or clustering algorithm.

Denoising and clustering are methods for correcting sequencing errors through grouping of similar sequence variants into a bin. This was originally done through OTU clustering, in which sequences are clustered based on a 97% similarity threshold. However, there are

several methods for implementing this threshold: closed-reference, open-reference and de novo clustering [275]. The de-novo method clusters reads against one another, based on the threshold, without a reference database, unlike the reference-based approaches. In contrast, the closed-reference method clusters reads against a database and excludes those sequences that do not align. Open-reference clustering also clusters against a database, but then clusters reads that do not align de novo. The most successful method is debatable [275,276], but may be dependent on the study design. Nonetheless, the quest for more reliable data has seen a shift away from OTU-clustering towards error modelling, which takes into account both abundance and error.

Denoisers, as seen in DADA2 [277] and deblur [278], generate error models learnt from the reads and use these models for sequence variant assignment with either ASVs (DADA2) or subOTUs (deblur). The error modelling approach allows for clustering down to the level of single-nucleotide differences in the sequence region, improving resolution, and allows consistently reproducible labels with intrinsic biological meaning [271]. These improvements allow researchers to distinguish between true sequences and those generated during PCR amplification and sequences, and result in comparable general community structure across different tools. However, some variability still exists, with differences in the number sequence variant produced and resulting alpha diversity [279], despite these improvements. These differences should be considered when making cross-study comparisons and during study design.

The most common bioinformatic pipeline for studies exploring the preterm infant gut microbiome is QIIME. QIIME's use of OTU clustering and its production of a large number of spurious OTUs and inflated alpha diversity [274] should be taken into account when considering older literature. However, QIIME was succeeded by QIIME2 in 2018 (first published in 2019) [172], which uses an updated error modelling approach, with either DADA2 or deblur plugins. Most studies on the preterm infant microbiome predate the release of QIIME2, and therefore will have used the older version. It is unclear if newer studies will make the transition, due to the limited number of papers released since the pipeline's publication. However, in order to produce more robust data, research in this field needs to move towards these new and improved methods.

Along with choosing the best bioinformatics tool, the reference database used is also an important consideration. Pipelines can use a homology based or Bayesian approach to match

sequence variants to sequences from the reference databases. This was originally achieved with a similarity threshold of >95% sequence match being considered to represent the same genus and >97% match for species level identification [280]. However, recent work suggests that these thresholds are too low for accurate assignment [281]. The reference databases contain FASTA files with reference sequences assigned to nodes of taxonomy. However, discrepancies exist in both nomenclature and lineages of taxonomy between databases [281,282], which has obvious implications for the taxa identified in a sample. The Ribosomal Database Project (RDP) is the database used most often for studies on the preterm infant microbiome. SILVA may have better recall and be more precise than the more commonly used RDP (*Figure 2.2*), when examining the human microbiome, based on a benchmarking paper by Almeida et al. [245]. However, more research on best practices, along with standardisation across databases is needed.

Bioinformatics – Contamination

Contamination is another way variability between studies can be introduced. DNA contamination from collection, extraction and sequencing protocols (and kits) can impact upon the interpretation of results [283,284]. Many of these contaminants could be considered normal inhabitants of the human gastro-intestinal tract, and so it is important that measures are taken to mitigate the risk of contamination, and that steps are taken to account for or remove this contamination. Negative controls, spike-in controls and microbial standards, when coupled with appropriate bioinformatic tools, are effective ways to account for both accuracy of techniques and contamination. Bioinformatic tools like microdecon [285] can be used to remove homogenous contamination. This is important for both study design and interpretation of the literature, as contamination may produce unusual or novel findings if the appropriate mitigation strategies are not in place [283].

Descriptive metrics

Analysing microbial data can provide a significant challenge due to the volume and complexity of the data. Additionally, it is difficult to provide a best practice approach for statistical analysis of microbiome data, as it is highly dependent on the hypotheses and objectives of the study. However, generally, there are three main metrics that are considered

in microbial analysis: alpha diversity, beta diversity and differential abundance. Unfortunately, finding a meaningful way to conduct these analyses can be a convoluted process.

Alpha diversity refers to the diversity within a sample, summarising the ecological structure with respect to either richness (number of taxonomic groups), evenness (distribution of abundances of the groups), or a combination of the two [286]. Alpha diversity can be represented by richness, the Chao 1 index, the Shannon-Weiner index, the Simpson index, Pielou's evenness or Faith's phylogenetic diversity. All these indices differ in what they represent, with richness and the Shannon-Weiner index common in the context of microbiome research. Richness is simply referring to the count of sequence variants, whereas the Shannon-Weiner index takes both richness and evenness into account. The Chao 1 index is a bias-corrected richness estimate [287] that has become less common with the advent of newer pipelines (e.g. DADA2), due to their handling of singletons, which is based on the assumption that many are spurious sequencing variants.

Beta diversity refers to microbial differences between samples or groups. There are several metrics for beta diversity that represent distances between samples, as there are in alpha diversity. Beta diversity metrics include the Jaccard distance, Bray-Curtis distances and UniFrac. Jaccard distance is the number of sequence variants shared by samples divided by the number shared. Bray-Curtis distances, one of the most widely used in microbiome research, builds on this by taking abundances into account as well, whereas UniFrac distances, either weighted or unweighted, represent the differences between samples based on phylogenetic differences. These distance matrices can all be represented with direct comparisons of distances between samples or groups, hierarchical clustering or ordination techniques. Ordination is the most common as it reduces the complex distance data to a 2D or 3D plot, making for easy interpretation.

Alpha and beta diversity, along with differential abundance, all require normalisation prior to analysis. Differential abundance is simply making comparisons for taxonomic abundance between samples or metadata. However, the process is more complicated than simply counting the number of reads per sample. This is because using read counts as a measure of abundance is flawed, as the number of reads is actually an artefact of sequencing, and is therefore not a good representation of abundance [288]. Alpha and beta diversity also take read counts into account, and so both are also sensitive to sequencing depth. Both diversity metrics

require an equal number of reads per sample for valid analysis. This means that if appropriate measures are not taken, library sizes can determine diversity results [289]. Thus, reads must be normalized to account for the differing number of reads per a sample prior to analysis. However, although normalisation is the solution, different methods are required for different analyses.

Older methods of normalisation include Total Sum Scaling/Normalisation (TSS) and rarefying. In TSS, data is transformed to proportions by dividing the reads for each sequence variant by the total number of reads, whereas rarefying adjusts for differences in library sizes by assigning a sequencing depth threshold, and subsequently subsampling samples with a depth above the threshold and discarding those below. However, both methods are poor options for differential abundance testing and can have high type 1 errors [290-293]. Additionally, TSS doesn't account for heteroskedasticity [130] and rarefying discards potentially useful data. As a result, other modern methods have begun to replace the old methods.

Newer methods include variance stabilising transformation with DESeq2 [294], upper quantile normalisation [290], CSS normalisation [295] and Trimmed Means of M-values (TMM) with EdgeR [296]. Methods like DESeq2 and EdgeR are generally favoured due to their performance across several comparative papers [290-292]. However, these comparisons are specific to standardisation of within-sample variance, the ability of data to cluster in ordinations, and their performance in differential abundance testing, but there are several important limitations [297]. These methods tend to focus on standardising within-sample variance across samples, as they were created for differential abundance testing. As a result, the newer methods do not guarantee equal number of reads across samples. They supress species evenness and overestimate the importance of low abundance taxa (through log transformations) [297]. The overestimation of low abundance taxa and suppression of evenness can contribute to an inaccurate representation of the community, and, along with non-equal read counts, can lead to inaccurate comparisons between samples.

So, although proportions, specifically TSS and rarefying, are not suitable for differential abundance testing, they are more suitable for diversity analysis, as they give a more accurate representation of the microbial communities while accounting for differences in read depths [293,297]. Additionally, methods like the variance stabilising transformation with DESeq2 are favoured for differential abundance testing. It is critical that researchers are aware of the

strengths and limitations of normalisation methods for accurate interpretation and robust study design and differences between the subsequent analyses.

Metagenomic Shotgun Sequencing (whole genome sequencing)

Shotgun metagenomics is another NGS approach that has been used a handful of times for characterising the gut microbiome of preterm infants. Shotgun metagenomic sequencing targets all DNA in a sample, in contrast to 16S rRNA sequencing, which targets a specific region/gene. The protocols differ from 16S rRNA sequencing slightly, although they use the same sequencing technology. Shotgun metagenomics does not require amplification, as there is no target region/gene, but it does require the removal of host DNA prior to mapping, as all extracted DNA in a sample is sequenced. This alternative NGS method provides greater taxonomic resolution and gene annotation, allowing more comprehensive analyses. As a result, studies using this technology are typically looking to link functional gene profiles or pathogen strains to disease.

Limitations (Pros and Cons)

Shotgun sequencing has its limitations, despite its obvious benefits. There are numerous experimental and computational approaches that can be carried out at each step, as in 16S rRNA sequencing [298]. DNA extraction methods have been shown to affect composition [299], due to kits and reagents containing microbes [300] and differences in lysis techniques [301]. Library preparation and sequencing can introduce errors through PCR amplification [298] and selection of platforms [302,303]. Furthermore, specifically metagenome profiling can cause protocol-associated variability, having several options for bioinformatics, as all metagenomic profiling techniques have their own limitations [298].

There are two approaches for metagenome-profiling: assembly-free methods and assembly-based methods. Assembly-free methods, also known as read-based profiling or ‘mapping’, make comparisons to reference databases that contain whole genomes, such as Kraken [304] or Centrifuge [305], or to selected marker genes, such as mOTU [306]. Alternatively, assembly-based analysis uses assemblers like Meta-IDBA [307] and SOAPdenonovo2 [308] to reconstruct genomes de novo. Assembly-based methods can construct multiple whole genomes and resolve novel organisms, but can be a significant computational burden and are limited in assessing complex communities. Alternatively, read-based analysis is computationally

efficient and can deal with more complex communities, assuming there are enough sequencing depth and genomes in the reference database. However, identification is limited to those microbes previously defined, and so community structure/function is limited. Both approaches have their strengths and weaknesses, and which is best may depend on the question being asked.

There are also pros and cons when comparing NGS shotgun approaches to other sequencing methods. Metagenomics has more reliable species identification and broader analyses potential relative to metabarcoding, but the bioinformatics is more involved, requiring more time, skill and computational power, and the sequencing is more expensive, as entire genomes are being sequenced instead of a single gene. As a result, older studies using shotgun approaches tend to have lower sample sizes [115] or only use the technique on a subset of the cohort [116]. Additionally, where fragments of bacterial genomes are mixed in with contamination from host species and other organisms, 16S rRNA amplicon sequencing has specificity for bacteria, does not require full reference genomes, and does not require large quantities of, nor high quality, DNA [309,310], as opposed to shotgun sequencing. However, the adoption of this technique will likely become more widespread as the price of shotgun metagenomic sequencing continues to drop, in combination with improved computational methods. However, 16S rRNA amplicon sequencing targeting variable regions continues to dominate studies in this field at the present time.

Third Generation Sequencing

Long read sequencing is another sequencing approach. Full length sequencing of the 16S gene was made possible with the advent of third generation sequencing technology, also known as long-read sequencing. This approach is possible with platforms like Oxford Nanopore Technologies Minion [311] and techniques like Pac Bio's Circular Consensus and Continuous Long Read Sequencing [312]. These technologies allow discrimination between millions of reads that may only differ by a single nucleotide [253], and have the capacity to produce reads in excess of 10,000 base pairs (bp) [313,314]. This allows the sequencing of the entire 1,500 bp 16S gene and increases the resolution in taxonomic profiling to species and strain level.

Third generation sequencing can produce these long reads because their design is distinct from previous sequencing methods. Nanopore technology produces long sequences by passing a single DNA molecule through a DNA pore, measuring changes in current across a membrane. The current passing through the membrane is dictated by the size of the base pairs in the sequence that is passed through the pore. Alternatively, PacBio's SMRT (Single Molecule, Real-Time) sequencing repeatedly passes a DNA molecule through a DNA polymerase attached to a well, with short sequences being read until there are enough overlapping reads to identify the entire sequence.

Long-read technologies can also be applied to whole genome sequencing (WGS) and shotgun metagenomics, as well as full length sequencing of the 16S rRNA gene. WGS or shotgun metagenomic approaches allow greater sequencing depth, meaning species level detection of the preterm microbiome can be achieved, like sequencing the entire 16S gene. This capacity was demonstrated by Legget et al., who took advantage of Oxford Nanopore's ability to produce near-real time data in developing a metagenomic screening platform for preterm infant microbiome samples [171]. Moreover, as shotgun metagenomic sequencing targets all genomic DNA in a sample, the data can be used for other analyses, like functional profiling and antibiotic resistance gene profiling. This provides a comprehensive investigation of microbial ecology.

Limitations (Pros and Cons)

The long reads produced from the two technologies are their major advantages. Nanopore can produce reads generally ranging from 10Kbp to 1Mbp, with the longest sequence produced being >2Mbp [315]. These longer reads, along with advances in the associated computational methods, allow for greater sequencing depth than short-read technologies, with potential for greater accuracy [316], as they can distinguish between sequencing artifacts and actual biological sequences [253]. However, high error rates [317-319] are still a problem in TGS, despite the claim of high accuracy. For Oxford Nanopore, this high error rate comes from using changes in current to identify base pairs [320]. In PacBio's Single Molecule, Real-Time sequencing no current technology can precisely capture the rate of information produced (DNA polymerase adds 100bp/s), which is one reason why the DNA must be passed through the enzyme multiple times to overcome this issue. So, although these technologies show promise, the high error rates and cost are still deterrents, which is probably why they have been used so seldom in studies on the preterm infant microbiome.

Conclusions

Variability of results will continue to be a limitation when investigating microbial populations in preterm infants until there is standardisation of protocols. This review aimed to describe the most common techniques used in researching the preterm infant microbiome, and their limitations. The objective was to provide those entering the field with a broad understanding, so that considerations can be taken for both literature interpretation and future study design. 16S rRNA amplicon sequencing is the most commonly used method, as it is cheaper than both long-read and shotgun metagenomic sequencing, more detailed than non-molecular techniques and allows the characterisation of taxa present across a wide range of samples. This approach, however, has several limitations that can introduce bias. Full length sequencing of the 16S gene or a shotgun metagenomics approach may provide better options, especially as accuracy continues to increase, along with a reduction in cost. However, until these options become more viable, 16S high throughput sequencing targeting a select number of hyper variable sub-regions will continue to dominate.

There are a number of options at different stages within 16S sequencing methods that can contribute to bias, and with the large number of tools and databases available, it can be a difficult task deciding on an optimal approach. In this work I briefly described the bias across methodologies, with emphasis on 16S techniques. The most commonly used techniques within 16S rRNA high throughput sequencing are sample storage at -80°C, QIAamp DNA Stool Kit for extraction, sequencing on the Roche 454 platform, targeting the V4 region, and using the QIIME or QIIME2 pipeline in combination with the Ribosomal Database Project reference database. However, the optimal combination for 16SrRNA sequencing would likely be storage at -80°C, an extraction kit that includes mechanical lysis, such as (IHMS) Protocol Q or the Power Faecal Pro (Qiagen), use of the Illumina platform, targeting of the V3/V4 regions, using the QIIME2 pipeline (or at least error modelling) in combination with the SILVA database. However, the research question, as well as reproducibility and consistency across studies should also be considered. To conclude, until standardisation of microbiome research is possible, significant consideration needs to be given to ensure correct interpretation of the literature and robust study design.

3. The bacterial gut microbiome of probiotic-treated very preterm infants – Changes from admission to discharge

Using best practice methods outlined in Chapter 2, this chapter addresses aims 1 and 2: exploring changes in the bacterial microbiome of very preterm probiotic-supplemented infants from admission to discharge, and the effect of clinical variables (both maternal and infant) on the developing very preterm probiotic-supplemented infant bacterial microbiome. The purpose of this chapter was to explore the gut microbiome of what is now a new normal in North Queensland clinics; the probiotic supplemented, very preterm infant. This chapter reinforces previous findings demonstrating a strong link between microbiome composition and both maternal and infant disease, but in a unique cohort. It is also the second, and largest, study to implicate the infant gut microbiome in ROP and identified a link between maternal preeclampsia and gut-microbiome perturbations in early infancy. Further work is needed to elucidate the role of the gut microbiome in ROP, and the impact of microbial alterations that occur in association with preeclampsia. These findings were published as "*The bacterial gut microbiome of probiotic-treated very preterm infants – Changes from admission to discharge*" in *Paediatric Research* (https://doi.org/10.1038/s41390-021-01738-6).

Abstract

Background:

Preterm birth is associated with the development of acute and chronic disease, potentially, through the disruption of normal gut microbiome development. Probiotics may correct for microbial imbalances and mitigate disease risk. Here amplicon sequencing was used to characterise the gut microbiome of probiotic-supplemented preterm infants. The aim was to identify and understand variation in bacterial gut flora from admission to discharge and in association with clinical variables.

Results:

Infants born <32 weeks gestation and/or <1500 g were recruited in North Queensland, Australia, with faecal samples collected at admission (n = 71) and discharge (n = 63). Univariate analyses showed significant changes in the gut flora from admission to discharge. Mixed effects modelling showed significantly lower alpha diversity in infants diagnosed with either sepsis or retinopathy of prematurity (ROP) and those fed formula. Additionally, chorioamnionitis, preeclampsia, sepsis, necrotising enterocolitis and ROP were also all associated with differential abundance of several taxa.

Conclusions:

The lower microbial diversity seen in infants with diagnosed disorders or formula-fed, as well as differing abundances of several taxa across multiple variables highlights the role of the microbiome in development of health and disease. This study supports the need for promoting healthy microbiome development in pre-term neonates.

Introduction

It is well known that preterm birth leads to retarded gut microbiome development and increased risk of acute and chronic disease in infants and adults [321]. The gut microbiome composition of preterm infants differs significantly to those born full term, and is characterised by lower diversity [69,322] and high inter-individual variation [71,72,77]. Additionally, despite high variability, preterm infants typically have fewer commensals like

Bifidobacterium [71,72] and *Lactobacillus* [71,73], and more potential pathogens like *Klebsiella pneumoniae* [74] and *Clostridium difficile* [72]. However, the gut microbiome is dynamic and changes significantly over time [323]. Although reduced levels of common commensal organisms and diversity can persist for months [42,43], maybe years [85], choreographed abrupt changes in composition [81,84] and increases in diversity [42] mean that eventually the preterm gut microbiome composition becomes more similar to that of full-term infants.

Shifts in the composition and organism dominance result from environmental changes and major colonising events. Colonisation occurs via different routes and is influenced by several factors, including delivery and diet. Delivery is the first major colonising event, contributing significantly to differences between individuals [8,37], including higher abundances of vaginally derived microbes in those born vaginally. This includes *Bacteroides* and *Lactobacillus* [37,39]. In contrast, caesarean born infants acquire greater abundances of skin dwelling microbes like *Staphylococcus* [37,42]. As for diet, breast milk and formula also produce distinct microbial communities [56,323] due to the presence of both microbes and human made oligosaccharides (HMOs) in breast milk [148]. Although maternal skin and vaginal microbes colonise infants during birth and feeding, these microbes may only be transient with maternal gut microbes, passed through birth or lactation proving to be more persistent [23].

As much of the microbial inoculation occurs through maternal-infant exchange, maternal health and medical interventions can also influence the developing infant microbiome. Interventions such as antibiotics [27] and diseases like chorioamnionitis [8], a bacterial infection occurring before or during labour, have been previously shown to influence the infant microbiome. Thus, other maternal microbiome-altering diseases, like type 2 diabetes [65] and preeclampsia [66], a pregnancy disorder characterised by high blood pressure, could also disrupt the infant microbiome. The resulting irregular infant microbiome could have severe consequences for infant health and development [324].

Disrupted microbial colonisation puts preterm infants at a high risk of acute infection [91,106], chronic disease [38,325] and developmental abnormalities [161,326]. The increased risk in disease is a consequence of the breakdown in the symbiotic relationship between infants and colonising microbes, with delayed colonisation of commensal microbes contributing to intolerances to normal flora [97,108]. Additive to this is an imbalance between commensals and pathogens that may induce intestinal inflammation and cytokine production [327]. These microbial imbalances contribute to higher rates of acute diseases like necrotising enterocolitis (NEC) and sepsis,

chronic diseases like asthma [120] and potentially, developmental disorders like retinopathy of prematurity (ROP) [328] in preterm infants. This disproportionate burden of disease leads to microbiome-altering supplementation with antibiotics, a staple in preterm neonatal care, and probiotics, an emerging preventative strategy. Antibiotics can disrupt microbial acquisition, resulting in reduced diversity and altered bacterial profiles [54], whilst probiotics have been shown to promote the growth of commensal microbes and increases in diversity [13,114,133], as well as reducing disease incidence [10].

As probiotic prophylaxis is now common for the most premature of infants, this prospective observational study using 16S rRNA high throughput analysis of faecal and meconium samples aimed to characterise the bacterial gut microbiome of probiotic preterm infants. Specifically, I set out to characterise changes in a probiotic-supplemented cohort of preterm infants from admission to discharge, and to examine the impact of several key variables on the microbiome. This includes assessing the reproducibility of past findings and exploring new potential associations through multivariant analyses.

Methods

Study population

16S rRNA high throughput sequencing was used to characterise the bacterial microbiome, down to genus, of infants receiving probiotic supplementation and born into the Townsville Hospital and Health Service's (THHS) Neonatal Intensive Care Unit (NICU). The THHS Neonatal intensive care unit (NICU) is the only level six tertiary referral unit outside southeast Queensland, Australia. Thus, all babies being born at <29 gestation weeks in North Queensland are referred here. North Queensland is affected disproportionately by preterm birth, with the North West experiencing the highest rate (12%) of pre-term births [329], and the Torres and Cape the highest proportion (11.7%) of low birth weight (LBW) infants [329]. North Queensland (NQLD) also has a large indigenous population, whose infants are more likely to be born prematurely (13%) and represent one out of ten preterm births in Queensland [329]. When considering the increasing prevalence of preterm birth in the NQLD, 5% over the last decade [329], the burden that preterm birth places on NQLD families and the healthcare system is significant.

Study design and ethics

Ethics was obtained from the Human Research Ethics Committee from the THHS, and recruitment commenced in October of 2017, and continued until October of 2018. Inclusion criteria was infants born <32 weeks' gestation and admitted to the NICU at the THHS. The exclusion criteria were no parental consent, gestational age of >32 weeks and contraindication to enteral feeds. One capsule of the probiotic Infloran® 330, containing *Lactobacillus acidophilus* (1×10^9 CFU) and *Lactobacillus bifidus* (*Bifidobacterium bifidum*) (1×10^9 CFU), is administered via enteral feeds and to all infants born <32 weeks gestations and <1500 g at the THHS NICU on a daily basis. Infloran® supplementation is commenced on the first day of feeding and ceased once the infant is > 34-36 weeks gestation. Recruitment was conducted by a neonatal nurse/research assistant who works at the NICU, and sample collection by NICU nurses using collection kits' (biohazard bag, sterile swab and storage container). Collection occurred at admission (meconium) and just prior to discharge (stool). However, it should be noted that these labels represent the general timing of collection, and as such, the admission samples may not have been the first stool passed and may already be under the influence of microbiome covariates. After collection, samples were sent via a pneumatic tube system to Pathology Queensland and stored at -80°C. Clinical information was also collected for downstream analysis. This included both maternal data – antenatal antibiotics, antenatal infections (clinically diagnosed), chorioamnionitis (clinically diagnosed), prolonged membrane rupture (clinically diagnosed), preeclampsia (clinically diagnosed), and diabetes (type 1 or 2, self-reported) and infant data – sex, mode of delivery (vaginal birth versus Caesarean section), diet, gestation at birth and collection, NEC (stage 2 or greater), sepsis (confirmed through culture), days and timing of antibiotics, death, ROP (stage 1 or greater), birth weight, nursery discharge weight and date of birth. A summary of this data can be found in *Table 3.1.*

Categorical Variables			
Variables	Levels	**Count**	**Percentage (%)**
Sex	Male	60	44.8
	Female	74	55.2
Diet	Formula	40	29.9

	Breastmilk	64	47.8
	Formula & Breastmilk	30	22.4
Delivery	Vaginal	45	33.6
	Caesarean	89	66.4
NEC	Yes	12	9.0
	No	122	91.0
Sepsis	Yes	8	6.0
	No	126	94.0
Died	Yes	7	5.2
	No	127	94.8
Antenatal antibiotics	Yes	90	67.2
	No	44	32.8
Neonatal antibiotics	Yes	126	94.0
	No	8	6.0
Chorioamnionitis	Yes	58	43.3
	No	76	56.7
Preeclampsia	Yes	20	14.9
	No	114	85.1
Maternal Diabetes	Yes	24	17.9
	No	110	82.1
Continuous Variables			
Variable	mean/median		
Gestational age at birth	28.3/28.1 weeks		
Gestational age at sample collection	Admission	29.3/29.3	
	Discharge	35.3/35.9	
Days on antibiotics prior to sample collection	Admission	3.3/3	
	Discharge	9.4/5	
Weight at birth	1193/1086g		
Weight at discharge	2448/2425g		

Table 3.1 Overview of the demographic data for the cohort. Annotation: NEC: necrotising enterocolitis, ROP: retinopathy of prematurity.

Sequencing and bioinformatics

In brief, the protocol used in this study included sample storage at -80°C [181], an extraction kit that includes mechanical lysis [206], use of the Illumina MiSeq platform [248], targeting of the V3/V4 regions [245] and use of the SILVA reference database [245].

DNA extraction was conducted using the Bioline ISOLATE Fecal DNA Kit [331], with modifications made in consultation with the manufacturer to optimise DNA yield. This included increased beta-mercaptoethanol (from 0.5 to 1% to increase DNA solubility and reduce secondary structure formation), addition of an extra wash step (to improve purity) and decreased elution buffer volume (to increase final DNA concentration). For library preparation I followed the Illumina metagenomics library preparation protocol [332], using the Index Kit v2 C [333], along with Platinum™ SuperFi™ PCR Master Mix [334]. The MiSeq Reagent Kit V3 [333] was used in combination with the Illumina MiSeq System, targeting the V3 and V4 regions with the 785F/800R primer combination for sequencing.

Pre-analytical bioinformatics were conducted in *R Studio* Version 3.6.1 [335] with a pipeline adapted from *Workflow for Microbiome Data Analysis: from raw reads to community analyses* [336], which along with the subsequent analyses can found under Supplementary Material. *DADA2* [337] was used for quality filtering and trimming, demultiplexing, denoising and taxonomic assignment (with the SILVA Database), and the *microDecon* package [285] used to remove homogenous contamination from samples using six blanks originating in extraction.

Statistical analysis

Exploring changes in composition and diversity from admission to discharge

For statistical analysis, a phyloseq object was created using the package *Phyloseq* [338], with taxa filtered by prevalence (threshold = 0.01) and agglomerated at the genus level. The data were then explored through Principle Coordinate Analysis (PCoA) plots using a Bray-Curtis dissimilarity matrix created from normalised (Total Sum Scaling) non-agglomerated data. Permutational analysis of variance (PERMANOVA) was then conducted for community-level comparisons between admission and discharge samples to observe group-level differences based on the Bray-Curtis dissimilarity matrix, using the *adnois()* function of the

package *Vegan* [339]. Alpha diversity indices, Shannon Index and Observed (richness), were then calculated on filtered, non-agglomerated data, and a comparison was made between admission and discharge samples using a Wilcoxon Rank Sum Test, with adjusted p-values accounting for False Discovery Rate using the Benjamini-Hochberg procedure [340]. To identify individual microbes whose abundance changed significantly from admission to discharge, data that were filtered and agglomerated at the genus level, but not transformed, were then normalised and modelled (negative-binomial) with *DESeq2* [294]. A Wald Test with the Benjamini-Hochberg multiple inference correction was then performed to determine significant differentially abundant taxa.

Exploring the effect of clinical variables on alpha diversity and taxonomic abundance

Lastly, associations between several clinical variables and community structure were explored. The relationship between clinical variables and both Shannon Diversity and taxonomic abundance were assessed using multivariant linear regression models. For exploring the relationship with Shannon diversity, a mixed effects linear regression model was created using the package *lme4* [341], with a gaussian distribution and using the restricted maximum likelihood estimation. Continuous predictors were scaled and cantered to avoid convergence issues and multicollinearity assessed using the *AED* package [342]. Collinear variables were removed from the model. Thirteen predictors: mode of delivery, feeding type, gestation, antenatal antibiotics, antenatal infections, NEC, sepsis, chorioamnionitis, neonatal antibiotics, death, prolonged membrane rupture, preeclampsia, diabetes and retinopathy of prematurity were included in the initial model. To control for high amounts of inter-individual variation in the microbiome of preterm infants [69], individual's identification (unique record number – *URN*) was included as a random factor. To assess the influence of clinical variables at both admission and discharge an interaction variable was included in the model (labelled Type). The resulting model *Shannon ~ (15 Parameters) * Type + (1|URN)*, assesses the effect of the 15 predictors on Shannon diversity for both types of samples, Admission and Discharge, whilst accounting for the individual, represented here by URN.

Backwards selection (69) was then implemented to simplify the model by comparing Akaike's Information Criterion (AIC) scores between regression models and removing predictors that were not contributing to the model. The process was repeated until the least

complex adequate model was identified. The covariates included in the final model were sepsis, antenatal antibiotics, gestational age at birth, gestational age at collection, diet, the mode of delivery, NEC, preeclampsia, ROP and days on antibiotics. The significance of the fixed effects variables in this final model was then assessed analysis of deviance (Type II Wald Chi-square test) from the *car* package [343], and post-hoc pairwise Tukey comparisons (correcting for multiple comparisons) from the *emmeans* package [344].

For differential taxonomic abundance, two negative binomial generalized linear models were created using the package DESeq2. A combination of previous literature and exploratory analysis, including PCoA plots, PCA and scatterplots, were used for model selection. Again, continuous predictors were scaled and cantered, and multicollinearity was assessed. Taxa were agglomerated at the genus level, due to the limited taxonomic depth of short amplicon sequencing. To reduce the number of false positives, two separate models were run; one each for admission and discharge samples. The resulting model assessed the effect of 11 independent predictors; sepsis, diet, chorioamnionitis, mode of delivery, gestation at birth, gestation at collection, NEC, preeclampsia, ROP and days on antibiotics prior to sample collection, on taxonomic abundance. Low abundance and low frequency taxa were then removed, and a Wald Test with the Benjamin-Hochberg multiple inference correction was then performed. More information on the analysis can be found in the Supplementary Material.

Results

Exploring changes in composition and diversity from admission to discharge

The study recruited 85 preterm infants born <32 weeks and <1500g from the THHS NICU. From these infants 134 stool samples were collected, of which 71 were from admission (meconium) and 63 from discharge (stool), with 44 infants have paired samples. 42 of the discharge samples were collected after the cessation of probiotic prophylaxis, with an average time since cessation of 23 days. Other cohort demographics can be observed in *Table 3.1*. Significant changes in genera were observed between admission and discharge (*Figure 3.1*), with *Staphylococcus* significantly higher at admission ($p<0.01$), and *Enterobacter* ($p<0.01$), *Lactobacillus* ($p<0.01$), *Clostridium sensu stricto 1* ($p<0.01$) and *Veillonella* ($p<0.05$) higher at discharge (*Figure 3.2C*). Although there was limited separation between admission and

discharge samples, the beta diversity showed a clustering pattern that resulted in a significant difference between the two groups (*Figure 3.2A*, PERMANOVA; $p<0.01$ & $R^2 = 0.06$, homogeneity of variance; $p = 0.85$). The average species diversity within samples (Observed and Shannon) increased from admission to discharge (*Figure 3.2B*), but not significantly.

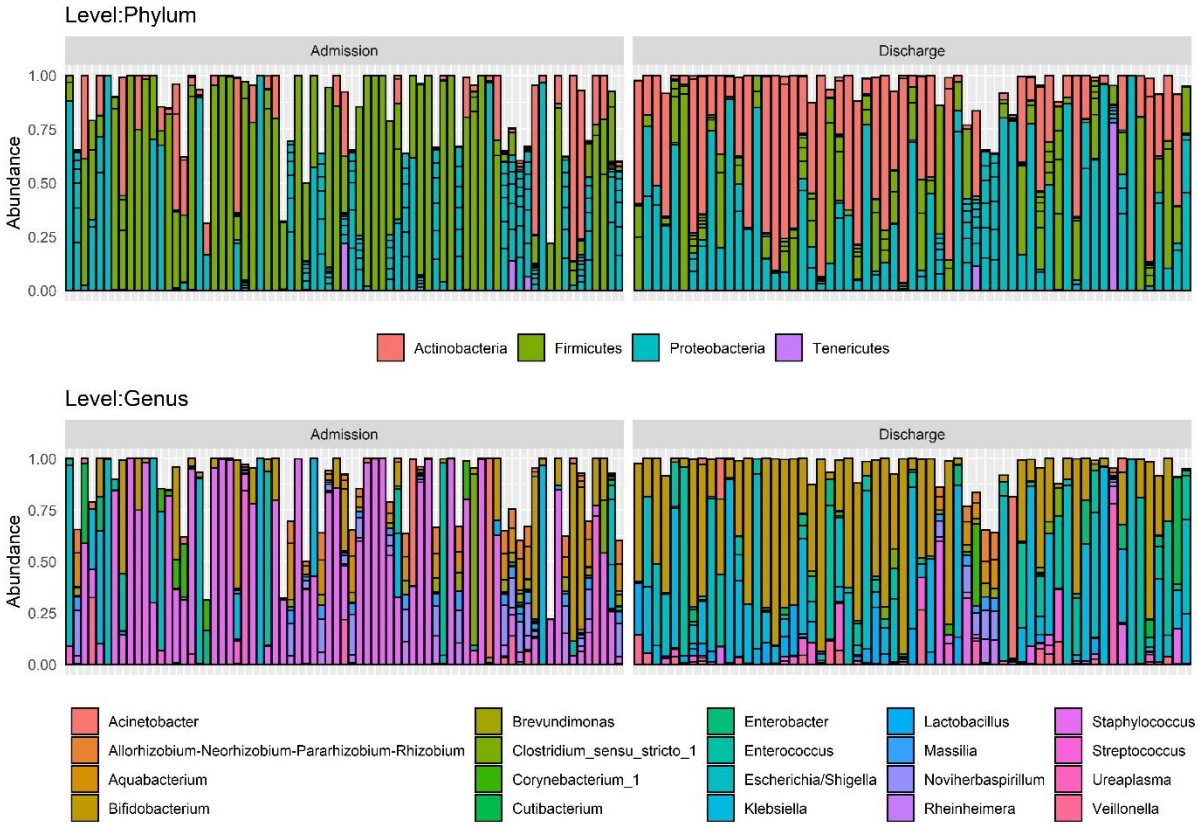
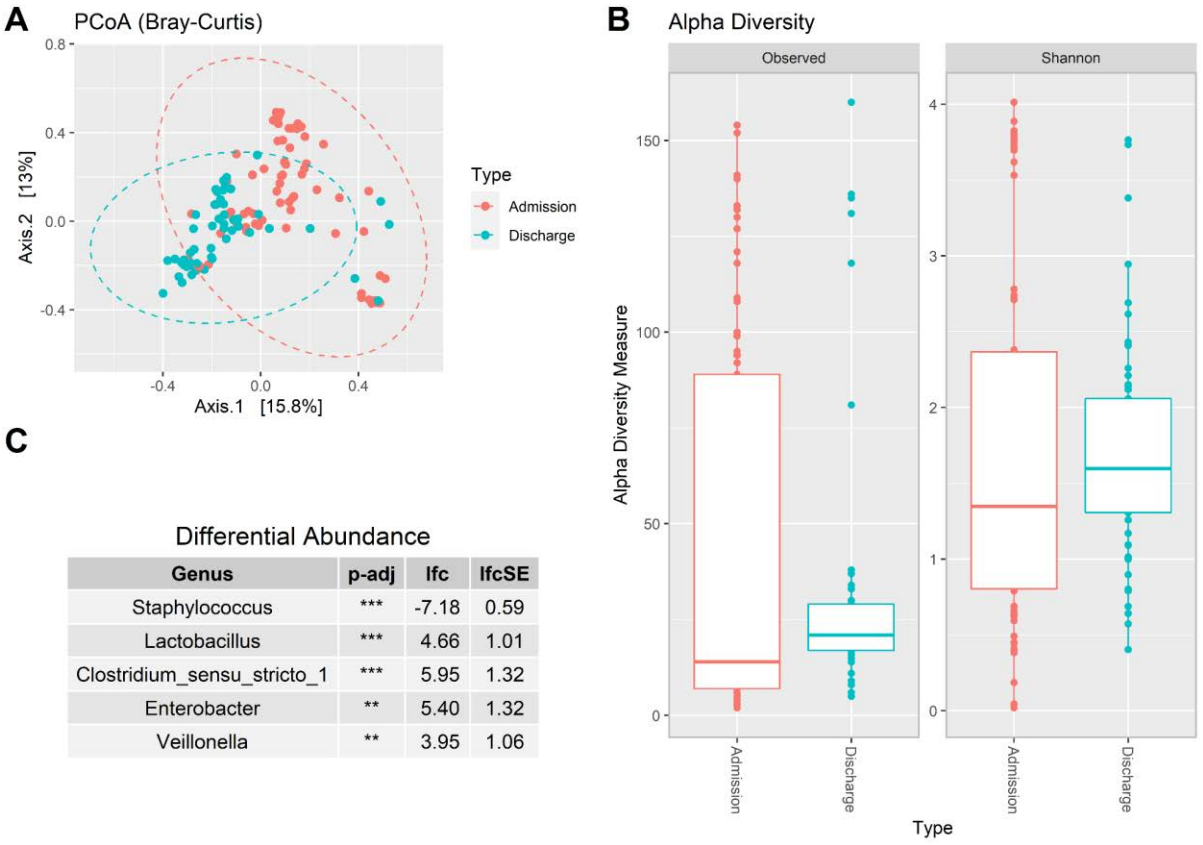


Figure 3.1 Histograms representing the distribution (top 20 taxa) of taxonomic relative abundance for admission and discharge samples at both phylum (A) and genus (B) levels.



Genus	p-adj	lfc	lfcSE
Staphylococcus	***	-7.18	0.59
Lactobacillus	***	4.66	1.01
Clostridium_sensu_stricto_1	***	5.95	1.32
Enterobacter	**	5.40	1.32
Veillonella	**	3.95	1.06

*Figure 3.2 A: Principle coordinate analysis plot for admission versus discharge based on Bray-Curtis dissimilarity matrix ($p<0.01$ & $R^2 = 0.06$), B: box plots of alpha diversity for admission versus discharge, C: table of differential abundance testing for admission versus discharge (base value is admission). Annotation: p-adj: Adjusted p value; lfc: log-fold change, & lfcSE: log-fold change standard error. $P<0.05$ = *, $P<0.01$ = **, $P<0.001$ = ***.*

Exploring the effect of clinical variables on alpha diversity and taxonomic abundance

Several maternal and infant variables were significantly associated with changes seen in the preterm infant gut microbiome. Mixed effects models show that several clinical and environmental variables were significantly associated with both the diversity and taxonomic composition within samples. Significant pairwise differences in diversity were observed for diet, sepsis and ROP (*Figure 3.3*), and chorioamnionitis, preeclampsia, sepsis, NEC, ROP and diet were all associated with changes in taxonomy (*Table 3.2*).

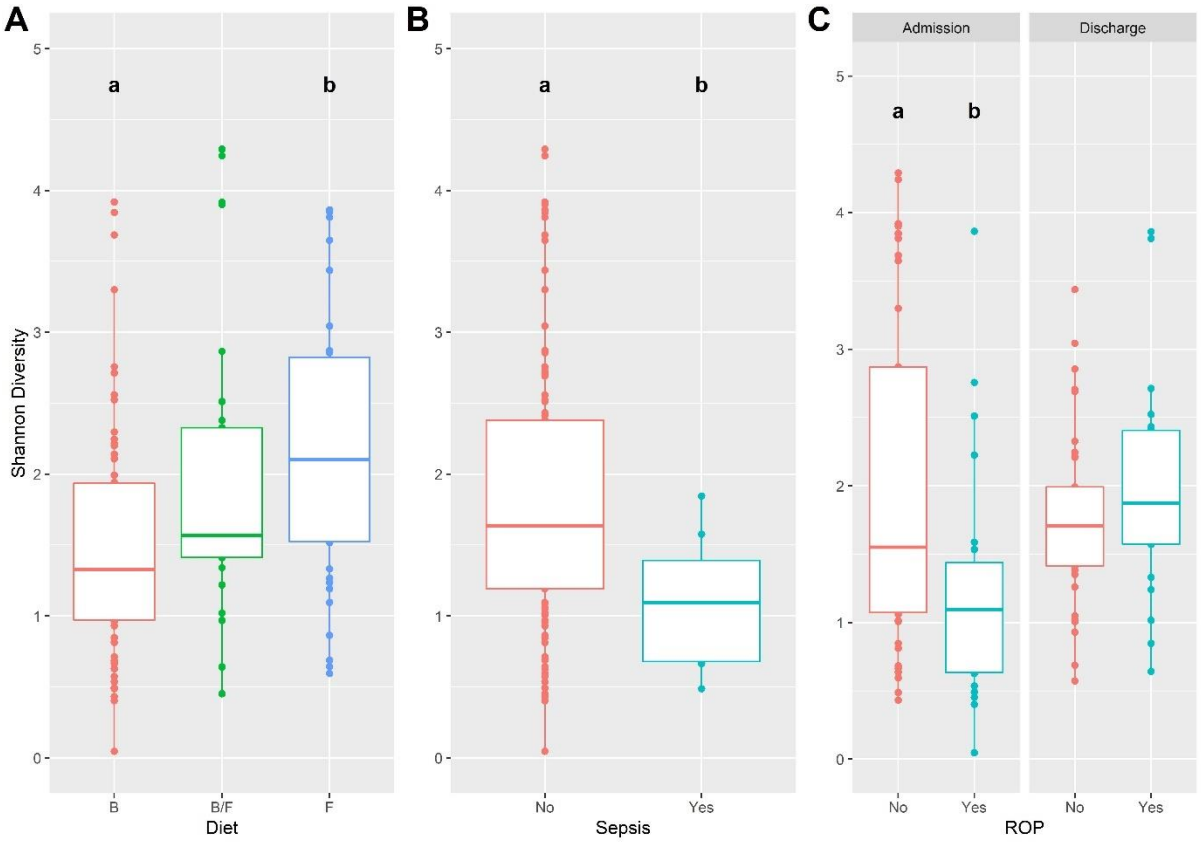


Figure 3.3 Boxplots of alpha diversity (Shannon Index) for significant Tukey's pairwise comparisons designated by lower case letters, (where a is significantly different from b) on a linear mixed effects model that used an interaction term to assess the effects both at and between admission and discharge, and was modified through backwards selection. Annotation for Diet; B: Breastmilk, B/F: Breastmilk and Formula & F: Formula. A: Box plot comparing alpha diversity at admission and discharge between different types of diet, B: Box plot comparing alpha diversity between sepsis diagnoses, C: Box plot comparing alpha diversity at admission and discharge between retinopathy of prematurity diagnoses.

log2FoldChange	***lfcSE***	***padj***	***Genus***	***Variable***	***Sample***
3.09	0.99	*	Staphylococcus	Chorioamnionitis:Yes	Admission
-17.58	3.22	***	Enhydrobacter	Sepsis:Yes	Admission
-15.38	3.95	**	Pseudomonas	Sepsis:Yes	Admission
10.33	2.84	**	Bifidobacterium	Sepsis:Yes	Admission
-11.62	2.31	***	Bifidobacterium	NEC:Yes	Admission
4.84	1.00	***	Staphylococcus	ROP:Yes	Admission

-27.65	2.64	***	Escherichia/Shigella	Preeclampsia:Yes	Discharge
-4.25	1.67	*	Veillonella	Diet:Breastmilk	Discharge
2.54	0.87	*	Bifidobacterium	Diet:Breastmilk	Discharge
3.64	1.46	*	Klebsiella	Diet:Breastmilk	Discharge
-5.52	1.91	*	Lactobacillus	Diet:Formula	Discharge

***P*<0.05 = *, *P*<0.01 = **, *P*<0.001 = ***.**

Table 3.2 The significant differentially abundant taxa at the genus level obtained from DESeq2 analysis, with log2FoldChage for the variable listed compared to the base value. Annotation: p-adj: Adjusted p value; lfc: log-fold change, lfcSE: log-fold change standard error, NEC: necrotising enterocolitis, ROP: retinopathy of prematurity.

Mode of delivery and diet

Only diet had a significant impact on the gut microbiome, with the mode of delivery not reaching significance for alpha diversity ($p = 0.057$, Supplementary Material) or any taxa. The type of milk the infant received had a significant effect on alpha diversity *(Figure 3.3A;* $\chi^2 = 13.5$, $df = 2$, $p<0.01$), with subsequent post-hoc pairwise comparisons finding a significant difference between formula-fed infants ($\bar{x} = 2.10 \pm 0.17$) and those that were breastfed ($\bar{x} = 1.56 \pm 0.11$) (*Figure 3.3A*; $p<0.01$). For differential abundance, infants who were fed only breastmilk had significantly higher abundances of both *Bifidobacterium* (*Table 3.2;* $p<0.05$) and *Klebsiella* (*Table 3.2;* $p<0.05$), and lower *Veillonella* (*Table 3.2;* $p<0.05$), relative to those that were only fed formula, but only at discharge. In addition, those fed only formula had significantly lower *Lactobacillus* at discharge (*Table 3.2;* $p<0.01$).

Pregnancy complications

Both preeclampsia and chorioamnionitis had a significant impact on the infant gut microbiome, with both conditions significantly associated with taxonomy (*Table 3.2*). In infants whose mothers were diagnosed with chorioamnionitis before or during labour, *Staphylococcus* was significantly higher at admission (*Table 3.2;* $p<0.05$). For infants whose mother was diagnosed with preeclampsia there were no differences at admission, but significantly lower *Escherichia*/*Shigella* (*Table 3.2;* $p<0.001$) at discharge.

Neonatal complications

Three neonatal complications, ROP, NEC and sepsis, were significantly associated with the developing preterm gut microbiome. Both sepsis (Fig 3B; $\chi^2 = 4.70, df = 1, p<0.05$) and ROP (Fig 3C; $\chi^2 = 10.98, df = 1, p = <0.001$) were significantly associated with diversity, with infants who were diagnosed with sepsis having significantly lower diversity ($\bar{x} = 1.10 \pm 0.17$) than infants who did not have the disease ($\bar{x} = 1.84 \pm 0.09$). For ROP, subsequent pairwise analysis suggests a significant association between diagnosis and differences in the microbiome at admission, suggesting early microbial perturbations could be implicated in later disease onset (*Figure 3.3C; p<0.01*).

NEC, sepsis and ROP were significantly associated with the abundances of several taxa at admission. Infants diagnosed with sepsis had significantly lower *Pseudomonas* ($p<0.01$) and *Enhydrobacter* ($p<0.01$), in combination with significantly enriched *Bifidobacterium* ($p<0.01$). *Bifidobacterium* was significantly lower in infants diagnosed with NEC ($p<0.01$), and *Staphylococcus* significantly enriched in infants diagnosed with ROP ($p<0.01$).

Discussion

The aim of this study was to identify and understand variation in gut microflora development in a cohort of probiotic-supplementation preterm infants from North Queensland, Australia. Specifically, I set out to assess the difference in bacterial microbiome between two time points while the infant was in hospital, between admission and discharge. I also sought to understand the effect of several clinical variables (both maternal and infant) on the development of the gut microbiome. To do so,16S rRNA gene high throughput sequencing was utilised. I then conducted univariate comparisons to examine the difference between the infant microbiome at admission and discharge, and mixed effects models to explore the influence of several clinical variables, including Sepsis, Feeding Type, Chorioamnionitis, Mode of Delivery, Gestation, NEC, Preeclampsia and ROP.

Exploring changes in composition and diversity from admission to discharge

Despite overlap, overall community structure was significantly different between admission and discharge faecal samples. *Staphylococcus*, commonly an early coloniser of the infant gut [345], was found in significantly higher abundance at admission. In healthy newborns, colonisation usually begins with oxygen-tolerant microbes [345] like *Staphylococcus*, that consume oxygen, shifting the environment from aerobic to anaerobic [60], allowing colonisation of strict anaerobes [345]. *Clostridium sensu stricto 1*, a genus of mostly strict anaerobes, along with the genera *Lactobacillus*, *Enterobacter* and *Veillonella* were found in significantly higher abundance at discharge. The significant presence of *Lactobacillus* at admission is surprising considering the delayed or limited colonisation of common commensals with *Lactobacillus* and *Bifidobacterium* normally seen in preterm infants [37,43,78]. Although not significant ($p = 0.11$), the presence of *Bifidobacterium* across 99 samples in such a young cohort is also noteworthy [78,79]. This is especially true considering their supplementation with Infloran®, which may explain the significant presence of both *Lactobacillus* and *Bifidobacterium* in such a cohort. Future work should apply a more robust sequencing methods, to see if the species present are those found within the probiotic.

Exploring the effect of clinical variables on alpha diversity and taxonomic abundance

Mode of delivery and diet

In contrast to previous studies, no significant pairwise differences in diversity or taxonomy between vaginally and caesarean delivered infants at admission or discharge was observed. Typically, caesarean born infants bypass the vaginal route of inoculation, resulting in greater diversity [8], with fewer or delayed colonisation of *Lactobacillus* [37], *Bifidobacterium* [37,43] and *Bacteroides* [44,46,346], coupled with higher than normal amounts of skin dwelling microbes. The inconsistency between the results of this study and the literature may be due to other confounding variables, such as prematurity itself or supplementation with probiotics, which has been demonstrated to alter *Bifidobacterium* and *Lactobacillus* populations in preterm infants [13]. If probiotic supplementation is driving the disparity between the results of and previous work, this would support previous work suggesting probiotic supplementation can correct for microbial differences seen in caesarean born infants [347].

Regarding the influence of diet on microbiome, there was significantly lower alpha diversity and higher abundances of *Bifidobacterium* and *Klebsiella* at discharge in breastfed infants, relative to those solely formula fed. The significant difference in *Bifidobacterium* supports previous work showing that breastfed infants have lower diversity [56] in combination with more commensal microbes [37,57], including different *Bifidobacterium* species [37]. The higher abundance of such microbes stems from the presence of both *Bifidobacterium* and HMOs in breastmilk [52,348,349]. Additionally, the higher abundance of Lactobacillus in infants the were fed a combination of formula and breastmilk, relative to those who only received formula, suggests that 'supplementing' formula feeding with some breastmilk may correct for some microbial imbalances associated with formula feeding. As for the differences in *Klebsiella*, the genus contains known pathogens such as *Klebsiella pneumoniae*, previously associated with NEC [153], and has been implicated in cases of sepsis. However, K. *pneumoniae* is a very diverse genus that is also part of normal flora.

Pregnancy complications

Maternal factors were also significantly associated with the composition of the probiotic-supplemented preterm infant microbiome. Associations were observed for both chorioamnionitis and preeclampsia. Infants whose mothers were diagnosed with chorioamnionitis had higher abundances of the genus *Staphylococcus*. Previous work has found microbes at different levels of taxonomy to be associated with chorioamnionitis, but not from the genus observed in this study [350]. As chorioamnionitis is a bacterial infection of the placenta and membrane surrounding the foetus, occurring before or during labour, what pathogens are translocated from the membrane to the foetus may dictate the associations found. Unfortunately, the translocation and resulting increased abundance of *Staphylococcus* may be why exposure to chorioamnionitis increases the risk of preterm infants to adverse neonatal outcomes [350], like sepsis, which has previously been associated with *Staphylococcus* [107,109].

For infants whose mothers were diagnosed with preeclampsia, *Escherichia*/*Shigella* was significantly lower at discharge. As preeclampsia can alter the maternal microbiome [66], the resulting dysbiosis, at least in part, may be being passed through a maternal route of inoculation. Previous work by Stewart et al., *The Environmental Determinants of Diabetes in*

the Young (TEDDY) study, has found preeclampsia contributes to significant difference at the species, but not genus, level [323]. However, the two cohorts are vastly different, with the TEDDY study including both full- and pre-term children, with samples from 3 months of age. In contrast, the cohort in this work was entirely preterm who at discharge may have only been 3 months old. Additionally, as preeclampsia is associated with preterm birth [351], the cohort in this study had a larger proportion of infants born to preeclamptic mothers (18% compared to 4%). Taken together, preeclampsia may have a greater impact on preterm infants, or may have more of an effect in the early months of life, when the mother is still the dominant colonising route for microbes. As to why Preeclampsia only has an effect at discharge is unclear. However, as the impact is occurring via the maternal route, it may be related to continued exposure to the mother. This continued exposure through the maternal route of transmission, either by touch or breastmilk, may compound the passing of irregular taxonomic profiles that resulting from continued preeclampsia-treatment post-delivery [352,353].

Neonatal complications

Sepsis was significantly associated with the abundance of *Bifidobacterium*, *Pseudomonas* and *Enhydrobacter*. Multi-omics approaches have previously linked sepsis to the gut microbiome [110], with other studies showing associations of sepsis with low diversity [81], as well as higher abundances of *Staphylococcus* [107,109], and lower abundances or absence of commensal microbes like *Bifidobacterium* [108,110]. Although I also observed differences in *Bifidobacterium*, the directional effect is counter to what was observed previously. However, it is worth noting that of the eight infants diagnosed with sepsis, only three had *Bifidobacterium* in their sample. So, despite reaching statistical significance, this finding may not be clinically relevant.

For NEC, significantly lower abundances of *Bifidobacterium* was observed, but in contrast to previous work, no enrichment of any taxa. As previously mentioned, *Bifidobacterium* is a common commensal microbe found in the probiotic Infloran®, that is uncommon in preterm infants born <33 weeks gestation [78], and has previously been shown to be protective against NEC [96]. Although my work does not support previous evidence of an associated pathogen, the plethora of microbes that have previously been associated with NEC [97,99,100] , in

combination with studies showing reduced commensal microbes [98,102] and diversity [94,140], suggests the aetiology is more complex than just the presence of a pathogen.

There was also significant enrichment of *Staphylococcus* (of the *Staphylococcaceae* family) and lower diversity at admission for infants diagnosed with ROP. An association between the gut microbiota and ROP has been explored once before, by Skondra et al. [354]. They observed significant enrichment of the family *Enterobacteriaceae* in preterm infants with the disease at 28 weeks postmenstrual age [354]. The discrepancy in my results is not necessarily a product of error, but rather, as seen with NEC, due to the complex aetiology characterised by more than just the presence of a particular group of taxa. This complexity makes it difficult to hypothesise the specific role that the microbiome could be playing in ROP. However, the associations observed suggest that microbial perturbations may be involved in the development of ROP, making it a candidate for targeted intervention. This should be the target of further research.

Limitations of this work include low taxonomic depth and only sampling in early infancy. The use of 16S rRNA gene metabarcoding limited detection power to the genus level, resulting in no identification of species or functional genes. Additionally, collecting samples only at admission and discharge means this work provides no insight into the longevity of the differences observed, this may impact the clinical significance. Future work will use a combination of 16S rRNA gene metabarcoding and shotgun metagenomic techniques to both characterize species allowing an exploration of the differences observed in this study, and others, and to investigate if these persist in the long-term.

This prospective observational study used 16S rRNA gene sequencing to characterise the bacterial microbiome of probiotic-supplemented infants. The study aimed to identify and understand variation in bacterial gut flora between two time points and as the result of several clinical variables. My study builds on previous research and supports other studies describing significant changes in the preterm microbiome over time and associations with several factors. The lower bacterial diversity seen in infants diagnosed with diseases or who were formula fed, as well as the differing abundances of several taxa across multiple variables reinforces the role of the microbiome in disease and supports the need for promoting healthy microbiome development. Additionally, the associations with maternal disease highlights the importance of maternal health to infant microbiome development, and in turn infant health.

4. To probiotic or not to probiotic: a metagenomic comparison of the discharge faecal microbiome of infants supplemented with probiotics in NICU and those who are not

Using the microbiome-covariates identified in Chapter 3, the studies in Chapter 4 utilise mixed effects modelling to parse out differences in the microbiome between probiotic-supplemented and non-supplemented infants. The purpose of this chapter is to address the current supplementation criteria for preterm infants in North Queensland, Australia, which is currently restricted to those born < 32 weeks and < 1,500 g. Although the targeted approach to probiotic prophylaxis may be justified, given the inverse relationship between NEC and gestational age, probiotics may afford benefits beyond prevention of this acute disease. If probiotics are inducing positive microbial changes during admission in those infants being supplemented, then all preterm infants may benefit. The objective of this chapter was to compare infants supplemented and not supplemented just prior to leaving hospital care. This chapter addresses aims three and four: investigating differences in the bacterial microbiomes of probiotic-supplemented and non-supplemented preterm infants during hospital admission, and whether the probiotic microbes colonise the infant gut.

The results of this chapter suggest that supplemented and non-supplemented infants have distinct microbiomes, with those not supplemented having lower alpha diversity and abundances of key taxa. These findings suggest that late-preterm infants may benefit from stability and protection provided through greater alpha diversity and abundance of beneficial microbes that results from probiotic prophylaxis. This chapter was published as "*To probiotic or not to probiotic: a metagenomic comparison of the discharge gut microbiome of infants supplemented with probiotics in NICU and those who are not*" in *Frontiers in Pediatrics* (https://doi.org/10.3389/fped.2022.838559).

Abstract

Background:

Preterm birth is associated with the development of both acute and chronic disease, and the disruption of normal gut microbiome development. Recent studies have sought to both characterise and understand the links between disease and the faecal microbiome. Probiotic prophylaxis may correct for these microbial imbalances and, in turn, mitigate disease. However, the criteria for probiotic supplementation in NICU's in North Queensland, Australia limits its usage to the most premature (< 32 weeks gestation) and small for gestational age infants (<1500 g). Here I use a combination of amplicon and shotgun metagenomic sequencing to compare the faecal microbiome of infants who fulfil the criteria for probiotic-prophylaxis and those who do not. The aim of this study was to determine if probiotic-supplemented preterm infants have significantly different taxonomic and functional profiles when compared to non-supplemented preterm infants at discharge.

Methods:

Preterm infants were recruited in North Queensland, Australia, with faecal samples collected just prior to discharge (36 ± 0.5 weeks gestation), to capture potential changes that could be probiotic induced. All samples underwent 16S rRNA gene amplicon sequencing, with a subset also used for shotgun metagenomics. Mixed effects models were used to assess the effect of probiotics on alpha diversity, beta diversity and taxonomic abundance, whilst accounting for other known covariates.

Results:

Mixed effects modelling demonstrated that probiotic supplementation had a significant effect on overall community composition (beta diversity), characterised by greater alpha diversity and differing abundances of several taxa, including *Bifidobacterium* and *Lactobacillus*, in supplemented infants.

Conclusions:

Late preterm-infants who go without probiotic-supplementation may be missing out on stabilising-effects provided through increased alpha diversity and the presence of commensal microbes, via the use of probiotics. These findings suggest that late-preterm infants may

benefit from probiotic supplementation. More research is needed to both understand the consequences of the differences observed and the long-term effects of this probiotic-protocol.

Introduction

The development of the gut microbiome is an important regulator of lifelong health [2-6]. However, being born preterm disrupts the gut microbiome's natural development [322]. Probiotics are increasingly used as supplements to support preterm infant development, particularly as adjunctive therapies to prolonged antibiotic treatment. In neonatal intensive care units (NICU) across Australia, probiotic supplementation is becoming the standard of care for the most premature (< 32 weeks gestation) and small for gestational age infants (< 1,500 g). This is in response to clinical trial validated evidence, that demonstrates effective probiotic supplementation against Necrotising Enterocolitis (NEC) and Late-onset sepsis (LoS) [10,11,355], in combination with an increased risk of acquiring the disease in very preterm infants [356]. It is now well recognised that there may be other wide ranging health benefits stemming from early and appropriate gut colonisation with bacterial probiotic-species. Although targeted probiotic prophylaxis of the most premature of infants may be justified, those who do not meet the supplementation-criteria may be missing out on potential health benefits [347,357]. Preterm infants not supplemented with probiotics may have a disadvantaged start to life [11-13].

The gut microbiome plays a critical role in the healthy development of the infant, particularly for immunological and metabolic programming [14,18,19]. The disruption of normal gut microbial colonisation caused by preterm birth can be associated with acute life-threatening diseases [86,178], such as NEC and sepsis [9,87,88]. Additionally, a growing body of evidence now suggests disrupted development of the gut microbiome is associated with chronic lifelong conditions, such as asthma [89], type 1 diabetes [90] and metabolic derangements [358]. These diseases are more common in those born prematurely, infants who harbour a gut microbiome characterised by low diversity [69] and commensal microbe abundance [71-73], in combination with the presence of a greater number of pathogens [72,74]. Probiotic prophylaxis may provide a solution for improving gut microbiome diversity and commensal microbe abundance, and, in turn, reduce the significant health burden placed on preterm infants.

Despite some heterogeneity between studies reported in the literature [122], probiotics have demonstrated efficacy in reducing the incidence of diseases, such as NEC [10,11,355], as well as positively modulating the infant gut microbiome [12,13], in the most premature infants. The heterogeneity observed could result from the use of different probiotic species [11], variability in the microbiome detection methods used [123] or the many confounding variables that influence the developing gut microbiome. Nonetheless, several countries, such as Japan and Australia, use probiotics as part of standard care for the most premature of infants, and those at high risk of NEC. Although supplementation protocols may vary between countries and neonatal units, here in North Queensland (NQLD) Australia, standard protocol dictates that all infants born < 32 weeks gestation and < 1500 g are supplemented with Infloran®, a probiotic containing *Bifidobacterium bifidum* and *Lactobacillus acidophilus*, as approved by the Therapeutic Goods Administration of Australia [359]. However, the specificity of this criterion means that preterm infants who fall outside of this criteria, infants who may also suffer from irregular microbial colonisation, go without probiotics. What significance this has for these non-supplemented preterm infants remains unclear.

Very little is known about the implications of limiting probiotics to the most premature for the developing microbiome of older preterm infants. Unfortunately, research exploring probiotic-supplementation in older preterm infants is lacking, with a 2017 meta-analysis showing the average age for clinical trials is < 33 weeks [11]. This is not unjustified when considering the previously mentioned inverse correlation of NEC with both gestational age and birth weight [360,361]. Thus, probiotics are targeted at this younger preterm demographic and, in turn, the research as well. Late-preterm infants could be missing out on benefits provided through probiotics.

This study was designed to investigate the effect of probiotic supplementation on the developing preterm infant gut microbiome, by comparing the gut microbiome of probiotic-supplemented (born < 32 weeks gestation) and non-supplemented (born >32 weeks & <37 weeks gestation) preterm infants. The aim of the study was to determine if these two groups have significantly different taxonomic and functional profiles when leaving care, at 36 weeks corrected gestational age. Additionally, I also collected data on known microbiome-covariates so that they could be controlled for using mixed effects modelling. Gut microbiome health at discharge is an important end goal to not leave the child at a lifelong disadvantage.

Materials and Methods

Study population

A combination of 16S rRNA gene amplicon and shotgun metagenomic sequencing was used to characterise the faecal microbiome of preterm infants from North Queensland (NQLD), Australia. 16S rRNA gene amplicon sequencing was applied to the entire cohort, and shotgun metagenomics to a small subset. NQLD is burdened disproportionately by preterm birth, with the North West experiencing the highest rate (12%) of pre-term births [329], and the Torres and Cape the highest proportion (11.7%) of low birth weight infants [329]. NQLD also has a large indigenous population, who are more likely to experience prematurity (13%), representing one in ten preterm births in Queensland [329]. As the prevalence of preterm birth in NQLD is increasing, 5% over the last decade [329], the burden that preterm birth places on the families and healthcare system in this region of Australia is significant.

Recruitment sites were the Townsville University Hospital's (TUH) Neonatal Intensive Care Unit (NICU) and Special Care Nurseries (SCN), as well as the Cairns and Hinterland Hospital and Health Service's (CHHHS) SCN. Samples from the probiotic-supplemented infants were all collected from the TUH NICU, as this is the only level six tertiary referral unit in NQLD, which is a specialised unit for dealing with complex pregnancies. All high risk preterm infants (<32 weeks gestation and/or <1500 g) received the probiotic Infloran® [330], containing *Lactobacillus acidophilus* (1×10^9 CFU) and *Bifidobacterium bifidum* (1×10^9 CFU) on a daily basis. Use of this probiotic is approved by the Therapeutic Goods Administration (TGA) of Australia. Prophylaxis with Infloran® is commenced on the first day of feeding and ceased once the infant is 34 weeks gestation.

Inclusion criteria for the cohort included: born <32 weeks' gestation and admitted to the NICU at the TUH for the probiotic-supplemented group, and <37 weeks but >32 and admitted to the SCN at the TUH or CHHHS for the non-supplemented group. The exclusion criteria were no parental consent, gestational age of >32 weeks and contraindication to enteral feeds for the probiotic-supplemented group, and no parental consent and gestational age of >37 weeks for the non-supplemented group. Ethics was obtained from the Townsville Hospital and Health Service Human Research Ethics, (HREC/17/QTHS/7). Recruitment and collection of faecal samples were conducted by neonatal nurses, who work in the nurseries, between October of 2017 and October of 2018.

Sample collection, storage, and DNA extraction

Faecal sample collection was carried out by neonatal nurses during routine nappy changes. This occurred just prior to discharge ($\bar{x}$ = 36 ± 0.5 weeks gestation) to capture potential probiotic-induced changes. Collection was conducted using collection kits (biohazard bag, sterile swab and storage container) with nurses collecting more than enough sample to provide the 150mg required for the Bioline ISOLATE Fecal DNA Kit protocol [331]. The collected faecal samples were then sent via pneumatic tube systems to Pathology Queensland and stored at -80°C for up to 6 months. DNA extraction for both the 16S rRNA gene amplicon and shotgun metagenomic sequencing was carried out using the Bioline ISOLATE Fecal DNA Kit, which includes mechanical bead-beating [331]. Modifications were made in consultation with the manufacturer to increase DNA yield. This included increased beta-mercaptoethanol (from 0.5 to 1% to increase DNA solubility and reduce secondary structure formation), addition of an extra wash step (to improve purity) and decreased elution buffer volume from 100µl to 50µl (to increase final DNA concentration), for overall increased DNA yield and purity. The extracted DNA was then stored frozen at -80°C. Clinical information was also collected for downstream analyses (*Table 4.1*).

Categorical Variables					
Variables	**Levels**	**Probiotic-supplemented**		**Non-supplemented**	
		n	**%**	**n**	**%**
Diet	Formula	23	36.5	0	0
	Breastmilk	26	41.3	12	38.7
	Formula & Breastmilk	14	22.2	19	61.3
Delivery	Vaginal	16	25.4	16	51.6
	Caesarean	47	74.6	15	48.3
NEC	Yes	5	7.9	0	0
	No	58	92.1	31	100
Sepsis	Yes	2	3.2	1	3.2
	No	61	96.8	30	96.8
Antenatal antibiotics	Yes	39	38.1	13	41.9
	No	24	61.9	18	58.1

Neonatal antibiotics	Yes	60	95.2	23	74.2
	No	3	4.8	8	25.8
Chorioamnionitis	Yes	27	42.9	1	3.2
	No	36	57.1	30	96.8
Preeclampsia	Yes	10	15.9	3	9.7
	No	53	84.1	28	90.3
Maternal Diabetes	Yes	12	19.0	7	22.6
	No	51	81.0	24	77.4
Died	Yes	2	3.2	0	0
	No	61	96.8	31	100

Continuous Variables

Variable	**Probiotic-supplemented**		**Non-supplemented**	
	Median	**IQR**	**Median**	**IQR**
Gestational age at birth (weeks)	28.1	3.9	35.6	2.7
Gestational age at collection (weeks)	36.7	4.8	38.7	3.6
Days since last antibiotic dose	38	44	9	23

Table 4.1 Overview of the demographic data for the preterm-infant cohort that underwent 16 rRNA gene amplicon sequencing. Days since last antibiotic dose are for those who are no longer being treated.

16S rRNA short amplicon sequencing

The Illumina metagenomics library preparation protocol was used for library preparation [332], using the Index Kit v2 C [333], along with Platinum™ SuperFi™ PCR Master Mix [334]. Sequencing was performed on the Illumina MiSeq system using the MiSeq Reagent Kit V3 600 cycles [333], targeting the V3 and V4 regions with the S-D-Bact-0431-b-S-17/S-D-Bact-0785-a-A-21primer combination[332]. Pre-analytical bioinformatics were conducted in *R Studio* Version 3.6.1 [335] with a pipeline adapted from *Workflow for Microbiome Data Analysis: from raw reads to community analyses* [336], which can found at

https://github.com/JacobAFW/SCN_vs_NICU_probiotic_study. *DADA2* [337] was used for quality filtering and trimming, demultiplexing, denoising and taxonomic assignment (using the SILVA Database), and the *microDecon* package [285] used to remove homogenous contamination from samples using blanks originating in extraction.

Shotgun metagenomics

A subset of the samples (n=6) was selected on the basis of suitability for shotgun metagenomics analysis (performed by Microba Life Sciences), with samples with the highest extracted DNA concentrations chosen. Six samples, three from probiotic-supplemented and three from non-supplemented infants, were chosen to make species-level and functional comparisons. Other demographic data specific to these infants can be found in the *Appendix*. These selected samples were shipped to Microba on dry ice. Sequencing was conducted on the Illumina NovaSeq6000 system with 300 bp, paired-end reads. Microba provided an end-to-end service, also conducting the bioinformatics and statistical analysis. This was done using Microba's Metagenomics Analysis Platform (MAP), which includes the Microba Genome Database, the Microba Community Profiler, and the Microba Gene and Pathway Profiler [362]. Microba's MAP produces taxonomic and functional profiles. Functional profiles include Enzyme Commission (EC) Number, Membrane Transport Proteins (TCDB) and MetaCyc (database) Pathways and MetaCyc Groups.

Statistical Analyses

Analysis of data obtained through 16S rRNA short amplicon sequencing

To assess the difference between the probiotic and non-supplemented preterm infants across the entire cohort, while accounting for known associates (*Table 1*) to the infant gut microbiome, I assessed alpha diversity, beta diversity and taxonomic abundance using mixed effects models. The scripts for these analyses can be found on the GitHub page previously mentioned. The covariates included; maternal antibiotics [363], maternal diabetes [364,365], chorioamnionitis [8], preeclampsia [366], maternal diabetes [364], mode of delivery [8,37], infant diet [56,323], gestational age, NEC [99,100], infant sepsis [107,109], neonatal antibiotics [54] and ROP [328]. For beta diversity, I performed an *EnvFit* analysis from the *Vegan* package [367], which compares the differences in the centroids relative to total variation. A Bray-Curtis dissimilarity matrix [339] based on data normalised through Total Sum Scaling (TSS) [297] was used for the *EnvFit*

analysis. The significance was based on 10,000 permutations and was transformed using the Benjamini-Hochberg (BH) procedure [340].

For alpha diversity (Shannon Index), I used the package *lme4* [341] to perform a generalised linear mixed effects model. Diversity was calculated at the ASV level. Multicollinearity was assessed using the *AED* package[342] and collinear variables removed. Backwards selection (69) was implemented to find the least complex, yet adequate, model. Significance was determined using an analysis of deviance (Type II Wald Chi-square test) from the *car* package [343], and subsequent post-hoc pairwise Tukey comparisons, correcting for multiple comparisons, using the *emmeans* package [344].

DESeq2 [294], which uses a negative binomial generalized linear model and variance stabilising transformation, was used for comparing taxonomic abundances between probiotic and non-supplemented groups. Taxa were agglomerated and assessed at the genus level. To identify taxa that were significantly differentially abundance, a Wald Test with the BH multiple inference correction was used. The statistical analyses can be found in the GitHub provided above.

Analysis of shotgun metagenomics data

To compare probiotic supplemented and non-supplemented infants in the subset of the cohort that underwent shotgun metagenomics, comparisons were again made for alpha diversity, beta diversity and taxonomic abundance. Standard t-tests were used for comparing alpha diversity (richness and the Shannon Index), Redundancy analysis (multiple linear regression) for beta diversity, and ALDEx2, with a Welch's t-test, for differential abundance. P values were corrected with the BH procedure.

Results

The aim of the study was to determine if probiotic-supplemented (born < 32 weeks gestation) and non-supplemented (born >32 weeks & <37 weeks gestation) preterm infants have significantly different taxonomic and functional profiles when leaving care. The study recruited 94 preterm infants, 63 of which were supplemented with probiotics and 31 not supplemented and collected 94 stool samples (one for each infant) at 35.3 and 39.7 weeks for supplemented and non-supplemented groups respectively (Table 4.1). Most infants across the

cohort received antibiotic therapy, with several infants from the supplemented (17) and non-supplemented (6) groups still receiving treatment at the time of collection. Additionally, 16 infants in each probiotic group were born vaginally, representing a greater proportion in the non-supplemented group, and most infants received either formula or a combination of formula and breastmilk, with a greater proportion of probiotic-supplemented infants being breastfed relative to non-supplemented infants. The average duration of probiotic prophylaxis was 33.4 days, with 42 samples collected after the cessation of prophylaxis, and an average time since cessation of 23 days. All samples underwent 16S rRNA gene amplicon sequencing, and a subset, 3 probiotic-supplemented and 3 non-supplemented, also underwent shotgun metagenomics.

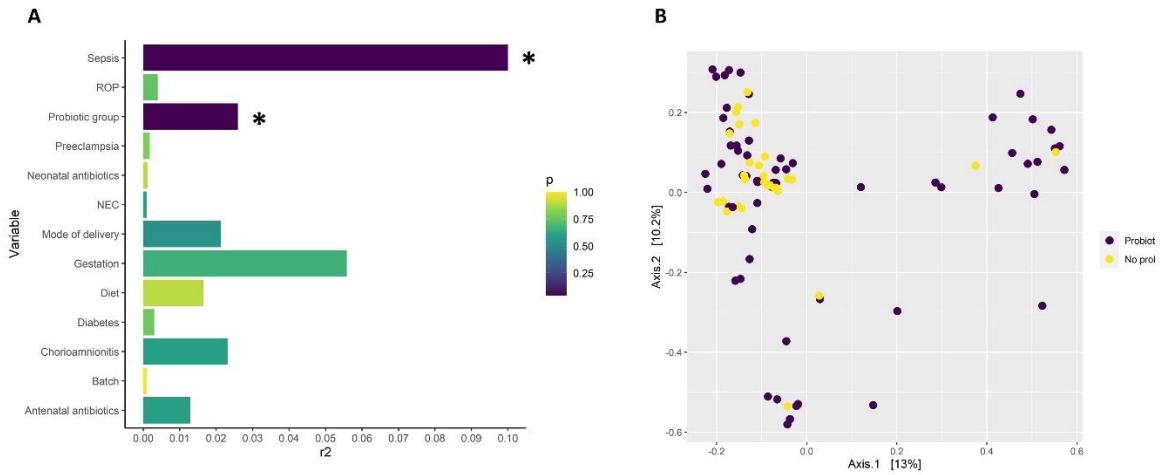
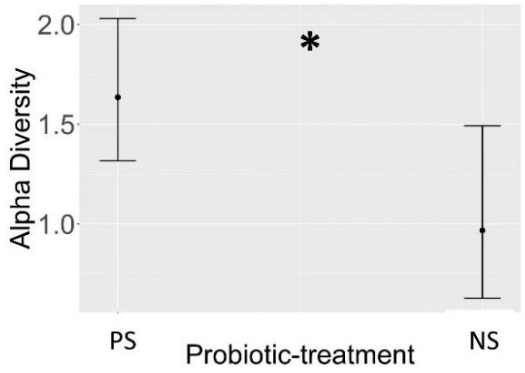


Figure 4.1 ***A:*** *Significance and the amount of variance in gut microbiome composition explained by several microbiome covariates modelled with EnvFit on an NMDS ordination based on Bray-Curtis distances from the 16S rRNA short amplicon sequencing data. The x axis describes the explained variance (r^2) and the colour the p value (adjusted for false discovery rate with the Benjamani-Hochberg method). Annotation for necrotising enterocolitis; NEC: retinopathy of prematurity; ROP. P < 0.05 = *. B: Principle coordinate analysis (PCoA) plot based on ASV level taxonomy obtained through 16S rRNA short amplicons sequencing describing the dissimilarity of probiotic-supplemented (n = 63) and non-supplemented groups (n = 31) based on taxonomy.*

A



B

lfc	lfcSE	padj	Genus
28.38	2.37	***	Enterobacter
21.13	3.44	***	Cronobacter
8.83	1.62	***	Klebsiella
9.93	2.03	***	Lactobacillus
6.31	2.02	**	Veillonella
-4.08	1.60	*	Streptococcus
3.27	1.48	*	Bifidobacterium
5.49	2.46	*	Clostridium_sensu_stricto_1

*Figure 4.2 A: Dot whisker plot of the estimates for the probiotic-supplementation covariate resulting from a generalised linear mixed effects regression model, exploring the effect of several known microbiome covariates on the Shannon diversity index derived from 16S rRNA short amplicon sequencing, and based on ASVs transformed through total sum scaling, B: Table describing significantly differentially abundant taxa between probiotic-supplemented (base-level) and non-supplemented infants, using 16S rRNA short amplicon sequencing, as determined by DESeq2 analysis, based on data transformed through DESeq2's variance stabilising transformation. Annotation for probiotic-supplemented; PS: non-supplemented; NS: p-adj: adjusted p value; lfc: log2-fold change; lfcSE: log2-fold change standard error; NT. $P < 0.05$ = *, $P < 0.01$ = **, $P < 0.001$ = ***. Sample sizes; probiotic supplemented = 63, and non-supplemented = 31.*

16S rRNA gene amplicon sequencing analysis

16S rRNA gene amplicon sequencing showed probiotic prophylaxis influences the preterm-infant faecal microbiome, having a significant effect across all three metrics measured: alpha and beta diversity, and taxonomic abundance. Probiotic supplementation, along with sepsis, were the only covariates found to have a significant association ($P < 0.05$) with ASV level bacterial profiles (*Figure 4.1A, Figure 4.1B*), with sepsis explaining more variation (*r2 = 0.33*). In addition, infants supplemented with probiotics had significantly higher alpha diversity ($P < 0.05$) than non-supplemented infants (*Figure 4.2A*), as well as significantly differential abundance of several taxa. This included higher abundance of *Enterobacter, Cronobacter, Klebsiella, Veillonella and Clostridium Sensu Stricto 1,* as well as the probiotic-genera *Bifidobacterium* and *Lactobacillus,* and lower abundances of *Streptococcus* (*Figure 4.2B*). *Bifidobacterium* and *Lactobacillus were* observed in 55 and 39 of the 63 discharge infants, in contrast to 10 and 6 in the non-supplemented group (*Supplementary File 1*).

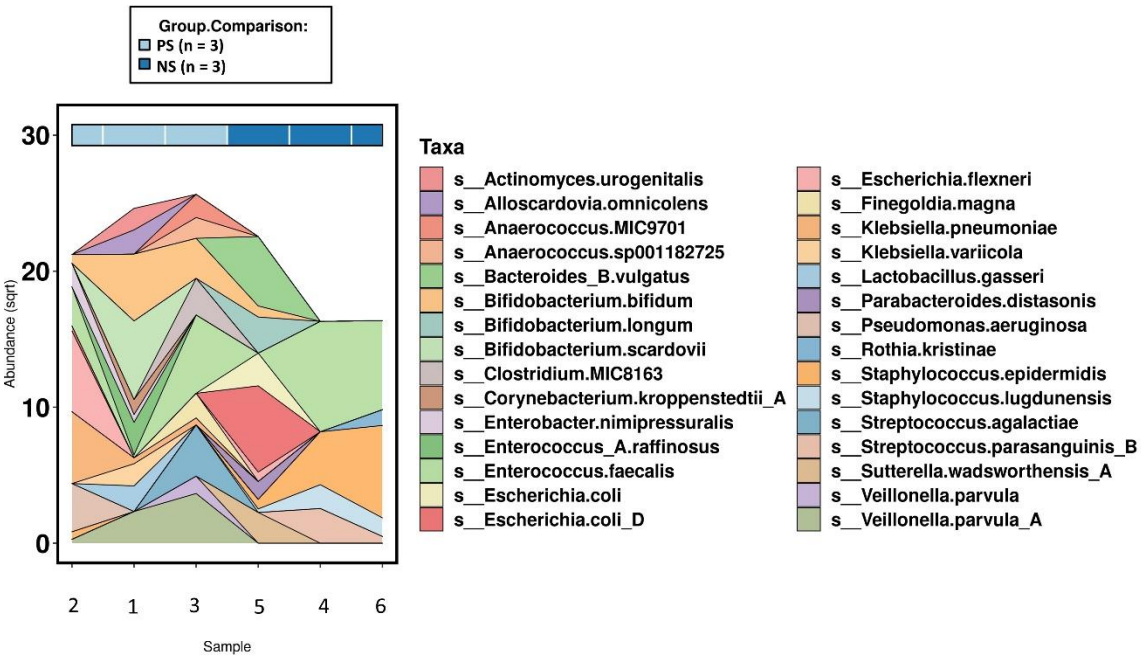


Figure 4.3 Area chart of species level abundances (top 30-most abundant) across the subset of samples that underwent shotgun metagenomics (n =6). Abbreviations: PS = probiotic-supplemented, NS = not supplemented, sqrt = square root transformation.

Shotgun metagenomic sequencing analysis

The results from the shotgun metagenomics showed that there was a high rate of colonisation with *Enterococcus faecalis*, as well as other aerobic species from the Proteobacteria phylum across all samples. Infants were also commonly colonised with skin dwelling microbes, such as *Streptococcus* spp., *Staphylococcus* spp., and *Veilonella* spp.. However, despite these cohort-wide trends, probiotic supplementation still appeared to have some effect on the faecal microbiome. Although, not significantly different, samples clustered by supplementation-group for species-level taxonomy, MetaCyc pathway, MetaCyc group and EC number profiles (Supplementary Material), suggesting distinct taxonomic and metabolic profiles. Due to the small sample size, I was unable to account for other covariates with the shotgun metagenomics analysis. However, no infants in this subset were diagnosed with sepsis, the only other significant beta diversity covariate identified through 16S metabarcoding. Alpha diversity metrics were supportive of what was observed at the genus level (Supplementary Material), with significantly higher species-level alpha diversity in probiotic supplemented infants (Shannon index, $P < 0.05$), which also translated into significantly higher MetaCyc pathway diversity (Richness, $P < 0.05$).

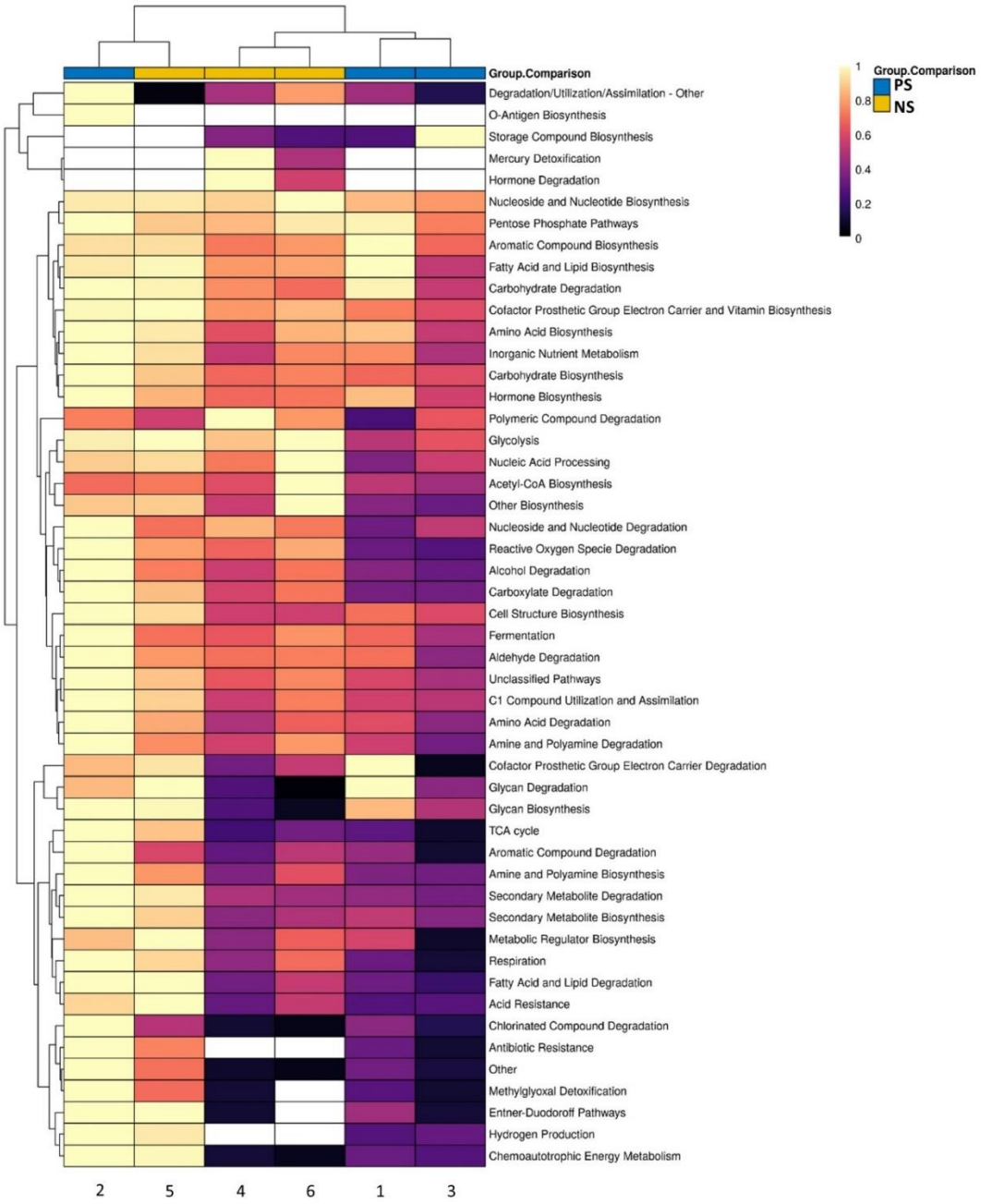


Figure 4.4 Scaled heatmap of MetaCyc groups for the subset of samples that underwent shotgun metagenomics across probiotic-supplemented (n = 3) and non-supplemented (n = 3). Colour scale shows not-detected (white), and abundances ranging from low (black) to high (yellow).

When comparing species and functional profile abundance between the probiotic-supplemented and non-supplemented groups, there were no significant differences when adjusting for multiple comparisons. However, there were several taxa that were only present in one group or the other, resulting in their ranking as top associations. The top associations

(by p value) in species level abundances were *Staphylococcus lugdunensis*, *Veilonella parvula* and *Klebsiella pneumoniae*. *S. lugdunensis* was only observed in non-supplemented infants and the latter two species only in probiotic-supplemented infants (*Figure 4.3*), supporting what was observed at the genus level. Additionally, probiotic-supplemented infants showed a different probiotic species colonisation pattern compared with non-supplemented infants. With the exception for *B. bifidum* and *B. longum* in a single infant, no *Lactobacillus* spp. or other *Bifidobacterium* spp. were observed in the non-supplemented infants. In contrast, *B. bifidum* was observed in all three of the probiotic-supplemented infants and one of the non-supplemented individuals. The species made up 9.8%, on average, of the total species relative abundance in the supplemented group and only 0.12% in the non-supplemented. *Lactobacillus acidophilus* was observed in only two of the probiotic-supplemented infants, but with only 0.23% of the total species abundance. Despite the different colonisation patterns between the two groups, neither univariate comparison resulted in a significant difference.

There were no significant differences between the supplemented and non-supplemented probiotic groups for functional genetic groups. However, there were several note-worthy differences observed within the MetaCyc group profile (*Figure 4.4*). This includes the presence of *Antibiotic Resistance* and *Hydrogen Production* groups in all three probiotic-supplemented infants relative to one non-supplemented infant, and the top associations of *Reactive Oxygen Species Degradation* (greater in probiotic-supplemented) and *Carboxylate Degradation* (greater in non-supplemented) (*Figure 4.4*).

Discussion

This work explored the faecal microbiome of preterm infants, comparing microbial populations between infants who received probiotic-supplementation and those who did not. Specifically, a combination of 16S rRNA gene amplicon (full cohort) and shotgun metagenomic sequencing (a subset of the cohort) was used to determine if differences exist in the faecal microbiome between probiotic-supplementation groups. The results suggest that a significant difference does exist in the bacterial profiles of probiotic-supplemented and non-supplemented preterm infants at discharge from the hospital (36 ± 0.5 weeks gestation), and that these differences in taxonomy may translate into differences in functional profiles. In addition, the probiotic-taxa contained within Infloran® may colonise most infants. Although

these findings may currently have limited direct translation in the clinic, they add weight to the argument for expanding the probiotic supplementation criteria in Australia, which is currently limited to those infants born < 32 weeks gestation and < 1500 g.

Probiotic groups have distinct microbiomes, characterised by greater alpha diversity in those supplemented

Probiotic supplementation may contribute to differences in gut microbiome diversity. The results suggest that there was significant variation in the alpha diversity of the faecal microbiome between the two groups, with probiotic supplemented infants having significantly greater alpha diversity. This suggests Infloran® may be contributing to the establishment of a more diverse, and in turn, healthier gut microbiome. In addition, these results run counter to what one would expect when comparing early- and late-preterm infants not supplemented with probiotics, as lower alpha diversity is typically associated with lower gestational age [69,74]. However, there is evidence to support increased diversity in response to probiotic supplementation in extremely preterm infants [368], which may also be compounded by the widespread use of antibiotics in the study cohort, and the ability of probiotics to correct for this [369]. With greater gestational age, microbial diversity increases, and the microbiome becomes more stable [370]. This increase in diversity is protective against instability [371,372], meaning protective against overgrowth by opportunistic pathogens, as seen in diseases such as NEC and LoS [81,88]. As a result, non-supplemented late-preterm infants may, therefore, be missing out on protection provided through higher diversity afforded via probiotic supplementation. However, it is worth noting that although higher diversity can be indicative of greater microbiome health, it may not always be the case. A prime example of this is the significant association previously shown between breastfeeding and low alpha diversity [56,373]. In addition, because of the links between greater age and alpha diversity, the greater diversity seen in the supplemented group may also reflect the greater post-menstrual age of non-supplemented infants at sample collection. Thus, caution should be taken when interpreting these results, and more broadly, when using alpha diversity metrics as a proxy for gut microbiome health.

Higher rate of detection and abundance of probiotic-taxa in those supplemented

Taken together, the 16S rRNA short amplicon and shotgun metagenomic sequencing suggest the probiotic-taxa, *Lactobacillus acidophilus* and *Bifidobacterium bifidum*, may colonise the

infant microbiome, but not consistently. This pattern of probiotic-species detection is supported by previous work [12,13]. The shotgun metagenomic sequencing, that was performed on a subset of the cohort, was able to identify *B. bifidum* across all three supplemented infants and *L. acidophilus* in one. The low level of *L. acidophilus* detection has been reported previously [12,13], and unfortunately, may result from poor product-quality. These quality assurance concerns are highlighted by an inability to produce robust quantification of *L. acidophilus* in the past [13], and 16S rRNA sequencing of the probiotic itself in my work, which found uneven proportions of taxa within the probiotic Infloran™ at the genus level, dominated by *Bifidobacterium* (Appendix Figure 1). The cause of these irregularities is unclear, however, this is not the first time irregularities in the microbial profiles of probiotic supplements have been observed [114].

With 16S rRNA short amplicon sequencing, both *Bifidobacterium* ($p < 0.05$) and *Lactobacillus* ($p < 0.001$) were in significantly greater abundance in the probiotic supplemented group and were observed in 55 and 39 of the supplemented infants respectively. In contrast, I identified only 10 infants with *Bifidobacterium* and 6 with *Lactobacillus* in the non-supplemented group. Although this does not provide direct evidence for widespread probiotic-species colonisation, the higher frequency and abundance of these genera suggests that supplementation with Infloran® promotes the growth of these commensals, which may aid in the fight against pathogenic infection and in immune and metabolic system development [374,375]. The significance of this greater presence of common commensal microbes in a very preterm demographic is compounded by the contrasting observations suggesting a negative relationship between birth gestational age and limited or delayed colonisation with *Lactobacillus* and *Bifidobacterium* [43,78]. However, not all supplemented infants had detectible *Lactobacillus* and *Bifidobacterium.* The limited detection of *Lactobacillus* may be due, at least in part, to issues with the probiotic outlined above. However, why colonisation with *Bifidobacterium* is not consistent remains unclear, as no clinical variable included in the analyses had a negative association with this genus, and all probiotic-supplemented infants that had samples collected at > 36 weeks gestation (post-supplementation) still had *Bifidobacterium* present (Supplementary Material). Whether *Bifidobacterium* colonises the probiotic-supplemented infant gut may be dependent on the complex interaction of multiple factors.

The greater abundance of *Bifidobacterium* colonisation may persist beyond probiotic prophylaxis. As previously mentioned, probiotic prophylaxis for infants at TUH ceases between 34- and 36-weeks gestational age. However, these results suggest that *Bifidobacterium* persists beyond this time point, as the genus was present in all 23 infants with samples collected > 36 weeks gestation. This supports previous studies that have observed long-term probiotic-species colonisation, at least with *Bifidobacterium* [13,376]. Thus, the probiotic species may continue to exert positive benefits beyond the supplementation period. This persistence could benefit moderate and late preterm infants, who despite having a lower relative risk of NEC and sepsis [360,361], are still vulnerable relative to those born full term [377]. Further work needs to be done to explore long-term differences between supplemented early- and non-supplemented moderate/late-preterm infants.

Probiotic supplementation associated with differences in non-probiotic taxa

Enterobacter, Cronobacter, Klebsiella, Veillonella and *Clostridium Sensu Stricto 1* were all higher in probiotic supplemented infants, whilst *Streptococcus* had a greater abundance in those not supplemented. The significance of such modulation is unclear, as despite several notable pathogens within these genera, many other species can be considered normal flora. As early-life microbial colonisation occurs concomitantly with development of the immune system, immune-system maturation is influenced by the presence of commensal microbes [378], therefore, fewer commensal microbes at this stage of life may be detrimental long-term. Although it is unclear whether these specific taxa play a role in morphological or functional development of the immune system [378], their presence will at least lead to preferential development of immune tolerance, reducing the likelihood of such taxa reaching their 'pathogen-potential' later in life. In addition, if these taxonomic differences persist, specifically reduced levels of *Veillonella*, non-supplemented infants may be at a greater risk of chronic diseases like asthma, which has previously been shown to be associated with such differences [89].

At the species level, there were notable differences in *Veilonella parvula*, *Klebsiella pneumoniae* and *Staphylococcus lugdunensis. S. lugdunensis* was found in all non-supplemented but not in probiotic-supplemented infants, and *V. parvula* and *K. pneumoniae* across all probiotic-supplemented but not in non-supplemented infants. Although this appears

to align, in part, to the difference in probiotic-supplemented at the genus level, these differences may be better explained by other variables. For instance, *K. pneumoniae* was one of the most abundant taxa in probiotic-supplemented infants. However, it is possible that this species was selected for through antibiotic treatment [379], which seems likely when considering all three of the probiotic-supplemented infants received antibiotics and that *K. pneumoniae* had the greatest abundance of ABR genes across all species (Supplementary Material). Unfortunately, the sample size of the shotgun analysis was too small to draw conclusions, and future work should apply shotgun methods to a greater sample size to elucidate why there are differences in given taxa.

Differences in functional profiles between probiotic groups

Although there were limited differences in taxonomy between supplemented and non-supplemented infants, this does not mean that there are no physiological consequences. From both an ecological and physiological perspective, several small changes in what may be critical taxa, may have significant consequences, especially if these differences are in taxa that harbour genes critical to key environmental processes. Although not significant, and as previously mentioned, ABR genes were in higher abundance across the probiotic-supplemented group, whilst only present in a single non-supplemented infant, with Hydrogen Production following the same pattern. The presence of ABR and Hydrogen Production genes in the probiotic-supplemented infants is also closely linked to specific species. The previously mentioned abundances of *K. pneumonia* and *V. parvula*, along with *E. flexneri* (in one infant), were the only species across all probiotic-supplemented infants to have these pathways present. Although this example may not have any significant implications, it highlights the functional importance the presence of a single species can have. Another example of this importance is highlighted by the lower abundance of 1.3-beta-galactosyl-N-acetylhexosamine phosphorylase in the non-supplemented group (Supplementary Material). The enzyme is a critical component of an enzymatic system within *Bifidobacterium spp.* that metabolises human milk oligosaccharides. Thus, without species like *B. bifidum*, the non-supplemented infants have less capacity to reap the benefits of breast feeding. Thus, although differences may be subtle and temporary for species like *B. bifidum*, these differences could have larger, long lasting physiological consequences.

Limitations

This study has several limitations. This includes the different ages in the supplementation-groups, the distribution of samples across two sequencing runs, the limited taxonomic depth provided through 16S rRNA gene amplicon sequencing, and the small samples size that underwent shotgun metagenomics. To mitigate the effect of the different ages between the probiotic groups, and the batch-effect introduced through multiple sequencing runs, both variables were included in all three of the mixed effects models. As for the limited taxonomic depth and limited sample size of the sub-cohort that underwent shotgun metagenomics, this could be overcome using shotgun metagenomics across the entire cohort. However, this technique, and others of similar resolution, are cost prohibitive at present [123].

Conclusion

There was a significant difference in overall faecal microbiome community composition between probiotic-supplemented and non-supplemented infants, with alpha diversity greater in the supplemented infants. Moderate to late preterm-infants who go without probiotic-supplementation may be missing out on stabilising-effects provided through probiotic-supplementation, which may help to prevent disease. These results suggest that there could be a role for probiotic supplementation in late preterm infants in North Queensland, Australia. However, caution should be taken when extrapolating from single-centre studies to other locations. In addition, rather than provide answers, the differences in taxonomy prompt more questions. Significant differences exist at the genus level, but what are the consequences of these differences? Additionally, differences observed at both the species and functional level highlight the power of shotgun metagenomic sequencing, and I suggest that as the cost of this technology continues to decrease, that future work should adopt this approach. Obtaining species-level and functional profiles in this cohort would provide us with a better understanding of the physiological and ecological consequences of withholding probiotic-prophylaxis from moderate to late preterm infants.

5. Exploring the long-term colonisation and persistence of probiotic-prophylaxis species on the gut microbiome of preterm infants: a pilot study.

In Chapter 4 I observed significant differences in microbial diversity and composition between probiotic-supplemented, very preterm infants and non-supplemented, moderate to late preterm infants just prior to discharge from the hospital. Chapter 5 builds on this by exploring this same comparison at an age of 18 months to 2 years, to determine if these differences persist post-discharge, as the complexity and stability of the gut microbiome increases. This chapter addresses the final aim of this thesis, which was to investigate differences in the bacterial microbiomes of probiotic-supplemented and non-supplemented preterm infants post-discharge.

This chapter was split into two separate manuscripts. In the first I present a validation of an at-home infant stool sample collection kit, the OMNIgene® GUT, for determining the faecal microbiome for use in longitudinal studies. The results suggest that the OMNIgene® GUT kit is an easy to use and robust and repeatable method for at-home sample collection when immediate freezing is not possible. This method was used for the subsequent manuscript. This work has been accepted for publication in the Australian Journal of Medical Science as "*A validation of at-home infant stool sample collection devices for determining the faecal microbiome*".

The second manuscript from this chapter addresses the aim stated above. The results of this study suggest that the species in the probiotic Infloran® do not persist 18 months – 2 years post discharge, and that the modulation observed at discharge does not persist. Rather, previously supplemented infants were observed to have lower diversity, in combination with differing abundances of three relatively understudied taxa. This component of Chapter 5 has been published in the *European Journal of Pediatrics* (https://doi.org/10.1007/s00431-022-04548-y).

Chapter 5.1: A validation of at-home infant stool sample collection devices for determining the faecal microbiome.

Abstract

In this study I present a validation of an at-home infant stool sample collection the OMNIgene® GUT for determining the faecal microbiome for use in longitudinal studies. Although developments in metagenomics and sequencing technologies have resulted in an explosion in microbiome research, studies exploring the long-term effects of probiotic modulation in infants are still relatively rare due in part to the challenges in the metagenomics methodology and standardised collection and storage techniques. To investigate the efficacy of the OMNIgene® GUT on microbial composition, I compared samples stored using the OMNIgene® GUT kit to the same samples collected in a standard/sterile collection tube and stored at - 80°C for 3 months prior to 16S rRNA metabarcoding. No significant differences in read depth or microbial composition of samples were found between the two methods of collection and storage. When coupled with clear guidelines for self-collection and shipping, the OMNIgene® GUT kit is a feasible option for sampling preterm infants at home.

Introduction

Developments in metagenomics and sequencing technologies have been a catalyst for an explosion in microbiome research. This explosion in research has led to an understanding of the importance of preterm infant gut microbiome development, and its emergence as a modifiable factor in neonatal intensive care [114,380,381]. The role of the gut microbiome in metabolic and immune system development means that positive modulation of bacterial populations in the gut could provide long term metabolic and immune benefits. This is especially important for those infants born preterm, where an immature gut microbiome has been linked to disease. A dysbiotic microbiome puts preterm infants at a greater risk of both acute diseases, such as necrotising enterocolitis (NEC) and sepsis [360,382], and chronic diseases, such as asthma and diabetes [383,384]. The introduction of probiotic supplementation has been shown to alleviate some of this disease burden [10,11], potentially through microbiome modulation [12,13]. Although this probiotic-associated microbiome modulation has been well characterised during their hospital stay, studies exploring the long-term effects of probiotic

modulation are still relatively rare. This sparseness is likely to be linked to challenges in the metagenomics methodology.

Metagenomic studies investigating the microbiome are complex, technically challenging, and vary between laboratories. Each stage of the project protocol can introduce biases that influence outcomes [128,129] and contribute to heterogeneity between studies [123]. Arguably the most important pre-analytical consideration for faecal sampling is storage conditions prior to analysis. This is a particularly important consideration during point of care or at home collection, where freezing is not always an option, and the protocol involves untrained individuals. Thus, the sample collection methods need to combine a simple method with robust storage. The reliability of downstream metagenomics analyses is highly dependent on appropriate storage conditions as this can influence both the stability of DNA and composition of the microbial communities [180].

Inadequate storage protocols may promote growth of specific taxa and can lead to DNA/RNA fragmentation in less than 24 hours at room temperature [180,181]. Commonly available storage methods include freezing or refrigeration, and the use of stabilizing buffers, with the optimal method dependent on the duration of storage [123]. Freezing at - 80°C is considered optimal for long-term storage, as it has been demonstrated to more consistently yield microbiota composition closely related to that of fresh samples [181,186]. However, when immediate freezing is not logistically possible, storing samples in a preservation buffer is preferred, as this can preserve genetic integrity for several weeks [192,193]. However, chemical storage is not without its issues as these storage buffers may result in lower diversity [199,200], and some have been shown to impede downstream DNA extraction and amplification [192]. These concerns may also be product specific.

The objective of this project was to explore the performance of the OMNIgene® GUT kit in preserving microbial communities during at home collection. The OMNIgene® GUT kit is an all-in-one system for self-collection and stabilisation of faecal samples containing microbial DNA [385]. The kit includes a collection tube with 200 μl of stabilising liquid. The sample is deposited into the tube and mixed with the stabilising liquid by shaking. Here I present a pilot study that compares storage of preterm infant stool samples with the OMNIgene® GUT kit against storage in sterile collection kits at - 80°C. I aimed to determine if at home sample

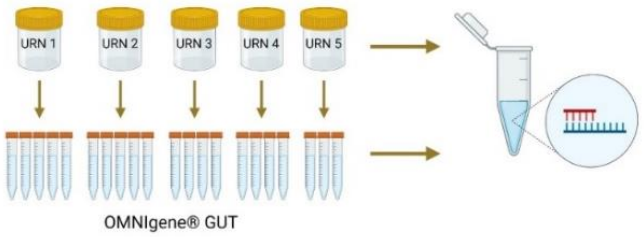
collection for future preterm infant studies, using the OMNIgene® GUT kit, is a feasible protocol.

Methods

Study Design

To investigate the effect of the OMNIgene® GUT on microbial composition, samples stored using the collection kit were compared to the same samples collected in a standard/sterile collection tube.. I chose a subset of samples (n = 5) from a previous study, using methods previously described [380], and aliquoted them into OMNIgene® GUT kits. Each sample had random proportions of stool added to mimic at home collection. The number of kits that were used for each sample was dependent on the amount of stool in the initial collection and the initial sample was used as the base level for the comparison. This resulted in a total of 26 technical replicates (21 stored with OMNIgene® GUT), from 5 different biological replicates (*Figure 5.1*), with all technical replicates stored at -80°C for 3 months, from the day they were prepared. As part of the previous study, samples were immediately frozen and stored for ~12 months, and then subsequently thawed for DNA extraction. Thus, a limitation of this study is that samples had been through one previous freeze-thaw cycle. The biological replicates are designated by the variable URN (unit record number), that has been de-identified from the original URN.

A



B

Biological replicates	Technical replicates	
URN	Original sample	OMNIgene® GUT
1	1	5
2	1	5
3	1	4
4	1	4
5	1	3

Figure 5.1 A; Visual description of the sampling method, where five original biological replicates were used to create 26 technical replicates by aliquoting the original replicates into OMNIgene® GUT collection kits. B; a table describing the number of original replicates, and resulting technical replicates that were used for the analyses.

16S rRNA short amplicon sequencing

The Bioline ISOLATE Fecal DNA Kit was used for DNA extraction [331], which involves mechanical lysis, with modifications made in consultation with the manufacturer to optimise DNA yield. Modifications included increased beta-mercaptoethanol from 0.5 to 1% (increasing DNA solubility and reducing secondary structure formation), addition of an extra wash step (improving purity) and decreased elution buffer volume from 100µl to 50µl (increasing final DNA concentration). Additionally, a modification was made for compatibility with OMNIgene® GUT. After consultation with the manufacturer, a volume of 150 µl was chosen for the initial sample volume, with 90-150 mg used for the original samples. All previously listed modifications were applied to both sample types.

The Illumina metagenomics library preparation protocol was used for library preparation [332], using the Index Kit v2 C [333], along with Platinum™ SuperFi™ PCR Master Mix [334]. 16S rRNA short amplicon sequencing was performed on the Illumina MiSeq system using the MiSeq Reagent Kit V3 [333], targeting the V3 and V4 regions with the S-D-Bact-0431-b-S-17/S-D-Bact-0785-a-A-21primer combination [332]. Pre-analytical bioinformatics were conducted in *R Studio* Version 3.6.1 [335] with a pipeline adapted from our previous work [380],

which can be found in the supplementary material. *DADA2* [337] was used for quality filtering and trimming, demultiplexing, denoising and taxonomic assignment (using the SILVA Database), and the *microDecon* package [285] was used to remove homogenous contamination from samples using blanks originating in extraction. The OMNIgene and original samples were processed, from extraction through to analysis, in parallel.

Statistical analysis

To compare storage methods, whilst accounting for the biological replicates, read depth, beta diversity, alpha diversity, and taxonomic abundance were assessed using mixed effects models. For beta diversity comparisons, I performed both a PERMANOVA and an *EnvFit* analysis from the *Vegan* package [367], which compare the differences in the centroids relative to total variation. Both analyses were applied to a Bray-Curtis dissimilarity matrix [339] based on data normalised through Total Sum Scaling (TSS) [297]. The significance was based on 10,000 permutations and was transformed using the Benjamini-Hochberg (BH) procedure [340].

For both alpha diversity and read depth, I performed generalised linear mixed effects models, using the package *lme4* [341]. For diversity, both richness and the Shannon Index were calculated at the ASV level. Multicollinearity was assessed using the *AED* package [342], and significance was determined using an analysis of deviance (Type II Wald Chi-square test) from the *car* package [343]. This was followed by subsequent post-hoc pairwise Tukey comparisons, to correct for multiple comparisons, using the *emmeans* package [344].

DESeq2 [294], which uses a negative binomial generalized linear model and variance stabilising transformation, was used for comparing taxonomic abundances between groups. Taxa were agglomerated at the genus level, due to the limited taxonomic depth of 16S-target technologies. A Wald Test with the BH multiple inference correction was performed to obtain taxa that were significantly differentially abundant among groups of interest. The pre-analytical bioinformatics and statistical analyses can be found in the GitHub link in the supplementary material.

Results

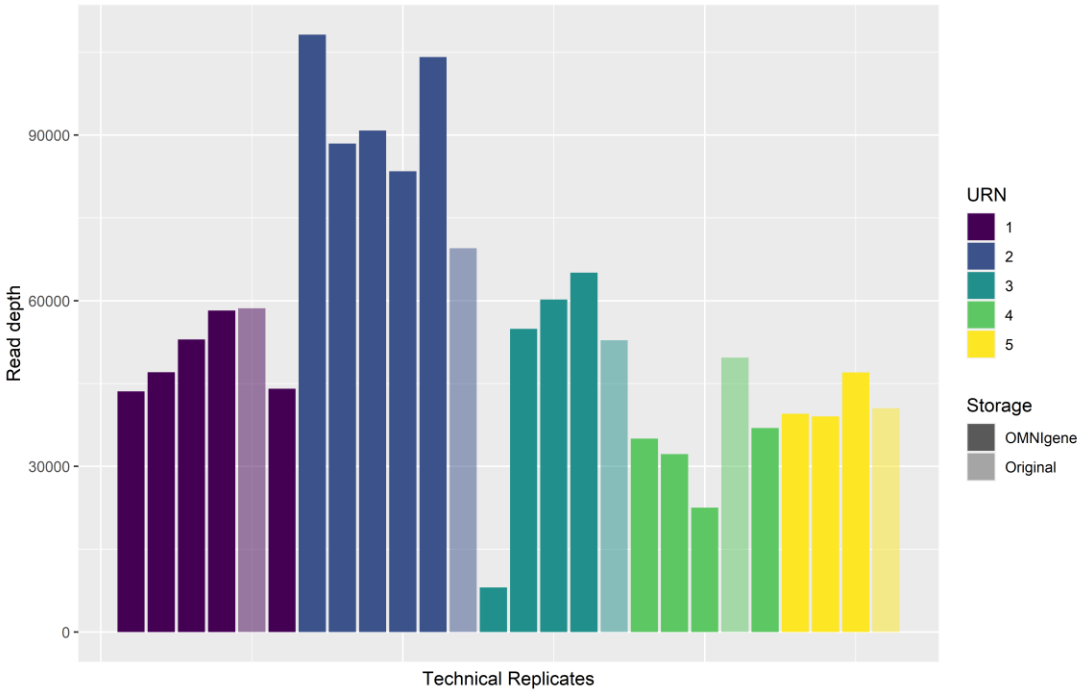
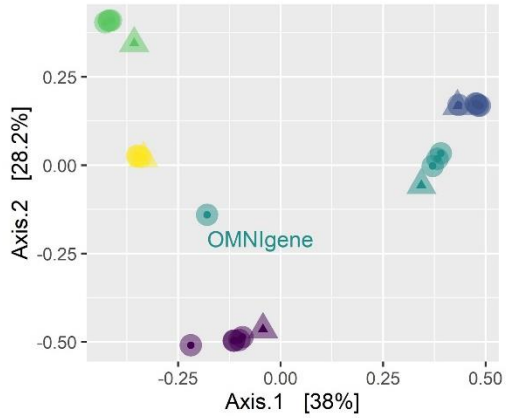


Figure 5.2 Column graph illustrating the similarities in read depth between technical replicates and coloured by biological replicates. Storage method had no significant effect on read depth (p = 0.87).

A



B

Analysis	Variable	r2	Adjusted-p
PERMANOVA	Storage	0.88	<0.01
	URN	<0.01	0.28
Homogeneity condition			0.90
envfit	Storage	<0.01	0.92
	URN	0.93	<0.01

Figure 5.3 A; Principle coordinate analysis plot based on Bray-Curtis distances describing the similarity/dissimilarity of samples based on taxonomic composition and demonstrating the clustering of samples based on URN (p < 0.01) and not the storage method (p = 0.24), with the exception of a single outlier (designated OMNIgene). B; Table describing the results from both a PERMANOVA performed with adonis2 (and the subsequent test for homogeneity of variance) and an envfit analysis performed on Bray-Curtis distances.

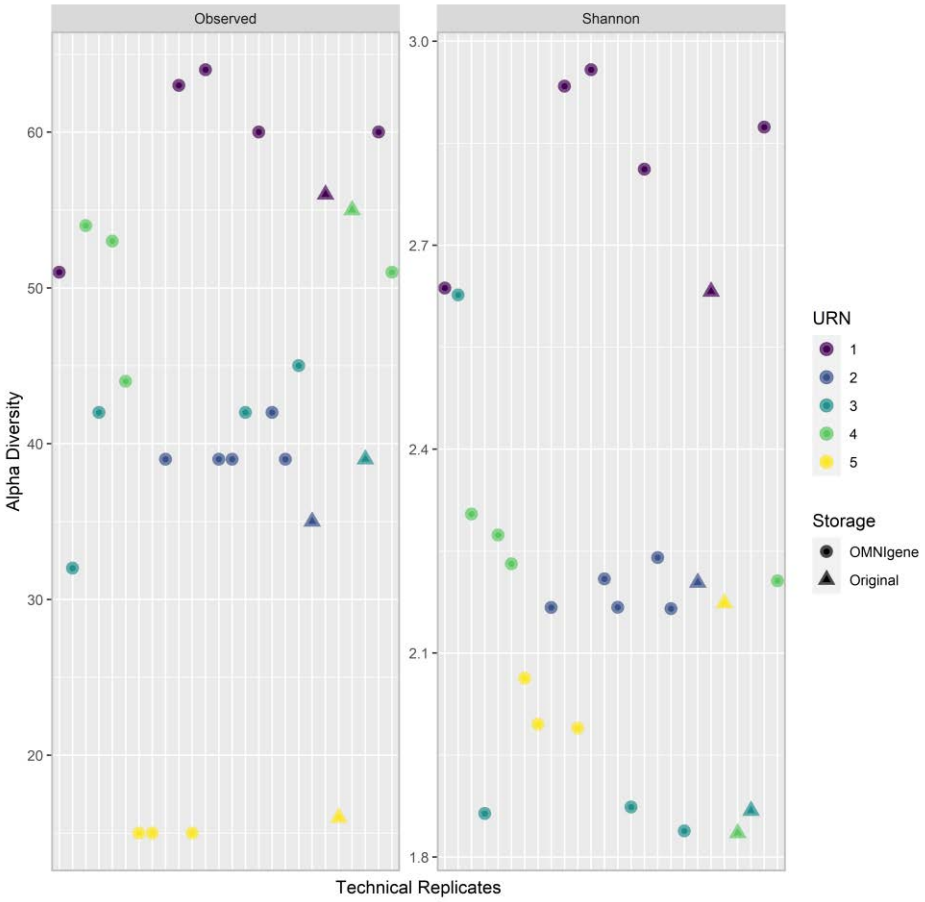


Figure 5.4 Dot plots, coloured by the URN and faceted by alpha diversity measure, representing the distribution of alpha diversity metrics across samples in the storage comparison. Alpha diversity was significantly associated with URN (Shannon: $p < 0.001$ & richness: $p < 0.001$) and not sample storage (Shannon: $p = 0.19$ & richness: $p = 0.74$).

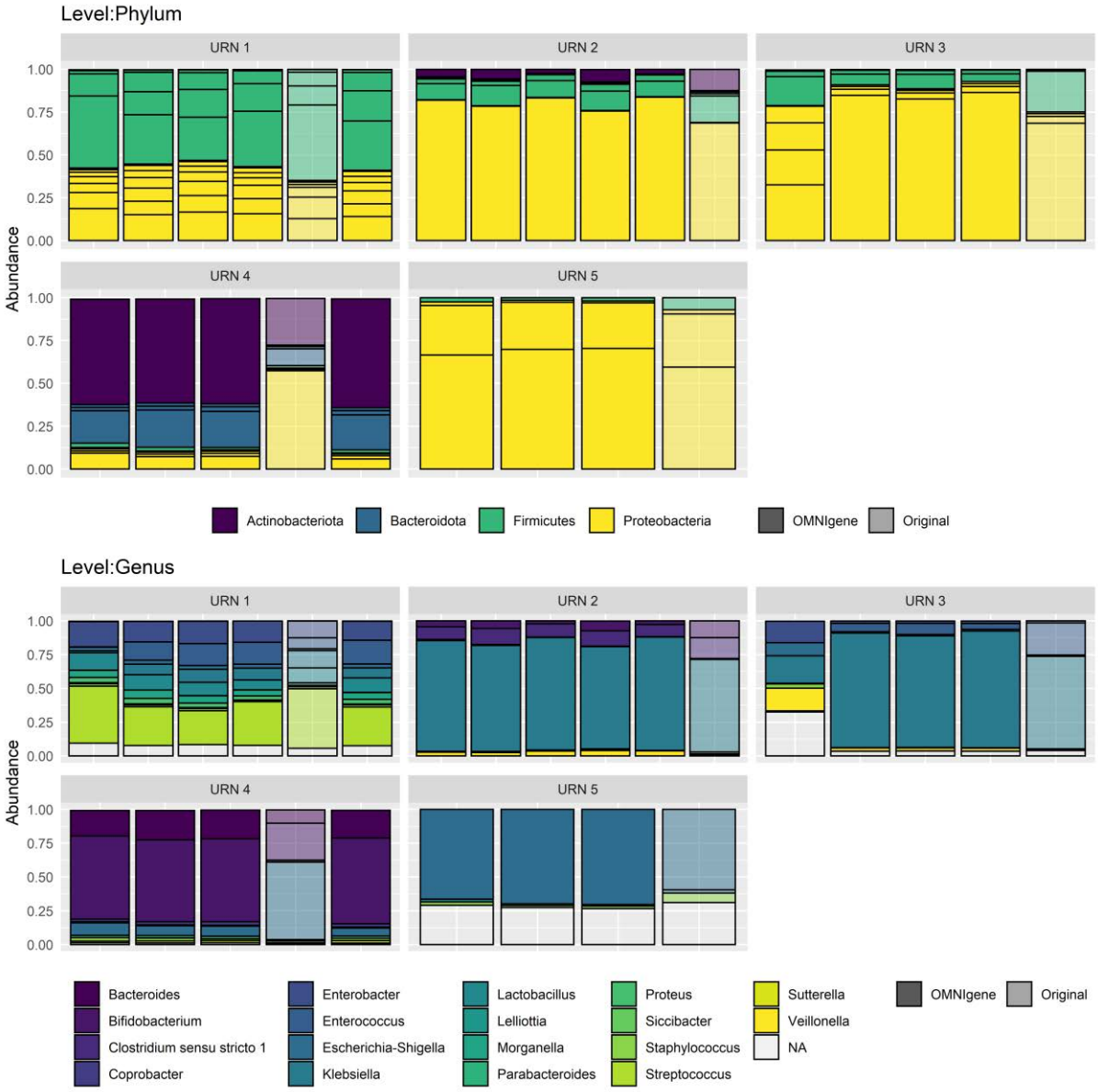


Figure 5.5 Bar charts comparing the distribution of the top 20 most abundant genera (colours) across technical replicates (individual bars), and within URN (facets) for the storage method comparison.

There was no significant different between storage with the OMNIgene® GUT collection kit and storage in sterile collection kits at - 80°C. The OMNIgene® GUT collection kit did not have a significant effect on read depth (*Figure 5.2*) or microbial composition (*Figures 5.3-5*). The results of the analyses show that the samples clustered by the biological replicates (URN), rather than storage type. URN had a significant association with overall community composition (*Figure 5.3A*) and explained most of the variation/similarity between samples (*Figure 5.3B*), as demonstrated by both PERMANVOA ($p < 0.05$) and envfit ($p < 0.05$)

mixed effects models. Despite some variation (*Figure 5.4, Shannon Index:* $\bar{x} = 2.27 \pm 0.35$, *Richness:* $\bar{x} = 42.5 \pm 14.74$), the storage method also had no significant effect on alpha diversity (*Shannon Index:* $p = 0.19$, *Richness:* $p = 0.74$). Lastly, when exploring taxonomic abundance, it did appear that the original sample had differences in taxonomy relative to those from the OMNIgene® GUT collection kit. However, differential abundance testing with DESeq2 demonstrated that despite this difference, that overall, the storage method had no significant effect on the abundance of taxa (*Figure 5.4*). Additionally, despite this variation in taxonomy between sample storage within URN 4, these samples still clustered based on taxonomic composition (*Figure 5.2A*). Using the OMNIgene® GUT collection kit is a viable option for at home collection when immediately freezing at -80°C is not possible.

Discussion

Optimal storage conditions are a key starting point of pre-analytical variation for all metagenomics studies. Without optimal storage conditions, samples may be compromised, and any conclusions drawn from the data unreliable. Previous work has found that storage on ice for up to 48 hours [184], or 4°C for 24 hours [185] is sufficient if samples are to be processed immediately. However, for long-term storage, freezing at -80°C is best as it inhibits bacterial growth and/or degradation, as demonstrated by similarities between long term storage of faecal samples at -80°C and fresh samples [181,186,187]. Thus, it is common for samples to be stored at -20°C or 4°C, until freezing at lower temperatures is possible. Alternatively, preservation buffers, like DNA/RNA Shield (Zymo Research) and RNAlater (Thermofisher), have been shown to preserve genetic integrity for weeks without refrigeration or freezing [192-198]. However, some preservation buffers may result in lower diversity, or impede downstream DNA extraction and amplification of target variable regions [192]. The objective of this study was to determine if storage with OMNIgene® GUT kit had a significant effect on metagenomic outcomes, and thus, was a feasible option for at home sample collection.

The DNA obtained from samples stored in the OMNIgene® GUT was sufficient for DNA extraction, amplification of variable regions and sequencing. The adequacy of the DNA extract can be assessed by estimating the read depth captured during sequencing. Sufficient read depth is necessary for capturing an accurate representation of microbiome composition and for producing accurate diversity metrics [297]. The results of this study show that there is no

significant variability between technical replicates belonging to the same biological replicate. The only outlier is a technical replicate within *URN 3* with low read depth, which is also evident in the PCoA. It is unclear as to why this sample has low read depth, relative to the other technical replicates. However, this could be due to an error in the aliquoting or library preparation protocol. Importantly, this sample was from an OMNIgene® GUT kit, the other three technical replicates stored in the same manner within this URN are relatively homogenous.

The similarities in read depth between storage methods translated into similarities in microbial composition. Similar to previous work by Hill et al. [200], samples in this study clustered significantly by the biological replicate, suggesting that the storage method has no effect on microbial composition, despite some taxonomic variation in URN 4. Other studies have demonstrated deficiencies in some preservation treatments can result in a reduction in alpha diversity and particular taxa, however, the microbial diversity and composition produced by the OMNIgene® GUT kit was similar to storage in a standard collection jar at - 80°C. Thus, despite the reliability of non-freezing preservation methods being called into question in the past [192,199,200], the counter narrative presented in this study is likely to be an effect of different products producing different results, which has also been noted previously [199].

Although the OMNIgene® GUT kit does not impact microbial communities, several limitations not assessed here should be noted. Firstly, the kits come at a significant cost, and if funds are limited then this will translate into a limited sample size. Secondly, the preservation method is specific to the genetic material, and thus can only be coupled with DNA extraction and associated downstream methods. Thus, the kit can limit both the size and scope of the analyses.

The present study shows that the OMNIgene® GUT kit as a reliable and repeatable storage option for faecal microbiome studies. This supports results from a recent study by Szopinska et al., who concluded that the OMNIgene® GUT kit is a participant-friendly collection method [386]. Not only is the kit reliable in its preservation of microbial communities, it is coupled with clear guidelines for self-collection [385]. The OMNIgene® GUT kit has no deleterious effects on preservation of microbial communities, and when coupled with clear guidelines for self-collection and shipping, is a feasible option for sampling preterm infants at home.

Conclusion

The current study demonstrates that the OMNIgene® GUT kit is an easy to use and robust and repeatable method for at-home sample collection when immediate freezing is not possible. The kit produces microbial populations that do not deviate significantly from what is considered the gold standard for long term storage. Using this method would allow for at-home sampling as part of an investigation into the gut microbiome of preterm infant, post-discharge. Although conducting such research is what instigated this investigation, these finds are translatable and support at-home sampling with the OMNIgene® GUT kit for other microbiome studies using human stool samples.

Chapter 5.2: Exploring the long-term colonisation and persistence of probiotic-prophylaxis species in the gut microbiome of preterm infants: a pilot study.

Abstract

Preterm infants suffer from a higher incidence of acute diseases such as necrotising enterocolitis and sepsis. This risk can be mitigated through probiotic prophylaxis during admission. This reduction in risk is likely the result of acute modulation of the gut microbiome induced by probiotic species, which has been observed to occur up until discharge. I aimed to determine if this modulation, and the associated probiotic species, persisted beyond discharge. I conducted both a cross-sectional analysis (n=18), at ~18 months of age, and a longitudinal analysis (n=6), from admission to 18 months of the gut microbiome of preterm infants using both shotgun metagenomics and 16S rRNA profiling. *Conclusion*: The metagenomics analyses suggest that the species from the probiotic Infloran®, as well as the positive modulatory effects previously associated with supplementation, may not persist beyond discharge and once prophylaxis has stopped. Despite the lack of long-term colonisation, the presence of probiotics during early neonatal life may still have modulatory effects on the microbiome assembly and immune system training.

Introduction

Preterm birth, defined by the World Health Organisation as < 37 weeks gestation [387], disrupts gut microbiome development [68]. The resulting preterm microbiome is characterised by low diversity and commensal microbe abundance, in combination with a greater number of pathogens [73,388]. This characteristic preterm microbiome has been linked to increased disease burden in these infants [86]. This includes acute diseases like necrotising enterocolitis (NEC) and late-onset sepsis (LOS), and chronic diseases like asthma and both type 1 and 2 diabetes, all of which have been linked to the microbiome [178]. However, probiotic prophylaxis can mitigate the risk of these acute diseases [11]. As a result, probiotic prophylaxis has now become the standard of care for the most premature (< 32 weeks gestation) and small for gestational age infants (< 1,500 g) in neonatal intensive care units (NICUs) across Australia.

Probiotic prophylaxis has been demonstrated to mitigate and treat several infectious and non-infectious diseases through modulation of the gut microbiome [389]. This includes *Helicobacter pylori* infection [390], rotavirus infection [391], obesity [392] and allergies [393]. In preterm infants,

probiotics have been shown to reduce the incidence of both NEC and LOS, with the benefits likely stemming from changes in the microbiome afforded by the presence of probiotic strains [389]. These strains, specifically from *Bifidobacterium* and *Lactobacillus*, have been shown to contribute to a *Bifidobacterium*-dominated microbiome, which, in turn, can positively modulate immune system activity and development [178,394]. To date, much of the research has been focused on improving survival for very and extremely preterm infants, who suffer a higher burden of disease and death in early life. As such, there is little data exploring the long-term impact of probiotics or their use in older preterm demographics.

Certain probiotic species have been shown to persist beyond discharge [13], possibly continuing to exert positive effects on the development of the infant gut microbiome. However, this observation of probiotic persistence is not consistent [13,376], and as infants have been shown to cluster in their microbial populations by NICU [113]; caution should be taken when extrapolating from these single unit studies to another unit. We have demonstrated in a previous study that probiotic-prophylaxis had a significant positive modulatory effect on very preterm infants over the course of their hospital admission [395]. We observed greater diversity in the gut microbiome of probiotic-supplemented preterm infants, relative to those not supplemented, at discharge from the hospital, suggesting that preterm infants who fall outside the criteria for probiotic prophylaxis (defined as <32 weeks gestation and/or <1500 g) may be missing out on the positive modulatory effects for healthy gut microbiome development. Our aim for this study was to investigate if these differences persist following discharge up to 1.5 - 2 years of age and conduct both a cross-sectional and longitudinal analysis of the gut microbiome of these preterm infants, using both shotgun metagenomics and 16S rRNA profiling. We were particularly interested to determine if the probiotic species, specifically *Lactobacillus acidophilus* and *B. bifidum* were persisting in the gut long-term. In addition, to determine if probiotics had a lasting modulatory effect on microbiome development, we compared these probiotic-supplemented infants to a group of infants who were born into the same nursery but did not receive probiotic supplementation. Lastly, we combined these newly acquired samples with data collected previously, to conduct a longitudinal examination of probiotic supplemented infants.

Methods

Study design

This observational study involves both a longitudinal and cross-sectional component. As the main objective of this project was to examine if probiotic prophylaxis during admission has a lasting effect, I performed a cross-sectional analysis of 18 infants using shotgun metagenomics and compared those who had received probiotics against those who had not. As previously mentioned, a subset of this cohort (n = 6) had samples collected as part of an earlier study, and so I also performed a longitudinal analysis of these probiotic-supplemented infants using 16S rRNA amplicon sequencing, as this was the technique used previously.

Study population

Infants recruited were previously admitted to the Townsville University Hospital's (TUH) Neonatal Intensive Care Unit (NICU) and Special Care Nurseries (SCN). The TUH NICU is the only level six tertiary referral unit in NQLD, which is a specialised unit for dealing with complex pregnancies. The criteria for probiotic prophylaxis at the TUH NICU dictates that all high risk preterm infants (defined as <32 weeks gestation and/or <1500 g) receive Infloran® [330], containing *Lactobacillus acidophilus* (1×10^9 CFU) and *Bifidobacterium bifidum* (1×10^9 CFU) on a daily basis, from the first day of feeding to > 34-36 weeks gestation. Inclusion criteria for the cohort included: born <32 weeks gestation and previously admitted to the NICU at the TUH for the probiotic group, and >32 weeks and admitted to the SCN at the TUH. The exclusion criteria were no parental consent, born >32 weeks and contraindication to enteral feeds for the probiotic group, and no parental consent for the non-supplemented group. Ethics was obtained from the Townsville Hospital and Health Service Human Research Ethics, (HREC/QTHS/65181 and HREC/17/QTHS/7). Informed consent was obtained from parents/legal guardians of all subjects through the signing of a Parental Information Sheet and Consent Form (PICF), which can be found in the supplementary material.

Recruitment and sample collection

Infants previously admitted to the TUH were recruited by a neonatal nurse between January and August of 2021. At recruitment, these infants were between 18 months and 2 years of

age. Parents/guardians of previously admitted preterm infants were contacted via the phone, and upon verbal approval, mailed out a Parental Information Sheet and Consent Form (PICF), and collection kit. The collection kit included:

- OMNIgene® GUT all in one system.
- Paid return postal package.
- Detailed instructions on sample collection and postage.
- Absorbent material and leak proof biohazard bag for postage requirements.
- Questionnaire.

The samples of recruited infants were stored in the OMNIgene® GUT collection tube, and once mailed back to the research team, the tubes were stored at -80°C, as recommended by the manufacturer.

Collection of metadata

Both clinical (during admission) and post-discharge metadata were also collected. For the clinical data, this included both maternal – antenatal antibiotics, chorioamnionitis (clinically diagnosed), preeclampsia (clinically diagnosed), and diabetes (type 1 & 2), and infant data – mode of delivery (vaginal birth versus caesarean section), diet, gestation at birth and collection, NEC (stage 2 or greater), sepsis (confirmed through culture), neonatal antibiotics and Retinopathy of Prematurity (ROP) (stage 1 or greater). The post-discharge information was collected through the previously mentioned, brief questionnaire.

Categorical Variables

Variables	**Levels**	**Probiotic-supplemented**		**Non-supplemented**	
		n	**%**	**n**	**%**
Probiotics post-discharge	Yes	9	64.3	0	0
	No	5	35.7	4	100
Diet during admission	Combination	8	57.1	3	75

	Breastmilk	6	42.6	1	25
Diet post-discharge	Combination	12	85.7	2	50
	Breastmilk	2	14.3	2	50
Mode of birth	Vaginal	4	28.6	1	25
	Caesarean	10	71.4	3	75
NEC	Yes	0	0	0	0
	No	14	100	4	100
Sepsis	Yes	2	14.3	0	0
	No	12	85.7	4	100
Antenatal antibiotics	Yes	8	57.1	0	0
	No	6	42.9	4	100
Neonatal antibiotics	Yes	14	100	2	50
	No	0	0	2	50
Chorioamnionitis	Yes	2	14.3	0	0
	No	12	85.7	4	100
Preeclampsia	Yes	2	14.3	0	0
	No	12	85.7	4	100
Maternal Diabetes	Yes	1	7.1	0	0
	No	13	92.9	4	100

Continuous Variables

Variable	**Probiotic-supplemented**		**Non-supplemented**	
	Median	**IQR**	**Median**	**IQR**
Gestational age at birth	27.7	2.7	35.6	1.8

Table 5.1 Demographic/clinical data of study population used for both the 16S metabarcoding and shotgun metagenomics.

16S rRNA short amplicon sequencing

The Bioline ISOLATE Fecal DNA Kit was used for DNA extraction [331]. Modifications were made in consultation with the manufacturer to optimise DNA yield, and included increased beta-mercaptoethanol from 0.5 to 1% (increasing DNA solubility and reducing secondary structure formation), addition of an extra wash step (improving purity) and decreased elution buffer volume from 100µl to 50µl (increasing final DNA concentration). After consultation with the manufacturer, 150 µl was chosen for the initial sample volume, in place of the usual 150 µg required by the kit, for compatibility with the OMNIgene® GUT kit. The Illumina metagenomics library preparation protocol was used for library preparation [332], using the Index Kit v2 C [333] and Platinum™ SuperFi™ PCR Master Mix [334]. Sequencing was performed on the Illumina MiSeq system using the MiSeq Reagent Kit V3 [333], targeting the V3 and V4 regions with the S-D-Bact-0431-b-S-17/S-D-Bact-0785-a-A-21primer combination [332]. Both the pre-analytical bioinformatics and statistical analyses were conducted in *R Studio* Version 3.6.1 [335] with a pipeline adapted from my previous work [380], which can be found in the supplementary material. *DADA2* [337] was used for quality filtering and trimming, demultiplexing, denoising and taxonomic assignment (SILVA Database). In addition, *microDecon* [285] was used to remove homogenous contamination from samples using extraction blanks.

Admission and discharge samples for longitudinal analyses

Data for a subset of individuals that had samples collected at both admission and just prior to discharge were obtained from previous work. As sample recruitment occurred at the same hospital, a small subset (n = 6) had samples collected at these previous time points, allowing us to make comparisons across these three time points, within the probiotic supplemented group. However, it should be noted, that for one infant we did not receive an admission sample. The recruitment, collection and sequencing protocols are as previously described [380].

Shotgun metagenomics

The shotgun metagenomics was performed by Microba Life Sciences [362]. Once samples had DNA extracted for 16S rRNA gene amplicon sequencing, the samples were then again stored

at -80°C, and soon after, shipped to Microba on dry ice. Sequencing was conducted on the Illumina NovaSeq6000 system with 300 bp, paired-end reads. This workflow was completed using Microba's patented Metagenomics Analysis Platform (MAP), which includes the Microba Genome Database, the Microba Community Profiler, and the Microba Gene and Pathway Profiler [362]. The MPA produces taxonomic and functional profiles.

Statistical analysis

For both the 16S rRNA short amplicon sequencing and shotgun metagenomics, beta diversity, alpha diversity, and taxonomic abundance were assessed using mixed effects models. For beta diversity comparisons, I performed both PERMANOVA and *EnvFit* analyses from the *Vegan* package [367], which compare the differences in the centroids relative to total variation. Both analyses were applied to Bray-Curtis dissimilarity matrices [339] based on data normalised through Total Sum Scaling (TSS) [297]. The significance was based on 10,000 permutations and was transformed based on the Benjamini-Hochberg (BH) procedure [340].

For alpha diversity comparisons, I performed generalised linear mixed effects models. The generalised linear mixed effects regression models were created using the package *lme4* [341]. Shannon diversity was calculated at the ASV level, and continuous predictors were scaled and centred. Multicollinearity was assessed using the *AED* package [342], and significance was using an analysis of deviance (Type II Wald Chi-square test) from the *car* package [343]. This was followed by subsequent post-hoc pairwise Tukey comparisons, to correct for multiple comparisons, using the *emmeans* package [344].

DESeq2 [294], which uses a negative binomial generalized linear model and variance stabilising transformation, was used for comparing taxonomic abundances between groups. For the 16S rRNA short amplicon sequencing, taxa were agglomerated at the genus level, due to the limited taxonomic depth of 16S-target technologies. A Wald Test with the BH multiple inference correction was performed to obtain taxa that were significantly differentially abundant. The pre-analytical bioinformatics and statistical analyses can be found in the GitHub link in the supplementary material.

Results

Changes in the gut microbiome of probiotic-supplemented infants over time

The data from the 16S amplicon sequencing revealed that the microbial composition of the microbiome of the probiotic-supplemented infants changed dramatically over time, with inter-individual variation reducing and stabilising at discharge (*Figure 5.6A*). Samples clustered significantly by the sampling time based on their taxonomic composition (*Figure 5.6A, $p < 0.01$*), coupled with a significant increase in alpha diversity post-discharge (*Figure 5.6B, admission and post-discharge: $p < 0.0001$, and discharge and post-discharge: $p < 0.0001$*), as taxa continued to colonise. The composition is dominated early on by the phylum Firmicutes, followed by Proteobacteria and Actinobacteriota at discharge, and then Bacteroidota and Firmicutes post-discharge (*Figure 5.7A).* At the genus level, *Streptococcus*, a facultative anaerobe, was observed to dominate at admission, followed by *Bifidobacterium* at discharge, and with subsequent maturation to a more diverse ecosystem post-discharge. The changes in *Streptococcus* over time were significant ($p < 0.01$), with it being in significantly greater abundance early on, compared with both discharge ($p < 0.001$) and post-discharge ($p < 0.01$) samples. A similar pattern was seen for *Bifidobacterium*, which was in significantly greater abundance at both admission ($p < 0.001$) and discharge ($p < 0.001$), relative to post-discharge samples.

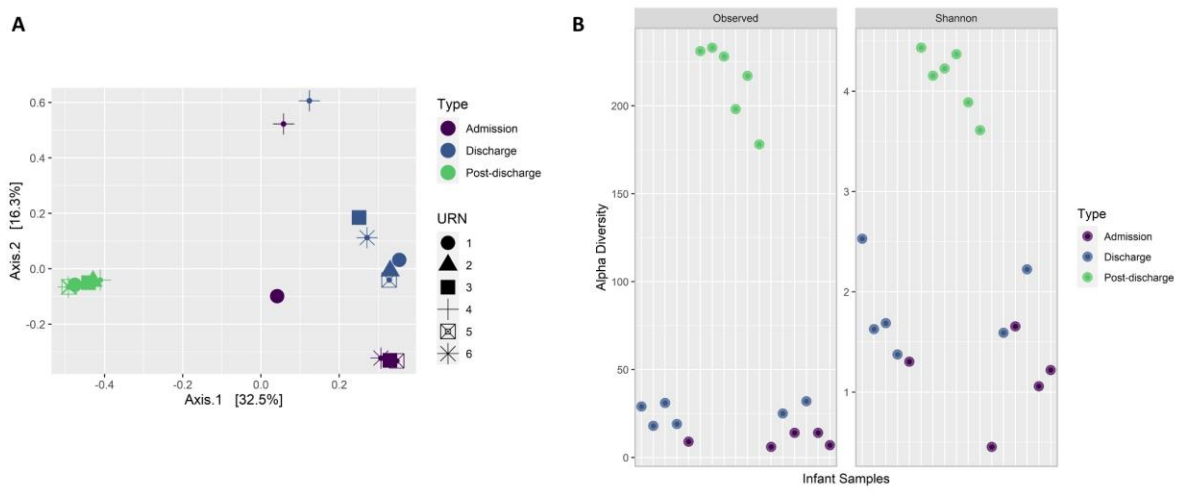


Figure 5.6 A: Principal coordinate analysis plot based on Bray-Curtis distances using ASV level taxonomy obtained through 16S rRNA short amplicons sequencing demonstrating the changes in gut microbial composition for the six infants tracked over time, with significant ($p < 0.01$) clustering of samples. B: dot plot representing the time-based increases in alpha diversity metrics for the same six infants tracked over time and based on transformed ASV

level taxa (16S amplicon sequencing), both observed (richness) and the Shannon Index, where pairwise comparisons found significant differences between admission and discharge samples (p =0.01), admission and post-discharge (p < 0.0001) and discharge and post-discharge (p < 0.0001).

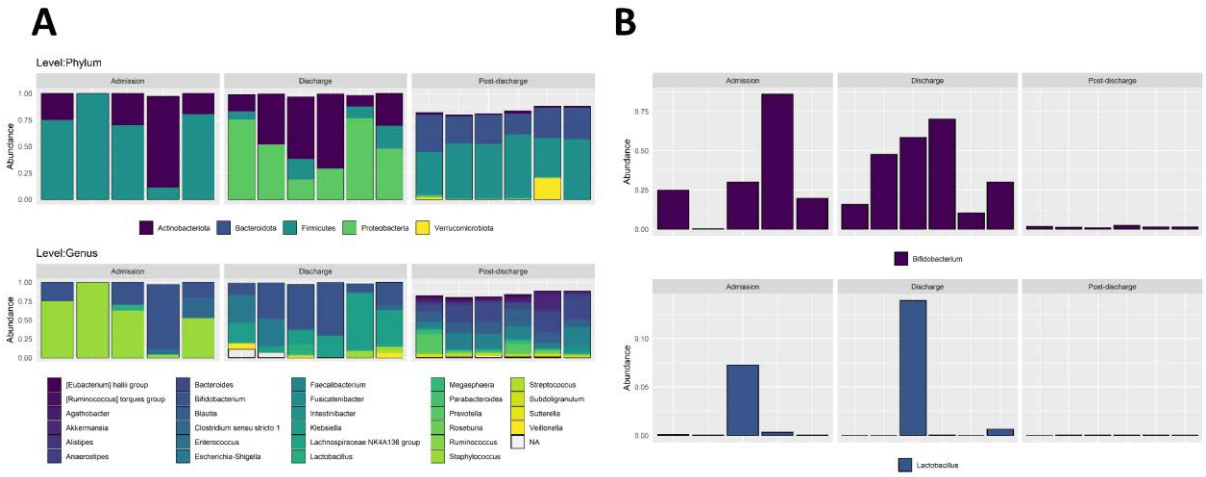


Figure 5.7 A: Changes in the proportions of taxa for the six infants tracked over time at both the phylum and genus levels (16S amplicon sequencing) across admission, just prior to discharge and post discharge, describing the significant (p <0.0001) reduction in Bifidobacterium abundance post-discharge relative to the first two time points. B: Changes in the proportions of both Bifidobacterium and Lactobacillus for the six infants tracked (16S amplicon sequencing) across admission, just prior to discharge, and post discharge, using 16S amplicon sequencing.

Persistence of probiotic species present at discharge up to 2 years of age

Bifidobacterium was present at admission, discharge, and post-discharge, however, *Lactobacillus* was sparse at all time points (*Figure 5.7B*). *Bifidobacterium* was at its greatest abundance at discharge, and thus, towards the end of supplementation. However, despite being present across all infants, there was a significant reduction in its abundance post-discharge. Using shotgun metagenomics, I observed that what remains of the genus post-discharge is mostly other *Bifidobacterium* species, with *B. bifidum* only present in 4/14 infants (*Table 5.2*). The other *Bifidobacterium* species present were *B. adolescentis, B. animalis, B. breve* and *B. longum,* with *B. longum* the most common and *B. breve* having the greatest mean relative abundance. *L. acidophilus*, was also not present post-discharge (*Table 5.2*). The only two species from this genus present post-discharge were *L. paracsei* and *L. rhmanosus*. The lack of long-term colonisation with *L. acidophilus* is consistent with previous work; however, *B. bifidum* has been observed to persist post-discharge at ~58 weeks

[13]. The scarcity of probiotic species present in the study cohort suggests transient colonisation.

Genus	***Species***	***Supplemented Infants (n = 14)***
Bifidobacterium	*bifidum*	*4*
	adolescentis	*5*
	animalis	*9*
	breve	*5*
	longum	*11*
Lactobacillus	*acidophilus*	*0*
	paracsei	*2*
	rhmanosus	*7*

Table 5.2 The number of infants that had species belonging to Bifidobacterium or Lactobacillus present post-discharge, determined through shotgun metagenomic sequencing.

Comparison of probiotic-supplemented infants and non-supplemented infants

Previously identified positive modulation of the gut microbiome associated with probiotic prophylaxis during hospital admission does not persist at 18 months to 2 years post-supplementation, however several other associations were observed. Overall community composition did not differ significantly between those who received probiotic prophylaxis and those who did not (Figure 3.A, *PERMANOVA: $p = 0.4$, envfit: $p = 0.88$*). However, differences in several taxa were oberved (Figure 4). Specifically, we observed greater abundance of *Clostridium_M sp001517625* ($p < 0.01$) and *Flavinofractor plauti* ($p < 0.01$), in combination with lower abundances of *Alistipes finegoldi* ($p < 0.01$), in those that received probiotic prophylaxis. *Clostridium_M sp001517625* (11/14 infants) and *Flavonifractor plauti* (13/14 infants) were only observed in the probiotic group. In contrast, *Alistipes finegoldi* was only found in half of those supplemented, but all of those who did not receive probiotics. Lastly, counter to what was expected, alpha diversity, both richness ($p < 0.05$) and the Shannon Index ($p < 0.05$), were significantly lower in those infants supplemented with probiotics (Figure 3.B). However, it is unclear whether this associated modulation is a result of probiotic prophylaxis or evidence of an inability of probiotics to exhibit lasting modulation beyond the supplementation period.

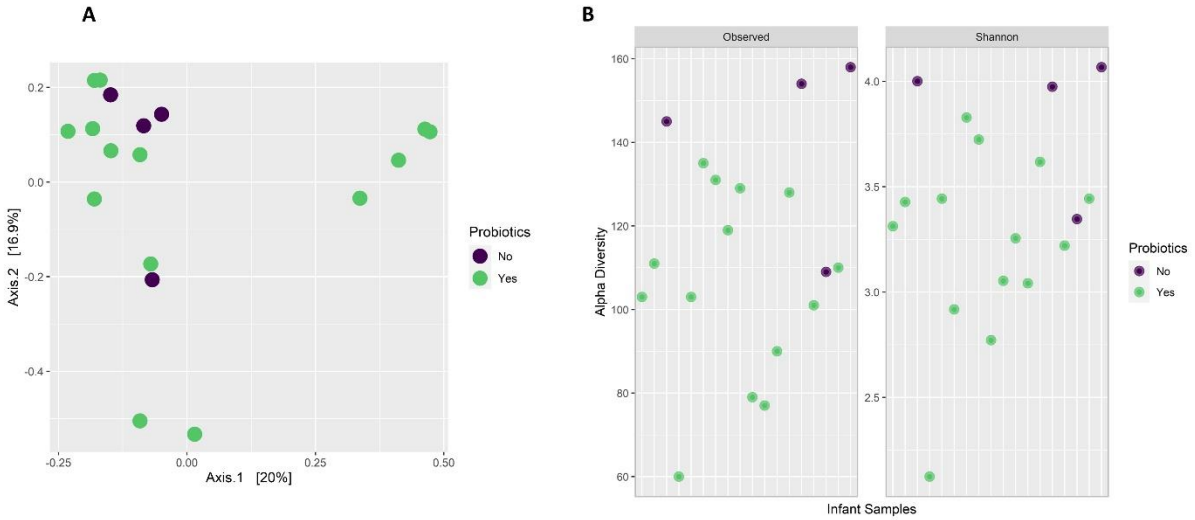
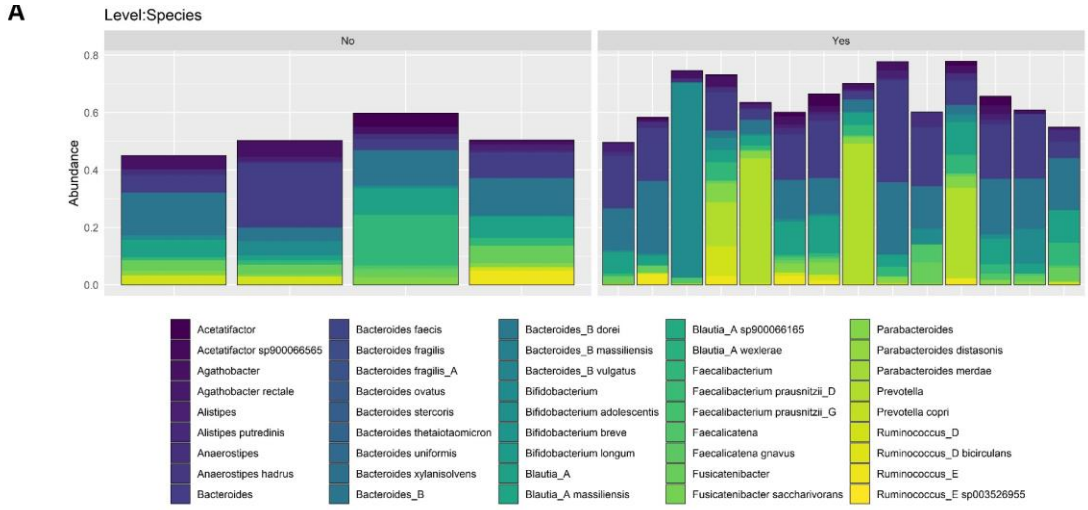


Figure 5.8A: Principal coordinate analysis plot based on Bray-Curtis distances exploring the clustering of samples post-discharge by probiotic-supplementation (coloured) using species level taxonomy obtained through shotgun metagenomics. B:Dot plots describing the significant difference in alpha diversity metrics post-discharge, both observed (richness) ($p < 0.05$) and the Shannon Index ($p < 0.05$), between probiotic supplementation groups and obtained through shotgun metagenomics.



B

Variable: base level	Lfc	Lfc SE	P-adj	Species
Probiotics: Yes	29.09	2.59	< 0.01	Clostridium_M sp001517625
Probiotics: Yes	18.66	1.95	< 0.01	Flavonifractor plautii
Probiotics: Yes	-16.02	2.53	< 0.01	Alistipes finegoldii

Figure 5.9 A: Bar plots comparing the relative distribution of the top 30 most abundant species identified through shotgun metagenomics, and across individuals and between

probiotic-supplementation status. B: Results of Wald-test on the probiotic-supplementation comparison from DESeq2 mixed effects modelling, that also accounted for diet, on species level taxonomy obtained through shotgun metagenomics.

Discussion

We have previously described the short-term positive modulatory impact that probiotic supplementation with Infloran® can have on the developing gut microbiome of preterm infants during their NICU admission [395], supporting an increased diversity and colonisation with beneficial taxa, such as *Bifidobacterium*. This study set out to investigate whether colonisation by these probiotic species affected the development of a healthy microbiome at 18 months to 2 years post discharge. Significant changes were observed in the microbiome of supplemented infants over time, with increases in alpha diversity and dynamic changes in taxonomic abundance, culminating in a reduction in heterogeneity between samples and stabilisation of the microbiome. However, these dynamic changes were coupled with a significant reduction in probiotic species. Several studies report persistence of the probiotic species in the faeces of preterm infants up to the time of hospital discharge [396-399], yet evidence for long-term colonisation with probiotic species is limited. The inability of probiotic species to colonise the infant gut may mean that probiotic associated modulation is short lived. However, the reported benefits of probiotic prophylaxis may extend beyond colonisation and include competitive pathogen exclusion, changes to intestinal barrier function and immune modulation [389]. These positive effects on the developing microbiome may prove important for pre-term infants.

Changes in probiotic-supplemented infants over time

Despite high levels of heterogeneity between individuals early in life, the infant microbiome generally follows a standardised colonisation process. This includes an increase in strict anaerobes across time, particularly Bifidobacterium [400]. However, in this work we observed a significant reduction in this genus after it rose to dominance at discharge. This drop in *Bifidobacterium*, as well as the lack of persistence of *B. bifidum*, may be the result of confounding factors. Delayed colonisation or reduced counts of *Bifidobacterium* have previously been linked to caesarean section [400], and greater colonisation to breastfeeding [380]. The link to breastfeeding largely stems from the presence of human milk oligosaccharides in

breastmilk, which selectively nourish specific microbes [401,402]. Without nourishment, microbes like *B. bifidum*, may not persist. This may also explain why our work does not align with that of Abdulkadir et. al, who observed persistence of *B. bifidum* following supplementation with Infloran® and post discharge [13]. An important distinction between this study and theirs, is that their entire cohort was breastfed, contrasting with only four infants in this study. Thus, diet may be an important factor for sustaining colonisation, and a modifiable factor that could encourage the long-term persistence of probiotic species in preterm infants.

L. acidophilus was not present post-discharge. Previous work has also struggled to isolate both the species and genus in supplemented infants [12,13,376]. However, unlike *B. bifidum*, this may not be a result of diet. *Yousuf et. al.* suggested that the limited presence of *Lactobacillus* was because it is a coloniser of the small intestine, and thus, less likely to be found in faecal samples [376]. In addition, we observed that the genus does not consistently establish itself in the gut during admission/prophylaxis, which is also supported by my previous work [395]. I previously suggested that this could be the result of poor probiotic integrity. However, more work still needs to be done to provide conclusive evidence. Taken together, it is likely that the persistence of probiotic species and even bacterial community succession over the long term is determined by multiple environmental factors.

Comparison of probiotic-supplemented infants and non-supplemented infants

Positive probiotic-associated modulation may not persist beyond discharge. Previous work suggests probiotic prophylaxis contributes to acute increases in bacterial diversity and abundance of known commensals, as well as a reduction in potential pathogens [12,13]. However, this modulation appears to be temporary and may result from the limited persistence of probiotic species. Durack et. al. observed that probiotics can temporarily correct for the delayed diversification associated with preterm infants, but that the inability of the probiotic species to engraft meant that these benefits are lost when probiotic prophylaxis is complete [403]. However, the cessation of probiotic prophylaxis may explain why we don't see previously observed modulation persist, but it does not explain the differences that were observed.

Probiotic supplementation was associated with both lower alpha diversity and a greater abundance of three taxa. However, despite the association between probiotics and lower

diversity, it is unclear if probiotics are the driver of this low diversity. This is especially true when one considers the lack of probiotic persistence. These results may suggest that probiotics cannot correct for the lower diversity common to the most premature of infants. However, if probiotics are the causative factor, the drop in diversity may be a result of the restructuring of the microbial ecology, where the probiotic supports growth of a few specific taxa. Either way, whether this lower diversity will have significant consequences for these infants is not known, and beyond providing stability, greater diversity may also have limited benefits. The primary benefit of probiotics may come through support of key taxa that possess invaluable functionality [22,404]. However, due to their implication in both positive and negative health outcomes, the taxa associated with supplementation in this study may not provide any significant benefit [404-408].

Despite the apparent limited long-term benefits in microbial modulation, the acute modulation observed previously during supplementation may provide lasting benefits. Microbial perturbations, including lower diversity, have been consistently associated with disease. This includes obesity [409], metabolic syndrome [410], Crohn's disease and ulcerative colitis [411-413], multiple sclerosis [414], and more. However, equally important to note is the effect on the development of both innate and adaptive immune function [415-417], as perturbations in the gut microbiome have also been shown to have long-lasting metabolic and immunological dysregulation [418,419], and are significantly associated with a atopy and asthma development in childhood [403,420,421]. Thus, the early-life gut microbiome, colonising a relatively sterile habitat, influencing the developing ecosystem, and in turn, immune and physiological conditions, may have the greatest impact on long-term health.

Limitations

This work has limited statistical power and was unable to account for all known microbial covariates due to its small sample size. As stated in the methods, the recruitment and collection protocol involved contacting parents/guardians at home and relying on their involvement for the collection and postage of stool samples. This proved too much of a burden for a demographic of people who have limited incentive to be involved in the project and are dealing with the stresses of being a new parent. We recommend that future studies take this into consideration during study design, and either have greater involvement in the collection process or target a larger group to ensure adequate sample size.

Conclusion

Probiotic supplementation in very preterm infants lowers the risk of several important neonatal outcomes, including NEC, LoS and all-cause mortality. However, this study suggests that the probiotic species from Infloran® may not persist beyond discharge. The implications of this are unclear. While probiotic-supplemented infants showed a healthier microbiome at discharge compared to other infants who did not receive probiotic supplementation, probiotic-supplemented infants had lower diversity in their gut microbiome at 18 months to 2 years of age. The small sample size reduces the certainty of this result. Nonetheless, with the emergence of a significant body of literature implicating the early gut microbiome in immune system development it is unclear if lower diversity at this age would have significant implications.

6. General Discussion

Preterm birth has been associated with the development of short and long-term complications, caused in part, through the disruption of normal gut microbiome development. Probiotic supplementation may correct for these microbial imbalances and mitigate disease risk. As a result, probiotic-supplementation has become part of standard care for all infants born into the Townsville University Hospital (TUH) at < 32 weeks gestation and/or < 1500g. Using a combination of 16S rRNA metabarcoding and shotgun metagenomics, this thesis explored the gut bacterial microbiome of this unique cohort, addressing the five aims outlined in the introduction. This includes characterising populations at admission and discharge, assessing the impact of potential covariates, and exploring differences between non-supplemented and supplemented preterm infants during admission and post-discharge. In brief, this thesis identified novel associations with the gut microbiome of a unique cohort of infants, that non-supplemented infants from the same hospital have significantly different microbial profiles, and that these differences do not persist 2 years following discharge. This chapter builds on previously discussed results, giving a more comprehensive synthesis of the data, and discussing translatability and direction for future research.

The gut microbiome of probiotic-supplemented very preterm infants from the TUH

The gut microbiome of probiotic-supplemented very preterm infants from TUH NICU follows similar patterns of colonisation to what is described in the literature but shares only a selection of previously described covariates. The significant differences between admission and discharge samples, highlighted by enrichment of *Staphylococcus* at admission and increases in *Clostridium sensu stricto 1*, *Lactobacillus*, *Enterobacter* and *Veillonella* at discharge, supports the previously described progression of oxygen-tolerant microbes that aid in a shift towards an anaerobic environment [60,345]. Additionally, although not significant, the presence of *Bifidobacterium* across 99 samples (53 of 99 were at discharge) in such a young cohort is also noteworthy, as very preterm infants can experience delayed or limited colonisation with *Bifidobacterium* and *Lactobacillus* [37,43,78,79]. The significant increase in *Lactobacillus* coupled with a high frequency of colonisation with *Bifidobacterium* in such a premature cohort suggests that prophylaxis with the probiotic Infloran® (containing

Lactobacillus and *Bifidobacterium*) is having a significant impact on the gut microbiome of this cohort during admission.

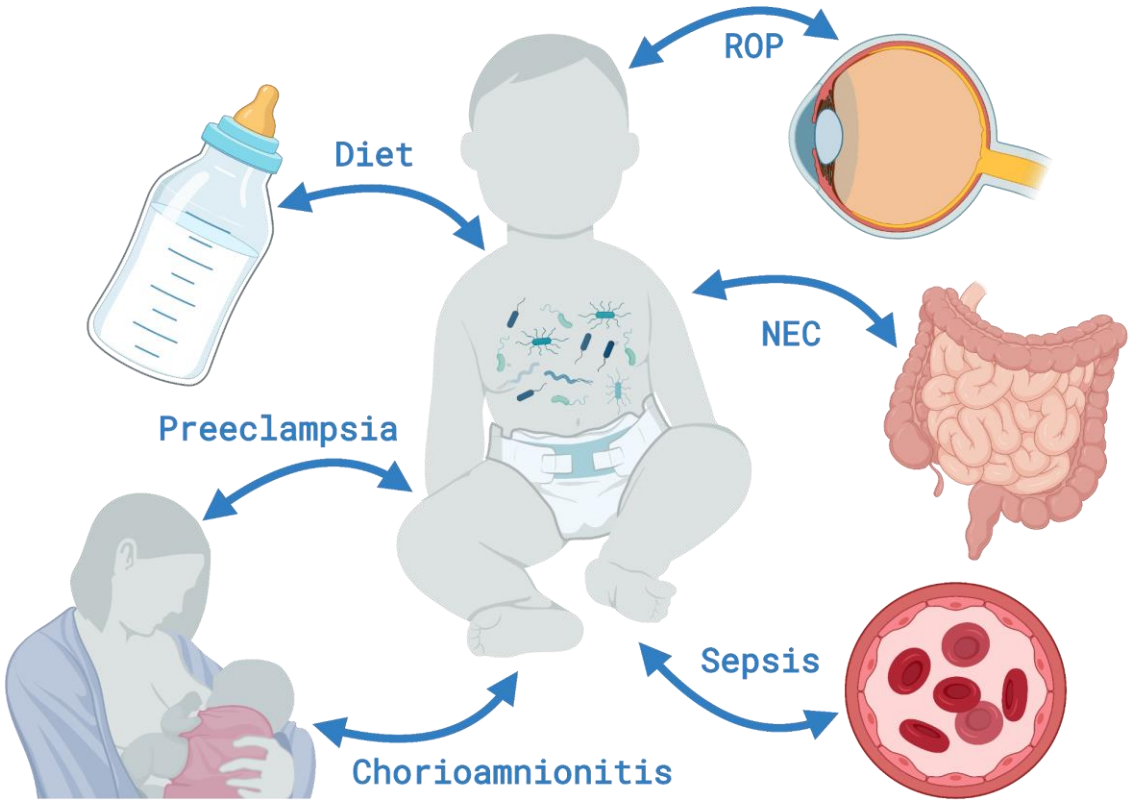


Figure 6.1 Covariates of the probiotic-supplemented very preterm infant gut microbiome, as determined by mixed effects exploring taxonomic abundance and alpha diversity. Annotation for necrotising enterocolitis; NEC: retinopathy of prematurity; ROP. Figure 1 created using www.biorender.com.

The lower bacterial diversity seen in infants diagnosed with unfavourable outcomes, as well as the differing abundances of several taxa across multiple variables reinforces the role of the gut microbiome in disease, the impact of maternal health on the microbiome, and supports the need for promoting healthy microbiome development (*Figure 6.1*). The association between breastfeeding and higher abundance of *Bifidobacterium* observed in this thesis and previous work[37,57], reinforces the importance of infant diet on health. As described in Chapter 4, *Bifidobacterium* species possess not only the ability to digest the HMOs in breastmilk [52,348,349], but also important immunological regulatory capacity [422-426]. This observation lends more credence to the importance of breastfeeding, and in turn, maternal health.

As much of the early microbial inoculation occurs vertically via a maternal route, maternal health has an important role to play in the developing infant microbiome [23]. In this thesis, both preeclampsia and chorioamnionitis had significant associations with taxa. The association between preeclampsia and lower *Escherichia*/*Shigella* at discharge may stem from continued exposure to the mother, who may continue preeclampsia-treatment post-delivery, compounding the passing of irregular microbial patterns over time [352,353]. Additionally, the increased abundances of certain *Staphylococcus* spp. in association with chorioamnionitis may partly explain why exposure to chorioamnionitis increases the risk of adverse neonatal outcomes [107,109,350]. Taken together, these findings support the idea that 'mum matters', and that maternal health plays a key role in shaping the developing infant microbiome. If the relationship between maternal health and the infant microbiome continues to be consistently replicated, considerations can be made around maternal health, and how that may have knock on effects for infant health.

As observed previously, both lower diversity and greater/lower abundances of specific taxa were associated with negative health outcomes [107,109,354]. However, instead of providing clarity the results add further heterogeneity to the literature. My findings of sepsis being associated with a greater abundance of *Bifidobacterium*, NEC not being linked to a greater abundance of any taxa, and the enrichment of *Staphylococcus* in infants diagnosed with ROP, all contrast previous work [97,98,108,110,354]. These contradictory findings could be the result of study design, including the study population or analyses used. However, another explanation discussed in Chapter 3, and highlighted by previous work [96], may be due to the non-uniform microbial structure and potential polymicrobial aetiology of NEC and late onset sepsis. These diseases are likely to be the synthesis of multiple microbial and physiological factors. This is further supported by the observation that NEC has very different gut microbial structures between different NICUs [104], which also differ more broadly in their microbiomes in general. However, despite the complex aetiology of NEC and sepsis, probiotics have been shown to reduce their incidence, and thus, the gut microbiome may still be a future target for interventions in ROP.

It is difficult to deduce how exactly the microbiome is implicated in the examined clinical variables, or vice versa. The gut microbiome is a complex ecological landscape where the presence or absence of a microbe may have significant consequences, or not. Both infant and maternal health have a significant association with the developing infant gut microbiome,

and, although probiotics may aid in microbiome development, correcting for some imbalances, they are not a wonder-drug that can correct for the influence of all covariates. However, the abnormally widespread presence of *Bifidobacterium* in the intestinal microbiome of these probiotic-supplemented infants suggests probiotics may elicit some positive modulatory effects. Although this impact may not correct for all microbial imbalances, the presence of this well-known commensal genus, could aid in colonisation of other commensal taxa and help to shape a more proficient immune system. However, if probiotics are indeed eliciting such a response, perhaps supplementation should be extended to older infants.

Differences in the microbiome of supplemented and non-supplemented preterm infants at discharge

Moderate to late preterm infants may be missing out on acute modulatory effects induced through probiotic prophylaxis. In the cohort from Chapter 4, probiotic supplementation corrected for associations between lower gestational age and low diversity [69,70], fewer commensal microbes [71-73] and greater pathogen abundance [72,74], producing a microbiome that is more similar to that of full-term infants. The targeted approach of probiotic supplementation in very preterm infants, as previously mentioned, is the result of NEC's inverse correlation with gestational age and birth weight [360,361], and the demonstrated efficacy of probiotics to reduce the risk of NEC[10,11]. However, the benefits of probiotics may extend beyond acute prevention of disease, as there is accumulating evidence suggesting that early microbial colonisation is a determinant for adult health [2-6]. This is not the first study to suggest probiotics positively modulate the infant gut [12,13], but this study builds on previous findings and suggests that infants born into the TUH (and other centres that use similar guidelines for probiotic supplementation) who do not receive probiotics are leaving care with lower microbial diversity and fewer key commensal microbes.

Moderate to late preterm infants could benefit from stabilising effects provided through the greater alpha diversity associated with probiotic supplementation [371,372]. The significant association between alpha diversity and probiotics suggest that the probiotic species contribute to the colonisation of more microbes [19]. Presence of a greater diversity of taxa could aid in protection against pathogenic species and contribute to the development of broader immune tolerance. This is because early colonisers of the microbiome play a critical

role in immune-system development, interacting with both epithelial and lymphoid tissue in a state of controlled inflammation [20]. A greater diversity of taxa early in life may contribute to a greater tolerance with age.

A greater abundance of key taxa could have significant benefits for those who receive probiotic prophylaxis. Reduced levels of *Bifidobacterium* in early life have also been linked to NEC, sepsis and atopic disease [98,325,427], including lower levels in 3-month old infants who go on to develop an autoimmune disorder at 2-4 years of age [420]. Some commensal microbes engage in crosstalk with human cells, aiding the development of innate immune defences, pathogen recognition, epithelial turnover, mucous synthesis, peristalsis, and antimicrobial secretions [21,22]. Members of the *Clostridium sensu stricto 1* cluster, such as C. *sutyricum*, also play a role in intestinal homeostasis through the production of metabolites like butyrate and other SCFAs [428]. So much so, that different *Clostridium* species, including C. *sutyricum*, have been suggested as potential probiotic-therapeutics [428]. *Bifidobacterium*, can *induce dendritic cell maturation* [422-424], mobilisation of *immune cells* [425,429] *and* modulation of *immune responses* [426,430]. Promotion of such immunological factors may be why probiotics reduce the incidence of NEC and sepsis [10,11]. The extent to which probiotic-induced modulation can mitigate other diseases in infants is understudied. However, because of the previously mentioned functional benefits, reduced levels of *Bifidobacterium*, and the other six taxa that are lower in non-supplemented infants, may have significant consequences for their immunity.

Despite probiotics appearing to drive significant changes in the gut microbiome, this work also highlighted concerns in probiotic integrity. Lower levels of *Lactobacillus*, and more specifically *L. acidophilus*, have also been observed previously [12,13,376]. Others have previously suggested that this is due to *Lactobacillus* being a coloniser of the small intestine [376]. However, sequencing of the probiotic itself, as discussed in Chapter 4, highlighted quality (composition) concerns, which are not uncommon in probiotics [114,124]. Unfortunately, the inclusion of the probiotic in the sequencing was done without careful consideration, and I should have included multiple probiotic samples to determine if this observation was consistent. The integrity of Infloran® should be pursued in future work.

Although questions remain regarding the consequences of the differences observed between groups, it is clear that even when accounting for known covariates, probiotics have a significant associated with the gut microbiome of preterm infants at discharge. As probiotic

supplementation is facilitating or instigating these changes, transforming the microbiome to one that mirrors that of full-term infants, then further consideration should be given to expansion of the probiotic supplementation criteria. As previously mentioned, the lack of research exploring the effect of probiotics in moderate to late preterm infants are an unfortunate consequence of a targeted approach to combating acute life-threatening infections in the most premature, however, if the benefits of probiotics go beyond prevention of acute disease more infants may benefit.

Differences between supplemented and non-supplemented infants at discharge do not persist

The previously mentioned changes in the gut microbiome associated with probiotic supplementation appear to be transient. The limited differences in the microbiomes of supplemented and non-supplemented infants post-discharge likely stems from an inability of the probiotic species to persist. As discussed in Chapter 5, the inability of *B. bifidum* to persist may be the result of other confounding factors, such as diet. Again, highlighting the importance of breastfeeding. The limited presence of *L. acidophilus* was no surprise, given what was observed at discharge and what previous work has described [12,13,376]. However, more work needs to be done to deduce the reason; whether the cause is low probiotic integrity or an inability of stool sampling to capture microbial populations in the upper gastrointestinal tract. Both outcomes would be concerning.

Despite limited persistence of the probiotic species, supplemented infants had lower alpha diversity in combination with a greater abundance of three taxa. With such a small sample size it is unclear if these differences are driven by the probiotic. However, a negative relationship between diversity and probiotic supplementation during prophylaxis has been observed previously [12]. Supplementation with probiotics may drive the persistence and growth of a few key taxa, decreasing diversity over time. This calls into question the use of greater alpha diversity as a metric for greater health, an error already highlighted by the effect of diet on the gut microbiome [380]. Unfortunately, the differences in taxonomic abundance associated with probiotic supplementation do not clearly support this hypothesis as there is conflicting evidence surrounding their benefits. Thus, it is unclear if probiotics have a positive long-term effect on the gut microbiome, but this does not mean that probiotic supplementation during admission does not have other long-lasting benefits.

Moderate to late preterm infants may still be afforded benefits from probiotic prophylaxis during hospital admission, despite changes in the microbiome not persisting. The aim of this chapter was built on the assumption that the long-term health benefits of probiotic supplementation would be provided through long-term modulation of the gut microbiome. However, moderate to late preterm infants may still benefit from acute disease mitigation, and probiotic species may only need to pass through transiently to exert long-term, positive health benefits. Correlative data suggests that infants colonised early in life with *Bifidobacterium* species are less likely to develop immune-mediated diseases later in life [48,431-433]. This relationship likely stems from the involvement of key taxa, like *Bifidobacterium*, in early-life immune system development [20-22,48,375]. Thus, infants who undergo probiotic-prophylaxis in early life may still have long-lasting immunological benefits, despite the probiotic taxa not persisting beyond the supplementation period. So, although the final chapter of my thesis suggests that the probiotic species do not persist beyond supplementation, moderate to late preterm infants may derive long-term health benefits from the transient presence of the probiotic strains during hospital admission.

Limitations

This thesis has several limitations that have proven to be informative learning experiences. The most notable include limitations associated with 16S rRNA gene amplicon sequencing, and its associated protocols, those that have arisen due to difficulties encountered during recruitment, and regarding the search methods/terms used in Chapter 2. Firstly, the power of 16S rRNA gene amplicon sequencing is limited to exploring high-level, group-based differences or associations. Due to its inability to provide species-level taxonomic depth and gene-associated functionality, it is difficult to extract physiological or clinical meaning from these high-level, group-based differences. In addition, the 16S method itself is biased and plagued with protocol heterogeneity, which can translate into greater heterogeneity between studies [123,434]. This limits the strength of inferences made when comparing the work in this thesis to that previously carried out. Secondly, in Chapter 5, although many guardians agreed to partake in the study, the number of samples that were returned was minimal, resulting in a limited sample size. Although I attempted to make the process as easy as possible, by providing everything needed for collection and postage, including simple instructions and a paid-return parcel, carrying out the collection and/or postage proved too much of a burden for

new parents or guardians. Lastly, the limited search terms used in Chapter 2 may not have been expansive or broad enough to capture the entire target literature under review. However, as previously mentioned, these limitations have proven to be good learning experiences.

These limitations will serve as a guide for future decision making in the context of my own study design. The limitation of using 16S rRNA gene amplicon sequencing could be solved with a switch to shotgun metagenomics, which as described in Chapter 2, can provide more detailed insights into the gut microbiome. In addition, using shotgun metagenomics would remove biases associated with primer selection. To address the sample size issue, I believe that greater involvement of the research team in the sample collection process would solve this problem. Restricting the recruitment to Townsville, and then going out and collecting the samples myself would have removed the significant reliance of the collection protocol on the parents or guardians of these infants. However, both these solutions are cost prohibitive, and provide their own logistical issues.

Future research and translation

The human microbiome is a rapidly evolving field, and this thesis is but a small contribution. However, I do believe the work here suggests that a randomised-controlled trial exploring the expansion of the inclusion criteria for probiotic supplementation to include moderate to late preterm infants is warranted. With the work discussed in this thesis, our research team now has a platform to build off and a track record to leverage for funding. In addition, such future work could be expanded to include other probiotic strains. With the reduction in costs of the rapidly evolving genomics technologies, more insight into the benefits of other strains may be uncovered, and the clinical setting should be ready to adapt accordingly. More broadly, as the cost of sequencing continues to drop, future work should harness the power of technologies to decipher the relationship, and underlying mechanisms, that exists between significant covariates and the microbiome, between the microbes and their host, and between the different species within the microbiome. Better understanding this complex ecological relationship may benefit the production of probiotic therapeutics and improve intervention strategies.

Conclusions

The overall aim of this thesis was to describe the faecal microbiome of a cohort of preterm infants born in North QLD, Australia, using a combination of metagenomic sequencing technologies and explore its relationship to a host of variables. Using mixed effects models, the results support previous work demonstrating a significant association between lower alpha diversity and unfavourable covariates, and suggests that probiotic taxa can colonise, persist, and positively modulate the gut microbiome of very preterm infants. These findings add further evidence to support the idea that 'mum matters', and that maternal health and infant microbiome development are crucially intertwined. Additionally, these findings suggest that those born moderate to late preterm could benefit from probiotic supplementation, despite limited long-term modulation in the gut microbiome, and supports the argument for expanding the current probiotic criteria.

The significant relationship between maternal and infant microbiome development was best depicted by the association between *Escherichia*/*Shigella* and preeclampsia at discharge. I hypothesize a possible reason that a difference was only observed at discharge may be due to the continued treatment for preeclampsia through anti-hypertensive drugs. The implications of this are twofold: firstly, this suggests that the mother continues to matter after the first major colonising event (birth), and secondly, that maternal treatments may have a negative impact on the developing infant microbiome. Whether this would be direct, through passing of the drug itself, or indirect, through passing of a negatively modulated maternal microbiome, is unclear. However, either way, if maternal interventions are having a significant impact on the infant microbiome, then consideration may need to be taken when prescribing treatment regimens to new mothers.

The relationship between enrichment of *Staphylococcus* and lower diversity for infants diagnosed with ROP was also an important observation. Previously explored in a single study, the sample size in Chapter 3 adds some much-needed power to the relationship between ROP and the preterm infant gut microbiome, providing more concrete evidence of this association. This work, and that of Skondra et al., not only highlights the role of the microbiome in development but provides a potential target for intervention in a disease that is particularly common in those born preterm. As mentioned in Chapter 3, if a role or mechanism can be identified in future work, the gut microbiome could become a target for intervention for this highly prevalent developmental disorder.

Evidence to support the expansion of probiotic supplementation to include all preterm infants is the most significant implication to come from this work. Despite lingering questions and concerns regarding probiotic-quality and long-term impacts, it is clear that probiotics can aid in the acute, positive modulation of the gut microbiome in preterm infants, possibly supporting metabolic and immune system development. This probiotic-induced positive microbial modulation renders a microbiome that is more like that of full-term infants, relative to those born moderate to late preterm, which is likely why clinical trials have shown reductions in NEC resulting from probiotic supplementation. Expanding the criteria for probiotics to include all infants born < 37 weeks could see a reduced health burden in preterm infants, through support of microbiome development, and in turn, metabolic and immune system development. However, more work still needs to be done before such translation can occur.

7. Abbreviations

16S-TRFLP: 16S terminal restriction fragment polymorphism

ABR: antibiotic resistance

ASV: amplicon sequence variant

B: Bifidobacterium

BH: Benjamini-Hochberg

CHHS: Cairns and Hinterland Hospital and Health Service

DGGE: Denaturing gradient gel electrophoresis

dHPLC: denaturing high-performance liquid chromatography

DNA: deoxyribonucleic acid

dNTP: deoxynucleoside triphosphate

E: Escherichia

EC: Enzyme Commission Number

emPCR: emulsion polymerase chain reaction

FISH: fluorescence in situ hybridization

HMO: human milk oligosaccharide

InviMag Kit: InviMag Stool DNA Kit

K: Klebsiella

L: Lactobacillus

LoS: Late-onset sepsis

MetaCyc: metabolic pathway database

MoBio Kit: MoBio Powersoil Bacterial DNA Kit

MPA: Metagenomics Analysis Platform

MPA: Metagenomics Analysis Platform

NEC: necrotising enterocolitis

NICU: neonatal intensive care unit

NMDS: Non-Metric Multidimensional Scaling

NQLD: North Queensland

NS: non-supplemented

NT: non-treated

OTU: operational taxonomic unit

PacBio SMRT: Pac Bio Single Molecule, Real-Time Sequencing

PCA: Principal component analysis

PCoA: Principal coordinate analysis

PCR: polymerase chain reaction

PERMANOVA: permutational analysis of variance

PFGE: pulse-field gel electrophoresis

PowerLyzer Kit: PowerLyzer PowerSoil Kit

PS: probiotic-supplemented

PT: probiotic-treated

QIAamp Kit: QIAamp DNA Stool Kit

QIIME: Quantitative Insights Into Microbial Ecology

qPCR: quantitative polymerase chain reaction

RAPD: random amplified polymorphic DNA

RDP: Ribosomal Database Project

RNA: ribonucleic acid

ROP: retinopathy of prematurity

rRNA: ribosomal ribonucleic acid

SCN: Special care nursery

Shotgun: Shotgun metagenomics

TCDB: Membrane Transport Proteins

TGGE: Temperature gradient gel electrophoresis

TGS: third generations sequencing

THHS: Townsville Hospital and Health Service

TSS: Total Sum Scaling

V: Veillonella

WGS: whole genome sequencing

8. Appendix

Scripts for bioinformatics and analyses

In the interest of both transparency and reproducibility, all scripts and most of the supplementary information for this thesis, and its associated manuscripts, can be found online at:

- Chapter 3: https://github.com/JacobAFW/NICU_Microbiome_Study
- Chapter 4: https://github.com/JacobAFW/SCN_vs_NICU_probiotic_study
- Chapter 5: https://github.com/JacobAFW/Long_term_effects_of_probiotics

Sequencing Data

The raw sequencing data for this thesis can be found at the International Nucleotide Sequence Database Collaboration at the *National Center for Biotechnology Information (NCBI)* repository (Home - BioProject - NCBI (nih.gov)) using the following ProjectID's:

- Chapter 3: PRJNA687291
- Chapter 4: PRJNA751712
- Chapter 5: PRJNA805057

Supplementary material

Spreadsheets and large files that were included as part of the supplementary material alongside publications have not been included in this thesis. However, these large files are available at the links to the publications provided below. The supplementary data included as part of this thesis is specific to outputs produced through data exploration and summarisation, and statistical analyses.

List of Appendix Figures

List of Appendix Tables

Chapter 3

All supplementary data for chapter 3 can be found at the link provided above, or published online (The bacterial gut microbiome of probiotic-treated very-preterm infants: changes from admission to discharge | Pediatric Research (nature.com)) as all significant outputs were included in the manuscript.

Chapter 4

The supplementary data below, for chapter 4, can also be found online in the *Supplementary Material* section (Frontiers | To Probiotic or Not to Probiotic: A Metagenomic Comparison of the Discharge Gut Microbiome of Infants Supplemented With Probiotics in NICU and Those Who Are Not | Pediatrics (frontiersin.org)) .

Categorical Variables

Variables	Levels	Count	%
Probiotic treatment	Yes	63	67.0
	No	31	33.0
Diet	Formula	23	24.5
	Breastmilk	38	40.4

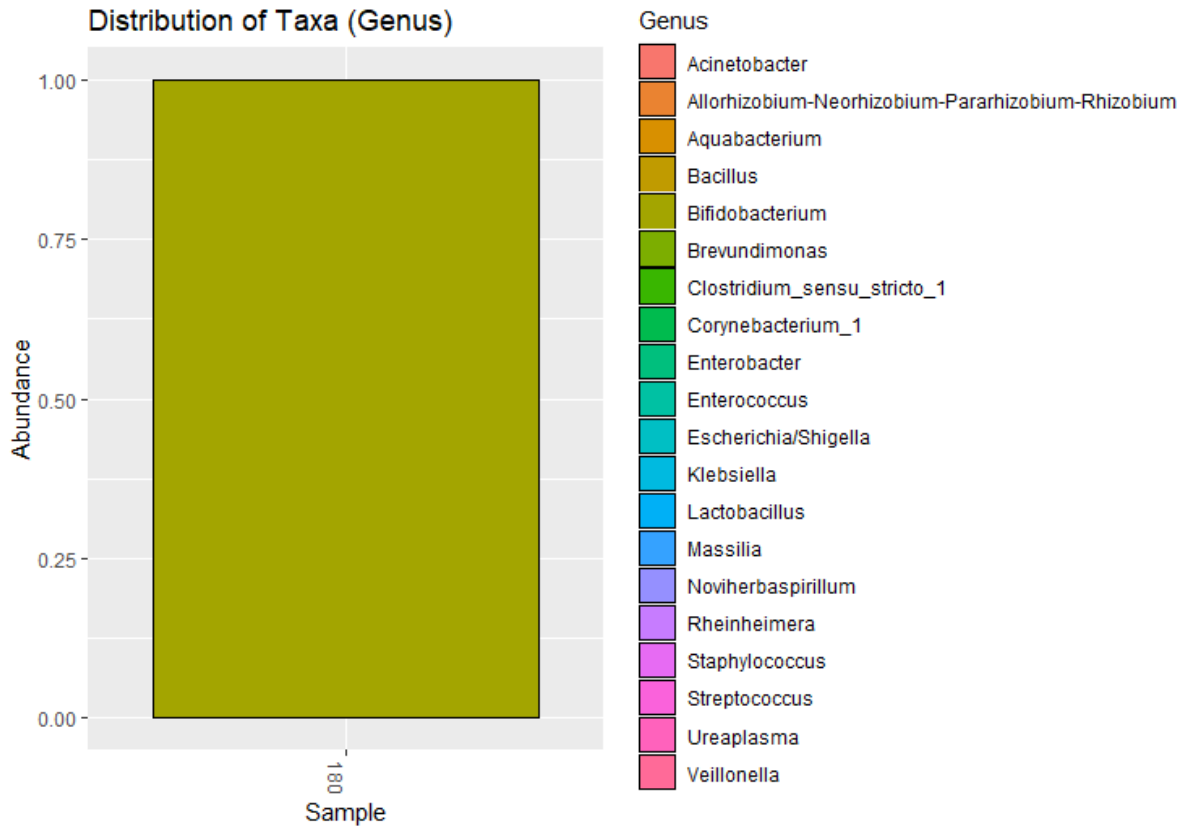
	Formula & Breastmilk	33	35.1
Delivery	Vaginal	32	34.0
	Caesarean	62	66.0
NEC	Yes	5	5.3
	No	89	94.7
Sepsis	Yes	3	3.2
	No	91	96.8
Antenatal antibiotics	Yes	52	55.3
	No	42	44.7
Neonatal antibiotics	Yes	83	88.3
	No	11	11.7
Chorioamnionitis	Yes	28	29.8
	No	66	70.2
Preeclampsia	Yes	13	13.8
	No	81	86.2
Maternal Diabetes	Yes	19	20.2
	No	75	79.8
Continuous Variables			
Variable	**mean/median**		
Gestational age at birth	30.8/30.1 weeks		
Gestational age at collection	36.0/36.0 weeks		

Appendix Table 1 Overview of the demographic data for the preterm-infant cohort that underwent 16 rRNA gene amplicon sequencing.

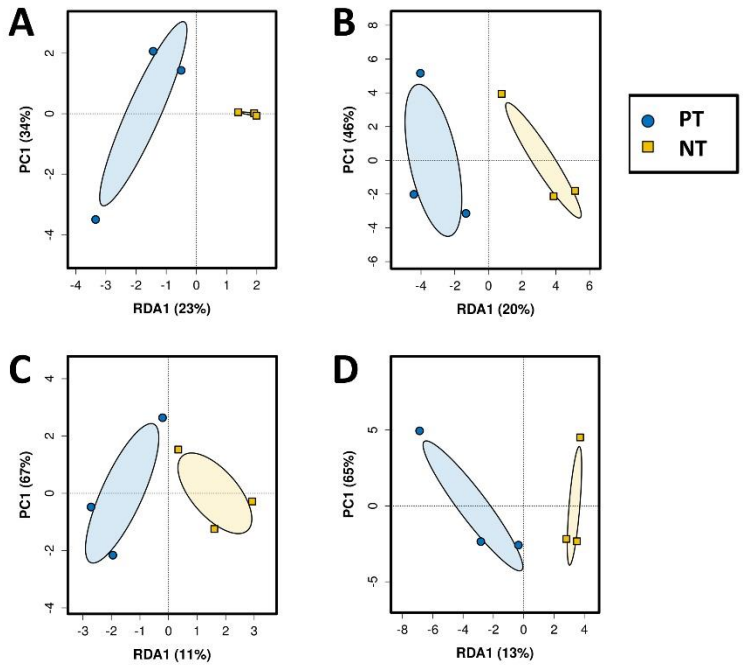
ID	**1**	**2**	**3**	**4**	**5**	**6**
Probiotics	Yes	Yes	Yes	No	No	No

Diet	Formula	Formula	Formula	Breastmilk	Breastmilk & Formula	Breastmilk
NEC	No	Yes	No	No	No	No
Sepsis	No	No	No	No	No	No
Delivery	Caesarean	Caesarean	Caesarean	Caesarean	Vaginal	Caesarean
Antenatal antibiotics	Yes	Yes	No	No	No	No
Neonatal antibiotics	Yes	Yes	Yes	No	No	No
Chorioamnionitis	No	Yes	Yes	No	No	No
Maternal diabetes	No	No	No	No	Yes	No
Preeclampsia	No	No	No	No	No	No
ROP	No	Yes	Yes	No	No	No
Gestational age at birth	26	24	25	34	37	34
Gestational age at collection	31	36	35	35	39	35

Appendix Table 2 Overview of the demographic data for the six preterm-infants who had samples that underwent shotgun metagenomic sequencing.

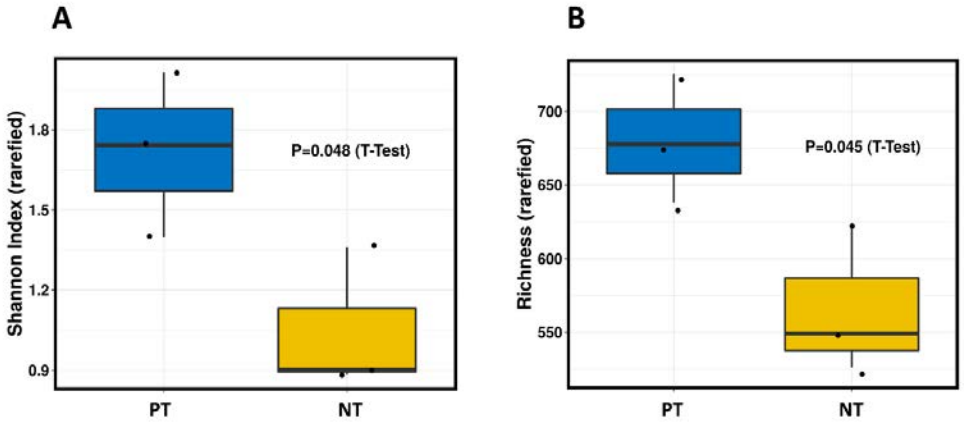


Appendix Figure 1 Bar plot of relative abundance of taxa at the genus level from the probiotic Infloran as determined by 16S rRNA amplicon sequencing.



Appendix Figure 2 Redundancy analysis of shotgun metagenomics outcomes. A: Redundancy analysis of clr transformed species-level taxonomic data, coloured by probiotic-treated and non-treated groups (P = 0.1), B: Redundancy analysis of clr transformed MetaCyc pathways

data coloured by probiotic-treated and non-treated groups (P = 0.6), C: Redundancy analysis of clr transformed MetaCyc groups data coloured by probiotic-treated and non-treated groups (P = 0.7), D: Redundancy analysis of clr transformed EC numbers data coloured by probiotic-treated and non-treated groups (P = 0.5). Annotation for probiotic-treated; PT: non-treated; NT.



Appendix Figure 3 Boxplots comparing diversity at discharge using shotgun metagenomics data. A: Boxplots comparing the Shannon Index based on species level abundances obtained through shotgun metagenomics for probiotic-treated and non-treated infants, B: Boxplots comparing the Richness based on species level abundances obtained through shotgun metagenomics for probiotic-treated and non-treated infants. Annotation for probiotic-treated; PT: non-treated; NT.

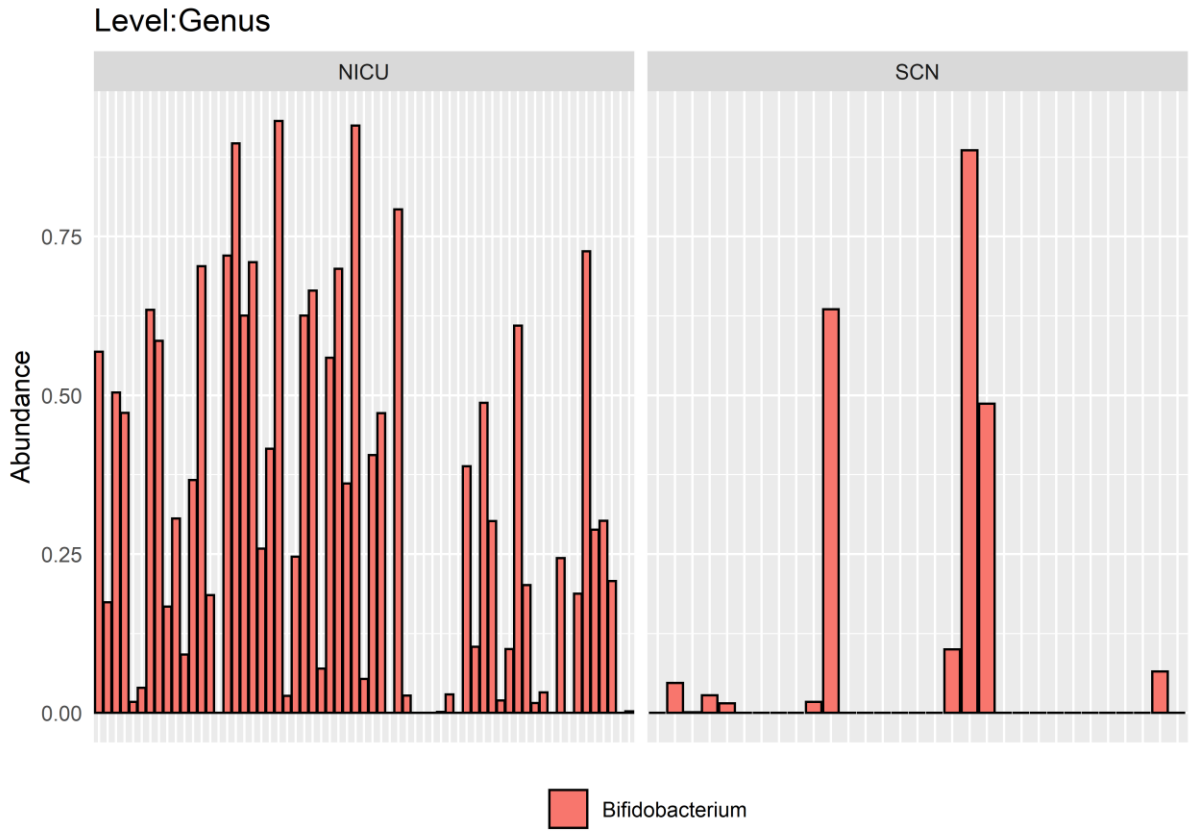
contrast	***estimate***	***SE***	***z.ratio***	***p.value***
Supplemented – Non-supplemented	0.52	0.25	2.12	0.03

Appendix Table 3 Tukey's pairwise comparison from a generalized linear mixed effects modelling on the Shannon Index, based on ASV level abundances obtained through 16S rRNA amplicon sequencing, and comparing probiotic-treated to non-treated infants.

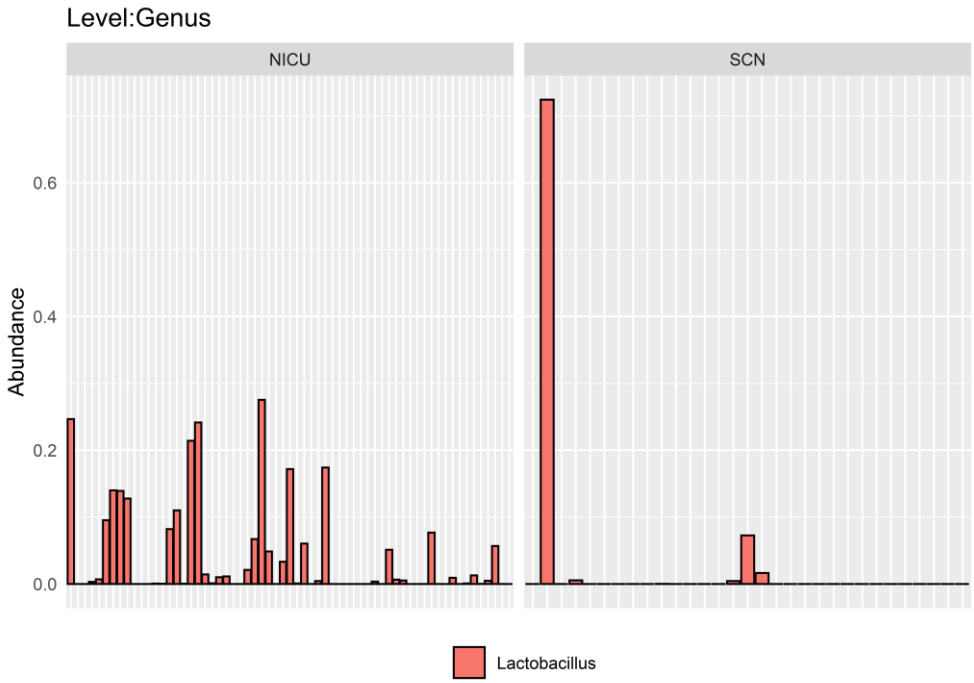
envfit		
Variable	*r2*	*p*
Gestational_Age_at_Birth	0.06	0.66
Probitoic_Treatment	**0.03**	**0.04**

Feeding_Type	0.02	0.90
NEC	< 0.01	0.59
Sepsis	**0.33**	**0.04**
Mode_of_Delivery	0.02	0.53
Neonatal_Antibiotics	< 0.01	0.90
Chorioamnionitis	0.02	0.59
Preeclampsia	< 0.01	0.80
ROP	< 0.01	0.74
Batch	< 0.01	1.00
Diabetes	< 0.01	0.77
Antenatal_Antibiotics	0.01	0.59

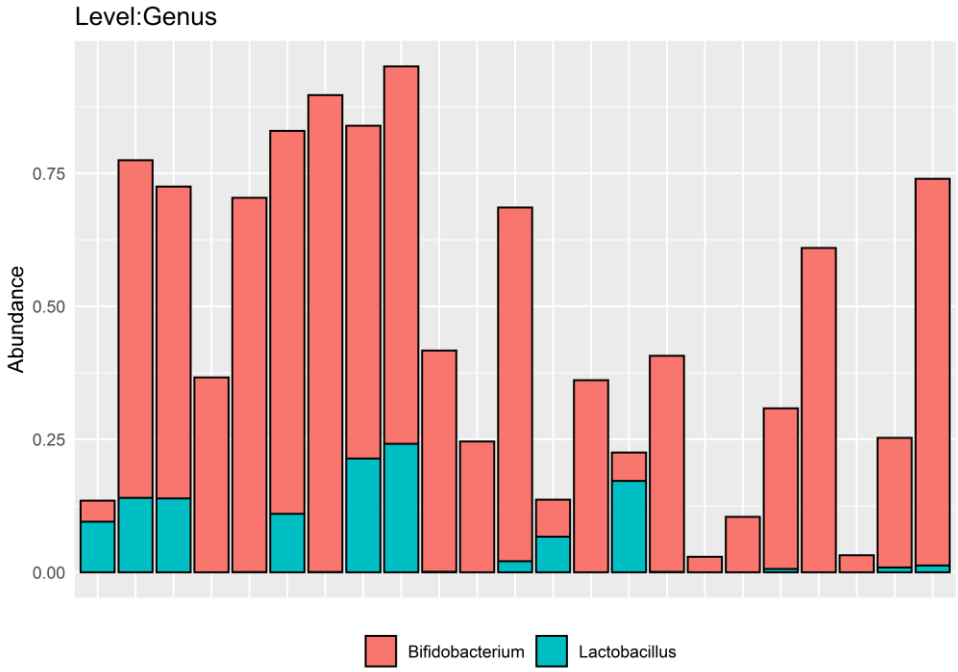
Appendix Table 4 Results (r2 and p value) of an envfit analysis showing the contribution of covariates to microbiome populations obtained through 16S rRNA amplicon sequencing.



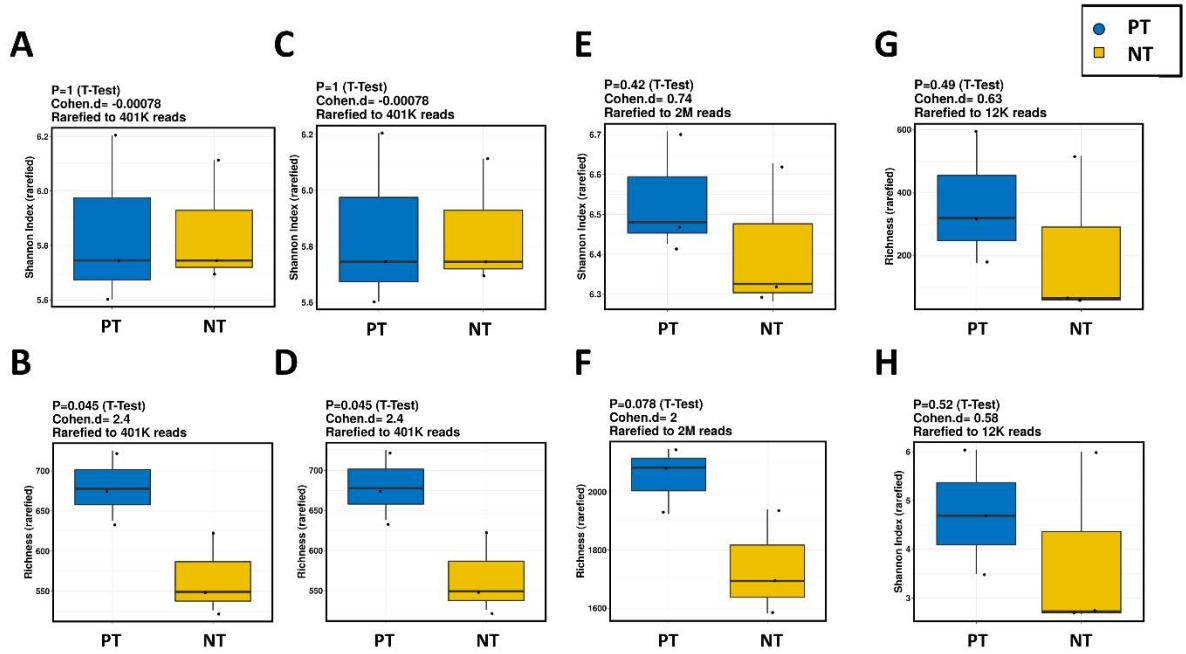
Appendix Figure 4 Bar chart representing the relative abundance of Bifidobacterium obtained through 16S rRNA amplicon sequencing across samples and between treatment groups, where NICU represents those treated with probiotics and SCN those not treated.



Appendix Figure 5 Bar chart representing the relative abundance of Lactobacillus obtained through 16S rRNA amplicon sequencing across samples and between treatment groups, where NICU represents those treated with probiotics and SCN those not treated.



Appendix Figure 6 Bar chart representing the relative abundance of both Bifidobacterium and Lactobacillus measured through 16S rRNA amplicon sequencing across samples collected at > 36 weeks gestation (post probiotic-treatment) in the treatment group.



Appendix Figure 7 Boxplots comparing diversity between probiotic groups based on shotgun metagenomics data. A: Boxplots comparing the Shannon Index for MetaCyc Groups between probiotic-treated and non-treated infants, B: Boxplots comparing the Shannon Index for MetaCyc Groups between probiotic-treated and non-treated infants, C: Boxplots comparing the Shannon Index for MetaCyc Pathways between probiotic-treated and non-treated infants, D: Boxplots comparing the Shannon Index for MetaCyc Pathways between probiotic-treated and non-treated infants, E: Boxplots comparing the Shannon Index for Enzyme Commission between probiotic-treated and non-treated infants, F: Boxplots comparing the Shannon Index for Enzyme Commission between probiotic-treated and non-treated infants, G: Boxplots comparing the Shannon Index for Membrane Transport Proteins between probiotic-treated and non-treated infants, H: Boxplots comparing the Shannon Index for Membrane Transport Proteins between probiotic-treated and non-treated infants; PT: non-treated; NT.

Chapter 5

As chapter 5 is two separate manuscripts, the supplementary material was divided accordingly.

A validation of at-home infant stool sample collection devices for determining the faecal microbiome.

	DF	***Sum of squares***	***R2***	***F***	***Pr(>F)***
Sample Type	*1*	*0.06*	*0.01*	*1.18*	*0.24*

URN	*4*	*7.61*	*0.88*	*38.84*	*< 0.01*
Residual	*20*	*0.98*	*0.11*	*NA*	*NA*
Total	*25*	*8.64*	*1*	*NA*	*NA*

Appendix Table 5 Results of a mixed effects model using PERMANOVA on ASV abundances obtained through 16S rRNA metabarcoding showing that storage method does not contribute to differences between groups/samples.

	R2	***Pr(>r)***
Sample Type	*0.00*	*0.92*
URN	*0.93*	*< 0.001*

Appendix Table 6 Results of a mixed effects model using envfit on ASV abundances obtained through 16S rRNA metabarcoding showing that storage method does not contribute to differences between groups/samples.

Variable	***Chisq***	***Df***	***Pr(>Chisq)***
Sample Type	*1.72*	*1*	*0.19*
URN	*53.23*	*4*	*< 0.0001*

Appendix Table 7 ANOVA results from a generalised linear mixed effects model on Shannon Index diversity, based on ASV abundances obtained through 16S rRNA metabarcoding and showing that storage method has no association with alpha diversity.

Variable	***Chisq***	***Df***	***Pr(>Chisq)***
Sample Type	*0.11*	*1*	*0.74*
URN	*565.91*	*4*	*< 0.0001*

Appendix Table 8 ANOVA results from a generalised linear mixed effects model on richness, based on ASV abundances obtained through 16S rRNA metabarcoding and showing that storage method has no association with alpha diversity.

Variable	***Chisq***	***Df***	***Pr(>Chisq)***
Sample Type	*0.3*	*1*	*0.87*

URN	*56.94*	*4*	*< 0.0001*

Appendix Table 9 ANOVA results from a generalised linear mixed effects model on read depth, based on ASV abundances obtained through 16S rRNA metabarcoding and showing that storage method has no association with read depth.

Exploring the long-term colonisation and persistence of probiotic-prophylaxis species on the gut microbiome of preterm infants: a pilot study.

Categorical Variables		
Variables	**Levels**	**Count**
Diet during admission	Combination	5
	Breastmilk	1
Diet post-discharge	Combination	5
	Breastmilk	1
Delivery	Vaginal	0
	Caesarean	6
NEC	Yes	0
	No	6
Sepsis	Yes	0
	No	6
Antenatal antibiotics	Yes	1
	No	5
Neonatal antibiotics	Yes	6
	No	0
Chorioamnionitis	Yes	0
	No	6

Preeclampsia	Yes	1
	No	5
Maternal Diabetes	Yes	0
	No	6
Continuous Variables		
Variable	**mean/median**	
Gestational age at birth	28.6 ± 0.5	

Appendix Table 10 Demographic/clinical data for infants that underwent 16S metabarcoding.

	DF	**Sum of squares**	**R2**	**F**	**Pr(>F)**
Timing of collection	2	2.26	0.35	4.45	<0.001
Diet	1	0.31	0.05	1.20	0.29
Residual	13	3.30	0.52	NA	NA
Total	16	6.38	1	NA	NA

Appendix Table 11 Results of a mixed effects model using PERMANOVA on ASV abundances obtained through 16S rRNA metabarcoding showing that the microbiome of this cohort differed significantly at different time points.

	R2	**Pr(>r)**
Timing of collection	0.85	<0.001
Diet	0.18	0.09

Appendix Table 12 Results of envfit analysis on ASV abundances obtained through 16S rRNA metabarcoding showing the time at which sample were collected was a significant covariate of the microbiome.

Variable	**Chisq**	**Df**	**Pr(>Chisq)**

Timing of collection	116.83	2	< 0.001
Diet	0.10	1	0.79

Appendix Table 13 ANOVA results from a generalised linear mixed effects model on the Shannon Index based on ASV abundances, obtained through 16S rRNA metabarcoding and showing that timing of sampling has a significant association with alpha diversity.

Contrast	**Estimate**	**SE**	**P**
Admission – Discharge	-0.71	0.25	< 0.01
Admission – Post-discharge	-3.02	0.29	<0.001
Discharge – post-discharge	-2.31	0.27	<0.001

Appendix Table 14 Tukey's pairwise comparison from a generalised linear mixed effects model on the Shannon Index based on ASV abundances, obtained through 16S rRNA metabarcoding and showing the significant pairwise differences between different time points in the infant's early life.

Variable	**Chisq**	**Df**	**Pr(>Chisq)**
Timing of collection	513.55	2	< 0.001
Diet	0.04	1	0.84

Appendix Table 15 ANOVA results from a generalised linear mixed effects model on richness based on ASV abundances, obtained through 16S rRNA metabarcoding and showing that timing of sampling has a significant association with alpha diversity.

Contrast	**Estimate**	**SE**	**P**
Admission – Discharge	-15.45	8.70	0.18
Admission – Post-discharge	-203.13	10.10	<0.001
Discharge – post-discharge	-187.66	9.21	<0.001

Appendix Table 16 Tukey's pairwise comparison from a generalised linear mixed effects model on the Shannon Index based on ASV abundances, obtained through 16S rRNA metabarcoding and showing a significant increase in richness post-discharge.

Comparison	Lfc	Lfc SE	P-adj	Species
Admission – Discharge	-6.80	1.72	<0.001	*Streptococcus*
Admission – Post-discharge	-7.28	1.96	<0.001	*Streptococcus*
Discharge – Post-discharge	4.02	0.91	<0.001	*Bifidobacterium*
Admission – Post-discharge	3.52	1.00	<0.001	*Bifidobacterium*

Appendix Table 17 Results of DESeq2 differential abundance testing showing the significant differences in Genera between different sampling time points.

Categorical Variables		
Variables	**Levels**	**Count**
Probiotics during admission	Yes	14
	No	4
Probiotics post-discharge	Yes	9
	No	9
Diet during admission	Combination	9
	Breastmilk	9
Diet post-discharge	Combination	15
	Breastmilk	3
Delivery	Vaginal	5
	Caesarean	13
NEC	Yes	0
	No	18
Sepsis	Yes	0

	No	18
Antenatal antibiotics	Yes	7
	No	11
Neonatal antibiotics	Yes	17
	No	1
Chorioamnionitis	Yes	2
	No	16
Preeclampsia	Yes	2
	No	0
Maternal Diabetes	Yes	1
	No	17
Continuous Variables		
Variable	**mean/median**	
Gestational age at birth	30.0 ± 1.3	

Appendix Table 18 Demographic/clinical data for infants included in the cross-sectional shotgun metagenomics analysis.

	DF	**Sum of squares**	**R2**	**F**	**Pr(>F)**
Probiotics	1	0.29	0.06	1.01	0.4
Delivery	2	0.65	0.13	1.14	0.24
Diet	1	0.24	0.05	0.85	0.6
Residual	13	3.71	0.75	NA	NA
Total	17	4.92	1	NA	NA

Appendix Table 19 Results of a mixed effects model using PERMANOVA on species level abundances obtained through shotgun metagenomics showing that probiotic supplementation does contribute to dissimilarity between groups post-discharge.

	R2	Pr(>r)
Probiotics	0.01	0.88
Delivery	0.11	0.88
Diet	0.02	0.88

Appendix Table 20 Results of a mixed effects model using envfit on species level abundances obtained through shotgun metagenomics showing that probiotic supplementation is not a significant covariate of the preterm infant microbiome post-discharge.

Variable	Chisq	Df	Pr(>Chisq)
Probiotics	5.28	1	< 0.05
Delivery	0.69	2	0.71
Diet	0.06	1	0.80

Appendix Table 21 ANOVA results from a generalised linear mixed effects model on the Shannon Index based on species level taxonomy, obtained through shotgun metagenomics, and showing that probiotic supplementation is significantly association with alpha diversity.

Contrast	Estimate	SE	P
Probiotics: No - Yes	0.63	0.27	< 0.05

Appendix Table 22 Tukey's pairwise comparison from a generalised linear mixed effects model on the Shannon Index based on species level taxonomy, obtained through shotgun metagenomics, and showing a significant negative association of probiotic prophylaxis with alpha diversity.

Variable	Chisq	Df	Pr(>Chisq)
Probiotics	6.53	1	< 0.05
Delivery	0.12	2	0.94
Diet	0.26	1	0.60

Appendix Table 23 ANOVA results from a generalised linear mixed effects model on richness based on species level taxonomy, obtained through shotgun metagenomics, and showing that probiotic supplementation is significantly association with alpha diversity.

Contrast	Estimate	SE	P
Probiotics: No - Yes	38	14.87	< 0.05

Appendix Table 24 Tukey's pairwise comparison from a generalised linear mixed effects model on richness based on species level taxonomy, obtained through shotgun metagenomics, and showing a significant negative association of probiotic prophylaxis with alpha diversity.

Variable: base level	Lfc	Lfc SE	P-adj	Species
Probiotics: Yes	29.09	2.59	< 0.01	Clostridium_M sp001517625
Probiotics: Yes	18.66	1.95	< 0.01	Flavonifractor plautii
Probiotics: Yes	-16.02	2.53	< 0.01	Alistipes finegoldii

Appendix Table 25 Results of DESeq2 differential abundance testing showing the significant differences in Genera between the probiotic supplementation groups.

9. References

1 Guinane, C. M. & Cotter, P. D. Role of the Gut Microbiota in Health and Chronic Gastrointestinal Disease: Understanding a Hidden Metabolic Organ. *Therap Adv Gastroenterol* **6**, 295-308 (2013).

2 Sommer, F. & Bäckhed, F. The Gut Microbiota--Masters of Host Development and Physiology. *Nature reviews Microbiology* **11**, 227-238 (2013).

3 Olszak, T. et al. Microbial Exposure During Early Life Has Persistent Effects on Natural Killer T Cell Function. *Science* **336**, 489-493 (2012).

4 Cho, I. et al. Antibiotics in Early Life Alter the Murine Colonic Microbiome and Adiposity. *Nature* **488**, 621-626 (2012).

5 Diaz Heijtz, R. Fetal, Neonatal, and Infant Microbiome: Perturbations and Subsequent Effects on Brain Development and Behavior. *Semin Fetal Neonatal Med* **21**, 410-417 (2016).

6 Cox, L. M. et al. Altering the Intestinal Microbiota During a Critical Developmental Window Has Lasting Metabolic Consequences. *Cell* **158**, 705-721 (2014).

7 Azad, M. B. et al. 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 and Gynaecology* **123**, 983-993 (2016).

8 Chernikova, D. A. et al. Fetal Exposures and Perinatal Influences on the Stool Microbiota of Premature Infants. *J Matern Fetal Neonatal Med* **29**, 99-105 (2016).

9 Liu, J. et al. Patterned Progression of Gut Microbiota Associated with Necrotizing Enterocolitis and Late Onset Sepsis in Preterm Infants: A Prospective Study in a Chinese Neonatal Intensive Care Unit. *PeerJ* **7**, e7310 (2019).

10 Sawh, S. C., Deshpande, S., Jansen, S., Reynaert, C. J. & Jones, P. M. Prevention of Necrotizing Enterocolitis with Probiotics: A Systematic Review and Meta-Analysis. *PeerJ* **4**, e2429 (2016).

11 Thomas, J. P., Raine, T., Reddy, S. & Belteki, G. Probiotics for the Prevention of Necrotising Enterocolitis in Very Low-Birth-Weight Infants: A Meta-Analysis and Systematic Review. *Acta Paediatr* **106**, 1729-1741 (2017).

12 Alcon-Giner, C. et al. Microbiota Supplementation with Bifidobacterium and Lactobacillus Modifies the Preterm Infant Gut Microbiota and Metabolome: An Observational Study. *Cell Reports Medicine* **1**, 100077 (2020).

13 Abdulkadir, B. et al. Routine Use of Probiotics in Preterm Infants: Longitudinal Impact on the Microbiome and Metabolome. *Neonatology* **109**, 239-247 (2016).

14 Belizario, J. E. & Napolitano, M. Human Microbiomes and Their Roles in Dysbiosis, Common Diseases, and Novel Therapeutic Approaches. *Front Microbiol* **6** (2015).

15 Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell* **148**, 1258-1270 (2012).

16 Brown, C. T. et al. Genome Resolved Analysis of a Premature Infant Gut Microbial Community Reveals a Varibaculum Cambriense Genome and a Shift Towards Fermentation-Based Metabolism During the Third Week of Life. *Microbiome* **1**, 30 (2013).

17 Human Microbiome Project Consortium. Structure, Function and Diversity of the Healthy Human Microbiome. *Nature* **486**, 207-214 (2012).

18 Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human Nutrition, the Gut Microbiome and the Immune System. *Nature* **474**, 327-336 (2011).

19 Moore, R. E. & Townsend, S. D. Temporal Development of the Infant Gut Microbiome. *Open biology* **9**, 190128 (2019).
20 Kaplan, J. L., Shi, H. N. & Walker, W. A. The Role of Microbes in Developmental Immunologic Programming. *Pediatr Res* **69**, 465-472 (2011).
21 Johansson, M. E., Sjovall, H. & Hansson, G. C. The Gastrointestinal Mucus System in Health and Disease. *Nat Rev Gastroenterol Hepatol* **10**, 352-361 (2013).
22 Adlerberth, I. & Wold, A. E. Establishment of the Gut Microbiota in Western Infants. *Acta Paediatr* **98**, 229-238 (2009).
23 Ferretti, P. et al. Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe* **24**, 133 (2018).
24 Van Daele, E., Knol, J. & Belzer, C. Microbial Transmission from Mother to Child: Improving Infant Intestinal Microbiota Development by Identifying the Obstacles. *Critical Reviews in Microbiology* **45**, 613-648 (2019).
25 Sonnenburg, E. D. et al. Diet-Induced Extinctions in the Gut Microbiota Compound over Generations. *Nature* **529**, 208-212 (2016).
26 Ward, D. V. et al. Metagenomic Sequencing with Strain-Level Resolution Implicates Uropathogenic E. Coli in Necrotizing Enterocolitis and Mortality in Preterm Infants. *Cell reports* **14**, 2912-2924 (2016).
27 Lemas, D. J. et al. Exploring the Contribution of Maternal Antibiotics and Breastfeeding to Development of the Infant Microbiome and Pediatric Obesity. *Seminars in Fetal and Neonatal Medicine* **21**, 406-409 (2016).
28 Kim, H., Sitarik, A. R., Woodcroft, K., Johnson, C. C. & Zoratti, E. Birth Mode, Breastfeeding, Pet Exposure, and Antibiotic Use: Associations with the Gut Microbiome and Sensitization in Children. *Current allergy and asthma reports* **19**, 22 (2019).
29 Perez-Muñoz, M. E., Arrieta, M. C., Ramer-Tait, A. E. & Walter, J. A Critical Assessment of the "Sterile Womb" and "in Utero Colonization" Hypotheses: Implications for Research on the Pioneer Infant Microbiome. *Microbiome* **5**, 48 (2017).
30 Theis, K. R. et al. No Consistent Evidence for Microbiota in Murine Placental and Fetal Tissues. *bioRxiv*, 2019.2012.2010.872275 (2019).
31 Aagaard, K. et al. The Placenta Harbors a Unique Microbiome. *Sci Transl Med* **6**, 237-265 (2014).
32 Stout, M. J. et al. Identification of Intracellular Bacteria in the Basal Plate of the Human Placenta in Term and Preterm Gestations. *Am J Obstet Gynecol* **208**, 226-227 (2013).
33 Jimenez, E. et al. Is Meconium from Healthy Newborns Actually Sterile? *Res Microbiol* **159**, 187-193 (2008).
34 Abrahamsson, T. R., Wu, R. Y. & Jenmalm, M. C. Gut Microbiota and Allergy: The Importance of the Pregnancy Period. *Pediatr Res* **77**, 214-219 (2015).
35 Gilbert, S. F. A Holobiont Birth Narrative: The Epigenetic Transmission of the Human Microbiome. *Frontiers in Genetics* **5**, 282 (2014).
36 de Goffau, M. C. et al. Human Placenta Has No Microbiome but Can Contain Potential Pathogens. *Nature* **572**, 329-334 (2019).
37 Itani, T. et al. Establishment and Development of the Intestinal Microbiota of Preterm Infants in a Lebanese Tertiary Hospital. *Anaerobe* **43**, 4-14 (2017).
38 Dietert, R. R. The Microbiome-Immune-Host Defense Barrier Complex (Microimmunosome) and Developmental Programming of Noncommunicable Diseases. *Reprod Toxicol* **68**, 49-58 (2017).

39 Dominguez-Bello, M. G. et al. Delivery Mode Shapes the Acquisition and Structure of the Initial Microbiota across Multiple Body Habitats in Newborns. *Proc Natl Acad Sci U S A* **107**, 11971-11975 (2010).
40 Aroutcheva, A. et al. Defense Factors of Vaginal Lactobacilli. *Am J Obstet Gynecol* **185**, 375-379 (2001).
41 Boskey, E. R., Cone, R. A., Whaley, K. J. & Moench, T. R. Origins of Vaginal Acidity: High D/L Lactate Ratio Is Consistent with Bacteria Being the Primary Source. *Hum Reprod* **16**, 1809-1813 (2001).
42 Stewart, C. J. et al. Preterm Gut Microbiota and Metabolome Following Discharge from Intensive Care. *Sci Rep* **5**, 17141 (2015).
43 Grzeskowiak, L. et al. Gut Bifidobacterium Microbiota in One-Month-Old Brazilian Newborns. *Anaerobe* **35**, 54-58 (2015).
44 Gregory, K. E. et al. Influence of Maternal Breast Milk Ingestion on Acquisition of the Intestinal Microbiome in Preterm Infants. *Microbiome* **4**, 68 (2016).
45 Costello, E. K., Carlisle, E. M., Bik, E. M., Morowitz, M. J. & Relman, D. A. Microbiome Assembly across Multiple Body Sites in Low-Birthweight Infants. *Mbio* **4** (2013).
46 Arboleya, S. et al. Intestinal Microbiota Development in Preterm Neonates and Effect of Perinatal Antibiotics. *Journal of Pediatrics* **166**, 538-544 (2015).
47 Arboleya, S. et al. Intestinal Microbiota and Weight-Gain in Preterm Neonates. *Front Microbiol* **8** (2017).
48 Arrieta, M. C., Stiemsma, L. T., Amenyogbe, N., Brown, E. M. & Finlay, B. The Intestinal Microbiome in Early Life: Health and Disease. *Front Immunol* **5**, 427 (2014).
49 Betran, A. P., Ye, J., Moller, A.-B., Souza, J. P. & Zhang, J. Trends and Projections of Caesarean Section Rates: Global and Regional Estimates. *BMJ Global Health* **6** (2021).
50 Ballard, O. & Morrow, A. L. Human Milk Composition: Nutrients and Bioactive Factors. *Pediatric clinics of North America* **60**, 49-74 (2013).
51 Martin, C. R., Ling, P. R. & Blackburn, G. L. Review of Infant Feeding: Key Features of Breast Milk and Infant Formula. *Nutrients* **8** (2016).
52 Bode, L. Human Milk Oligosaccharides: Prebiotics and Beyond. *Nutr Rev* **67 Suppl 2**, 183-191 (2009).
53 Zimmermann, P. & Curtis, N. Breast Milk Microbiota: A Review of the Factors That Influence Composition. *The Journal of infection* **81**, 17-47 (2020).
54 Dardas, M. et al. The Impact of Postnatal Antibiotics on the Preterm Intestinal Microbiome. *Pediatr Res* **76**, 150-158 (2014).
55 Meinzen-Derr, J. et al. Role of Human Milk in Extremely Low Birth Weight Infants' Risk of Necrotizing Enterocolitis or Death. *J Perinatol* **29**, 57-62 (2009).
56 Mshvildadze, M. et al. Intestinal Microbial Ecology in Premature Infants Assessed with Non-Culture-Based Techniques. *J Pediatr* **156**, 20-25 (2010).
57 Mastromarino, P. et al. Correlation between Lactoferrin and Beneficial Microbiota in Breast Milk and Infant's Feces. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine* **27**, 1077-1086 (2014).
58 Houghteling, P. D. & Walker, W. A. Why Is Initial Bacterial Colonization of the Intestine Important to Infants' and Children's Health? *Journal of Pediatric Gastroenterology and Nutrition* **60**, 294-307 (2015).
59 Clarke, G., O'Mahony, S. M., Dinan, T. G. & Cryan, J. F. Priming for Health: Gut Microbiota Acquired in Early Life Regulates Physiology, Brain and Behaviour. *Acta Paediatr* **103**, 812-819 (2014).

60 Jost, T., Lacroix, C., Braegger, C. P. & Chassard, C. New Insights in Gut Microbiota Establishment in Healthy Breast Fed Neonates. *PLoS One* **7**, e44595 (2012).
61 Voreades, N., Kozil, A. & Weir, T. L. Diet and the Development of the Human Intestinal Microbiome. *Frontiers in microbiology* **5**, 494 (2014).
62 Yang, I. et al. The Infant Microbiome: Implications for Infant Health and Neurocognitive Development. *Nurs Res* **65**, 76-88 (2016).
63 Timmerman, H. M. et al. Intestinal Colonisation Patterns in Breastfed and Formula-Fed Infants During the First 12 Weeks of Life Reveal Sequential Microbiota Signatures. *Sci Rep* **7**, 8327 (2017).
64 Parra-Llorca, A. et al. Preterm Gut Microbiome Depending on Feeding Type: Significance of Donor Human Milk. *Frontiers in microbiology* **9**, 1376 (2018).
65 Qin, J. et al. A Metagenome-Wide Association Study of Gut Microbiota in Type 2 Diabetes. *Nature* **490**, 55-60 (2012).
66 Lv, L. J. et al. Early-Onset Preeclampsia Is Associated with Gut Microbial Alterations in Antepartum and Postpartum Women. *Frontiers in cellular and infection microbiology* **9**, 224 (2019).
67 Arboleya, S. et al. Impact of Prematurity and Perinatal Antibiotics on the Developing Intestinal Microbiota: A Functional Inference Study. *Int J Mol Sci* **17** (2016).
68 Forsgren, M., Isolauri, E., Salminen, S. & Rautava, S. Late Preterm Birth Has Direct and Indirect Effects on Infant Gut Microbiota Development During the First Six Months of Life. *Acta Paediatr* **106**, 1103-1109 (2017).
69 Hill, C. J. et al. Evolution of Gut Microbiota Composition from Birth to 24 Weeks in the Infantmet Cohort. *Microbiome* **5**, 4 (2017).
70 Rouge, C. et al. Investigation of the Intestinal Microbiota in Preterm Infants Using Different Methods. *Anaerobe* **16**, 362-370 (2010).
71 Barrett, E. et al. The Individual-Specific and Diverse Nature of the Preterm Infant Microbiota. *Arch Dis Child Fetal Neonatal Ed* **98**, 334-340 (2013).
72 Arboleya, S. et al. Establishment and Development of Intestinal Microbiota in Preterm Neonates. *FEMS Microbiol Ecol* **79**, 763-772 (2012).
73 Chang, J. Y., Shin, S. M., Chun, J., Lee, J. H. & Seo, J. K. Pyrosequencing-Based Molecular Monitoring of the Intestinal Bacterial Colonization in Preterm Infants. *Journal of Pediatric Gastroenterology and Nutrition* **53**, 512-519 (2011).
74 Schwiertz, A. et al. Development of the Intestinal Bacterial Composition in Hospitalized Preterm Infants in Comparison with Breast-Fed, Full-Term Infants. *Pediatr Res* **54**, 393-399 (2003).
75 Patel, A. L. et al. Longitudinal Survey of Microbiota in Hospitalized Preterm Very-Low-Birth-Weight Infants. *J Pediatr Gastroenterol Nutr* **62**, 292-303 (2016).
76 Moles, L. et al. Bacterial Diversity in Meconium of Preterm Neonates and Evolution of Their Fecal Microbiota During the First Month of Life. *Plos One* **8** (2013).
77 Magne, F. et al. Low Species Diversity and High Interindividual Variability in Faeces of Preterm Infants as Revealed by Sequences of 16s Rrna Genes and Pcr-Temporal Temperature Gradient Gel Electrophoresis Profiles. *FEMS Microbiol Ecol* **57**, 128-138 (2006).
78 Butel, M. J. et al. Conditions of Bifidobacterial Colonization in Preterm Infants: A Prospective Analysis. *Journal of Pediatric Gastroenterology and Nutrition* **44**, 577-582 (2007).
79 Campeotto, F. et al. A Fermented Formula in Pre-Term Infants: Clinical Tolerance, Gut Microbiota, Down-Regulation of Faecal Calprotectin and up-Regulation of Faecal Secretory Iga. *Br J Nutr* **105**, 1843-1851 (2011).

80 Morowitz, M. J. et al. Strain-Resolved Community Genomic Analysis of Gut Microbial Colonization in a Premature Infant. *Proc Natl Acad Sci U S A* **108**, 1128-1133 (2011).
81 Xiong, W., Brown, C. T., Morowitz, M. J., Banfield, J. F. & Hettich, R. L. Genome-Resolved Metaproteomic Characterization of Preterm Infant Gut Microbiota Development Reveals Species-Specific Metabolic Shifts and Variabilities During Early Life. *Microbiome* **5**, 72 (2017).
82 Gomez, M. et al. Early Gut Colonization of Preterm Infants: Effect of Enteral Feeding Tubes. *J Pediatr Gastroenterol Nutr* **62**, 893-900 (2016).
83 Arboleya, S., Solis, G., Fernandez, N., de los Reyes-Gavilan, C. G. & Gueimonde, M. Facultative to Strict Anaerobes Ratio in the Preterm Infant Microbiota: A Target for Intervention? *Gut microbes* **3**, 583-588 (2012).
84 La Rosa, P. S. et al. Patterned Progression of Bacterial Populations in the Premature Infant Gut. *Proc Natl Acad Sci U S A* **111**, 12522-12527 (2014).
85 Gomez, M. et al. Bacteriological and Immunological Profiling of Meconium and Fecal Samples from Preterm Infants: A Two-Year Follow-up Study. *Nutrients* **9** (2017).
86 Tamburini, S., Shen, N., Wu, H. C. & Clemente, J. C. The Microbiome in Early Life: Implications for Health Outcomes. *Nat Med* **22**, 713-722 (2016).
87 Shaw, A. G. et al. Late-Onset Bloodstream Infection and Perturbed Maturation of the Gastrointestinal Microbiota in Premature Infants. *PLoS One* **10** (2015).
88 Dobbler, P. T. et al. Low Microbial Diversity and Abnormal Microbial Succession Is Associated with Necrotizing Enterocolitis in Preterm Infants. *Front Microbiol* **8** (2017).
89 Hufnagl, K., Pali-Schöll, I., Roth-Walter, F. & Jensen-Jarolim, E. Dysbiosis of the Gut and Lung Microbiome Has a Role in Asthma. *Seminars in Immunopathology* **42**, 75-93 (2020).
90 Siljander, H., Honkanen, J. & Knip, M. Microbiome and Type 1 Diabetes. *EBioMedicine* **46**, 512-521 (2019).
91 Barron, L. K. et al. Independence of Gut Bacterial Content and Neonatal Necrotizing Enterocolitis Severity. *Journal of pediatric surgery* **52**, 993-998 (2017).
92 Gregory, K. E., Deforge, C. E., Natale, K. M., Phillips, M. & Van Marter, L. J. Necrotizing Enterocolitis in the Premature Infant: Neonatal Nursing Assessment, Disease Pathogenesis, and Clinical Presentation. *Adv Neonatal Care* **11**, 155-164 (2011).
93 Neu, J. & Pammi, M. Pathogenesis of Nec: Impact of an Altered Intestinal Microbiome. *Seminars in perinatology* **41**, 29-35 (2017).
94 Wang, Y. et al. 16s Rrna Gene-Based Analysis of Fecal Microbiota from Preterm Infants with and without Necrotizing Enterocolitis. *ISME J* **3**, 944-954 (2009).
95 Athalye-Jape, G., Deshpande, G., Rao, S. & Patole, S. Benefits of Probiotics on Enteral Nutrition in Preterm Neonates: A Systematic Review. *Am J Clin Nutr* **100**, 1508-1519 (2014).
96 Stewart, C. J. et al. Temporal Bacterial and Metabolic Development of the Preterm Gut Reveals Specific Signatures in Health and Disease. *Microbiome* **4**, 67 (2016).
97 Mai, V. et al. Fecal Microbiota in Premature Infants Prior to Necrotizing Enterocolitis. *Plos One* **6** (2011).
98 Pammi, M. et al. Intestinal Dysbiosis in Preterm Infants Preceding Necrotizing Enterocolitis: A Systematic Review and Meta-Analysis. *Microbiome* **5**, 31 (2017).

99 Cassir, N. et al. Clostridium Butyricum Strains and Dysbiosis Linked to Necrotizing Enterocolitis in Preterm Neonates. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **61**, 1107-1115 (2015).
100 Sim, K. et al. Dysbiosis Anticipating Necrotizing Enterocolitis in Very Premature Infants. *Clinical Infectious Diseases* **60**, 389-397 (2015).
101 Itani, T. et al. Preterm Infants with Necrotising Enterocolitis Demonstrate an Unbalanced Gut Microbiota. *Acta Paediatr* **107**, 40-47 (2018).
102 Warner, B. B. et al. Gut Bacteria Dysbiosis and Necrotising Enterocolitis in Very Low Birthweight Infants: A Prospective Case-Control Study. *Lancet* **387**, 1928-1936 (2016).
103 Sharif, S., Meader, N., Oddie, S. J., Rojas-Reyes, M. X. & McGuire, W. Probiotics to Prevent Necrotising Enterocolitis in Very Preterm or Very Low Birth Weight Infants. *Cochrane Database of Systematic Reviews* (2020).
104 Granger, C. et al. Necrotising Enterocolitis, Late-Onset Sepsis and Mortality after Routine Probiotic Introduction in the Uk. *Archives of Disease in Childhood - Fetal and Neonatal Edition* (2021).
105 Berkhout, D. J. C. et al. Detection of Sepsis in Preterm Infants by Fecal Volatile Organic Compounds Analysis: A Proof of Principle Study. *J Pediatr Gastroenterol Nutr* **65**, 47-52 (2017).
106 Bizzarro, M. J., Raskind, C., Baltimore, R. S. & Gallagher, P. G. Seventy-Five Years of Neonatal Sepsis at Yale: 1928-2003. *Pediatrics* **116**, 595-602 (2005).
107 Stewart, C. J. et al. The Preterm Gut Microbiota: Changes Associated with Necrotizing Enterocolitis and Infection. *Acta Paediatr* **101**, 1121-1127 (2012).
108 Mai, V. et al. Distortions in Development of Intestinal Microbiota Associated with Late Onset Sepsis in Preterm Infants. *PLoS One* **8**, 52876 (2013).
109 Madan, J. C. et al. Gut Microbial Colonisation in Premature Neonates Predicts Neonatal Sepsis. *Arch Dis Child Fetal Neonatal Ed* **97**, 456-462 (2012).
110 Stewart, C. J. et al. Longitudinal Development of the Gut Microbiome and Metabolome in Preterm Neonates with Late Onset Sepsis and Healthy Controls. *Microbiome* **5**, 75 (2017).
111 Sherman, M. P. New Concepts of Microbial Translocation in the Neonatal Intestine: Mechanisms and Prevention. *Clin Perinatol* **37**, 565-579 (2010).
112 Sharma, R., Young, C. & Neu, J. Molecular Modulation of Intestinal Epithelial Barrier: Contribution of Microbiota. *Journal of biomedicine & biotechnology* **2010**, 305879 (2010).
113 Taft, D. H. et al. Intestinal Microbiota of Preterm Infants Differ over Time and between Hospitals. *Microbiome* **2** (2014).
114 Underwood, M. A. et al. A Randomized Placebo-Controlled Comparison of 2 Prebiotic/Probiotic Combinations in Preterm Infants: Impact on Weight Gain, Intestinal Microbiota, and Fecal Short-Chain Fatty Acids. *J Pediatr Gastroenterol Nutr* **48**, 216-225 (2009).
115 Raveh-Sadka, T. et al. Evidence for Persistent and Shared Bacterial Strains against a Background of Largely Unique Gut Colonization in Hospitalized Premature Infants. *Isme j* **10**, 2817-2830 (2016).
116 Stewart, C. J. et al. Development of the Preterm Gut Microbiome in Twins at Risk of Necrotising Enterocolitis and Sepsis. *Plos One* **8** (2013).
117 Rosberg-Cody, E. et al. Mining the Microbiota of the Neonatal Gastrointestinal Tract for Conjugated Linoleic Acid-Producing Bifidobacteria. *Appl Environ Microbiol* **70**, 4635-4641 (2004).

118 Moles, L. et al. Preterm Infant Gut Colonization in the Neonatal Icu and Complete Restoration 2 Years Later. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **21**, 936 (2015).
119 Kuppala, V. S., Meinzen-Derr, J., Morrow, A. L. & Schibler, K. R. Prolonged Initial Empirical Antibiotic Treatment Is Associated with Adverse Outcomes in Premature Infants. *J Pediatr* **159**, 720-725 (2011).
120 Russell, S. L. et al. Early Life Antibiotic-Driven Changes in Microbiota Enhance Susceptibility to Allergic Asthma. *Embo Rep* **13**, 440-447 (2012).
121 Deshpande, G., Rao, S. & Patole, S. Probiotics in Neonatal Intensive Care - Back to the Future. *The Australian & New Zealand journal of obstetrics & gynaecology* **55**, 210-217 (2015).
122 Costeloe, K. et al. Bifidobacterium Breve Bbg-001 in Very Preterm Infants: A Randomised Controlled Phase 3 Trial. *Lancet* **387**, 649-660 (2016).
123 Westaway, J. A. F. et al. Methods for Exploring the Faecal Microbiome of Premature Infants: A Review. *Maternal Health, Neonatology and Perinatology* **7**, 11 (2021).
124 Lewis, Z. T. et al. Validating Bifidobacterial Species and Subspecies Identity in Commercial Probiotic Products. *Pediatr Res* **79**, 445-452 (2016).
125 Repa, A. et al. Probiotics (Lactobacillus Acidophilus and Bifidobacterium Bifidum) Prevent Nec in Vlbw Infants Fed Breast Milk but Not Formula. *Pediatr Res* **77**, 381-388 (2015).
126 Frese, S. A. et al. Persistence of Supplemented Bifidobacterium Longum Subsp. Infantis Evc001 in Breastfed Infants. *mSphere* **2** (2017).
127 Dinan, T. G. & Cryan, J. F. Gut Instincts: Microbiota as a Key Regulator of Brain Development, Ageing and Neurodegeneration. *J Physiol* **595**, 489-503 (2017).
128 Lozupone, C. A. et al. Meta-Analyses of Studies of the Human Microbiota. *Genome Research* **23**, 1704-1714 (2013).
129 Mao, D. P., Zhou, Q., Chen, C. Y. & Quan, Z. X. Coverage Evaluation of Universal Bacterial Primers Using the Metagenomic Datasets. *BMC Microbiol* **12**, 66 (2012).
130 Arboleya, S. et al. Assessment of Intestinal Microbiota Modulation Ability of Bifidobacterium Strains in in Vitro Fecal Batch Cultures from Preterm Neonates. *Anaerobe* **19**, 9-16 (2013).
131 Bjorkstrom, M. V. et al. Intestinal Flora in Very Low-Birth Weight Infants. *Acta Paediatr* **98**, 1762-1767 (2009).
132 Carl, M. A. et al. Sepsis from the Gut: The Enteric Habitat of Bacteria That Cause Late-Onset Neonatal Bloodstream Infections. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **58**, 1211-1218 (2014).
133 Chrzanowska-Liszewska, D., Seliga-Siwecka, J. & Kornacka, M. K. The Effect of Lactobacillus Rhamnosus Gg Supplemented Enteral Feeding on the Microbiotic Flora of Preterm Infants-Double Blinded Randomized Control Trial. *Early Hum Dev* **88**, 57-60 (2012).
134 Ishizeki, S., Sugita, M., Takata, M. & Yaeshima, T. Effect of Administration of Bifidobacteria on Intestinal Microbiota in Low-Birth-Weight Infants and Transition of Administered Bifidobacteria: A Comparison between One-Species and Three-Species Administration. *Anaerobe* **23**, 38-44 (2013).
135 Andriantsoanirina, V., Teolis, A. C., Xin, L. X., Butel, M. J. & Aires, J. Bifidobacterium Longum and Bifidobacterium Breve Isolates from Preterm and Full Term Neonates: Comparison of Cell Surface Properties. *Anaerobe* **28**, 212-215 (2014).

136 Arboleya, S. et al. Deep 16s Rrna Metagenomics and Quantitative Pcr Analyses of the Premature Infant Fecal Microbiota. *Anaerobe* **18**, 378-380 (2012).
137 Ardissone, A. N. et al. Meconium Microbiome Analysis Identifies Bacteria Correlated with Premature Birth. *PLoS One* **9**, 90784 (2014).
138 Gupta, R. W. et al. Histamine-2 Receptor Blockers Alter the Fecal Microbiota in Premature Infants. *J Pediatr Gastroenterol Nutr* **56**, 397-400 (2013).
139 Leach, S. T. et al. Multiple Opportunistic Pathogens, but Not Pre-Existing Inflammation, May Be Associated with Necrotizing Enterocolitis. *Dig Dis Sci* **60**, 3728-3734 (2015).
140 McMurtry, V. E. et al. Bacterial Diversity and Clostridia Abundance Decrease with Increasing Severity of Necrotizing Enterocolitis. *Microbiome* **3** (2015).
141 Millar, M. et al. The Microbiome of Infants Recruited to a Randomised Placebo-Controlled Probiotic Trial (Pips Trial). *EBioMedicine* **20**, 255-262 (2017).
142 Normann, E., Fahlen, A., Engstrand, L. & Lilja, H. E. Intestinal Microbial Profiles in Extremely Preterm Infants with and without Necrotizing Enterocolitis. *Acta Paediatr* **102**, 129-136 (2013).
143 Poroyko, V. et al. Diet Creates Metabolic Niches in the "Immature Gut" That Shape Microbial Communities. *Nutricion hospitalaria* **26**, 1283-1295 (2011).
144 Said, M. B. et al. Gut Microbiota in Preterm Infants with Gross Blood in Stools: A Prospective, Controlled Study. *Early Hum Dev* **90**, 579-585 (2014).
145 Sherman, M. P., Sherman, J., Arcinue, R. & Niklas, V. Randomized Control Trial of Human Recombinant Lactoferrin: A Substudy Reveals Effects on the Fecal Microbiome of Very Low Birth Weight Infants. *Journal of Pediatrics* **173**, 37-42 (2016).
146 Taft, D. H. et al. Center Variation in Intestinal Microbiota Prior to Late-Onset Sepsis in Preterm Infants. *PLoS One* **10**, 0130604 (2015).
147 Underwood, M. A. et al. Digestion of Human Milk Oligosaccharides by Bifidobacterium Breve in the Premature Infant. *J Pediatr Gastroenterol Nutr* **65**, 449-455 (2017).
148 Underwood, M. A. et al. Human Milk Oligosaccharides in Premature Infants: Absorption, Excretion, and Influence on the Intestinal Microbiota. *Pediatr Res* **78**, 670-677 (2015).
149 Young, G. R. et al. Reducing Viability Bias in Analysis of Gut Microbiota in Preterm Infants at Risk of Nec and Sepsis. *Frontiers in cellular and infection microbiology* **7**, 237 (2017).
150 Zeber-Lubecka, N. et al. Effect of Saccharomyces Boulardii and Mode of Delivery on the Early Development of the Gut Microbial Community in Preterm Infants. *PLoS One* **11**, 0150306 (2016).
151 Zhou, Y. J. et al. Longitudinal Analysis of the Premature Infant Intestinal Microbiome Prior to Necrotizing Enterocolitis: A Case-Control Study. *Plos One* **10** (2015).
152 Morrow, A. L. et al. Early Microbial and Metabolomic Signatures Predict Later Onset of Necrotizing Enterocolitis in Preterm Infants. *Microbiome* **1**, 13 (2013).
153 Sim, K. et al. Dysbiosis Anticipating Necrotizing Enterocolitis in Very Premature Infants. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **60**, 389-397 (2015).
154 iHo, T. T. B. et al. Dichotomous Development of the Gut Microbiome in Preterm Infants. *Microbiome* **6**, 157 (2018).
155 Aujoulat, F. et al. Temporal Dynamics of the Very Premature Infant Gut Dominant Microbiota. *Bmc Microbiol* **14** (2014).

156 Ferraris, L. et al. Clostridia in Premature Neonates' Gut: Incidence, Antibiotic Susceptibility, and Perinatal Determinants Influencing Colonization. *PLoS One* **7**, 30594 (2012).
157 Rouge, C. et al. Oral Supplementation with Probiotics in Very-Low-Birth-Weight Preterm Infants: A Randomized, Double-Blind, Placebo-Controlled Trial. *American Journal of Clinical Nutrition* **89**, 1828-1835 (2009).
158 de la Cochetière, M.-F. et al. Early Intestinal Bacterial Colonization and Necrotizing Enterocolitis in Premature Infants: The Putative Role of Clostridium. *Pediatr Res* **56**, 366-370 (2004).
159 Smith, B. et al. Investigation of the Early Intestinal Microflora in Premature Infants with/without Necrotizing Enterocolitis Using Two Different Methods. *Pediatr Res* **71**, 115-120 (2012).
160 Millar, M. R. et al. Application of 16s Rrna Gene Pcr to Study Bowel Flora of Preterm Infants with and without Necrotizing Enterocolitis. *J Clin Microbiol* **34**, 2506-2510 (1996).
161 Van Den Berg, J. P., Westerbeek, E. A. M., Bröring-Starre, T., Garssen, J. & Van Elburg, R. M. Neurodevelopment of Preterm Infants at 24 Months after Neonatal Supplementation of a Prebiotic Mix: A Randomized Trial. *Journal of Pediatric Gastroenterology and Nutrition* **63**, 270-276 (2016).
162 Westerbeek, E. A. M. et al. The Effect of Enteral Supplementation of Specific Neutral and Acidic Oligosaccharides on the Faecal Microbiota and Intestinal Microenvironment in Preterm Infants. *Eur J Clin Microbiol* **32**, 269-276 (2013).
163 Underwood, M. A. et al. A Comparison of Two Probiotic Strains of Bifidobacteria in Premature Infants. *J Pediatr* **163**, 1585-1591 (2013).
164 Underwood, M. A. et al. Prebiotic Oligosaccharides in Premature Infants. *Journal of Pediatric Gastroenterology and Nutrition* **58**, 352-360 (2014).
165 Abdulkadir, B. et al. Stool Bacterial Load in Preterm Infants with Necrotising Enterocolitis. *Early Hum Dev* **95**, 1-2 (2016).
166 Aly, H. et al. Medically Graded Honey Supplementation Formula to Preterm Infants as a Prebiotic: A Randomized Controlled Trial. *Journal of Pediatric Gastroenterology and Nutrition* **64**, 966-970 (2017).
167 Arboleya, S. et al. In Vitro Evaluation of the Impact of Human Background Microbiota on the Response to Bifidobacterium Strains and Fructo-Oligosaccharides. *Br J Nutr* **110**, 2030-2036 (2013).
168 Gregory, K. E., LaPlante, R. D., Shan, G., Kumar, D. V. & Gregas, M. Mode of Birth Influences Preterm Infant Intestinal Colonization with Bacteroides over the Early Neonatal Period. *Adv Neonatal Care* **15**, 386-393 (2015).
169 Hickey, L., Garland, S. M., Jacobs, S. E., O'Donnell, C. P. & Tabrizi, S. N. Cross-Colonization of Infants with Probiotic Organisms in a Neonatal Unit. *The Journal of hospital infection* **88**, 226-229 (2014).
170 Luoto, R. et al. Gross Blood in Stools of Premature Neonates, a Clinical and Microbiological Follow-up Study. *Acta Paediatr* **102**, 486-491 (2013).
171 Leggett, R. M. et al. Rapid Minion Profiling of Preterm Microbiota and Antimicrobial-Resistant Pathogens. *Nat Microbiol* (2019).
172 Bolyen, E. et al. Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using Qiime 2. *Nat Biotechnol* **37**, 852-857 (2019).
173 Schloss, P. D. et al. Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microb* **75**, 7537-7541 (2009).

174 Bertani, G. Studies on Lysogenesis. I. The Mode of Phage Liberation by Lysogenic Escherichia Coli. *J Bacteriol* **62**, 293-300 (1951).
175 Macconkey, A. Lactose-Fermenting Bacteria in Faeces. *J Hyg (Lond)* **5**, 333-379 (1905).
176 Zajc-Satler, J. & Gragas, A. Z. Xylose Lysine Deoxycholate Agar for the Isolation of Salmonella and Shigella from Clinical Specimens. *Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene Erste Abteilung Originale Reihe A: Medizinische Mikrobiologie und Parasitologie* **237**, 196-200 (1977).
177 Lagier, J. C. et al. Culture of Previously Uncultured Members of the Human Gut Microbiota by Culturomics. *Nat Microbiol* **1**, 16203 (2016).
178 Berrington, J. E., Stewart, C. J., Embleton, N. D. & Cummings, S. P. Gut Microbiota in Preterm Infants: Assessment and Relevance to Health and Disease. *Archives of Disease in Childhood: Fetal and Neonatal Edition* **98**, 286-290 (2013).
179 Amann, R. I., Ludwig, W. & Schleifer, K. H. Phylogenetic Identification and in-Situ Detection of Individual Microbial-Cells without Cultivation. *Microbiol Rev* **59**, 143-169 (1995).
180 Cardona, S. et al. Storage Conditions of Intestinal Microbiota Matter in Metagenomic Analysis. *BMC Microbiol* **12**, 158 (2012).
181 Carroll, I. M., Ringel-Kulka, T., Siddle, J. P., Klaenhammer, T. R. & Ringel, Y. Characterization of the Fecal Microbiota Using High-Throughput Sequencing Reveals a Stable Microbial Community During Storage. *Plos One* **7** (2012).
182 Roesch, L. F. et al. Influence of Fecal Sample Storage on Bacterial Community Diversity. *Open Microbiol J* **3**, 40-46 (2009).
183 Gorzelak, M. A. et al. Methods for Improving Human Gut Microbiome Data by Reducing Variability through Sample Processing and Storage of Stool. *Plos One* **10** (2015).
184 Wu, G. D. et al. Sampling and Pyrosequencing Methods for Characterizing Bacterial Communities in the Human Gut Using 16s Sequence Tags. *Bmc Microbiol* **10** (2010).
185 Tedjo, D. I. et al. The Effect of Sampling and Storage on the Fecal Microbiota Composition in Healthy and Diseased Subjects. *Plos One* **10** (2015).
186 Hale, V. L., Tan, C. L., Knight, R. & Amato, K. R. Effect of Preservation Method on Spider Monkey (Ateles Geoffroyi) Fecal Microbiota over 8 Weeks. *J Microbiol Meth* **113**, 16-26 (2015).
187 Shaw, A. G. et al. Latitude in Sample Handling and Storage for Infant Faecal Microbiota Studies: The Elephant in the Room? *Microbiome* **4**, 40 (2016).
188 Walker, A. W. et al. 16s Rrna Gene-Based Profiling of the Human Infant Gut Microbiota Is Strongly Influenced by Sample Processing and Pcr Primer Choice. *Microbiome* **3**, 26 (2015).
189 Hang, J. et al. 16s Rrna Gene Pyrosequencing of Reference and Clinical Samples and Investigation of the Temperature Stability of Microbiome Profiles. *Microbiome* **2** (2014).
190 Maukonen, J., Simoes, C. & Saarela, M. The Currently Used Commercial DNA-Extraction Methods Give Different Results of Clostridial and Actinobacterial Populations Derived from Human Fecal Samples. *Fems Microbiol Ecol* **79**, 697-708 (2012).
191 Bahl, M. I., Bergstrom, A. & Licht, T. R. Freezing Fecal Samples Prior to DNA Extraction Affects the Firmicutes to Bacteroidetes Ratio Determined by Downstream Quantitative Pcr Analysis. *Fems Microbiol Lett* **329**, 193-197 (2012).

192 Nechvatal, J. M. et al. Fecal Collection, Ambient Preservation, and DNA Extraction for Pcr Amplification of Bacterial and Human Markers from Human Feces. *J Microbiol Meth* **72**, 124-132 (2008).
193 Han, M. et al. A Novel Affordable Reagent for Room Temperature Storage and Transport of Fecal Samples for Metagenomic Analyses. *Microbiome* **6**, 43 (2018).
194 Anderson, E. L. et al. A Robust Ambient Temperature Collection and Stabilization Strategy: Enabling Worldwide Functional Studies of the Human Microbiome. *Sci Rep* **6**, 31731 (2016).
195 Song, S. J. et al. Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies. *mSystems* **1**, 21 (2016).
196 Choo, J. M., Leong, L. E. X. & Rogers, G. B. Sample Storage Conditions Significantly Influence Faecal Microbiome Profiles. *Sci Rep-Uk* **5** (2015).
197 Flores, R. et al. Collection Media and Delayed Freezing Effects on Microbial Composition of Human Stool. *Microbiome* **3**, 33 (2015).
198 Dominianni, C., Wu, J., Hayes, R. B. & Ahn, J. Comparison of Methods for Fecal Microbiome Biospecimen Collection. *Bmc Microbiol* **14** (2014).
199 Menke, S., Gillingham, M. A. F., Wilhelm, K. & Sommer, S. Home-Made Cost Effective Preservation Buffer Is a Better Alternative to Commercial Preservation Methods for Microbiome Research. *Front Microbiol* **8** (2017).
200 Hill, C. J. et al. Effect of Room Temperature Transport Vials on DNA Quality and Phylogenetic Composition of Faecal Microbiota of Elderly Adults and Infants. *Microbiome* **4**, 19 (2016).
201 GenoHub. *Whole Genome Sequencing and Re-Sequencing Guide*, <https://genohub.com/dna-seq-library-preparation/#:~:text=DNA%20Sample%20Submission%2D%20Typically%20100,10%20ng%20of%20input%20material> (2019).
202 Mackenzie, B. W., Waite, D. W. & Taylor, M. W. Evaluating Variation in Human Gut Microbiota Profiles Due to DNA Extraction Method and Inter-Subject Differences. *Front Microbiol* **6** (2015).
203 Costea, P. I. et al. Towards Standards for Human Fecal Sample Processing in Metagenomic Studies. *Nat Biotechnol* **35**, 1069-1076 (2017).
204 Salonen, A. et al. Comparative Analysis of Fecal DNA Extraction Methods with Phylogenetic Microarray: Effective Recovery of Bacterial and Archaeal DNA Using Mechanical Cell Lysis. *J Microbiol Meth* **81**, 127-134 (2010).
205 Claassen, S. et al. A Comparison of the Efficiency of Five Different Commercial DNA Extraction Kits for Extraction of DNA from Faecal Samples. *J Microbiol Methods* **94**, 103-110 (2013).
206 Fiedorova, K. et al. The Impact of DNA Extraction Methods on Stool Bacterial and Fungal Microbiota Community Recovery. *Frontiers in microbiology* **10**, 821 (2019).
207 Vebo, H. C., Karlsson, M. K., Avershina, E., Finnby, L. & Rudi, K. Bead-Beating Artefacts in the Bacteroidetes to Firmicutes Ratio of the Human Stool Metagenome. *J Microbiol Meth* **129**, 78-80 (2016).
208 Thomas, V., Clark, J. & Doré, J. Fecal Microbiota Analysis: An Overview of Sample Collection Methods and Sequencing Strategies. *Future microbiology* **10**, 1485-1504 (2015).
209 Lim, M. Y., Song, E.-J., Kim, S. H., Lee, J. & Nam, Y.-D. Comparison of DNA Extraction Methods for Human Gut Microbial Community Profiling. *Systematic and Applied Microbiology* **41**, 151-157 (2018).
210 Hamady, M. & Knight, R. Microbial Community Profiling for Human Microbiome Projects: Tools, Techniques, and Challenges. *Genome Res* **19**, 1141-1152 (2009).

211 Frueh, F. W. & Noyer-Weidner, M. The Use of Denaturing High-Performance Liquid Chromatography (Dhplc) for the Analysis of Genetic Variations: Impact for Diagnostics and Pharmacogenetics. *Clin Chem Lab Med* **41**, 452-461 (2003).
212 Lee, P. Y., Costumbrado, J., Hsu, C.-Y. & Kim, Y. H. Agarose Gel Electrophoresis for the Separation of DNA Fragments. *J Vis Exp*, 3923 (2012).
213 Viglasky, V. Polyacrylamide Temperature Gradient Gel Electrophoresis. *Methods in molecular biology (Clifton, NJ)* **1054**, 159-171 (2013).
214 Muyzer, G. & Smalla, K. Application of Denaturing Gradient Gel Electrophoresis (Dgge) and Temperature Gradient Gel Electrophoresis (Tgge) in Microbial Ecology. *Antonie van Leeuwenhoek* **73**, 127-141 (1998).
215 Liesack, W. & Dunfield, P. F. in *Environmental Microbiology: Methods and Protocols* (Walker, J. M., Spencer, J. F. T. & Ragout de Spencer, A. L. eds.) 23-37 (Humana Press, 2004).
216 Osborn, A. M., Moore, E. R. B. & Timmis, K. N. An Evaluation of Terminal-Restriction Fragment Length Polymorphism (T-Rflp) Analysis for the Study of Microbial Community Structure and Dynamics. *Environ Microbiol* **2**, 39-50 (2000).
217 Bokulich, N. A. & Mills, D. A. Differentiation of Mixed Lactic Acid Bacteria Communities in Beverage Fermentations Using Targeted Terminal Restriction Fragment Length Polymorphism. *Food Microbiol* **31**, 126-132 (2012).
218 O'Donovan, M. C. et al. Blind Analysis of Denaturing High-Performance Liquid Chromatography as a Tool for Mutation Detection. *Genomics* **52**, 44-49 (1998).
219 Xiao, W. & Oefner, P. J. Denaturing High-Performance Liquid Chromatography: A Review. *Hum Mutat* **17**, 439-474 (2001).
220 Hannachi-M'Zali, F., Ambler, J. E., Taylor, C. F. & Hawkey, P. M. Examination of Single and Multiple Mutations Involved in Resistance to Quinolones in Staphylococcus Aureus by a Combination of Pcr and Denaturing High-Performance Liquid Chromatography (Dhplc). *Journal of Antimicrobial Chemotherapy* **50**, 649-655 (2002).
221 Domann, E. et al. Culture-Independent Identification of Pathogenic Bacteria and Polymicrobial Infections in the Genitourinary Tract of Renal Transplant Recipients. *J Clin Microbiol* **41**, 5500-5510 (2003).
222 Lueders, T. & Friedrich, M. W. Evaluation of Pcr Amplification Bias by Terminal Restriction Fragment Length Polymorphism Analysis of Small-Subunit Rrna and ≪Em≫Mcra≪/Em≫ Genes by Using Defined Template Mixtures of Methanogenic Pure Cultures and Soil DNA Extracts. *Appl Environ Microb* **69**, 320 (2003).
223 Hamady, M. & Knight, R. Microbial Community Profiling for Human Microbiome Projects: Tools, Techniques, and Challenges. *Genome Research* **19**, 1141-1152 (2009).
224 Engelbrektson, A. et al. Experimental Factors Affecting Pcr-Based Estimates of Microbial Species Richness and Evenness. *ISME J* **4**, 642-647 (2010).
225 Fierer, N. & Jackson, R. B. The Diversity and Biogeography of Soil Bacterial Communities. *Proc Natl Acad Sci U S A* **103**, 626-631 (2006).
226 Paliy, O. & Agans, R. Application of Phylogenetic Microarrays to Interrogation of Human Microbiota. *Fems Microbiol Ecol* **79**, 2-11 (2012).
227 Govindarajan, R., Duraiyan, J., Kaliyappan, K. & Palanisamy, M. Microarray and Its Applications. *J Pharm Bioallied Sci* **4**, S310-S312 (2012).
228 in *The Human Microbiota* 75-84.
229 Rajilic-Stojanovic, M. et al. Development and Application of the Human Intestinal Tract Chip, a Phylogenetic Microarray: Analysis of Universally Conserved

Phylotypes in the Abundant Microbiota of Young and Elderly Adults. *Environ Microbiol* **11**, 1736-1751 (2009).
230 Wilson, K. H. et al. High-Density Microarray of Small-Subunit Ribosomal DNA Probes. *Appl Environ Microbiol* **68**, 2535-2541 (2002).
231 He, Z. et al. Geochip: A Comprehensive Microarray for Investigating Biogeochemical, Ecological and Environmental Processes. *ISME J* **1**, 67-77 (2007).
232 Claesson, M. J. et al. Comparative Analysis of Pyrosequencing and a Phylogenetic Microarray for Exploring Microbial Community Structures in the Human Distal Intestine. *PLoS One* **4**, 6669 (2009).
233 van den Bogert, B., de Vos, W. M., Zoetendal, E. G. & Kleerebezem, M. Microarray Analysis and Barcoded Pyrosequencing Provide Consistent Microbial Profiles Depending on the Source of Human Intestinal Samples. *Appl Environ Microbiol* **77**, 2071-2080 (2011).
234 DeSantis, T. Z. et al. High-Density Universal 16s Rrna Microarray Analysis Reveals Broader Diversity Than Typical Clone Library When Sampling the Environment. *Microb Ecol* **53**, 371-383 (2007).
235 Deepak, S. et al. Real-Time Pcr: Revolutionizing Detection and Expression Analysis of Genes. *Curr Genomics* **8**, 234-251 (2007).
236 Wong, W., Farr, R., Joglekar, M., Januszewski, A. & Hardikar, A. Probe-Based Real-Time Pcr Approaches for Quantitative Measurement of Micrornas. *J Vis Exp*, 52586 (2015).
237 Wallner, G., Amann, R. & Beisker, W. Optimizing Fluorescent in Situ Hybridization with Rrna-Targeted Oligonucleotide Probes for Flow Cytometric Identification of Microorganisms. *Cytometry* **14**, 136-143 (1993).
238 Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A. & Korpela, K. Quantitative Pcr Provides a Simple and Accessible Method for Quantitative Microbiota Profiling. *PloS one* **15**, 0227285-0227285 (2020).
239 Vázquez, L., Guadamuro, L., Giganto, F., Mayo, B. & Flórez, A. B. Development and Use of a Real-Time Quantitative Pcr Method for Detecting and Quantifying Equol-Producing Bacteria in Human Faecal Samples and Slurry Cultures. *Front Microbiol* **8**, 1155-1155 (2017).
240 Sanger, F. & Coulson, A. R. A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase. *Journal of Molecular Biology* **94**, 441-448 (1975).
241 Churko, J. M., Mantalas, G. L., Snyder, M. P. & Wu, J. C. Overview of High Throughput Sequencing Technologies to Elucidate Molecular Pathways in Cardiovascular Diseases. *Circ Res* **112**, 1613-1623 (2013).
242 Liu, Z., DeSantis, T. Z., Andersen, G. L. & Knight, R. Accurate Taxonomy Assignments from 16s Rrna Sequences Produced by Highly Parallel Pyrosequencers. *Nucleic Acids Res* **36**, 120 (2008).
243 Liu, Z., Lozupone, C., Hamady, M., Bushman, F. D. & Knight, R. Short Pyrosequencing Reads Suffice for Accurate Microbial Community Analysis. *Nucleic Acids Res* **35**, 120 (2007).
244 Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian Classifier for Rapid Assignment of Rrna Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol* **73**, 5261-5267 (2007).
245 Almeida, A., Mitchell, A. L., Tarkowska, A. & Finn, R. D. Benchmarking Taxonomic Assignments Based on 16s Rrna Gene Profiling of the Microbiota from Commonly Sampled Environments. *Gigascience* **7** (2018).

246 Schloss, P. D., Gevers, D. & Westcott, S. L. Reducing the Effects of Pcr Amplification and Sequencing Artifacts on 16s Rrna-Based Studies. *PLoS One* **6**, 27310 (2011).
247 Klindworth, A. et al. Evaluation of General 16s Ribosomal Rna Gene Pcr Primers for Classical and Next-Generation Sequencing-Based Diversity Studies. *Nucleic Acids Res* **41**, 1 (2013).
248 Tremblay, J. et al. Primer and Platform Effects on 16s Rrna Tag Sequencing. *Front Microbiol* **6** (2015).
249 Baker, G. C., Smith, J. J. & Cowan, D. A. Review and Re-Analysis of Domain-Specific 16s Primers. *J Microbiol Methods* **55**, 541-555 (2003).
250 Wang, Y. & Qian, P. Y. Conservative Fragments in Bacterial 16s Rrna Genes and Primer Design for 16s Ribosomal DNA Amplicons in Metagenomic Studies. *PLoS One* **4**, 7401 (2009).
251 Tringe, S. G. & Hugenholtz, P. A Renaissance for the Pioneering 16s Rrna Gene. *Curr Opin Microbiol* **11**, 442-446 (2008).
252 Sim, K. et al. Improved Detection of Bifidobacteria with Optimised 16s Rrna-Gene Based Pyrosequencing. *PLoS One* **7**, 32543 (2012).
253 Johnson, J. S. et al. Evaluation of 16s Rrna Gene Sequencing for Species and Strain-Level Microbiome Analysis. *Nature Communications* **10** (2019).
254 Yang, B., Wang, Y. & Qian, P. Y. Sensitivity and Correlation of Hypervariable Regions in 16s Rrna Genes in Phylogenetic Analysis. *BMC Bioinformatics* **17**, 135 (2016).
255 Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the Miseq Illumina Sequencing Platform. *Appl Environ Microbiol* **79**, 5112-5120 (2013).
256 Rausch, P. et al. Comparative Analysis of Amplicon and Metagenomic Sequencing Methods Reveals Key Features in the Evolution of Animal Metaorganisms. *Microbiome* **7**, 133 (2019).
257 Thompson, L. R. et al. A Communal Catalogue Reveals Earth's Multiscale Microbial Diversity. *Nature* **551**, 457-463 (2017).
258 Luo, C., Tsementzi, D., Kyrpides, N., Read, T. & Konstantinidis, K. T. Direct Comparisons of Illumina Vs. Roche 454 Sequencing Technologies on the Same Microbial Community DNA Sample. *PLOS ONE* **7**, 30087 (2012).
259 Margulies, M. et al. Genome Sequencing in Microfabricated High-Density Picolitre Reactors. *Nature* **437**, 376-380 (2005).
260 Quince, C. et al. Accurate Determination of Microbial Diversity from 454 Pyrosequencing Data. *Nat Methods* **6**, 639-641 (2009).
261 Loman, N. J. et al. Performance Comparison of Benchtop High-Throughput Sequencing Platforms. *Nat Biotechnol* **30**, 434 (2012).
262 Gomez-Alvarez, V., Teal, T. K. & Schmidt, T. M. Systematic Artifacts in Metagenomes from Complex Microbial Communities. *The ISME journal* **3**, 1314-1317 (2009).
263 Lam, H. Y. K. et al. Performance Comparison of Whole-Genome Sequencing Platforms. *Nat Biotechnol* **30**, 78-82 (2012).
264 Quail, M. A. et al. A Tale of Three Next Generation Sequencing Platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina Miseq Sequencers. *Bmc Genomics* **13**, 341 (2012).

265 Salipante, S. J. et al. Performance Comparison of Illumina and Ion Torrent Next-Generation Sequencing Platforms for 16s Rrna-Based Bacterial Community Profiling. *Appl Environ Microbiol* **80**, 7583-7591 (2014).
266 Erlich, Y., Mitra, P. P., delaBastide, M., McCombie, W. R. & Hannon, G. J. Alta-Cyclic: A Self-Optimizing Base Caller for Next-Generation Sequencing. *Nat Methods* **5**, 679-682 (2008).
267 Dolan, P. C. & Denver, D. R. Tileqc: A System for Tile-Based Quality Control of Solexa Data. *BMC Bioinformatics* **9**, 250 (2008).
268 Nakamura, K. et al. Sequence-Specific Error Profile of Illumina Sequencers. *Nucleic Acids Res* **39**, 90 (2011).
269 Schröder, J., Bailey, J., Conway, T. & Zobel, J. Reference-Free Validation of Short Read Data. *PLOS ONE* **5**, e12681 (2010).
270 Caporaso, J. G. K., J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K. . *Qiime Allows Analysis of High-Throughput Community Sequencing Data* 335-336 (Nat Methods, 2010).
271 Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis. *The ISME journal* **11**, 2639-2643 (2017).
272 Matias Rodrigues, J. F., Schmidt, T. S. B., Tackmann, J. & von Mering, C. Mapseq: Highly Efficient K-Mer Search with Confidence Estimates, for Rrna Sequence Analysis. *Bioinformatics* **33**, 3808-3810 (2017).
273 Edgar, R. C. Search and Clustering Orders of Magnitude Faster Than Blast. *Bioinformatics* **26**, 2460-2461 (2010).
274 Prodan, A. et al. Comparing Bioinformatic Pipelines for Microbial 16s Rrna Amplicon Sequencing. *PloS one* **15**, 0227434-0227434 (2020).
275 Westcott, S. L. & Schloss, P. D. De Novo Clustering Methods Outperform Reference-Based Methods for Assigning 16s Rrna Gene Sequences to Operational Taxonomic Units. *PeerJ* **3**, 1487-1487 (2015).
276 Rideout, J. R. et al. Subsampled Open-Reference Clustering Creates Consistent, Comprehensive Otu Definitions and Scales to Billions of Sequences. *PeerJ* **2**, 545 (2014).
277 Callahan, B. J. et al. Dada2: High Resolution Sample Inference from Amplicon Data. *bioRxiv*, 024034 (2015).
278 Amir, A. et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems* **2**, e00191-00116 (2017).
279 Nearing, J. T., Douglas, G. M., Comeau, A. M. & Langille, M. G. I. Denoising the Denoisers: An Independent Evaluation of Microbiome Sequence Error-Correction Approaches. *PeerJ* **6**, e5364 (2018).
280 Schloss, P. D. & Handelsman, J. Introducing Dotur, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. *Appl Environ Microb* **71**, 1501-1506 (2005).
281 Edgar, R. C. Updating the 97% Identity Threshold for 16s Ribosomal Rna Otus. *Bioinformatics* **34**, 2371-2375 (2018).
282 Balvočiūtė, M. & Huson, D. H. Silva, Rdp, Greengenes, Ncbi and Ott - How Do These Taxonomies Compare? *Bmc Genomics* **18**, 114 (2017).
283 Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B. & Chiodini, R. J. Inherent Bacterial DNA Contamination of Extraction and Sequencing Reagents May Affect Interpretation of Microbiota in Low Bacterial Biomass Samples. *Gut Pathogens* **8**, 24 (2016).

284 Salter, S. J. et al. Reagent and Laboratory Contamination Can Critically Impact Sequence-Based Microbiome Analyses. *BMC Biol* **12**, 87-87 (2014).
285 McKnight, D. et al. Microdecon: A Highly Accurate Read-Subtraction Tool for the Post-Sequencing Removal of Contamination in Metabarcoding Studies. *Environmental DNA*, 14:25 (2019).
286 Willis, A. D. Rarefaction, Alpha Diversity, and Statistics. *Front Microbiol* **10** (2019).
287 Chao, A. Nonparametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics* **11**, 265-270 (1984).
288 Vandeputte, D. et al. Quantitative Microbiome Profiling Links Gut Community Variation to Microbial Load. *Nature* **551**, 507-511 (2017).
289 Lande, R., DeVries, P. J. & Walla, T. R. When Species Accumulation Curves Intersect: Implications for Ranking Diversity Using Small Samples. *Oikos* **89**, 601-605 (2000).
290 Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of Statistical Methods for Normalization and Differential Expression in Mrna-Seq Experiments. *BMC Bioinformatics* **11**, 94 (2010).
291 Dillies, M. A. et al. A Comprehensive Evaluation of Normalization Methods for Illumina High-Throughput Rna Sequencing Data Analysis. *Brief Bioinform* **14**, 671-683 (2013).
292 McMurdie, P. J. & Holmes, S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLOS Computational Biology* **10**, 1003531 (2014).
293 Weiss, S. et al. Normalization and Microbial Differential Abundance Strategies Depend Upon Data Characteristics. *Microbiome* **5**, 27 (2017).
294 Love, M. I., Huber, W. & Anders, S. Moderated Estimation of Fold Change and Dispersion for Rna-Seq Data with Deseq2. *Genome Biology* **15**, 550 (2014).
295 Paulson, J. N., Stine, O. C., Bravo, H. C. & Pop, M. Differential Abundance Analysis for Microbial Marker-Gene Surveys. *Nature methods* **10**, 1200-1202 (2013).
296 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. Edger: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **26**, 139-140 (2010).
297 McKnight, D. T. et al. Methods for Normalizing Microbiome Data: An Ecological Perspective. *Methods in Ecology and Evolution* **10**, 389-400 (2019).
298 Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun Metagenomics, from Sampling to Analysis. *Nat Biotechnol* **35**, 833-844 (2017).
299 Wesolowska-Andersen, A. et al. Choice of Bacterial DNA Extraction Method from Fecal Material Influences Community Structure as Evaluated by Metagenomic Analysis. *Microbiome* **2** (2014).
300 Tanner, M. A., Goebel, B. M., Dojka, M. A. & Pace, N. R. Specific Ribosomal DNA Sequences from Diverse Environmental Settings Correlate with Experimental Contaminants. *Appl Environ Microb* **64**, 3110-3113 (1998).
301 Yuan, S. Q., Cohen, D. B., Ravel, J., Abdo, Z. & Forney, L. J. Evaluation of Methods for the Extraction and Purification of DNA from the Human Microbiome. *Plos One* **7** (2012).
302 Nelson, M. C., Morrison, H. G., Benjamino, J., Grim, S. L. & Graf, J. Analysis, Optimization and Verification of Illumina-Generated 16s Rrna Gene Amplicon Surveys. *Plos One* **9** (2014).
303 Baym, M. et al. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes (Vol 10, E0128036, 2015). *Plos One* **10** (2015).
304 Wood, D. E., Lu, J. & Langmead, B. Improved Metagenomic Analysis with Kraken 2. *Genome Biology* **20**, 257 (2019).

305 Kim, D., Song, L., Breitwieser, F. P. & Salzberg, S. L. Centrifuge: Rapid and Sensitive Classification of Metagenomic Sequences. *Genome Res* **26**, 1721-1729 (2016).
306 Milanese, A. et al. Microbial Abundance, Activity and Population Genomic Profiling with Motus2. *Nature Communications* **10**, 1014 (2019).
307 Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. Meta-Idba: A De Novo Assembler for Metagenomic Data. *Bioinformatics (Oxford, England)* **27**, i94-i101 (2011).
308 Li, R. et al. De Novo Assembly of Human Genomes with Massively Parallel Short Read Sequencing. *Genome Res* **20**, 265-272 (2010).
309 Salipante, S. J. et al. Rapid 16s Rrna Next-Generation Sequencing of Polymicrobial Clinical Samples for Diagnosis of Complex Bacterial Infections. *PLoS One* **8**, 65226 (2013).
310 Sunagawa, S. et al. Metagenomic Species Profiling Using Universal Phylogenetic Marker Genes. *Nature Methods* **10**, 1196-1199 (2013).
311 Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore Minion: Delivery of Nanopore Sequencing to the Genomics Community. *Genome Biology* **17**, 239 (2016).
312 Hebert, P. D. N. et al. A Sequel to Sanger: Amplicon Sequencing That Scales. *Bmc Genomics* **19** (2018).
313 Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of Age: Ten Years of Next-Generation Sequencing Technologies. *Nat Rev Genet* **17**, 333-351 (2016).
314 Levy, S. E. & Myers, R. M. Advancements in Next-Generation Sequencing. *Annu Rev Genomics Hum Genet* **17**, 95-115 (2016).
315 Payne, A., Holmes, N., Rakyan, V. & Loose, M. Whale Watching with Bulkvis: A Graphical Viewer for Oxford Nanopore Bulk Fast5 Files. *bioRxiv*, 312256 (2018).
316 Callahan, B. J. et al. High-Throughput Amplicon Sequencing of the Full-Length 16s Rrna Gene with Single-Nucleotide Resolution. *Nucleic Acids Research* **47** (2019).
317 Jain, M. et al. Improved Data Analysis for the Minion Nanopore Sequencer. *Nat Methods* **12**, 351-356 (2015).
318 Jain, M. et al. Nanopore Sequencing and Assembly of a Human Genome with Ultra-Long Reads. *Nat Biotechnol* **36**, 338-345 (2018).
319 Carneiro, M. O. et al. Pacific Biosciences Sequencing Technology for Genotyping and Variation Discovery in Human Data. *Bmc Genomics* **13**, 375 (2012).
320 Rang, F. J., Kloosterman, W. P. & de Ridder, J. From Squiggle to Basepair: Computational Approaches for Improving Nanopore Sequencing Read Accuracy. *Genome Biology* **19**, 90 (2018).
321 Shreiner, A. B., Kao, J. Y. & Young, V. B. The Gut Microbiome in Health and in Disease. *Curr Opin Gastroenterol* **31**, 69-75 (2015).
322 Dahl, C. et al. Preterm Infants Have Distinct Microbiomes Not Explained by Mode of Delivery, Breastfeeding Duration or Antibiotic Exposure. *International Journal of Epidemiology* **47**, 1658-1669 (2018).
323 Stewart, C. J. et al. Temporal Development of the Gut Microbiome in Early Childhood from the Teddy Study. *Nature* **562**, 583-588 (2018).
324 Obermajer, T. et al. Microbes in Infant Gut Development: Placing Abundance within Environmental, Clinical and Growth Parameters. *Sci Rep-Uk* **7** (2017).
325 Kalliomaki, M., Collado, M. C., Salminen, S. & Isolauri, E. Early Differences in Fecal Microbiota Composition in Children May Predict Overweight. *Am J Clin Nutr* **87**, 534-538 (2008).

326 Dinan, T. G. & Cryan, J. F. Gut Instincts: Microbiota as a Key Regulator of Brain Development, Ageing and Neurodegeneration. *J Physiol* (2016).
327 Torrazza, R. M. & Neu, J. The Altered Gut Microbiome and Necrotizing Enterocolitis. *Clinics in Perinatology* **40**, 93-108 (2013).
328 Soleimani, F., Zaheri, F. & Abdi, F. Long-Term Neurodevelopmental Outcomes after Preterm Birth. *Iran Red Crescent Med J* **16**, 17965-17965 (2014).
329 Queensland Health. The Health of Queenslanders 2018. (Brisbane, 2018).
330 Evidence Based Probiotics. *Infloran*, <https://www.infloran.com.au/?gclid=CjwKCAiA-_L9BRBQEiwA-bm5fjBoxiUHkDF7r40k4SgIjF7M_MDTTVue4HDOB6QFbsX1XD_WgJICshoCPY8QAvD_BwE> (2019).
331 Meridian. *Meridian Bioscience*, <https://www.bioline.com/> (2020).
332 Illumina Inc. *16s Metagenomic Sequencing Library Preparation*, <https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf> (2018).
333 Illumina Inc. *Illumina*, <https://www.illumina.com/index-d.html> (2020).
334 ThermoFisher Scientific. *Thermofisher Scientific*, <https://www.google.com/search?q=platinum+superfi+pcr+master+mix&oq=platinum+superfi&aqs=chrome.1.69i57j0l7.3863j0j4&sourceid=chrome&ie=UTF-8> (2020).
335 Rstudio: Integrated Development for R. Rstudio (PBC, Boston, MA 2020).
336 Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J. & Holmes, S. P. Bioconductor Workflow for Microbiome Data Analysis: From Raw Reads to Community Analyses. *F1000Research* **5**, 1492 (2016).
337 Callahan, B. J. et al. Dada2: High-Resolution Sample Inference from Illumina Amplicon Data. *Nature methods* **13**, 581-583 (2016).
338 McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, 61217 (2013).
339 Vegan: Community Ecology Package. v. 2.5-7 (2020).
340 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57**, 289-300 (1995).
341 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using Lme4. *2015* **67**, 48 (2015).
342 Alain F. Zuur, E. N. L., Neil Walker, Anatoly A. Saveliev, Graham M. Smith. *Mixed Effects Models and Extensions in Ecology with R* (Springer Nature, 2009).
343 Fox, J. & Weisberg, S. *An R Companion to Applied Regression* Third edn(2019).
344 Searle, S. R., Speed, F. M. & Milliken, G. A. Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. *The American Statistician* **34**, 216-221 (1980).
345 Mackie, R. I., Sghir, A. & Gaskins, H. R. Developmental Microbial Ecology of the Neonatal Gastrointestinal Tract. *Am J Clin Nutr* **69**, 1035S-1045S (1999).
346 Bennet, R. & Nord, C. E. Development of the Faecal Anaerobic Microflora after Caesarean Section and Treatment with Antibiotics in Newborn Infants. *Infection* **15**, 332-336 (1987).
347 Korpela, K. et al. Probiotic Supplementation Restores Normal Microbiota Composition and Function in Antibiotic-Treated and in Caesarean-Born Infants. *Microbiome* **6**, 182 (2018).

348 Díaz-Ropero, M. P. et al. Two Lactobacillus Strains, Isolated from Breast Milk, Differently Modulate the Immune Response. *Journal of applied microbiology* **102**, 337-343 (2007).
349 Martín, R. et al. Isolation of Bifidobacteria from Breast Milk and Assessment of the Bifidobacterial Population by Pcr-Denaturing Gradient Gel Electrophoresis and Quantitative Real-Time Pcr. *Appl Environ Microbiol* **75**, 965-969 (2009).
350 Puri, K. et al. Association of Chorioamnionitis with Aberrant Neonatal Gut Colonization and Adverse Clinical Outcomes. *PloS one* **11**, 0162734-0162734 (2016).
351 Davies, E. L., Bell, J. S. & Bhattacharya, S. Preeclampsia and Preterm Delivery: A Population-Based Case–Control Study. *Hypertension in Pregnancy* **35**, 510-519 (2016).
352 Ishimwe, J. A. Maternal Microbiome in Preeclampsia Pathophysiology and Implications on Offspring Health. *Physiological reports* **9**, 14875 (2021).
353 Choi, M. S., Yu, J. S., Yoo, H. H. & Kim, D. H. The Role of Gut Microbiota in the Pharmacokinetics of Antihypertensive Drugs. *Pharmacological research* **130**, 164-171 (2018).
354 Skondra, D. et al. The Early Gut Microbiome Could Protect against Severe Retinopathy of Prematurity. *Journal of American Association for Pediatric Ophthalmology and Strabismus* (2020).
355 Meyer, M. P. et al. Probiotics for Prevention of Severe Necrotizing Enterocolitis: Experience of New Zealand Neonatal Intensive Care Units. *Front Pediatr* **8**, 119 (2020).
356 Henry, M. C. W. & Moss, R. L. Neonatal Necrotizing Enterocolits. *Seminars in Pediatric Surgery* **17**, 98-109 (2008).
357 Korpela, K. & de Vos, W. M. Early Life Colonization of the Human Gut: Microbes Matter Everywhere. *Current Opinion in Microbiology* **44**, 70-78 (2018).
358 Rastelli, M., Knauf, C. & Cani, P. D. Gut Microbes and Health: A Focus on the Mechanisms Linking Microbes, Obesity, and Related Disorders. *Obesity (Silver Spring)* **26**, 792-800 (2018).
359 Administration, T. T. G. *Australian Register of Therapeutic Goods*, <https://www.tga.gov.au/sites/default/files/foi-276-1213-02.pdf> (2013).
360 Sharma, R. & Hudak, M. L. A Clinical Perspective of Necrotizing Enterocolitis: Past, Present, and Future. *Clinics in Perinatology* **40**, 27-51 (2013).
361 Choi, Y. Y. Necrotizing Enterocolitis in Newborns: Update in Pathophysiology and Newly Emerging Therapeutic Strategies. *Clin Exp Pediatr* **57**, 505-513 (2014).
362 Donovan H. Parks, F. R., Lutz Krause, Kaylyn Tousignant, Alena L. Pribyl, Philip Hugenholtz, Gene W. Tyson, David L.A. Wood. Microba's Community Profiler Enables Precise Measurement of the Gut Microbiome. (Microba Life Sciences, Microba Life Sciences, 2020).
363 Grech, A. et al. Maternal Exposures and the Infant Gut Microbiome: A Systematic Review with Meta-Analysis. *Gut microbes* **13**, 1-30 (2021).
364 Hu, J. et al. Diversified Microbiota of Meconium Is Affected by Maternal Diabetes Status. *PLoS One* **8**, 78257 (2013).
365 Wang, J. et al. Dysbiosis of Maternal and Neonatal Microbiota Associated with Gestational Diabetes Mellitus. *Gut* **67**, 1614-1625 (2018).
366 Gohir, W., Ratcliffe, E. M. & Sloboda, D. M. Of the Bugs That Shape Us: Maternal Obesity, the Gut Microbiome, and Long-Term Disease Risk. *Pediatr Res* **77**, 196-204 (2015).
367 Dixon, P. Vegan, a Package of R Functions for Community Ecology. *Journal of Vegetation Science* **14**, 927-930 (2003).

368 Martí, M. et al. Effects of Lactobacillus Reuteri Supplementation on the Gut Microbiota in Extremely Preterm Infants in a Randomized Placebo-Controlled Trial. *Cell Rep Med* **2**, 100206-100206 (2021).
369 Grazul, H., Kanda, L. L. & Gondek, D. Impact of Probiotic Supplements on Microbiome Diversity Following Antibiotic Treatment of Mice. *Gut microbes* **7**, 101-114 (2016).
370 Yatsunenko, T. et al. Human Gut Microbiome Viewed across Age and Geography. *Nature* **486**, 222 (2012).
371 Jeffery, I. B., Lynch, D. B. & O'Toole, P. W. Composition and Temporal Stability of the Gut Microbiota in Older Persons. *The ISME journal* **10**, 170-182 (2016).
372 Koenig, J. E. et al. Succession of Microbial Consortia in the Developing Infant Gut Microbiome. *Proceedings of the National Academy of Sciences of the United States of America* **108 Suppl 1**, 4578-4585 (2011).
373 Roger, L. C., Costabile, A., Holland, D. T., Hoyles, L. & McCartney, A. L. Examination of Faecal Bifidobacterium Populations in Breast- and Formula-Fed Infants During the First 18 Months of Life. *Microbiology* **156**, 3329-3341 (2010).
374 Ruiz, L., Delgado, S., Ruas-Madiedo, P., Sánchez, B. & Margolles, A. Bifidobacteria and Their Molecular Communication with the Immune System. *Frontiers in microbiology* **8**, 2345 (2017).
375 Henrick, B. M. et al. Bifidobacteria-Mediated Immune System Imprinting Early in Life. *Cell* **184**, 3884-3898 (2021).
376 Yousuf, E. I. et al. Persistence of Suspected Probiotic Organisms in Preterm Infant Gut Microbiota Weeks after Probiotic Supplementation in the Nicu. *Frontiers in microbiology* **11**, 574137 (2020).
377 Karnati, S., Kollikonda, S. & Abu-Shaweesh, J. Late Preterm Infants - Changing Trends and Continuing Challenges. *Int J Pediatr Adolesc Med* **7**, 36-44 (2020).
378 Gensollen, T., Iyer, S. S., Kasper, D. L. & Blumberg, R. S. How Colonization by Microbiota in Early Life Shapes the Immune System. *Science* **352**, 539-544 (2016).
379 Effah, C. Y., Sun, T., Liu, S. & Wu, Y. Klebsiella Pneumoniae: An Increasing Threat to Public Health. *Annals of Clinical Microbiology and Antimicrobials* **19**, 1 (2020).
380 Westaway, J. A. F. et al. The Bacterial Gut Microbiome of Probiotic-Treated Very-Preterm Infants: Changes from Admission to Discharge. *Pediatr Res* (2021).
381 Deshpande, G. C., Rao, S. C., Keil, A. D. & Patole, S. K. Evidence-Based Guidelines for Use of Probiotics in Preterm Neonates. *Bmc Med* **9** (2011).
382 Collins, A., Weitkamp, J. H. & Wynn, J. L. Why Are Preterm Newborns at Increased Risk of Infection? *Arch Dis Child Fetal Neonatal Ed* **103**, 391-394 (2018).
383 Zhang, J. et al. Is Preterm Birth Associated with Asthma among Children from Birth to 17 Years Old? -a Study Based on 2011-2012 Us National Survey of Children's Health. *Italian journal of pediatrics* **44**, 151 (2018).
384 Crump, C., Sundquist, J. & Sundquist, K. Preterm Birth and Risk of Type 1 and Type 2 Diabetes: A National Cohort Study. *Diabetologia* **63**, 508-518 (2020).
385 DNA Genotek Inc. Omnigene Gut, for Microbiome. Report No. Report, 1 (Inc., D. G., 2019).
386 Szopinska, J. W. et al. Reliability of a Participant-Friendly Fecal Collection Method for Microbiome Analyses: A Step Towards Large Sample Size Investigation. *BMC Microbiol* **18**, 110 (2018).
387 World Health Organisation. *Preterm Birth*, <https://www.who.int/news-room/fact-sheets/detail/preterm-birth#:~:text=Preterm%20is%20defined%20as%20babies%20born%20alive%20befor

e,planned%20before%2039%20completed%20weeks%20unless%20medically%20indicated.> (2018).
388 Jacquot, A. et al. Dynamics and Clinical Evolution of Bacterial Gut Microflora in Extremely Premature Patients. *J Pediatr* **158**, 390-396 (2011).
389 Raheem, A., Liang, L., Zhang, G. & Cui, S. Modulatory Effects of Probiotics During Pathogenic Infections with Emphasis on Immune Regulation. *Front Immunol* **12** (2021).
390 Zuo, F., Appaswamy, A., Gebremariam, H. G. & Jonsson, A.-B. Role of Sortase a in Lactobacillus Gasseri Kx110a1 Adhesion to Gastric Epithelial Cells and Competitive Exclusion of Helicobacter Pylori. *Front Microbiol* **10** (2019).
391 Saitoh, T. et al. A20 Is a Negative Regulator of Ifn Regulatory Factor 3 Signaling. *The Journal of Immunology* **174**, 1507 (2005).
392 Pei, R., Martin, D. A., DiMarco, D. M. & Bolling, B. W. Evidence for the Effects of Yogurt on Gut Health and Obesity. *Critical Reviews in Food Science and Nutrition* **57**, 1569-1583 (2017).
393 Saliganti, V., Kapila, R. & Kapila, S. Consumption of Probiotic Lactobacillus Rhamnosus (Mtcc: 5897) Containing Fermented Milk Plays a Key Role in Development of the Immune System in Newborn Mice During the Suckling–Weaning Transition. *Microbiology and Immunology* **60**, 261-267 (2016).
394 van Zyl, W. F., Deane, S. M. & Dicks, L. M. T. Molecular Insights into Probiotic Mechanisms of Action Employed against Intestinal Pathogenic Bacteria. *Gut microbes* **12**, 1831339 (2020).
395 Westaway, J. A. F. et al. To Probiotic or Not to Probiotic: A Metagenomic Comparison of the Discharge Gut Microbiome of Infants Supplemented with Probiotics in Nicu and Those Who Are Not. *Front Pediatr* **10** (2022).
396 Li, Y. et al. Effects of Bifidobacterium Breve Supplementation on Intestinal Flora of Low Birth Weight Infants. *Pediatr Int* **46**, 509-515 (2004).
397 Mohan, R. et al. Effects of Bifidobacterium Lactis Bb12 Supplementation on Intestinal Microbiota of Preterm Infants: A Double-Blind, Placebo-Controlled, Randomized Study. *J Clin Microbiol* **44**, 4025-4031 (2006).
398 Plummer, E. L. et al. Gut Microbiota of Preterm Infants Supplemented with Probiotics: Sub-Study of the Proprems Trial. *BMC Microbiol* **18**, 184 (2018).
399 Horigome, A. et al. Colonization of Supplemented Bifidobacterium Breve M-16v in Low Birth Weight Infants and Its Effects on Their Gut Microbiota Weeks Post-Administration. *Frontiers in Microbiology* **12** (2021).
400 Dogra, S. et al. Dynamics of Infant Gut Microbiota Are Influenced by Delivery Mode and Gestational Duration and Are Associated with Subsequent Adiposity. *Mbio* **6** (2015). <http://europepmc.org/abstract/MED/25650398>.
401 Su, P., Henriksson, A. & Mitchell, H. Prebiotics Enhance Survival and Prolong the Retention Period of Specific Probiotic Inocula in an in Vivo Murine Model. *Journal of applied microbiology* **103**, 2392-2400 (2007).
402 Panigrahi, P. et al. Long-Term Colonization of a Lactobacillus Plantarum Synbiotic Preparation in the Neonatal Gut. *J Pediatr Gastroenterol Nutr* **47**, 45-53 (2008).
403 Durack, J. et al. Delayed Gut Microbiota Development in High-Risk for Asthma Infants Is Temporarily Modifiable by Lactobacillus Supplementation. *Nat Commun* **9**, 707 (2018).
404 Lopetuso, L. R., Scaldaferri, F., Petito, V. & Gasbarrini, A. Commensal Clostridia: Leading Players in the Maintenance of Gut Homeostasis. *Gut Pathog* **5**, 23 (2013).
405 Ruuskanen, M. O. et al. Links between Gut Microbiome Composition and Fatty Liver Disease in a Large Population Sample. *Gut microbes* **13**, 1-22 (2021).

406 Ogita, T. et al. Oral Administration of Flavonifractor Plautii Strongly Suppresses Th2 Immune Responses in Mice. *Front Immunol* **11**, 379 (2020).
407 Gupta, A. et al. Association of Flavonifractor Plautii, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India. *mSystems* **4** (2019).
408 Parker, B. J., Wearsch, P. A., Veloo, A. C. M. & Rodriguez-Palacios, A. The Genus Alistipes: Gut Bacteria with Emerging Implications to Inflammation, Cancer, and Mental Health. *Front Immunol* **11**, 906 (2020).
409 Winer, D. A., Luck, H., Tsai, S. & Winer, S. The Intestinal Immune System in Obesity and Insulin Resistance. *Cell metabolism* **23**, 413-426 (2016).
410 Richard, C. et al. Individuals with Obesity and Type 2 Diabetes Have Additional Immune Dysfunction Compared with Obese Individuals Who Are Metabolically Healthy. *BMJ open diabetes research & care* **5**, 000379 (2017).
411 Sokol, H. et al. Fungal Microbiota Dysbiosis in Ibd. *Gut* **66**, 1039-1048 (2017).
412 Sokol, H. et al. Faecalibacterium Prausnitzii Is an Anti-Inflammatory Commensal Bacterium Identified by Gut Microbiota Analysis of Crohn Disease Patients. *Proc Natl Acad Sci U S A* **105**, 16731-16736 (2008).
413 Kudelka, M. R. et al. Cosmc Is an X-Linked Inflammatory Bowel Disease Risk Gene That Spatially Regulates Gut Microbiota and Contributes to Sex-Specific Risk. *Proc Natl Acad Sci U S A* **113**, 14787-14792 (2016).
414 Kallaur, A. P. et al. Genetic, Immune-Inflammatory, and Oxidative Stress Biomarkers as Predictors for Disability and Disease Progression in Multiple Sclerosis. *Molecular neurobiology* **54**, 31-44 (2017).
415 Schirmer, M. et al. Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* **167**, 1125-1136.e1128 (2016).
416 Hooper, L. V. et al. Molecular Analysis of Commensal Host-Microbial Relationships in the Intestine. *Science* **291**, 881-884 (2001).
417 Levy, M. et al. Microbiota-Modulated Metabolites Shape the Intestinal Microenvironment by Regulating Nlrp6 Inflammasome Signaling. *Cell* **163**, 1428-1443 (2015).
418 Ruiz, V. E. et al. A Single Early-in-Life Macrolide Course Has Lasting Effects on Murine Microbial Network Topology and Immunity. *Nature Communications* **8**, 518 (2017).
419 Lynn, M. A. et al. Early-Life Antibiotic-Driven Dysbiosis Leads to Dysregulated Vaccine Immune Responses in Mice. *Cell Host Microbe* **23**, 653-660 (2018).
420 Fujimura, K. E. et al. Neonatal Gut Microbiota Associates with Childhood Multisensitized Atopy and T Cell Differentiation. *Nat Med* **22**, 1187-1191 (2016).
421 Stokholm, J. et al. Maturation of the Gut Microbiome and Risk of Asthma in Childhood. *Nat Commun* **9**, 141 (2018).
422 Nicola, S. et al. Searching for the Perfect Homeostasis: Five Strains of Bifidobacterium Longum from Centenarians Have a Similar Behavior in the Production of Cytokines. *Journal of clinical gastroenterology*, S126-s130 (2016).
423 López, P., Gueimonde, M., Margolles, A. & Suárez, A. Distinct Bifidobacterium Strains Drive Different Immune Responses in Vitro. *International journal of food microbiology* **138**, 157-165 (2010).
424 Medina, M., Izquierdo, E., Ennahar, S. & Sanz, Y. Differential Immunomodulatory Properties of Bifidobacterium Logum Strains: Relevance to Probiotic Selection and Clinical Applications. *Clinical and experimental immunology* **150**, 531-538 (2007).
425 Mouni, F. et al. Effect of Bifidobacterium Bifidum Dsm 20082 Cytoplasmic Fraction on Human Immune Cells. *Immunological investigations* **38**, 104-115 (2009).

426 Tan, T. G. et al. Identifying Species of Symbiont Bacteria from the Human Gut That, Alone, Can Induce Intestinal Th17 Cells in Mice. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 8141-8150 (2016).
427 Kalliomäki, M. et al. Distinct Patterns of Neonatal Gut Microflora in Infants in Whom Atopy Was and Was Not Developing. *Journal of Allergy and Clinical Immunology* **107**, 129-134 (2001).
428 Guo, P., Zhang, K., Ma, X. & He, P. Clostridium Species as Probiotics: Potentials and Challenges. *Journal of Animal Science and Biotechnology* **11**, 24 (2020).
429 Laparra, J. M., Olivares, M., Gallina, O. & Sanz, Y. Bifidobacterium Longum Cect 7347 Modulates Immune Responses in a Gliadin-Induced Enteropathy Animal Model. *PloS one* **7**, 30744-30744 (2012).
430 Herfel, T. M. et al. Dietary Supplementation of Bifidobacterium Longum Strain Ah1206 Increases Its Cecal Abundance and Elevates Intestinal Interleukin-10 Expression in the Neonatal Piglet. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association* **60**, 116-122 (2013).
431 Arrieta, M. C. et al. Associations between Infant Fungal and Bacterial Dysbiosis and Childhood Atopic Wheeze in a Nonindustrialized Setting. *The Journal of allergy and clinical immunology* **142**, 424-434 (2018).
432 Arrieta, M. C. et al. Early Infancy Microbial and Metabolic Alterations Affect Risk of Childhood Asthma. *Science Translational Medicine* **7**, 152 (2015).
433 Vatanen, T. et al. Variation in Microbiome Lps Immunogenicity Contributes to Autoimmunity in Humans. *Cell* **165**, 842-853 (2016).
434 Abellan-Schneyder, I. et al. Primer, Pipelines, Parameters: Issues in 16s Rrna Gene Sequencing. *mSphere* **6**, 01202-01220 (2021).