Dynamics of faecal bacterial populations in early infancy as determined by massively parallel sequencing

by

Shantelle Claassen
(CLSSHA002)

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In fulfilment of the requirements for the degree

MSc in Medical Microbiology

March 2015

Department of Clinical and Laboratory Sciences
Faculty of Health Sciences
University of Cape Town
South Africa

This dissertation is lovingly dedicated in memory of my father

Pieter Johannes Claassen

His boundless love for family inspired my life and continues to inspire me today

Supervisors

Prof. Mark. P. Nicol (MBBCh, MMed (Med Microbiol), SA FCPath (Microbiol), PhD)[1,2,3]

[1]*Division of Medical Microbiology, Department of Clinical Laboratory Sciences, University of Cape Town, Cape Town, South Africa.*

[2]*Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa.*

[3]*National Health Laboratory Service of South Africa, Groote Schuur Hospital, Cape Town, South Africa.*

Dr. Mamadou Kaba (MD, MSc, PhD)[1,2]

[1]*Division of Medical Microbiology, Department of Clinical Laboratory Sciences, University of Cape Town, Cape Town, South Africa.*

[2]*Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa.*

Declaration

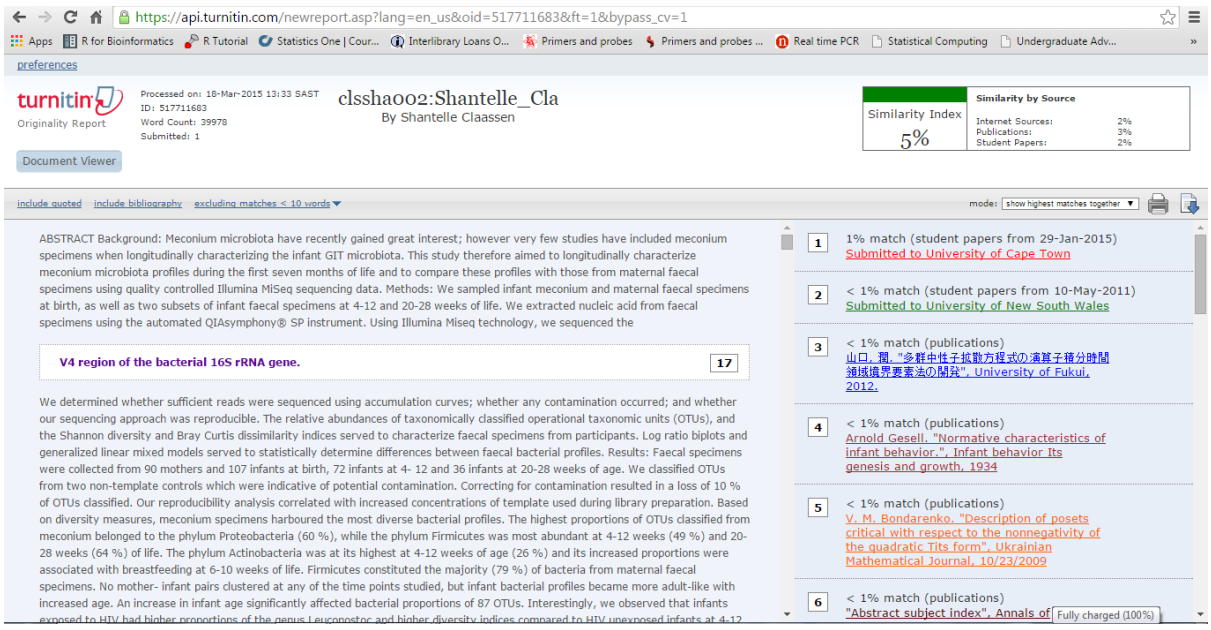
I, *Shantelle Claassen,* hereby declare that the work on which this dissertation is based on my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signed by candidate

Signature removed

18 March 2015



Acknowledgements

I wish to express my sincere appreciation to my supervisors Prof. Mark Nicol and Dr. Mamadou Kaba whom constantly supported and encouraged me throughout my Masters degree. You have provided me with the guidance needed to excel in my own expectations and to grow as a young scientist in the field I adore most.

Many thanks also to Prof. Heather Zar for providing me with the opportunity to study specimens from the Drakenstein Child Health Study. I would also like to thank everyone at the Division of Medical Microbiology for being such a warm and pleasant team to work with, I really appreciate it.

Many thanks to Dr. William Nierman's group at the J. Craig Venter Institute, United States of America, for the internship offered in Next Generation Sequencing. Special thanks to Stephanie Harris Mounaud and to Dr. Jyoti Shankar for your constant support and willingness to help during and after completing my internship. I would also like to thank Gerrit Botha at the Computational Biology Group, University of Cape Town, for your bio-informatics support.

My sincerest gratitude goes to A/Prof. Sugnet Lubbe from the Department of Statistical Sciences, University of Cape Town for your undivided attention and constant support during data analysis. Thank you for your patience and willingness to help, no matter what type of analysis I suggested. I dearly appreciate the fact that you were always available and ready to answer any questions or solve any queries I had (especially after hours).

Personally, I would like to honour and thank my family and friends for their unconditional love, prayers and motivation throughout the past two years. A special thanks to my mother for your constant support and love every day of my life. To my husband, Stefan Weitz, thank you for your love, support and for being my rock every day. I am grateful that my Heavenly Father bestowed His mercy, strength, wisdom and love on me.

Lastly, I acknowledge the National Research Foundation (Innovation and Scarce Skills Master's Scholarship) and the Division of Medical Microbiology, University of Cape Town for their financial assistance.

Preface

This dissertation is submitted for the degree of Master of Science in Medicine (MSc Med) in Medical Microbiology at the Division of Medical Microbiology, Department of Clinical Sciences, University of Cape Town, South Africa. The study was approved by the Human Research Ethics Committee (HREC) of the University of Cape Town, South Africa (742/2013); and supported by the Bill and Melinda Gates Foundation Global Health Grant (OPP1017641), the National Research Foundation (South Africa), the Carnegie Corporation of New York (United States of America) and the Wellcome Trust (United Kingdom). The work reported in this dissertation resulted from a collaborative effort between the Department of Paediatrics and Child Health; Department of Clinical and Laboratory Sciences, University of Cape Town, South Africa; J. Craig Venter Institute (JCVI), Maryland, United States of America; and the Department of Statistical Sciences, University of Cape Town, South Africa.

The aim of this dissertation was to provide a detailed description of the experimental and computational approaches involved in generating Illumina MiSeq sequencing data for the purpose of studying the dynamics of the human stool microbiota from birth to seven months of age. The first chapter presents a detailed overview of the literature in context of the project. The second chapter mainly focusses on the experimental and computational approaches used to generate Illumina MiSeq sequencing data from faecal specimens. Work reported in the second chapter resulted from the following collaborative efforts. The MSc candidate performed nucleic acid extraction sections at the Division of Medical Microbiology, University of Cape Town, based on the published work by Claassen et al. (2013) in the Journal of Microbiological Methods. The MSc candidate prepared the DNA library for sequencing under guidance of staff from Dr. William Nierman's group at JCVI. JCVI's sequencing team performed the Illumina Miseq run. Dr. Jyoti Shankar from Dr. William Nierman's laboratory performed the bioinformatics workflow to correct for sequencing artifacts, and to construct and classify operational taxonomic units. Further quality checks (such as contamination and reproducibility assessment) were designed by the MSc candidate and A/Prof. Sugnet Lubbe from the Department of Statistical Sciences, University of Cape Town. The final chapter of this dissertation focusses on infant faecal bacterial dynamics from birth up until seven months of life. This chapter uses the quality controlled data generated from Chapter 2 to study the infant faecal microbiota from maternal and faecal specimens sampled at birth as well as a subset of infants sampled up until seven months of age. The data analysis system for Chapter 3 was designed by the MSc candidate and A/Prof. Sugnet Lubbe from the Department of Statistical Sciences, University of Cape Town. The submitted material is the work of the MSc candidate, unless stated otherwise by acknowledgments.

Table of content

Abbreviations and acronyms

BSH	bile acid hydrolases
bp	base pair
CR	colonization resistance
DC	dendritic cell
DGGE	denaturing gradient gel electrophoresis
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
EDTA	ethylenediaminetetra-acetic acid
EtBr	ethidium bromide
FISH	fluorescence in situ hybridization
GALT	GIT associated lymphoid tissue
GIT	gastrointestinal tract
GPCR	G protein-coupled receptor
IL	interleukin
Kb	kilo base
m	minute(s)
mg	milligram(s)
MALDI-TOF	matrix-assisted laser desorption ionization time-of-flight
MAMP	microbe-associated molecular pattern
NGS	next generation sequencing
NLR	NOD-like receptor
NCR	natural cytotoxicity receptor
OTU	operational taxonomic unit
PCR	polymerase chain reaction
PSA	polysaccharide A
pM	picomolar
RNA	ribonucleic acid
rRNA	ribosomal RNA
rpm	revolutions per minute
SCFA	short chain fatty acids
sec	second(s)
SFB	segmented filamentous bacteria
TAE	tris-acetate-EDTA
TLR	toll-like receptor
Th cell	T helper cell
T-RFLP	terminal restriction fragment length polymorphism
Treg	T regulatory
Tris-HCl	tris-hydrochloride
µl	microliter
V	volts
%	percent
˚C	degree(s) Celsius

Glossary

Absolute abundance	: Total number (count) of a specific type of organism or OTU present in a specimen[1]
Relative abundance	: Measurement of the number of organisms, sequences or OTUs detected in a specimen, in relation to all others in that specimen[1]
Chimera	: Hybrid PCR product, resulting from an aborted extension product acting as a primer in a subsequent PCR cycle. When this product prime DNA synthesis of an improper template, the resulting hybrid products may be falsely interpreted as novel organisms[2]
Commensal	: In this study commensals refer to GIT microbiota inhabiting the GIT without harming to the host. These organisms benefit from the host and may provide beneficial functions (mutualistic properties) to the human host[3]
Coverage	: The number of times a gene or genome is sequenced[3] or the number of sequences obtained per specimen in a single sequencing run[1]
Diversity	: Is a combination of richness (the number of different organisms) and evenness (a measure of the skew in abundance of community members) within a specimen[4]
Dysbiosis	: Changes in the normal or healthy state microbiota composition[1]
Enterocytes	: Epithelial cells participating in mucosal barrier function[5]
Evenness	: A measure of how evenly all organisms are represented within a community[4]
Goblet cells	: A subset of epithelial cells that produce mucin and other major components of mucus[6]
Meconium	: The newborn's first intestinal discharge[7]
Microbiome	: Defined as the collection of genes from trillions of microbial genomes colonizing the human body[8]
Microfold (M) cells	: Specialised epithelial cells presenting antigens to immune cells[9]
Mock community	: A mixture of bacteria, representative of the specimen analysed. The mock community allows for evaluation of the reproducibility and efficacy of DNA extraction methods by excluding biological variation inherent in clinical specimens[10]
Mycobiome	: Primarily refers to fungal biota in an environment[11]
Operational taxonomic unit	: A group of organisms with 16S rRNA gene sequences grouped together based on their level of sequence identity.[3] These clusters roughly represent taxa at phylogenetic levels defined by user-defined sequence similarity cut-offs.[12] Generally, sequences with 95 % similarity are assigned to the same genus and those with 97 % similarity to the same species.[13]

Overdispersion	:	The incident where a dataset contains greater variance than expected in a statistical model[14]
Paired-end sequencing	:	Obtaining sequence information from opposite ends of long templates, providing data on both ends of the fragment of interest[15]
Paneth cells	:	Epithelial cells responsible for the protection of other epithelial cells via secretion of a variety of bactericidal molecules such as lysozymes, defensins, cathelicidins and c-type lectins such as RegIIIc)[6]
Phylotypes	:	Assignment of clustered 16S rRNA sequences to a taxonomic group based on sequences from a known reference database[1]
Q score (Phred score)	:	Determines how accurately each of the nucleotides is called during Illumina sequencing; and are therefore essential for identification and removal of low quality reads from raw sequencing data.[16–18]
Read	:	A short stretch nucleotide sequence produced from sequencing a targeted region of a single DNA fragment[3]
Richness	:	The number of different types of organisms present within a specimen[4]
Segmented filamentous bacteria	:	Commensal bacteria sharing strong similarities with the genus Clostridium[19]
Virome	:	The collection of all viruses residing in or on humans (i.e. the viral component of the human microbiome)[20]

References

1. Tyler AD, Smith MI, Silverberg MS. Analyzing the Human Microbiome: A “How To” guide for Physicians. *Am J Gastroenterol*. 2014. doi:10.1038/ajg.2014.73.

2. Haas BJ, Gevers D, Earl AM, et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res*. 2011;21(3):494-504. doi:10.1101/gr.112730.110.

3. Weinstock GM. Genomic approaches to studying the human microbiota. *Nature*. 2012;489(7415):250-6. doi:10.1038/nature11553.

4. Cox MJ, Cookson WOCM, Moffatt MF. Sequencing the human microbiome in health and disease. *Hum Mol Genet*. 2013;22(1):R88-94. doi:10.1093/hmg/ddt398.

5. Tlaskalová-Hogenová H, Štěpánková R, Kozáková H, et al. The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer : contribution of germ-free and gnotobiotic animal models of human diseases. *Cell Mol Immunol*. 2011;8:110-120. doi:10.1038/cmi.2010.67.

6. Goto Y, Kiyono H. Epithelial barrier : an interface for the cross-communication between gut flora and immune system. *Immunol Rev*. 2012;245:147-163.

7. Gosalbes MJ, Llop S, Vallès Y, Moya A, Ballester F, Francino MP. Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clin Exp Allergy*. 2012;43(2):198-211. doi:10.1111/cea.12063.

8. Turnbaugh PJ, Gordon JI. The core gut microbiome, energy balance and obesity. *J Physiol*. 2009;587(17):4153-58. doi:10.1113/jphysiol.2009.174136.

9. Kucharzik T, Lügering N, Rautenberg K, Lügering A. Role of M Cells in Intestinal Barrier Function. *Ann N Y Acad Sci*. 2000;915(1):171-183.

10. Abusleme L, Hong B-Y, Dupuy AK, Strausbaugh LD, Diaz PI. Influence of DNA extraction on oral microbial profiles obtained via 16S rRNA gene sequencing. *J Oral Microbiol*. 2014;6:1-7. doi:10.3402/jom.v6.23990.

11. Cui L, Morris A, Ghedin E. The human mycobiome in health and disease. *Genome Med*. 2013;5(7):63. doi:10.1186/gm467.

12. Tanaseichuk O, Borneman J, Jiang T. Phylogeny-based classification of microbial communities. *Bioinformatics*. 2014;30(4):449-56. doi:10.1093/bioinformatics/btt700.

13. Ye Y. Identification and Quantification of Abundant Species from Pyrosequences of 16S rRNA by Consensus Alignment. *Proceedings (IEEE Int Conf Bioinformatics Biomed)*. 2011;2010:153-157. doi:10.1109/BIBM.2010.5706555.

14. Yu P, Shaw CA. An efficient algorithm for accurate computation of the Dirichlet-multinomial log-likelihood function. *Bioinformatics*. 2014;30(11):1547-54. doi:10.1093/bioinformatics/btu079.

15. Holt RA, Jones SJM. The new paradigm of flow cell sequencing. *Genome Res*. 2008;18(6):839-46. doi:10.1101/gr.073262.107.

16. Kircher M, Heyn P, Kelso J. Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics*. 2011;12(1):382. doi:10.1186/1471-2164-12-382.

17. Ewing B, Hillier L, Wendl MC, Green P. Base-Calling of Automated Sequencer Traces Using Phred . I . Accuracy Assessment. *Genome Res*. 1998;8:175-185.

18. Ewing B, Green P. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Res*. 1998;8:186-194.

19. Sczesnak A, Segata N, Qin X, et al. The genome of Th17 cell-inducing segmented filamentous bacteria reveals extensive auxotrophy and adaptations to the intestinal environment. *Cell Host Microbe*. 2011;10(3):260-72. doi:10.1016/j.chom.2011.08.005.

20. Wylie KM, Weinstock GM, Storch GA. Emerging View of the Human Virome. *Transl Res*. 2012;160(4):283-290. doi:10.1016/j.trsl.2012.03.006.

ABSTRACT

Background: Meconium microbiota have recently gained great interest; however very few studies have included meconium specimens when longitudinally characterizing the infant GIT microbiota. This study therefore aimed to longitudinally characterize meconium microbiota profiles during the first seven months of life and to compare these profiles with those from maternal faecal specimens using quality controlled Illumina MiSeq sequencing data.

Methods: We sampled infant meconium and maternal faecal specimens at birth, as well as two subsets of infant faecal specimens at 4-12 and 20-28 weeks of life. We extracted nucleic acid from faecal specimens using the automated QIAsymphony® SP instrument. Using Illumina Miseq technology, we sequenced the V4 region of the bacterial 16S rRNA gene. We determined whether sufficient reads were sequenced using accumulation curves; whether any contamination occurred; and whether our sequencing approach was reproducible. The relative abundances of taxonomically classified operational taxonomic units (OTUs), and the Shannon diversity and Bray Curtis dissimilarity indices served to characterize faecal specimens from participants. Log ratio biplots and generalized linear mixed models served to statistically determine differences between faecal bacterial profiles.

Results: Faecal specimens were collected from 90 mothers and 107 infants at birth, 72 infants at 4-12 and 36 infants at 20-28 weeks of age. We classified OTUs from two non-template controls which were indicative of potential contamination. Correcting for contamination resulted in a loss of 10 % of OTUs classified. Our reproducibility analysis correlated with increased concentrations of template used during library preparation. Based on diversity measures, meconium specimens harboured the most diverse bacterial profiles. The highest proportions of OTUs classified from meconium belonged to the phylum Proteobacteria (60 %), while the phylum Firmicutes was most abundant at 4-12 weeks (49 %) and 20-28 weeks (64 %) of life. The phylum Actinobacteria was at its highest at 4-12 weeks of age (26 %) and its increased proportions were associated with breastfeeding at 6-10 weeks of life. Firmicutes constituted the majority (79 %) of bacteria from maternal faecal specimens. No mother-infant pairs clustered at any of the time points studied, but infant bacterial profiles became more adult-like with increased age. An increase in infant age significantly affected bacterial proportions of 87 OTUs. Interestingly, we observed that infants exposed to HIV had higher proportions of the genus Leuconostoc and higher diversity indices compared to HIV unexposed infants at 4-12 weeks of age.

Conclusion: Our study highlights that reproducibility may be worsened by the use of low template concentrations during library preparation, which may also skew diversity measures. We conclude that meconium is not sterile and that infant faecal bacterial profiles become more adult-like with increased age.

GENERAL INTRODUCTION

The human gastrointestinal tract (GIT) is home to a complex community composed of approximately 100 trillion microbial cells and 3.3 million microbial genes.[1–3] These microbial inhabitants perform various functions essential for the host's health,[4–6] and have even been described as a vital "organ" due to their homeostatic properties.[7–9] Despite these findings, studies have not yet been able to define a core GIT microbial profile across human hosts.[10,11] Not only are these microbial profiles highly diverse,[12–14] but the composition of these profiles are also influenced by a number of external factors, including medication, diet, age, genetics and environment.[15]

The first two years of life, have been considered as the most critical period during which the GIT microbiome is established.[16] Since early life microbial profiles contribute to the establishment of successive GIT microbiota profiles, and have been shown to contribute to health and disease,[16–22] they are key profiles to study when investigating the complexity of faecal bacterial profiles later in life. Recently, studies have made interesting observations with regards to the conventionally considered "sterile" uterus and now suggest an even earlier colonization stage of the infant GIT referred to as in-utero colonization.[23] Since studies have recently identified microbiota from the previously considered sterile meconium,[24–27] studies may need to also investigate the role of prenatal factors when studying the dynamics of early life microbial profiles. In support of this, a number of influencing factors, such as prematurity, prolonged rupture of fetal membranes, intrapartum antibiotic administration and maternal diabetes status, have already been associated with patterns of colonization of meconium microbiota.[28–30] Such factors may provide an even earlier influence on successive bacterial colonization and resultant health and disease states.[31–34]

The first chapter of this dissertation provides an updated overview of the literature focussing on the function of the GIT microbiome, techniques used to study it, questions around a "core" microbiome and the development of the early life GIT microbiota. The second chapter describes the experimental and computational approaches used to the generate Illumina Miseq sequencing data for studying the GIT bacterial profiles from infants and their mothers. The third and final chapter provides quality controlled high-throughput sequencing information on the bacterial composition of meconium specimens in a large cohort. It also compares bacterial profiles from meconium specimens to those of maternal faecal specimens. In addition, Chapter 3 provides bacterial profiles of a subset of infants at two later time points, 4-12 weeks and 20-28 weeks of age, and determines how these profiles relate to maternal faecal bacterial profiles. Finally, we also address a number of external factors which may influence the above-mentioned bacterial profiles.

References

1. Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*. 2006;124(4):837-48. doi:10.1016/j.cell.2006.02.017.

2. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. *Science*. 2005;307(5717):1915-20. doi:10.1126/science.1104816.

3. Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464(7285):59-65. doi:10.1038/nature08821.

4. Sekirov I, Russell SL, Antunes LCM, Finlay BB. Gut microbiota in health and disease. *Physiol Rev*. 2010;90(3):859-904. doi:10.1152/physrev.00045.2009.

5. Fraher MH, O'Toole PW, Quigley EMM. Techniques used to characterize the gut microbiota: a guide for the clinician. *Nat Rev Gastroenterol Hepatol*. 2012;9(6):312.

6. Lagier J-C, Million M, Hugon P, Armougom F, Raoult D. Human gut microbiota: repertoire and variations. *Front Cell Infect Microbiol*. 2012;2(November):136. doi:10.3389/fcimb.2012.00136.

7. Evans JM, Morris LS, Marchesi JR. The gut microbiome: the role of a virtual organ in the endocrinology of the host. *J Endocrinol*. 2013;218(3):R37-47. doi:10.1530/JOE-13-0131.

8. Possemiers S, Bolca S, Verstraete W, Heyerick A. The intestinal microbiome: a separate organ inside the body with the metabolic potential to influence the bioactivity of botanicals. *Fitoterapia*. 2011;82(1):53-66. doi:10.1016/j.fitote.2010.07.012.

9. Foxman B, Goldberg D, Murdock C, Xi C, Gilsdorf JR. Conceptualizing human microbiota: from multicelled organ to ecological community. *Interdiscip Perspect Infect Dis*. 2008;2008:613979. doi:10.1155/2008/613979.

10. Lagier J-C, Armougom F, Million M, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect*. 2012;18(12):1185-93. doi:10.1111/1469-0691.12023.

11. Bäckhed F, Fraser CM, Ringel Y, et al. Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications. *Cell Host Microbe*. 2012;12(5):611-22. doi:10.1016/j.chom.2012.10.012.

12. Huse SM, Ye Y, Zhou Y, Fodor AA. A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS One*. 2012;7(6):e34242. doi:10.1371/journal.pone.0034242.

13. Caporaso JG, Lauber CL, Costello EK, et al. Moving pictures of the human microbiome. *Genome Biol*. 2011;12(5):R50. doi:10.1186/gb-2011-12-5-r50.

14. Scott KP, Duncan SH, Louis P, Flint HJ. Nutritional influences on the gut microbiota and the consequences for gastrointestinal health. *Biochem Soc Trans*. 2011;39(4):1073-8. doi:10.1042/BST0391073.

15. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012;489(7415):220-30. doi:10.1038/nature11550.

16. Wopereis H, Oozeer R, Knipping K, Belzer C, Knol J. The first thousand days - intestinal microbiology of early life: establishing a symbiosis. *Pediatr allergy Immunol*. 2014. doi:10.1111/pai.12232.

17. Aidy S El, Hooiveld G, Tremaroli V, Bäckhed F, Kleerebezem M. The gut microbiota and mucosal homeostasis: Colonized at birth or at adulthood , does it matter ? *Gut Microbes*. 2013;4(2):118-124. doi:10.4161/gmic.23362.

18. Abrahamsson TR, Jakobsson HE, Andersson AF, Björkstén B, Engstrand L, Jenmalm MC. Low gut microbiota diversity in early infancy precedes asthma at school age. *Clin Exp allergy*. 2013;44(6):842-850. doi:10.1111/cea.12253.

19. Ismail IH, Oppedisano F, Joseph SJ, et al. Reduced gut microbial diversity in early life is associated with later development of eczema but not atopy in high-risk infants. *Pediatr allergy Immunol*. 2012;23(7):4674-81. doi:10.1111/j.1399-3038.2012.01328.x.

20. Isolauri E. Development of healthy gut microbiota early in life. *J Paediatr Child Health*. 2012;48 Suppl 3:1-6. doi:10.1111/j.1440-1754.2012.02489.x.

21. Vael C, Vanheirstraeten L, Desager KN, Goossens H. Denaturing gradient gel electrophoresis of neonatal intestinal microbiota in relation to the development of asthma. *BMC Microbiol*. 2011;11(1):68. doi:10.1186/1471-2180-11-68.

22. Penders J, Thijs C, Van Den Brandt PA, et al. Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. *Gut*. 2007;56(5):661-667. doi:10.1136/gut.2006.100164.

23. Funkhouser LJ, Bordenstein SR. Mom knows best: the universality of maternal microbial transmission. *PLoS Biol*. 2013;11(8):e1001631. doi:10.1371/journal.pbio.1001631.

24. Jiménez E, Marín ML, Martín R, et al. Is meconium from healthy newborns actually sterile? *Res Microbiol*. 2008;159(3):187-93. doi:10.1016/j.resmic.2007.12.007.

25. Makino H, Kushiro A, Ishikawa E, et al. Transmission of intestinal Bifidobacterium longum subsp. longum strains from mother to infant, determined by multilocus sequencing typing and amplified fragment length polymorphism. *Appl Environ Microbiol*. 2011;77(19):6788-93. doi:10.1128/AEM.05346-11.

26. Tsuji H, Oozeer R, Matsuda K, et al. Molecular monitoring of the development of intetinal microbiota in Japanese infants. *Benef Microbes*. 2012;3(2):113-25. doi:10.3920/BM2011.0038.

27. Gosalbes MJ, Llop S, Vallès Y, Moya A, Ballester F, Francino MP. Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clin Exp Allergy*. 2012;43(2):198-211. doi:10.1111/cea.12063.

28. Moles L, Gómez M, Heilig H, et al. Bacterial diversity in meconium of preterm neonates and evolution of their fecal microbiota during the first month of life. *PLoS One*. 2013;8(6):e66986. doi:10.1371/journal.pone.0066986.

29. Hu J, Nomura Y, Bashir A, et al. Diversified microbiota of meconium is affected by maternal diabetes status. *PLoS One*. 2013;8(11):e78257. doi:10.1371/journal.pone.0078257.

30. Mshvildadze M, Neu J, Shuster J, Theriaque D, Li N, Mai V. Intestinal microbial ecology in premature infants assessed with non-culture-based techniques. *J Pediatr*. 2010;156(1):20-5. doi:10.1016/j.jpeds.2009.06.063.

31. Ardissone AN, De La Cruz DM, Davis-Richardson AG, et al. Meconium microbiome analysis identifies bacteria correlated with premature birth. *PLoS One*. 2014;9(3):e90784. doi:10.1371/journal.pone.0090784.

32. Rautava S, Luoto R, Salminen S, Isolauri E. Microbial contact during pregnancy, intestinal colonization and human disease. *Nat Rev Gastroenterol Hepatol*. 2012;9(10):565-76.

33. Madan JC, Salari RC, Saxena D, et al. Gut microbial colonisation in premature neonates predicts neonatal sepsis. *Arch Dis childhood Fetal neonatal Ed*. 2012;97(6):F456-62. doi:10.1136/archdischild-2011-301373.

34. Stewart CJ, Marrs ECL, Magorrian S, et al. The preterm gut microbiota: changes associated with necrotizing enterocolitis and infection. *Acta Paediatr*. 2012;101(11):1121-7. doi:10.1111/j.1651-2227.2012.02801.x.

CHAPTER 1

THE HUMAN GASTROINTESTINAL TRACT MICROBIOME EARLY IN LIFE - A LITERATURE REVIEW

Summary

The gastrointestinal tract (GIT) is home to 10^{14} microbial cells which have been referred to as a "virtual organ" essential for the host's health. Despite the wide range of techniques available to study our GIT inhabitants, it still proves difficult to describe what a "healthy" GIT microbiota profile should look like and whether a common core microbiota exists for all humans. Reasons for this have been ascribed due to considerable inter-individual variation, fluctuation over time, and a wide range of factors (such as age, diet, geography, genetics and medication) influencing microbial profiles throughout life. Although a number of factors have been shown to influence adult GIT microbiota profiles; early life colonizers seem to act as pioneers for colonization later in life. Conventionally, it has been accepted that early life GIT colonization occurs directly after birth and that factors such as mode of delivery, feeding practices, weaning, geography and the use of antibiotics primarily shape these profiles. Recent studies, however, have now suggested that colonization actually occurs prior to the process of birth and that in-utero colonization of the GIT may be important in establishing GIT microbial profiles later in life.

This review explores some of the beneficial functions the GIT microbiota provides to its host which includes maintenance of the GIT-barrier function, protection from pathogen colonization, immune system maturation and metabolic functions. It also discusses a number of techniques commonly used to study the GIT microbiota. These techniques range from conventional culture and molecular approaches which target known bacterial components within samples; to high-throughput culture and molecular techniques which provide comprehensive insights of both the known and unknown bacterial components of our GITs. This review highlights the challenges around identifying a core microbiome and emphasises the importance of early-life colonizers as pioneers for bacterial colonization later in life as well as their role in health and disease states during both early and later life. Furthermore, it accentuates that the importance of in-utero colonization is becoming apparent due to its influence on successive bacterial colonisation; its association with immune system maturation; and possible contributions to both early and later life disease. Based on the associations between specific meconium profiles and neonatal sepsis, infant mucus congestion, and even premature birth; insight into meconium microbiota profiles may be of great value for manipulation of in-utero or early life bacterial profiles to ensure a healthy start to life.

1.1 The human gastrointestinal microbiota is a vital 'organ'

Humans have been described as "superorganisms" harbouring 100-fold more microbial genes in the GIT compared to human genes.[1] These GIT inhabitants are essential for the host's health[2–6] and have even been referred to as a "virtual organ".[3,7,8] Functions performed by this 'virtual organ' include maintenance of the GIT-barrier function, protection from pathogen colonization, immune system maturation, bile acid metabolism, fermentation of undigested polysaccharides and vitamin production.[9] The following section describes in detail the biological functions of the GIT microbiota for the host.

1.1.1 The epithelial barrier function of the GIT microbiota

The GIT barrier system is comprised of two main layers known as the epithelial cell and mucus layers.[10] GIT epithelial cells generally function to absorb nutrients and to retain water and electrolytes.[11] However, together with the GIT mucus, it also functions as a barrier preventing undesired antigens from entering the host.[12] Malfunctioning of this defence system may result in infection and inflammation.[10]

The intestinal epithelial layer functions as a physiological barrier by sampling antigens, secreting anti-microbial peptides, producing digestive enzymes, shaping mucosal immune responses and by providing essential barrier functions for prevention of microbial infection.[13] However, in order to exert these functions, the epithelial cell layer relies on a network of specialized epithelial cells, such as endocrine cells, Microfold (M) cells, Goblet cells, Paneth cells and columnar epithelial cells.[10] The GIT microbiota have been shown to influence the development of some of these cells.[12] Maturation of enterocytes, for example, is enhanced during intestinal colonization with *Escherichia coli* strains,[14] while M cells have been shown to increase after the introduction of *Salmonella typhimurium* to the GIT.[15,16] A reduction in Goblet cells has been observed in germ-free animals and animals with specific bacterial profiles.[17–21] Finally, Paneth cells have been shown to express bactericidal proteins after being stimulated by Gram-positive and Gram-negative bacteria as well as bacterial components such as lipopolysaccharides and lipoteichoic acid.[22,23] This in turn also influence the mucus layer (essential for protection of the epithelial cell layer against pathogens), since Goblet and Paneth cells secrete mucins and antimicrobial peptides which contribute to the mucosa's viscous and protective properties.[24,25]

1.1.2 Colonization resistance against enteric pathogens

Colonization resistance (CR) is a term used to describe the role of the GIT microbiota in protection against colonization and invasion by enteric pathogens.[26] This concept is supported by studies showing an increase in susceptibility to enteric infections as a result of disruption of GIT microbiota during antibiotic administration.[27,28] Three mechanisms have been proposed to describe colonization resistance elicited by GIT microbiota.[26] The first being "direct inhibition" suggesting that bacteria can impede the growth of other bacterial species by releasing bacterial toxins or inhibitory metabolites such as acetate and butyrate.[29–31] The second mechanism is described as "nutrient depletion" which suggests that efficient utilization of nutrients by commensals makes nutrient niches inaccessible for pathogens.[32] This can be explained by the intestinal "food-web" during which several primary and secondary bacterial fermenters efficiently deplete high-energy nutrients resulting in limited nutrient sources for incoming pathogens.[32,33] The third mechanism, described as "indirect inhibition", is explained by the microbiota's ability to stimulate the host to act upon enteric pathogens. Stimulation of the host's specialized epithelial cells results in the production of antimicrobial peptides, immunoglobulin A (IgA) and secretion of mucin, which mediate pathogen clearance from the GIT lumen.[19,34–36]

1.1.3 Maturation of the host's immune system is modulated by GIT microbiota

Studies in mice have shown that both the innate and adaptive immune systems are negatively impacted in the absence of commensal GIT bacteria.[37,38] This can be explained by the fact that immune responses are elicited during recognition of bacterial molecules via innate immune receptors.[39] Toll-like receptors (TLRs), for example, detect various microbe-associated molecular patterns (MAMPs). Lipoproteins and lipoteichoic acids are detected by TLR2, lipopolysaccharides (LPS) by TLR4, flagellin by TLR5 and CpG deoxyribonucleic acid (DNA) by TLR9.[39,40] These, and other receptors known as NOD-like receptors (NLRs) and G protein-coupled receptors (GPCRs) receptors[39] are present on epithelial and immunological cells throughout the entire gastrointestinal tract, and upon stimulation, coordinates a cascade of innate and adaptive immune responses to control infection.[41,42] Two main functions of commensal bacteria in stimulating the host's immune responses are to prevent bacterial infection and to elicit systemic immune responses (discussed in Sections 1.1.3.1 and 1.1.3.2).

1.1.3.1 *Immune-related effects of the GIT microbiota in preventing bacterial infection*

Immunological protection against bacterial infections involves a variety of innate and adaptive immune cells of both epithelial and lymphoidal origin[37] (Figure 1). Epithelial cells are the first set of host cells which pathogens encounter and play an essential role in ensuring minimal

bacterial contact and invasion of the epithelial cell layer. LPS from Gram positive bacteria stimulate enterocytes via TLR4 to produce the antibacterial lectin RegIIIg which spatially segregates the luminal microbiota from the epithelial layer by maintaining an approximately 50 micrometer zone between the host and the microbiota.[23,43,44] Gram negative LPS and flagellin A, as well as Gram positive lipoteichoic acids are bacterial products that modulate mucin production.[45] These bacterial products activate pathways known to have a strong secretagogue effect on goblet cells, increasing mucin secretion.[46] Increased mucin secretion, in turn results in a larger mucus barrier which protects epithelial cells against pathogens.[46] Bacterial molecules are recognised by TLRs on Paneth cells and MyD88 signalling is activated.[34] This results in Paneth cells secreting antimicrobial peptides which kill luminal pathogens and prevent penetration of the host epithelial layer.[34,47,48]

The second set of immune cells participating in innate defence against luminal pathogens are innate lymphoid cells (ILCs).[49] These cells are grouped into three main groups, which are all regulated indirectly via epithelial, myeloid and dendritic cell responses to the GIT microbiota. The first group, T-bet+ILCs, produce cytokines to limit epithelial cell permeability.[49] The second group, GATA3+ILCs, also produce cytokines, however their role is not yet determined in microbial regulation.[49] Finally, the third group, RORgt+ ILCs promote antimicrobial protein and mucin production by epithelial cells.[49] Both T-bet+ILCs and RORgt+ILCs can also be directly influenced by microbial signals via immune receptors such as TLRs and natural cytotoxicity receptors (NCRs).

In addition to the innate immune responses, adaptive immune responses are also influenced by GIT microbiota.[50,51] Macrophages and dendritic cells (DCs) are mononuclear phagocytes that sample bacteria via M cells within Peyer's patches, where they interact with B and T cells.[37,52] Plasma cells are generated from B cells and secrete IgA in the lamina propria which is transcytosed to the intestinal lumen where it binds to bacteria and prevents translocation across the epithelial barrier.[53]

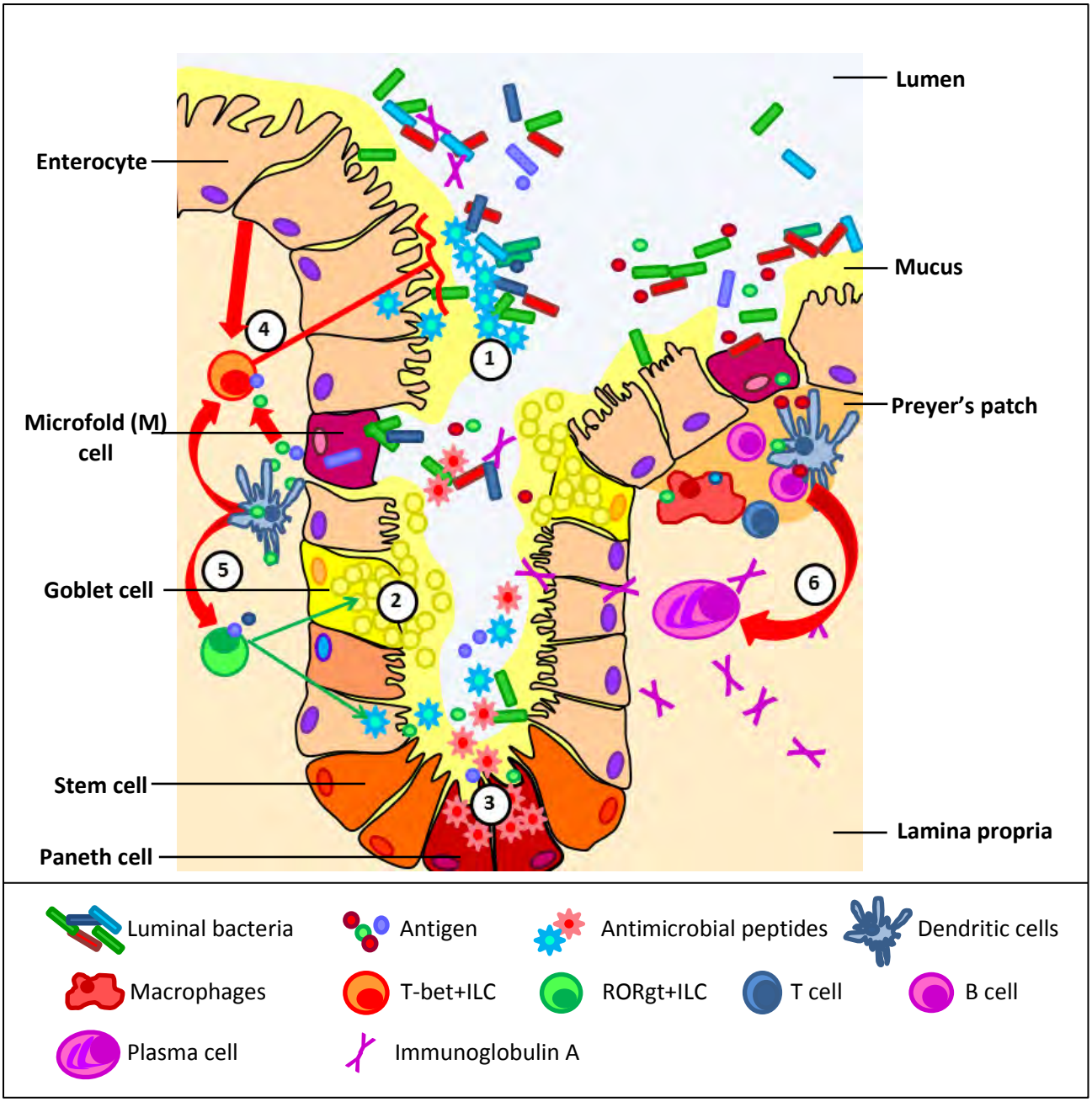


Figure 1. Schematic representation of immunological protection against bacterial infections by gastrointestinal microbiota

1. Gram positive lypopolysaccharides (LPS) stimulate enterocytes via toll like receptor (TLR) 4. This results in antimicrobial protein secretion and spatial segregation of luminal microbiota from the epithelial layer. **2.** Increased mucin secretion via Goblet cells following stimulation from gram negative LPS and flagellin A, as well as gram positive lipoteichoic acids. **3.** Paneth cells secrete antimicrobial peptides as a result of MyD88 signalling when TLRs recognise bacterial molecules. **4.** T-bet+ILCs are a group of innate lymphoid cells (ILCs) limiting epithelial cell permeability by the production of cytokines. **5.** Another group of ILCs (RORgt+ ILCs) contribute to antimicrobial protein and mucin production by stimulating epithelial cells. Both these groups of ILCs receive bacterial stimulation indirectly via other immunological cells such as dendritic cells (DCs) or directly via TLRs and natural cytotoxicity receptors (NCRs). **6.** Bacteria entering Peyer's patches via Microfold (M) cells are sampled by macrophages and DCs, where they interact with B and T cells. Activated B cells differentiate into IgA secreting plasma cells. IgA is then trancytosed into the intestinal lumen and prevents bacterial translocation across the epithelial barrier.

1.1.3.2 *Complex bacterial profiles in the GIT also shape systemic immunity.*

Immunological responses elicited by GIT microbiota do not only protect the host from bacterial infections, but also promote and regulate systemic immune responses[54] (Figure 2). Ivanov et al. (2009), while studying the role of commensal microbiota on the host's immune system in a murine model, showed that segmented filamentous bacteria (SFB) stimulate DCs which promote CD4+ T cell differentiation into T helper 17 (Th17) cells.[55] Th17 cells predominantly protect the host against extracellular pathogens by secreting interleukin-17 (IL-17), IL-17F and IL-22.[56] Such inflammatory responses, have however also been implicated as a primary mediator of autoimmune disease.[57] Another murine model study by Mazmanian et al. (2005), on the immunomodulatory activities of a bacterial molecule in germ-free and conventionally raised mice, supports the role of commensal bacteria in the development of the host's immune system.[58] The former study showed that *Bacteroides fragilis* (a common colonizer of the lower intestinal tract of mammals) produces a T-cell dependent antigen known as polysaccharide A (PSA) which stimulates CD4+ T cell maturation. It has been proposed that *B. fragilis* and/or PSA activates DCs in the GIT associated lymphoid tissue (GALT), whereafter they migrate to lymphoid organs to signal Th1 cell cytokine production via an IL-12 pathway.[58] Not only do these Th1 cells protect against intracellular pathogens,[57] they also reduce allergic disease by down-regulating systemic Th2 and Th17 responses.[42,59] In addition, Pang et al. (2013) observed suppressed allergic airway responses in mice which were treated with non-pathogenic *E. coli.*[60] This study observed an increase in Th1 cytokines and IL-10-secreting immunosuppressive T regulatory (Treg) cells in the para-tracheal lymph nodes of the treated mice, supporting the role of these common colonizers in preventing allergic diseases. Finally, Atarashi and colleagues[61] showed that 17 non-pathogenic strains of Clostridial clusters IV, XIVa and XVIII act as a community to promote growth and differentiation of immunosuppressive Treg cells in germ-free mice. Oral administration of these strains also alleviated allergic diseases such as colitis and allergic diarrhoea in adult mice.[61]

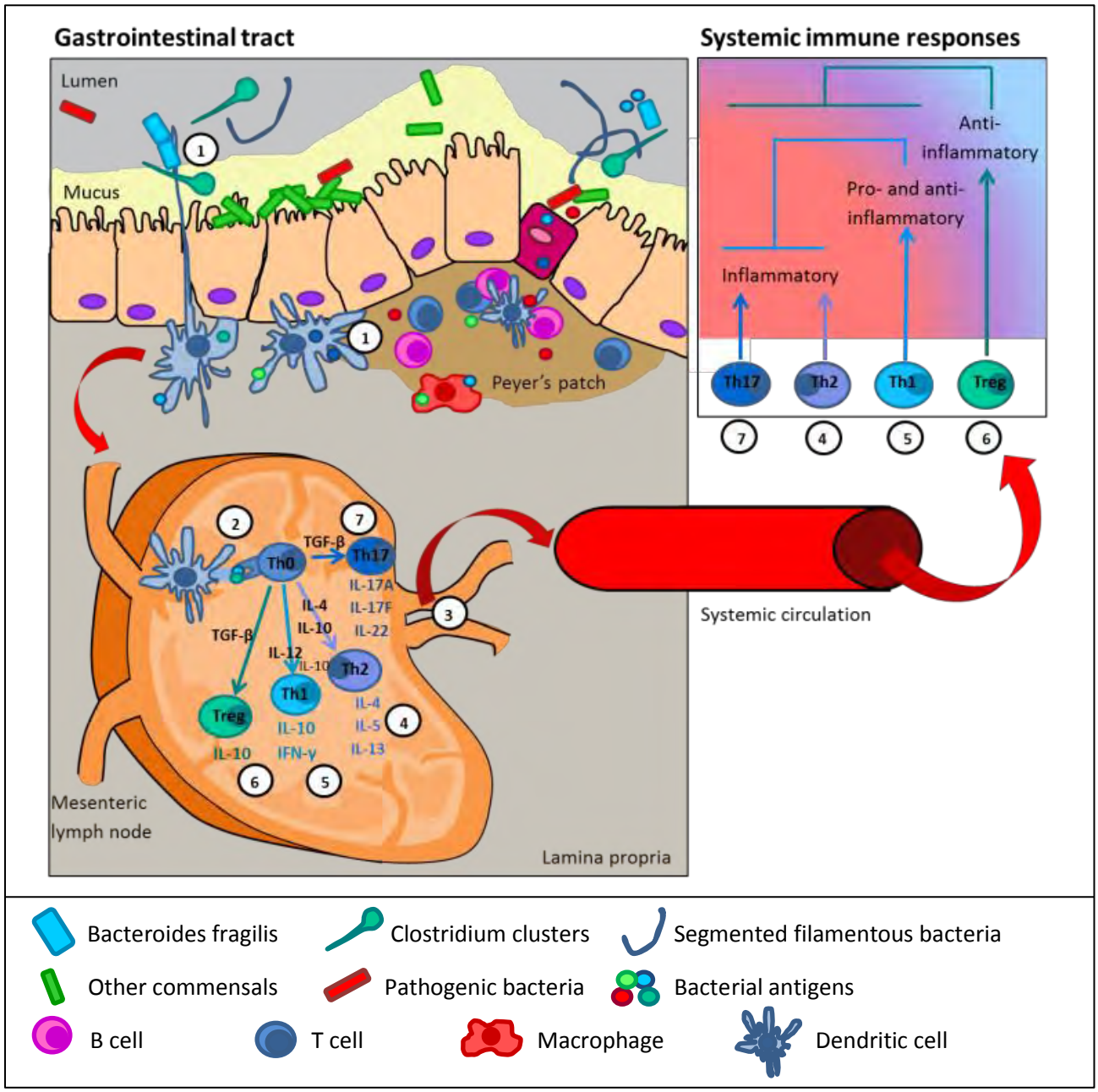


Figure 2. Schematic representation of gastrointestinal microbiota shaping systemic immunity

1. Dendritic cells (DCs) and macrophages sample antigen in the lamina propria and Peyer's patches of the small intestine. DCs also sample antigen by extending their dendrites into the intestinal lumen. **2.** The interactions between DCs and microbial associated molecular patterns (MAMPs) allow DCs to present antigen to naïve T cells in the mesenteric lymph nodes. DCs present epitopes together with major histocompatibility complex (MHC) class II and specific immunomodulatory cytokines to naïve $CD4^+$ T (Th0) cells. **3.** After proliferation and activation of various T cell subets, they enter systemic circulation via the efferent lymph, and home to mucosal surfaces inside and outside of the gut. **4.** IL-4 and IL10 cytokines are responsible for Th2 cell differentiation. Overproduction of Th2 cytokines and subsequent IgE production contribute to inflammatory responses (such as allergic airway responses) via mast cell degranulation, eosinophilia and mucus hypersecretion in the respiratory tract. **5.** *Bacteroides fragilis* produces polysaccharide A which signals Th1 cell cytokine production via an IL-12 pathway. Th1 cytokines, IFN-γ and IL-10, have been shown to supress Th2 and Th17 cytokines involved in asthma pathogenesis and pro-inflammatory responses, however, may also contribute to pro-inflammatory responses such as irritable bowel disease. **6.** Clostridium clusters IV, XIVa and XVIII stimulate T regulatory (Treg) cell differentiation via the TGF-β cytokine. Treg cells downregulate Th2, Th1 and Th17 pro-inflammatory responses. **7.** Segmented filamentous bacteria stimulate the differentiation of Th17 cells via the TGF-β, which produce cytokines involved in pro-inflammatory responses and allergic diseases.

1.1.4 Metabolic functions of GIT microbiota

GIT microbiota have been shown to impact on their host's health by a range of metabolic functions.[62] The following section will focus on two specific functions: The production of short chain fatty acids (SCFAs) and the metabolism of bile acids.

SCFAs are well known bacterial metabolites produced by intestinal bacteria.[63] These metabolites are produced by fermentation from undigested dietary carbohydrates and mainly consists of butyrate, acetate and propionate.[64] Specific bacteria, such as anaerobes and members of the Firmicutes and Bacteroides phyla have been shown to be involved in these fermentation processes.[65] SCFAs, especially butyrate, have been associated with increased intestinal barrier function and protection against diet-induced obesity and insulin resistance.[66–68] In addition, SCFAs also play a crucial role in promoting anti-inflammatory responses.[69] Low levels of SCFAs have been associated with a number of inflammatory disease states such as inflammatory bowel disease (ulcerative colitis, Crohn's disease)[70–72] and allergic airway disease.[73]

Bile acid metabolism is another important metabolic function in which GIT microbiota are involved.[74] Bile acids are associated with a range of biological activities such as emulsification of lipids and lipid-soluble compounds, regulation of cholesterol metabolism, activation of cell signalling pathways during cell apoptosis, energy metabolism, inflammatory responses, glucose metabolism, and protection of the intestinal mucosa from microbial invasion.[75] GIT microbiota modifies the bile acid pool circulated throughout the body by a range of biotransformations.[76] One of these transformations, known as deconjugation, is essential for further modifications of bile acids into secondary bile acids.[77] Deconjugation is dependent on bile acid hydrolase (BSH) enzymes which are produced by Gram-positive bacterial species, belonging to genera such as Lactobacillus, Bifidobacterium, Enterococcus and Clostridium as well as two Gram-negative Bacteroides strains.[77] BSH activity and resultant deconjugated bile acids may play an important role in controlling weight gain and lowering cholesterol in the host.[78] It may also, however, compromise normal lipid digestion, promote gastrointestinal diseases such as diarrhea, inflammation and colon cancer, and may also cause gallstones.[77]

Ultimately, the GIT microbiota forms a vital component of the human body and its functions are undoubtedly crucial for the host's immune system development and metabolic responses. As a "healthy GIT microbiota" could be delineated by its functional core,[79] bacterial profiles from the GIT may possibly be used as markers for health and disease.[71,80–83] Consequently, studies focussing on the GIT microbiome as a baseline for health and disease need to use optimal laboratory tools and downstream analyses to ensure detection of both high and low abundance bacterial species.[84–89] The following section therefore focusses on techniques used to study the GIT microbiota, highlighting their advantages as well as their limitations.

1.2 How to study our microbial inhabitants?

Microbiological studies date back as far as 300 years when Antonie van Leuwenhoek made his first observation through a handcrafted microscope.[90–93] Since then, our bacterial inhabitants have become a topic of great interest, especially during the last twenty years when their role in disease became more apparent.[94–97] Studying the GIT microbiome, however, has proven to be challenging due to the vast number of diverse bacterial species inhabiting the GIT.[98] This section describes the tools and techniques used to study the GIT microbiome, highlighting their strengths and weaknesses (Table 1), and emphasising the benefits of high-throughput technology for both culture-dependent and -independent techniques.

1.2.1 Culture-dependent techniques

1.2.1.1 *Conventional culture*

Ever since Pasteur and Koch introduced culture-based detection of bacteria during the 1800s, also referred to as the "golden age of microbiology"; culture has been the primary contributor to our microbial knowledge until the 1990s.[95,99] The primary aims of classical microbiology have been to isolate microbes using differential media and to identify them by their biochemical properties.[92] The major concern with regards to culture, however, has been its inability to cultivate the majority of GIT microbiota under standard laboratory conditions;[100–102] resulting from a gap in the knowledge of GIT microbiota's optimal environmental conditions as well as practical considerations regarding time and cost.[96] Conversely, recent advances in culture techniques have resulted in rephrasing the term "unculturable" to "as-yet uncultured" after recognising that the ability to culture is in actual fact dependent on the determination and innovation of the microbiologist.[103,104] "Culturomics" is a term introduced by Lagier et al. (2012) which is characterised as a high-throughput approach to culture what is now considered the "as-yet uncultured" components of the microbiota.[105]

1.2.1.2 *High-throughput culture or "culturomics"*

Despite the concern around unculturable bacteria, microbial culturing is still regarded as a valuable tool in microbiology. One of the advantages of microbial culturing over molecular techniques is that it allows for demonstrating mechanistic and experimental links between biological aspects of viable microbes (such as growth, metabolism and pathogenesis) and their host's physiology.[106,107] In addition to the information provided on the ecological role of microbial communities and their interactions on host health and disease states; culture-based methods are

also useful for obtaining complete microbial genome sequences used for analysing their genetic traits and evolution.[99,106,107]

Recognising the importance of culture, researchers are now making use of a new concept called "culturomics" for the extensive assessment of microbial profiles by high-throughput culture.[108] This high-throughput approach uses comprehensive culture conditions mimicking natural environments for large-scale isolation of GIT microbiota; together with matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry and/or 16S rRNA/*rpoB* gene amplification or sequencing to rapidly and accurately identify microorganisms.[109] In the face of it being labour intensive, this approach allows microbiologists to isolate functional and viable aerobic and anaerobic microbiota at large scale; and to discover new species never described previously in the human GIT.[105,107,110] The study by Lagier et al. (2012) identified 174 species which have not previously been described in the human gut and 31 species never described prior to culturomics.[105] In addition, high-throughput culture methods have also been able to demonstrate the impact of antibiotic pressure on gut microbiota.[110] Despite these findings, however, studies have shown discrepancies between bacteria identified using high-throughput sequencing and high-throughput culturing; and therefore these techniques are ideal complementary approaches in order to better understand our GIT microbiota repertoires.[105,111]

1.2.2 Culture-independent techniques

1.2.2.1 *Targeted approach*

The limitations that have been associated with culture greatly encouraged researchers to develop culture-independent techniques to study what were previously considered to be "unculturable" bacteria. Culture-independent techniques mainly rely on sequence variation within a specific gene which is amplified using polymerase chain reaction (PCR). The most commonly used bacterial target is the 16S rRNA gene which is highly conserved amongst all bacteria, however contains enough variation to distinguish species at lower taxonomic levels.[112,113] Despite its enormous contribution in the field of microbiology[114–116] and in particular GIT microbiology,[93,117,118] bacterial diversity analysis based on specific variable regions of the 16S rRNA gene are not without its own limitations. Some of the major areas of concern when studying the GIT microbiota and using 16S rRNA-based techniques are whether optimal DNA extraction was performed; which areas of the 16S rRNA gene should be targeted; and whether any PCR bias occurred.[119]

The importance of effective DNA isolation in whole community analysis has been demonstrated using mock communities as well as a range of specimens such as supragingival plaque, bronchoalveolar lavage, drinking water biofilms, colonic biopsies and faecal specimens.[84,120–123] Isolation of GIT microbiota from faecal specimens has important considerations to take into account

such as immensely variable microbial loads, high bacterial diversity (which includes both the Gram-positive and Gram-negative bacteria with hard and easy to lyse cell walls), inhibitory compounds such as complex carbohydrates, bile salts and bilirubins, as well as undigested material.[119] Also of concern when using culture-independent techniques is selection of the 16S rRNA hypervariable region (V1 - V9). These species-specific sequences constitute useful targets for microbial studies, however, whole community analyses is somewhat limited when targeting a single region due to their varying degrees of sequence diversity and inability to differentiate amongst all bacteria.[124] Moreover, the selection of primers targeting the V1-V2 regions of the 16S rRNA gene, for example, led Palmer et al. (2007) to the erroneous conclusion that Bifidobacteria were only minor components of infants studied throughout the first year of life.[125] Since PCR forms the basis of 16S rRNA-based molecular techniques, and bias in the PCR reaction may result in incorrect microbial population data, PCR bias also needs to be addressed. PCR biases may result from the presence of PCR inhibitors, variation in accessibility of the priming site, primer universality, the number of copies of the target gene, the formation of chimeras and decreased amplification rates for abundant PCR products in later PCR cycles.[119,126]

A wide range of 16S rRNA-based techniques have been used to study the GIT microbiota.[80,93,100,127–133] Sanger sequencing (which is still as a gold standard sequencing technique), now also referred to as “first-generation sequencing”, was used to generate a large proportion of the sequence data publically available for microbial identification.[134] Since its initiation in 1977 by Frederick Sanger,[135] this technology has been very successful in sequencing 16S rRNA genes of up to 1500 base pairs in length which enables researchers to study genera and species not identified using culture-based techniques.[134] However, despite its success in sequence-based studies, this technology is limited by the fact that purified DNA from a single species needs to be cloned into vectors and transformed into competent bacterial cells prior to sequencing. This makes the technology prone to cloning-biases, contributes to its expense and limits throughput. As the interest in microbiome projects grew, the demand for more time- and cost-effective techniques with higher throughput (known as high-throughput technologies) soon arose. The following section solely focuses on high-throughput sequencing technologies which has greatly contributed to our current understanding of GIT microbiota in human health and disease.[93]

High-throughput sequencing (second and third generation sequencing), also known as Next generation sequencing (NGS) or massively parallel sequencing, is responsible for an enormous increase in the number of 16S rRNA genes sequenced.[92] Compared to first generation sequencing, high-throughput sequencing does not require sub-cloning of bacterial DNA.[136–138] These technologies also have the ability to simultaneously sequence DNA from diverse environments containing various

bacterial species during a single sequencing reaction. By increasing the scale of operations and resultantly the number of nucleotides and copies of each nucleotide sequenced per run, it allows for a more cost and time effective approach to sequencing.[92,138] Second generation sequencing (SGS) primarily make use of three platforms.[139,140] The 454 GS FLX Pyrosequencer from Roche was the first NGS platform commercially introduced in 2004.[139–141] This sequencing-by-synthesis technology applies emulsion PCR[142] to beads containing a single template molecule which is followed by pyrosequencing.[138,143,144] The second SGS platform introduced commercially in 2006 is known as Illumina from the company Illumina (formerly known as Solexa).[139–141] This sequencing-by-synthesis technology applies bridge amplification[145] to single template molecules captured on a glass surface. This is followed by sequencing-by-synthesis during which nucleotide incorporation is detected by the use of reversible dye-labelled terminators.[136] The third SGS platform introduced commercially in 2007 was SOLiD from the company Life Technologies (formerly known as Applied Biosystems).[139–141] This sequencing-by-ligation technology uses the same bead and emulsion chemistry as 454 GS FLX pyrosequencing, however sequencing is initiated by a DNA ligase instead of polymerase.[136,146]

During the past four years, both 454 GS FLX pyrosequencing and Illumina sequencing of the 16S rRNA gene has been successfully used to sequence complex bacterial profiles from the human GIT.[132,147–154] These two techniques have provided great insight of our GIT bacterial communities, especially regarding the impact of dysbiosis on human health. Although both these techniques have their own advantages and limitations, Illumina sequencing seems to be less expensive, produces a lower sequencing error rate[155] and provides much better coverage which is essential for studying the highly diverse and complex GIT environment.[156] One limitation of Illumina sequencing (which has been its inability to target more than one hypervariable region)[157] has recently been resolved.[158] Indeed, recent studies showed that 454 pyrosequencing primers, targeting two hypervariable regions, can be modified to be compatible with Illumina technology.[158] In addition, a study by Ong et al. (2013) successfully developed a shotgun short-read sequencing approach for constructing and assembling 16S rRNA amplicon sequences spanning the V3-V6 hypervariable regions using Illumina technology.[159] Although a single hypervariable region (in spite of the recent advances in Illumina technology) has been and still is widely used when perfoming sequencing on the Illumina platform; Illumina technology is still the preferred choice due to its high accuracy and throughput when compared to 454 pyrosequencing.[158] However, to date, no consensus has been reached with regards to which hypervariable region to target using Illumina Miseq technologies.[160] Mizrahi-Man et al. (2013) recommended the use of the V3 and V4 regions for studying bacterial communities after performing *in silico* analysis of seven different amplicon designs,[160] while Ong et al. (2013) strongly recommends targeting the V3-V6 regions using a shotgun approach.[159] Of note, the V4 region is

regarded one of the more robust 16S rRNA hypervariable regions in terms of reliable taxonomic classification.[161] The V4 region also provides species richness estimates similar to that of full length sequences;[162,163] and has been targeted by a number of studies assessing the GIT microbiota by way of Illumina sequencing.[88,163–167]

1.2.2.2 *Non-targeted approach*

Third generation NGS technology involves high-throughput sequencing of all microbial genes without prior amplification of the target gene.[140] Unlike second generation, third generation sequencing (TGS) do not rely on base-specific signalling and imaging resulting from numerous enzymatic reaction cycles.[168] TGS relies on direct reading of a single molecule of DNA, also referred to as single molecule sequencing (SMS).[169] This has allowed for circumventing some of the major limitations of second generation technologies such as PCR bias, sequencing errors and short reads related to PCR amplification and/or "wash-and-scan" steps. Other advantages of TGS technologies include higher throughput, shorter run times, reduced cost and the use of small amounts of starting material.[169] These TGS or SMS technologies can be grouped into three categories, each providing novel approaches to the field of DNA sequencing.[168,169] The SMRT sequencer from Pacific Biosciences, for example, uses a sequencing-by-synthesis approach (as SGS does); however the SMRT sequencer synthesizes single molecules of DNA which results in single molecules of DNA polymerase being detected. Another example is the Nanopore sequencing technologies by Oxford Nanopore Technologies, which detects individual nucleotides as single molecules of DNA are positioned in the vicinity of a nanopore or threaded through one.[170] A major limitation of the latter, however, is the anticipated error rate versus throughput tradeoff.[171]

Another non-targeted approach to sequencing is whole genome shotgun (WGS) sequencing which has been recognized as the most comprehensive method for metagenomics investigations.[172] Advantages of WGS sequencing approaches include no prior amplification of a target gene (producing less biased results compared to PCR depended approaches);[173] as well as the ability to survey an entire genome (providing a complete profile of the whole community studied). Moreover, WGS sequencing approaches not only provides taxonomic information but also offers direct information regarding the community's functions.[173] Despite these advances, limitations of shotgun sequencing experiments include increased cost and the need for greater computational resources due to the number of sequences generated.[174] Typically, 2-10 Gb of sequences are generated per sample when performing WGS on paired-end generating platforms such as the Illumina HiSeq, potentially costing hundreds of dollars per specimen.[174] However, performing WGS sequencing on Illumina sequencing platforms offers the great advantage of identification of species-level taxonomy.[172]

Table 1. Advantages and limitations associated with culture-dependent and -independent techniques used to study the gastrointestinal tract microbiota

Category	Techniques / technology used to study GIT microbiota	16S rRNA based	Description	Advantages		Limitations	
				Shared between techniques	Specific to technique	Shared between techniques	Specific to technique
Culture dependent							
Conventional culture	Culture[175]	No[94]	Selective media is used to cultivate specific bacteria[95,104]	Provides the opportunity to study functional characteristics[94,99] Provides the opportunity to obtain complete genome sequences[99] Detects viable bacteria[99,108]	Inexpensive[94] Limited equipment needed[176]	Culture results are dependent on the conditions used[108]	Large number of bacteria is considered as "unculturable" using standard laboratory procedures[94]
High-throughput culture	Culturomics[105]	No[94]	High-throughput isolation of bacteria using various selective and/or enrichment culture conditions[105] and identification using MALDI-TOF or 16S rRNA/*rpoB* amplification/sequencing[109]		Cultivation of previously considered "unculturable" species[108] Potential to detect low abundant species[105]		Very labour intensive and time consuming[108]
Culture independent							
Polymerase chain reaction	Quantitative real-time PCR[127]	Yes[94]	Amplification and quantification of specific bacterial groups/species by targeting specific regions of the 16S rRNA gene. The concentration of the amplicon is measured as amplification occurs[95,177]	Detection of specific bacterial groups/species from complex communities[94]	Reliably quantifies the amount of DNA targeted[94,95] Phylogenetic identification[95]	Primer- or probe-based targeting which results in detection of previously identified bacteria only[94,95] No community-wide assessment of all bacteria present in a specimen[94] Optimal DNA extraction is needed as differential lysis of bacterial cells will influence the microbial composition detected[95] PCR bias[95]	
Fingerprinting techniques	DGGE[80]	Yes[94]	16S rRNA amplicons are denatured based on their nucleotide sequences using a denaturing gradient gel[95,178]		Inexpensive tool for diversity studies[94] Semi-quantitative[95] Communities from different treatment groups are easily compared[94]		Poor reproducibility[94] Lack of direct phylogenetic identification[95]
	T-RFLP[128,179]	Yes[94]	Fluorescently labelled 16S rRNA amplicons are digested using a restriction endonuclease, followed by digital detection of fluorescently labelled fragments on a gel[95,180]		Inexpensive tool for diversity studies[94] Semi-quantitative[95] Communities from different treatment groups are easily compared[94] Reproducible[94]		Lack of direct phylogenetic identification[95]

Table 1. Advantages and limitations associated with culture-dependent and -independent techniques used to study the gastrointestinal tract microbiota (continued)

Category	Techniques / technology used to study GIT microbiota	16S rRNA based	Description	Advantages		Limitations	
				Shared between techniques	Specific to technique	Shared between techniques	Specific to technique
Culture independent (continued)							
Probe hybridization	FISH[129]	Yes[94]	Fluorescently labelled oligonucleotide probes are designed to hybridize target bacteria 16S rRNA sequences[95]	Detection of specific bacterial groups/species from complex communities[94]	Semi-quantitative[95] Phylogenetic identification[95]	Primer- or probe-based targeting resulting in detection of previously identified bacteria only[94,95] No community-wide assessment of all bacteria present in a specimen[94] Optimal DNA extraction is needed as differential lysis of bacterial cells will influence the microbial composition detected[95] PCR bias[95]	Cross-hybridization[94,95] Difficult to detect low-level species[95] Only detects sequences contained on the chip (uncharacterized phylotypes are not detected)[94]
	DNA microarrays[130]	Yes[94]	Fluorescence is detected during hybridization of fluorescently labelled 16S rRNA amplicons with oligonucleotide probes spotted on a glass slide[95,181]		High-throughput detection of specific bacterial groups/species from complex communities[95] Semi-quantitative[95] Phylogenetic identification[95] Commonly used to compare microbiota between different populations[95]		
Sequencing (First generation)	Sanger sequencing[100]	Yes[94]	Cloning of the full length 16S rRNA amplicon occurs prior to sequencing, followed by detection of nucleotide sequences using capillary electrophoresis[95,136]		Increased taxonomic resolution from the 16S rRNA gene[94]	Optimal DNA extraction is needed as differential lysis of bacterial cells will influence the microbial composition detected[95] PCR bias[95] Shorter read lengths compared to Third generation sequencing[169]	Higher cost (compared to other high-throughput sequencing technologies)[92] A clone library insert is needed to sequence the full length 16S rRNA gene[92]
High-throughput sequencing (Second generation)	454 GS FLX pyro-sequencing[131]	Yes[94]	Emulsion PCR of 16S rRNA gene and pyrosequencing[95,136,138,146]	High throughput and low cost per base (compared to First and Third generation sequencing)[169]	Longer read lengths, high numbers of bases aligned and lowest number of duplicate reads (compared to Illumina and SOLiD)[182] Shorter run time and bigger insert size (compared to Illumina)[92] Up to three hypervariable regions per read[157]		Lower number of reads and bases per run compared to other Second generation technologies[92,182] Higher error rates compared to Illumina sequencing[92] Extensive bioinformatics analysis[94,95]

Table 1. Advantages and limitations associated with culture-dependent and -independent techniques used to study the gastrointestinal tract microbiota (continued)

Category	Techniques / technology used to study GIT microbiota	16S rRNA based	Description	Advantages		Limitations	
				Shared between techniques	Specific to technique	Shared between techniques	Specific to technique
Culture independent (continued)							
High-throughput sequencing (Second generation)	Illumina (GAIIx, HiSeq and MiSeq)[132]	Yes[94]	Bridge amplification of 16S rRNA gene using reversible dye-labelled terminators[136,146]	High throughput (compared to First and Third generation sequencing)[169] Low cost per base (compared to First and Third generation sequencing)[169]	Higher number of reads and bases per run (compared to 454 GS FLX pyrosequencing)[92] Lower cost factor and error rates (compared to 454 GS FLX pyrosequencing)[92]	Optimal DNA extraction is needed as differential lysis of bacterial cells will influence the microbial composition detected[95] PCR bias[95] Extensive bioinformatics analysis[94,95] Shorter read lengths (compared to Third generation sequencing)[169]	Lower number of reads and bases per run (compared to SOLiD)[92,182] Shorter read lengths and insert size (compared 454 GS FLX pyrosequencing)[92] Longer run time (compared to 454 GS FLX pyrosequencing)[92] Only one hypervariable region per read[157]
	SOLiD[154]	Yes[94]	Emulsion amplification of 16S rRNA gene followed by sequencing initiated by a DNA ligase instead of polymerase[136,146]		Higher number of reads and bases per run (compared to other Second generation technologies)[182]		Shorter read lengths (compared to other Second generation technologies)[182] Lowest number of reads and bases aligned (compared to other Second generation technologies)[182] Highest number of duplicate reads (compared to other Second generation technologies)[182]

Table 1. Advantages and limitations associated with culture-dependent and -independent techniques used to study the gastrointestinal tract microbiota (continued)

Category	Techniques / technology used to study GIT microbiota	16S rRNA based	Description	Advantages		Limitations	
				Shared between techniques	Specific to technique	Shared between techniques	Specific to technique
Culture independent (continued)							
High-throughput sequencing (Third generation)	SMRT sequencer from Pacific Biosystems	No[168]	Uses an enzymatic template replication system (as second generation sequencing does), however it resolves the problem of detecting the incorporation of a single nucleotide against a large pool of potential nucleotides by using zero-mode waveguide technology[168,169]	High throughput community-wide assessment of bacteria present in a specimen[94] Longer reads and bigger insert size (compared to First and Second generation sequencing)[92,169] Shorter run time (compared to Second generation technologies)[92] No PCR bias[168] Lower cost per run (compared to Second generation technologies)[169]	No PCR bias[139] Real-time detection of base incorporation against a large pool of potential nucleotides[169] Long read lengths[169]	Optimal DNA extraction is needed as differential lysis of bacterial cells will influence the microbial composition detected[95] Moderate raw read accuracy (compared to First and Second generation sequencing)[169]	High error rates[169] Lower throughput (compared to Second generation technologies)[169] May require more advanced data analyses[169]
	Oxford Nanopore Technologies	No[168]	A single strand of DNA is passed through the nanopore which results in the development of a residual ionic current. This current is dependent on the nucleotide passing through the nanopore at that time. This results in the sequence of bases in the DNA molecule being captured by the change in ionic current[183]		No PCR bias[183] Label free single-molecule approach which can be scaled for high-throughput analysis[183] Long read lengths[169]		Temporal resolution for detecting individual nucleotides with high sensitivity is problematic for currently available optical and electrical technologies[183] Trade-offs agains error rates and throughput[171]
High-throughput sequencing (Whole genome shotgun)	Illumina (second generation) or Helicos (third generation)[172]	No[173]	The entire genome of a microbe of interest is randomly sheared into millions of DNA fragments which are sequenced. These fragments are then aligned against one another to form continueous sequences called contigs and mapped to a reference genome[174]		No gene restrictions; less bias compared to target-based approaches; more comprehensive insight into the genomic and organismal composition (including gene and species information and metabolic potential); and can be screened for target genes[174]		Undersampling of low-abundance organisms may result in incomplete assemblies of genomes[174] High costs for sequencing, computational analysis and storage[174]

DGGE: Denaturing gradient gel electrophoresis; DNA: deoxyribonucleic acid; FISH: Fluorescence in situ hybridization; MALDI-TOF: Matrix-assisted laser desorption ionization time-of-flight mass spectrometry; PCR: Polymerase chain reaction; rRNA: ribosomal ribonucleic acid; T-RFLP: Terminal restriction fragment length polymorphism

1.3 Is there a core gastrointestinal microbiome?

As discussed above, the GIT microbiota can be considered as a "virtual organ" due to its role in human physiology and health,[6] its host-dependent metabolic functions,[3] its role in energy consumption and redistribution,[184] and its ability to maintain and repair itself.[185] However, beneficial properties of this "virtual organ" rely on the composition of the microbiota, since changes in intestinal microbiota profiles have been associated with disease states.[186,187] Understanding the composition of a "healthy" GIT microbiota, as well as factors influencing its composition, may aid in the design of therapies to modulate and maintain healthy profiles.[98]

Despite the promising prospects in understanding the composition of a "healthy" GIT microbiota, studying these communities are rather complex. Not only does the adult human GIT harbour around 10^{14} microbial cells,[3,85,184,188,189] but considerable inter-individual variation has also been described.[100,190,191] This, together with fluctuation over time[191] and a wide range of factors influencing microbial profiles (such as age, diet, genetics, environment as well as personalised medicines),[98] have made it difficult to define a "healthy GIT microbiota profile". In addition to the complexity of microbial profiles, bacterial cells are not the sole inhabitants of the GIT. The GIT virome (constituting around 10^9 viral particles per gram of faeces),[192] for example, also reflects high inter-individual diversity and is influenced by factors such as diet.[193] These viruses my transmit genes in the form of bacteriophages to their bacterial hosts, conferring antibiotic resistance, increased pathogenicity and perhaps even new metabolic capacity.[193–196] Interestingly, recent studies are also suggesting beneficial effects of enteric viruses for the host.[197,198] The study by Kernbauer and colleagues,[199] for example, clearly showed that the murine norovirus (a mammalian enteric virus) restored abnormal intestinal morphology and aberrant immune development of lymphocytes in germ-free and antibiotic treated mice. This may affect outcomes when defining inter-relationships between microbial community structure and function, and physiological states of the host.[196] In addition, interactions have also been reported between the GIT microbiome and the GIT mycobiome (constituting of 10^6 micro-organisms per gram of faeces), which impacts on the pathophysiology of GIT inflammatory diseases.[200] Despite its complexity, however, studies have postulated that a core intestinal microbiome does exist, representing a conserved set of health-related symbionts which have coevolved with humans.[87,201]

A core intestinal microbiome has been described as "the stable part of the microbiota present within the vast majority of human beings"[202,203] as well as "the number and types of bacteria that are shared among different individuals".[201] Microbiome studies using high-throughput sequencing and micro-array analyses have demonstrated that two major butyrate-producing phyla,

the Firmicutes and the Bacteroidetes, are found across healthy adult intestinal tracts and represents around 90 % of the gut bacterial population (Figure 3).[1,79,87,100,191,204,205] Other phyla also found in the intestinal tracts of healthy adults, but with reduced abundance, are Actinobacteria, Proteobacteria and Verrucomicrobia.[87,100,202,205] Even though these findings suggest a shared phylogenetic core between healthy individuals at phylum-level; challenges in defining a common core at lower taxonomic levels may be explained by the fact that humans harbour more than 1,000 species-level phylotypes. In addition, each of these phylotypes consists of clusters of sequences with high 16S rRNA diversity.[98]

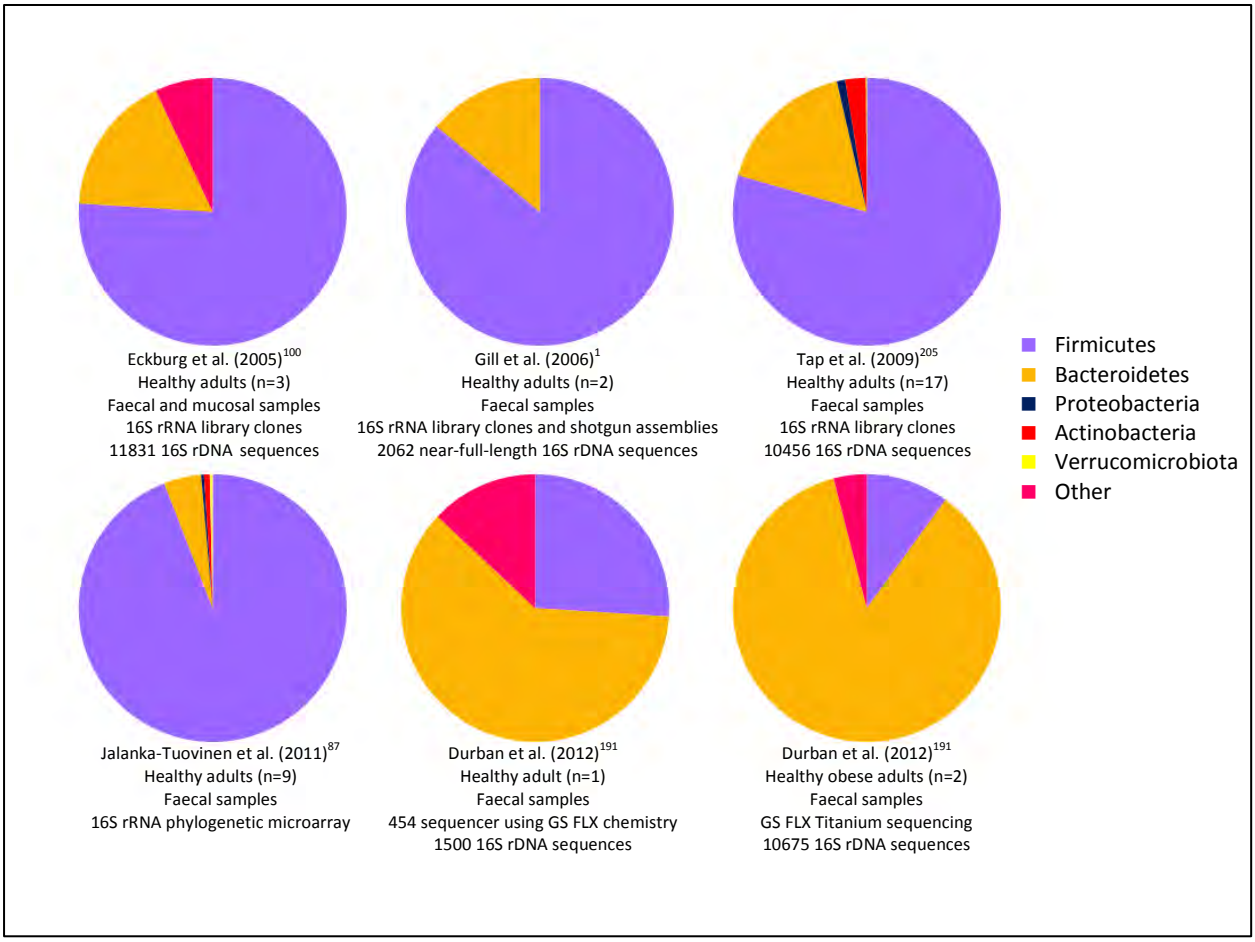


Figure 3. Major gastrointestinal phyla identified by using different techniques such as high-throughput 16S rRNA gene sequencing and phylogenetic microarray analysis

Despite the high inter-individual diversity observed at species-level; individuals may in actual fact share a small subset of species commonly detected in healthy adult faecal specimens.[152,205] Studies have also shown that species-level phylotypes have unique functional properties and could be crucial role players in the microbial community and may potentially contribute to the host's health or diseases states.[71,80–83,206] It is clear from previous sections of this review that specific

bacterial species or members from specific genera or phyla are responsible for protection against pathogens,[20,27,28,207] elicit systemic immune responses[55,58,60,61] and provide metabolic functions of the host.[208] Early life GIT colonizers are regarded as essential pioneers of these processes,[209] and studies have shown that the development of GIT microbiota during infancy may hold profound consequences for health later in life.[210–215] The following section describes the GIT profiles during early life; how these profiles are influenced and developed; and their importance in health and disease.

1.4 The gastrointestinal microbiota early in life

The development of the infant GIT microbiota is influenced by a number of factors such as mode of delivery, feeding practices, host genetics, the use of pre-, pro- and antibiotics, geographic location and lifestyle.[216,217] Since the GIT has been considered sterile in utero,[218,219] studies have focused primarily on factors influencing these early profiles from the time of birth.[220] Recent studies now suggest in-utero colonization of the GIT[221] which could be a key contributor to microbial profiles of the infant GIT later in life.[222,223] In support of this, studies have suggested that initial colonizers prepare the environment for later colonizers such as obligate anaerobes.[224] The following sections will describe current knowledge around the composition of the infant GIT microbiota, factors associated with its development and findings regarding in-utero colonization and the meconium microbiota.

1.4.1 The composition of the GIT microbiota in infants

An understanding of the GIT microbiota early in life has been established by both culture-dependent and -independent studies. Conventional culture-based studies on healthy breast- and bottle-fed infants during the 20th century have shown that aerobic and facultative anaerobic bacteria are predominantly the first colonizers of the infant GIT.[225,226] *Enterococcus*, *Staphylococcus* and *Streptococcus* spp. within the phylum Firmicutes; as well as *Enterobacter* spp. within the phylum Proteobacteria have been identified as the pioneers in GIT colonization during the first week of life.[225,227] Anaerobes, such as *Bifidobacterium* spp. within the phylum Actinobacteria; *Bacteroides* spp. within the phylum Bacteroidetes; and *Clostridium* spp. within to the phylum Firmicutes are soon to follow during the first and second week of life.[225,227] Bifidobacteria rapidly reach high levels during this time and have been considered the most prevalent genus at the age of one month.[226,228]

Culture-independent studies have shown similar findings with Bifidobacterium, Clostridium and Bacteroides as the dominant anaerobes.[224]

Even though bacterial diversity seems to be low shortly after birth with only a select group of facultative bacteria and anaerobes identified, a study using a 16S rRNA microarray approach showed that bacterial profiles quickly diversify after the first few days of life.[229] Up until six months of age, infant intestinal bacteria cluster by individuals and inter-individual diversity is much greater compared to changes observed over time.[229] Although still distinct, by the end of the first year of life, bacterial profiles had transitioned to an "adult-like" profile, probably due to the introduction of solid foods.[229] These profiles were largely comprised of the main phyla, Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria and Verrucomicrobiota, commonly inhabiting the adult GIT.

Despite these findings, conflicting data have also been reported around the composition of GIT microbiota in infants.[226,228–234] A good example is the controversy around the colonisation of the genus Bifidobacteria (well known for the health benefits it provides to the host).[230] For many years, culture-based and species-specific DNA targeted approaches have suggested that Bifidobacteria constitute the dominant component of the infant GIT microbiota.[226,228,231–233] However, this does not always seem to be the case as considerable variation in Bifidobacterial colonization has been reported by other studies.[229,234] For example, using microarray technology a study by Palmer et al. (2007) concluded that Bifidobacteria constituted only minor components of faecal microbiota at any age ranging from birth to adulthood.[229] In contrast with the findings from Palmer and colleagues,[229] a study using GS-FLX Titanium pyrosequencing reported a predominance in Bifidobacteria from the infantile gut.[234] These results indicate how the use of different techniques could result in variation in study outcomes, as highlighted by reports on optimised protocols for Bifidobacterial detection.[235,236] In addition to the impact of techniques used to study Bifidobacterial colonization, studies have also shown that environmental factors such as breast- and formula-feeding influence Bifidobacterial profiles in infants.[232,237,238] A second example of conflicting data around the composition of GIT microbiota in infants was sparked by Jost et al. (2012).[239] This study used more than one screening method (culture and high-throughput) when concluding that anaerobes outnumbered facultative anaerobes in the first days of life, thereby interrogating the "colonization dogma".[239] A third example around conflicting data on the composition of GIT microbiota results from studies reporting different ages at which an adult-like microbiome is achieved, ranging from the end of the first year of life, to around two and a half years of age.[229,240,241]

Nonetheless, an important concept to consider around conflicting data on GIT microbiota in infants is the issue of external factors which may influence these profiles (as mentioned for Bifidobacteria above). Controversy around GIT microbiota in infants may be explained by early life

maternal and environmental factors, such as mode of delivery, early life feeding practices, prematurity, hospitalisation, antibiotic use and older siblings.[218] Some of these influential factors will be described in more detail in the following section.

1.4.2 Early life determinants of the infant GIT microbiota

Although a number of factors have been associated with changes in GIT microbiota profiles, the first encounters with bacteria that colonize the GIT occurs in mainly four stages.[125,242] Recent data suggests that initial colonization begins even before birth; referred to as in utero-colonization.[221,243,244] This is followed by the second colonization stage; referred to as the process of birth (previously considered as the initial colonization stage).[219,223,245–249] The third colonization stage is referred to as the period of exclusive milk-feeding (previously considered the second colonization stage).[250] Finally, the fourth stage of colonization starts when solid food is introduced to the diet.[250] Each of these stages has been associated with changes in early life GIT microbiota profiles which may have consequences for health and disease (Table 2).

Table 2. Factors influencing the composition of the gastrointestinal tract microbiota and their relation with health and disease

Colonization stage	Influencing factor	Impact on GIT microbiota of infants	Impact on health and disease
Colonization during birth	Caesarean section delivery	↓ Microbial diversity[251] ↓ Bacteroides[251]	↓ Th1 response[251]
		↑ Clostridium cluster I[252]	↑ Risk for atopic dermatitis[252]
		↑ *Clostridium difficile*[246,253]	↑ *C. difficile* colonization is associated with increased risk for eczema, recurrent wheeze, atopic dermatitis[215]
		↑ *C. difficile*[254]	↑ Risk for wheeze, eczema and asthma[254]
			Caesarean section delivery results in an increased risk for asthma[255]
		↓ Bifidobacteria[246,253,256]	↓ Bifidobacteria colonization is associated with increased risk for irritable bowel disease[257,258]
		↑ *Escherichia coli*[253]	↑ *E. coli* colonization is associated with increased risk for Crohn's Disease[257] and atopic sensitization[259]
Colonization during exclusive milk feeding	Formula feeding	↑ Rumnicoccus[260] ↓ Lactobacillus[260]	↑ Formula feeding was associated with and increase in weight gain and higher serum insulin[260]
		↑ Serratia and Lactococcus[261]	↑ Formula feeding may result in an increased vulnerability to develop disease[261]
		↑ *Bacerteroides fragilis*, *Clostridium coccoides*, *E. coli* and *Bifidobacterium lactis*[262]	↑ Staphylococcus and *B. fragilis* colonization results in an increased risk for coeliac disease[262]
		↑ *C. difficile*[263,264]	↑ Risk for asthma[215,254]
		↓ Bifidobacteria[233,263]	↓ Bifidobacteria colonization is associated with increased risk for inflammatory immune responses[265]
Colonization during weaning	Introduction of solid food	↑ Bacteroidetes[241]	↑ Bacteroidetes colonization is associated with increased production of short chain fatty acids and decreased risk for obesity[241]

Table 2. Factors influencing the composition of the gastrointestinal tract microbiota and their relation with health and disease (continued)

Colonization stage	Influencing factor	Impact on GIT microbiota of infants	Impact on health and disease
Other influential factors	Antibiotics	↓ Microbial diversity[266]	↓ Microbial diversity is associated with an increased risk for asthma[266]
		↓ Micorbial diversity[267]	↓ Microbial diversity is associated with an increased risk for necrotizing enterocolitis[267]
		↑ Enterococci (overgrowth)[268]	↑ Enterococci colonization is associated with an increased risk for necrotizing enterocolitis[269]
		↓ Commensal bacteria[263] ↑ *C. difficile*[263]	↑ Risk for asthma[263]
	Probiotics	↑ Specific Lactobacillus strains[270,271]	↑ Weight gain and loss[270] ↑ Th1 allergic responses[271]
		↑ *Lactobacillus rhamnosus* and *Bifidobacterium lactis*[272]	↓ Allergic sensitization and airway disease[272]
		↑ Commensal bacteria and specific probiotic strains[273]	↑ Intestinal host defence by maturing intestinal barrier function and reducing inflammatory signals[273]
		↓ *Clostridium perfringens*[207]	↓ Risk for necrotising enterocolitis[207]
		↑ Lactobacilli and Bifidobacteria[274]	↓ Eczema and atopic eczema[274]

GIT: gastrointestinal tract; TH1: Type 1 T-helper cells; ↓: Descrease; ↑: Increase

1.4.2.1 *In-utero colonization*

Interestingly, studies conducted over the past decade have now suggested that bacterial colonization of the infant GIT occurs even before birth. This is referred to as the in-utero colonization stage during which colonization of the infant GIT microbiota occurs in the previously considered "sterile" maternal womb.[243] Pioneer species colonising the GIT before birth may play an essential role in shaping the GIT environment for successive bacterial colonization[224,244] and may also play a role in much earlier immune system development than previously considered.[275,276] The suggestion of an in-utero bacterial colonization stage originated from microbial detection in meconium specimens, placenta, umbilical cord blood, amniotic fluid and fetal membranes.[244,275–279] New perspectives on GIT colonization prior to birth, with emphasis on the meconium microbiota, will be the main focus of Section 1.4.3.

1.4.2.2 *Colonization during birth*

Mode of delivery has been identified as a key determinant factor of the newborn's GIT microbiota. Vaginally delivered infants are colonized with microbes similar to those in the maternal vaginal microbiota.[248] These include *Bifidobacterium*, *Bacteroides*, *Lactobacillus*, *Prevotella* and *Atopobium* spp..[247,248,280,281] In contrast, Ceasarean delivered infants have slower diversifying microbiota.[282] GIT microbial communities in Caesarean section delivered infants resemble those of the maternal skin with *Corynebacterium*, *Staphylococcus* and *Propionibacterium* spp. predominating.[248] In addition, high levels of *Clostridium difficile* are characteristic in these infants, together with low levels of *Bifidobacterium, Lactobacillus* and *Bacteroides* spp..[247,264,280,283] Of note, birth by Caesarean section has been associated with a number of disease states such as allergy and obesity.[213,284,285]

1.4.2.3 Colonization during exclusive milk feeding

Breast milk is a rich source of diverse bacteria[286] and several studies have shown mother-to-infant transfer of breast milk bacterial strains.[287–290] Despite breast milk harbouring a diverse collection of bacteria, a "core" breast milk microbiome has been identified by the use of pyrosequencing technology.[291] Members from the genera Bifidobacterium, Lactobacillus, Bacteroides, Prevotella, Staphylococcus and Streptococcus are common colonizers of breastfed infants, however Bifidobacterium is considered as the dominant GIT inhabitant of these infants.[232,260,292,293] Breast milk bacteria have been suggested to contribute to important functions in the infant GIT such as immune system modulation[294] and protection against infection.[295,296] In addition, breast milk oligosaccharides act as prebiotics which is of benefit to specific intestinal bacteria.[297,298] In contrast to breastfed infants, the GIT microbiota of formula fed infants is less

dominated by Bifidobacteria.[232,292,293,299,300] In addition, formula-fed infants have a more diverse microbiome[280] with higher Enterococcus,[301] Clostridia,[232,301] Bacteroides[232,300] and Enterobacteriaceae[292,299] counts compared to infants being breastfed. Formula-fed induced changes in intestinal microbiota have been hypothesized to contribute to adulthood diseases such as inflammatory diseases, obesity, cardiovascular disease and insulin resistance.[260,302]

1.4.2.4 *Colonization during the introduction of solid foods*

The introduction of solid food is another factor which adds to the succession of the infant's GIT microbiota. Solid foods have been shown to promote shifts from the highly variable infant GIT microbiota profiles towards more stable adult-like profiles.[229,303] A species-specific probe-based study showed a significant decrease in Bifidobacteria after the introduction of solid foods, as well as an increase in the overall microbial diversity.[304] Another shift towards adult-like microbiota profiles is evident from the increase in strictly anaerobic Clostridial proportions, while a significant reduction is evident for facultative anaerobes (conventionally considered the first colonizers of the infant's GIT).[304] Bifidobacteria and Bacteroides, however, still dominated the GIT of infants, which is in agreement with results from previous studies.[304] A metagenomic study showed a sharp, but sustained, increase in the abundance of the Bacteroidetes phylum after the introduction of solid foods.[241] This sharp increase in Bacteroidetes, their ability to break down complex plant polysaccharides,[305] and the GIT microbiome's rapid response to altered diets,[306] may explain the role of GIT microbiota in preparing the infant for adult-like solid foods.

Despite the changes in the proportion of bacteria during weaning, many of the primary influences on GIT microbiota profiles are still evident following the introduction of solid foods. Fallani et al. (2011) reported that the influence of mode of delivery, geographical region and pre-weaning feeding persisted after weaning.[304] This study, for example, highlighted the effect of exclusive milk feeding on subsequent faecal microbial profiles by comparing the GIT microbiota profiles of formula-fed infants to breastfed infants. Moreover, Fallani and colleagues[304] reported that compared to formula-fed infants, breastfed infants displayed a rapid reduction of *C. perfringens* and *C. difficile* species, together with a lagging increase of *C. leptum* after the introduction of solid foods. In addition to these findings, the effect of breastfeeding prior to weaning was reported by another study showing a sharp increase in the counts of *Bacteroides spp*., Clostridia, Enterobacteria, Enterococci and Streptococci in breastfed infants receiving solid food.[307] These changes were not observed for infants receiving formula.[307] Another example of primary influential factors on faecal microbial profiles later in life is geographical origin. Fallani et al. (2012) showed that bacterial profiles associated with geographical origin, prior to weaning, still persisted four weeks after the introduction of solid foods.[304] These findings highlight that changes in the microbiota are indeed

associated with weaning; however initial factors such as exclusive milk feeding and geographical origin seem to be key determinants in faecal microbial profiles later in life.

1.4.3 New perspectives on GIT microbiota profiling: The meconium microbiota

The concept of a sterile womb and fetus has drastically changed with recent studies describing the isolation of bacteria from infant meconium specimens, which were conventionally considered to be sterile.[244] This paradigm shift from infants "being born sterile" to "being born dirty"[308] has been supported by bacteria detected from the placenta, amniotic fluid, fetal membranes and umbilical cord blood.[275–279] Interestingly, all of these specimen types have been shown to contain bacteria common to the GIT, which further suggests in-utero colonization of the fetal GIT. In addition, murine models have confirmed in-utero mother-to-offspring bacterial transfer. Jimenez et al. (2005) inoculated pregnant mice with a genetically labelled bacterial strain and isolated the labelled strain from the amniotic fluid of the inoculated mice.[279] This group conducted the same experiment a few years later, however this time isolated the labelled bacterial strain from the meconium of pups born to inoculated mice.[244] In both these studies, this labelled bacterial strain could not be detected in non-inoculated control groups, supporting the in-utero transfer of bacteria from the maternal GIT to the GIT of the developing infant.

Despite the evidence of in-utero bacterial transfer, the process by which this occurs is still not clear (Figure 4).[242] One hypothesis is that ascending microbiota of the vagina and cervics colonize the "sterile" infantile gut.[309] This is supported by studies on preterm birth, showing microbial invasion of amniotic fluid by vaginal bacteria.[310,311] Another hypothesis for mother-to fetus transfer, not necessarily associated with preterm birth, is bacterial dissemination in the maternal bloodstream followed by transplacental invasion.[309,312] Hematogenous transfer has been suggested to occur when maternal oral bacteria enters the bloodstream during periodontal disease states.[312–315] Another possibility of hematogenous transfer is that bacteria from the maternal GIT lumen enter the maternal systemic circulation as a result of bacterial translocation.[276,311,316–318]

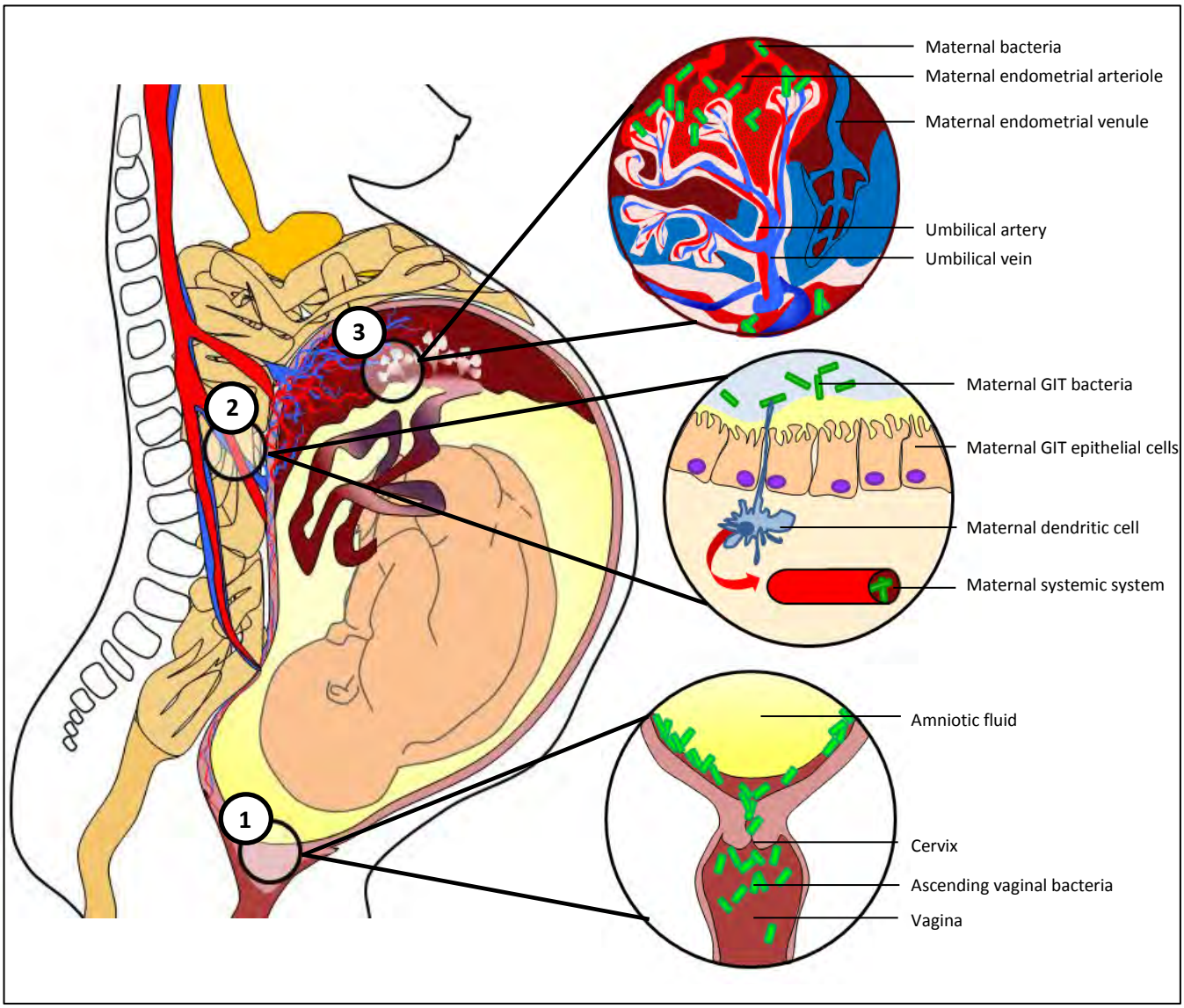


Figure 4. Schematic representation of proposed in-utero mother-to-fetus efflux of maternal microbiota
1. In-utero gastrointestinal colonization may result from vaginal and cervical microbiota migrating into the amniotic fluid. These ascending bacteria cross the cervix into the amniotic fluid which is swallowed by the fetus. Maternal bacteria may also disseminate in the maternal bloodstream which is followed by transplacental invasion. **2.** One possibility of hematogenous bacterial transfer is when maternal gastrointestinal (GIT) bacteria enter the maternal systemic circulation. Maternal GIT microbiota may be transported from the intestinal lumen via extending dendrites of the dendritic cells (DCs). These antigen presenting cells transfer luminal bacteria into the maternal systemic circulation where it can be transferred transplacentally to the fetus. **3.** Maternal GIT bacteria are transferred transplacentally from the endometrial arterioles to the umbilical arterioles and successively reach the fetus.

Interestingly, both facultative and strict anaerobes were identified in meconium of newborns.[256] This is in contrast with the "colonization dogma" which suggests that facultative anaerobes are the pioneer bacteria of the newborn GIT. Despite these contradictory findings, the findings by Tsuji and colleagues[256] are in support of the findings by Jost and colleagues[239] which also interrogated the "colonization dogma" by showing that anaerobes outnumbered facultative anaerobes in the first days of life.[239] In addition, strictly anaerobic bacterial strains from the bacterium *Bifidobacterium longum subsp. longum* have also been isolated from meconium specimens.[317] In support of the "colonization dogma", however; a study using high-throughput sequencing identified predominantly Proteobacteria and Firmicutes in the meconium of healthy neonates.[222] Proteobacteria were represented by the family Enterobacteriaceae of which the genera *Escherichia* and *Shigella* were most abundant. Firmicutes were mainly represented by lactic acid bacteria of which the most abundant genera were *Leuconostoc*, *Enterococcus*, *Lactococcus*, *Staphylococcus* and *Streptococcus*.[222] Of note, differences in meconium microbiota profiles have been shown for infants born at a lower gestational age; infants born to mothers using antibiotics or probiotics during pregnancy; infants born to diabetic mothers; and infants exposed to prolonged rupture of membranes in-utero.[269,275,319,320] These changes may influence successive bacterial colonisation, immune system maturation and may also contribute to both early and later life disease.[277,321–323] Knowledge of meconium microbiota profiles may allow in-utero manipulation in order to provide a newborn with the healthiest start to life.[275,308]

References

1. Gill SR, Pop M, Deboy RT, et al. Metagenomic analysis of the human distal gut microbiome. *Science*. 2006;312(5778):1355-9. doi:10.1126/science.1124234.
2. Smith K, Mccoy KD, Macpherson AJ. Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Semin Immunol*. 2007;19:59-69. doi:10.1016/j.smim.2006.10.002.
3. Evans JM, Morris LS, Marchesi JR. The gut microbiome: the role of a virtual organ in the endocrinology of the host. *J Endocrinol*. 2013;218(3):R37-47. doi:10.1530/JOE-13-0131.
4. Quigley EMM. Gut bacteria in health and disease. *Gastroenterol Hepatol (N Y)*. 2013;9(9):560-9.
5. Quigley EMM. " Gut Microbes - Importance in Health and Disease " World Digestive Health Day 2014. *South African Gastroenterol Rev*. 2014;(April):14-18.
6. Kinross JM, Darzi AW, Nicholson JK. Gut microbiome-host interactions in health and disease. *Genome Med*. 2011;3:14.
7. Possemiers S, Bolca S, Verstraete W, Heyerick A. The intestinal microbiome: a separate organ inside the body with the metabolic potential to influence the bioactivity of botanicals. *Fitoterapia*. 2011;82(1):53-66. doi:10.1016/j.fitote.2010.07.012.
8. Foxman B, Goldberg D, Murdock C, Xi C, Gilsdorf JR. Conceptualizing human microbiota: from multicelled organ to ecological community. *Interdiscip Perspect Infect Dis*. 2008;2008:613979. doi:10.1155/2008/613979.
9. Prakash S, Rodes L, Coussa-Charley M, Tomaro-Duchesneau C. Gut microbiota: next frontier in understanding human health and development of biotherapeutics. *Biologics*. 2011;5:71-86. doi:10.2147/BTT.S19099.
10. Goto Y, Kiyono H. Epithelial barrier : an interface for the cross-communication between gut flora and immune system. *Immunol Rev*. 2012;245:147-163.
11. Leser TD, Mølbak L. Better living through microbial action : the benefits of the mammalian gastrointestinal microbiota on the host. *Environ Microbiol*. 2009;11(9):2194-2206. doi:10.1111/j.1462-2920.2009.01941.x.
12. Gill N, Wlodarska M, Finlay BB. Roadblocks in the gut: barriers to enteric infection. *Cell Microbiol*. 2011;13(5):660-9. doi:10.1111/j.1462-5822.2011.01578.x.
13. Artis D. Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. *Nat Rev Immunol*. 2008;8(6):411-20. doi:10.1038/nri2316.
14. Kozakova H, Kolinska J, Lojda Z, Rehakova Z, Sinkora J. Effect of bacterial monoassociation on brush-border enzyme activities in ex-germ-free piglets : comparison of commensal and pathogenic Escherichia coli strains. *Microbes Infect*. 2006;8:2629-39. doi:10.1016/j.micinf.2006.07.008.
15. Savidge TC, Smith MW, James PS, Aldred P. Salmonella-induced M-cell Formation in Germ-free Mouse Peyer ' s Patch Tissue. *Am J Pathol*. 1991;139(1):177-184.
16. Smith MW, James PS, Tivey DR. M Cell Numbers Increase After Transfer of SPF Mice to a Normal Animal House Environment. *Am J Pathol*. 1987;128(3):385-389.
17. Han K, Balan P, Hong H, et al. Function Korean ginseng modulates the ileal microbiota and mucin gene expression in the growing rat. *Food Funct*. 2014. doi:10.1039/c4fo00087k.
18. Cheled-Shoval S, Gamage N, Amit-Romach E, et al. Differences in intestinal mucin dynamics between germ-free and conventionally reared chickens after mannan-oligosaccharide supplementation. *Poult Sci*. 2014;93(3):636-44.
19. Bergström A, Kristensen MB, Bahl MI, et al. Nature of bacterial colonization influences transcription of mucin genes in mice during the first week of life. *BMC Res Notes*. 2012;5:402. doi:10.1186/1756-0500-5-402.
20. Caballero-Franco C, Keller K, Simone C, Chadee K. The VSL # 3 probiotic formula induces mucin gene expression and secretion in colonic epithelial cells. *Am J Physiol Gastrointest Liver Physiol*. 2007;292:315-322. doi:10.1152/ajpgi.00265.2006.
21. Deplancke B, Gaskins HR. Microbial modulation of innate defense : goblet cells and the intestinal mucus layer. *Am J Clin Nutr*. 2001;73(Suppl):1131S-41S.
22. Ayabe T, Satchell DP, Wilson CL, Parks WC, Selsted ME, Ouellette AJ. Secretion of microbicidal α-defensins by intestinal Paneth cells in response to bacteria. *Nat Immunol*. 2000;1:113-18.

23. Brandl K, Plitas G, Mihu CN, et al. Vancomycin-resistant enterococci exploit antibiotic-induced innate immune deficits. *Nature*. 2008;455(7214):804-7. doi:10.1038/nature07250.

24. McGuckin MA, Lindén SK, Sutton P, Florin TH. Mucin dynamics and enteric pathogens. *Nat Rev Microbiol*. 2011;9(4):265-278. doi:10.1038/nrmicro2538.

25. Johansson ME V, Phillipson M, Petersson J, Velcich A, Holm L, Hansson GC. The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *PNAS*. 2008;105(39):15064-69. doi:10.1073/pnas.0803124105.

26. Stecher B, Hardt W-D. Mechanisms controlling pathogen colonization of the gut. *Curr Opin Microbiol*. 2011;14(1):82-91. doi:10.1016/j.mib.2010.10.003.

27. Sekirov I, Tam NM, Jogova M, et al. Antibiotic-Induced Perturbations of the Intestinal Microbiota Alter Host Susceptibility to Enteric Infection Antibiotic-Induced Perturbations of the Intestinal Microbiota Alter Host Susceptibility to Enteric Infection. *Infect Immun*. 2008;76(10):4726-36. doi:10.1128/IAI.00319-08.

28. Croswell A, Amir E, Teggatz P, Barman M, Salzman NH. Prolonged Impact of Antibiotics on Intestinal Microbial Ecology and Susceptibility to Enteric Salmonella Infection Prolonged Impact of Antibiotics on Intestinal Microbial Ecology and Susceptibility to Enteric Salmonella Infection. *Infect Immun*. 2009;77(7):2741-53. doi:10.1128/IAI.00006-09.

29. Rea MC, Alemayehu D, Casey PG, et al. Bioavailability of the anti-Clostridial Bacteriocin Thuricin CD in Gastrointestinal Tract. *Microbiology*. 2014;160(Pt 2):439-45. doi:10.1099/mic.0.068767-0.

30. Rea MC, Sit CS, Clayton E, Connor PMO, Whittal RM, Zheng J. Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile. *PNAS*. 2010;107(20):9352-57. doi:10.1073/pnas.0913554107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.0913554107.

31. Gantois I, Ducatelle R, Pasmans F, et al. Butyrate Specifically Down-Regulates Salmonella Pathogenicity Island 1 Gene Expression. *Appl Environ Microbiol*. 2006;72(1):946-9. doi:10.1128/AEM.72.1.946.

32. Kamada N, Kim Y-G, Sham HP, et al. Regulated Virulence Controls the Ability of a Pathogen to Compete with the Gut Microbiota. *Science*. 2012;336(6086):1325-1329. doi:10.1126/science.1222195.

33. Stecher B, Hardt W. The role of microbiota in infectious disease. *Trends Microbiol*. 2008;16(3):107-114. doi:10.1016/j.tim.2007.12.008.

34. Vaishnava S, Behrendt CL, Ismail AS, Eckmann L, Hooper L V. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *PNAS*. 2008;105(52):20858-63. doi:10.1073/pnas.0808723105.

35. Tsuji M, Suzuki K, Kinoshita K, Fagarasan S. Dynamic interactions between bacteria and immune cells leading to intestinal IgA synthesis. *Semin Immunol*. 2008;20(1):59-66. doi:10.1016/j.smim.2007.12.003.

36. Endt K, Stecher B, Chaffron S, et al. The Microbiota Mediates Pathogen Clearance from the Gut Lumen after Non-Typhoidal Salmonella Diarrhea. *Plos Pathog*. 2010;6(9):e1001097. doi:10.1371/journal.ppat.1001097.

37. Hooper L V, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science*. 2012;336(6086):1268-73. doi:10.1126/science.1223490.

38. Lee YK, Mazmanian SK. Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science*. 2010;330(6012):1768-1773. doi:10.1126/science.1195568.

39. Hill DA, Artis D. Intestinal bacteria and the regulation of immune cell homeostasis. *Annu Rev Immunol*. 2010;28:623-67. doi:10.1146/annurev-immunol-030409-101330.

40. Wallace TC, Guarner F, Madsen K, et al. Human gut microbiota and its relationship to health and disease. *Nutr Rev*. 2011;69(7):392-403. doi:10.1111/j.1753-4887.2011.00402.x.

41. Cario E. Bacterial interactions with cells of the intestinal mucosa: Toll-like receptors and NOD2. *Gut*. 2005;54(8):1182-93. doi:10.1136/gut.2004.062794.

42. Round JL, Lee SM, Li J, et al. The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Sci Signal*. 2011;332(6032):974. doi:10.1126/science.1206095.

43. Vaishnava S, Yamamoto M, Severson KM, et al. The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. *Science*. 2011;334(6053):255-8. doi:10.1126/science.1209791.

44. Cash HL, Whitham C V, Behrendt CL, Hooper L V. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science*. 2006;313(5790):1126-30. doi:10.1126/science.1127119.

45. Dharmani P, Srivastava V, Kissoon-Singh V, Chadee K. Role of intestinal mucins in innate host defense mechanisms against pathogens. *J Innate Immun*. 2009;1(2):123-35. doi:10.1159/000163037.

46. Barnett AM, Roy NC, McNabb WC, Cookson AL. The interactions between endogenous bacteria, dietary components and the mucus layer of the large bowel. *Food Funct*. 2012;3(7):690-9. doi:10.1039/c2fo30017f.

47. Ouellette AJ. Paneth cells and innate mucosal immunity. *Curr Opin Gastroenterol*. 2010;26(6):547-53. doi:10.1097/MOG.0b013e32833dccde.

48. Stappenbeck TS. Paneth cell development, differentiation, and function: new molecular cues. *Gastroenterology*. 2009;137(1):30-3. doi:10.1053/j.gastro.2009.05.013.

49. Sonnenberg GF, Artis D. Innate lymphoid cell interactions with microbiota: implications for intestinal health and disease. *Immunity*. 2012;37(4):601-10. doi:10.1016/j.immuni.2012.10.003.

50. Macpherson AJ, Uhr T. Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria. *Science*. 2004;303(5664):1662-5. doi:10.1126/science.1091334.

51. Macpherson AJ. A Primitive T Cell-Independent Mechanism of Intestinal Mucosal IgA Responses to Commensal Bacteria. *Science*. 2000;288(5474):2222-2226. doi:10.1126/science.288.5474.2222.

52. Wells JM, Rossi O, Meijerink M, Van Baarlen P. Epithelial crosstalk at the microbiota-mucosal interface. *PNAS*. 2011;108(Suppl 1):4607-14. doi:10.1073/pnas.1000092107.

53. Salzman NH. Microbiota–immune system interaction: an uneasy alliance. *Curr Opin Microbiol*. 2011;14(1):99-105. doi:10.1016/j.mib.2010.09.018.

54. Geuking M, Köller Y, Rupp S, McCoy K. The interplay between the gut microbiota and the immune system. *Gut Microbes*. 2014;5(3). doi:10.4161/gmic.29330.

55. Ivanov II, Atarashi K, Manel N, et al. Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell*. 2009;139(3):485-98. doi:10.1016/j.cell.2009.09.033.

56. Aujla SJ, Dubin PJ, Kolls JK. Th17 cells and mucosal host defense. *Semin Immunol*. 2007;19(6):377-82. doi:10.1016/j.smim.2007.10.009.

57. Bettelli E, Oukka M, Kuchroo VK. T(H)-17 cells in the circle of immunity and autoimmunity. *Nat Immunol*. 2007;8(4):345-50. doi:10.1038/ni0407-345.

58. Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*. 2005;122(1):107-18. doi:10.1016/j.cell.2005.05.007.

59. McLoughlin RM, Mills KHG. Influence of gastrointestinal commensal bacteria on the immune responses that mediate allergy and asthma. *J Allergy Clin Immunol*. 2011;127(5):1097-1107. doi:10.1016/j.jaci.2011.02.012.

60. Pang W, Wang H, Shi L, et al. Immunomodulatory effects of Escherichia coli ATCC 25922 on allergic airway inflammation in a mouse model. *PLoS One*. 2013;8(3):e59174. doi:10.1371/journal.pone.0059174.

61. Atarashi K, Tanoue T, Oshima K, et al. Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature*. 2013;500(7461):232-6. doi:10.1038/nature12331.

62. Joyce SA, Gahan CGM. The gut microbiota and the metabolic health of the host. *Curr Opin Gastroenterol*. 2014;30(2):120-7. doi:10.1097/MOG.0000000000000039.

63. Tan J, McKenzie C, Potamitis M, Thorburn AN, Mackay CR, Macia L. The role of short-chain fatty acids in health and disease. In: *Advances in immunology*.Vol 121. 1st ed. Elsevier Inc.; 2014:91-119. doi:10.1016/B978-0-12-800100-4.00003-9.

64. Matsuki T, Tanaka R. Function of the human gut microbiota. In: *The human microbiota and microbiome*.; 2014:90-107.

65. Russell WR, Hoyles L, Flint HJ, Dumas M-E. Colonic bacterial metabolites and human health. *Curr Opin Microbiol*. 2013;16(3):246-54. doi:10.1016/j.mib.2013.07.002.

66. Brahe LK, Astrup A, Larsen LH. Is butyrate the link between diet, intestinal microbiota and obesity-related metabolic diseases? *Obes Rev*. 2013;14(12):950-9. doi:10.1111/obr.12068.

67. Lin H V, Frassetto A, Kowalik EJ, et al. Butyrate and propionate protect against diet-induced obesity and regulate gut hormones via free fatty acid receptor 3-independent mechanisms. *PLoS One*. 2012;7(4):e35240. doi:10.1371/journal.pone.0035240.

68. Vrieze A, Van Nood E, Holleman F, et al. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology*. 2012;143(4):913-6.e7. doi:10.1053/j.gastro.2012.06.031.

69. Meijer K, De Vos P, Priebe MG. Butyrate and other short-chain fatty acids as modulators of immunity: what relevance for health? *Curr Opin Clin Nutr Metab Care*. 2010;13(6):715-21. doi:10.1097/MCO.0b013e32833eebe5.

70. Huda-Faujan N, Abdulamir AS, Fatimah AB, et al. The impact of the level of the intestinal short chain Fatty acids in inflammatory bowel disease patients versus healthy subjects. *Open Biochem J*. 2010;4:53-8. doi:10.2174/1874091X01004010053.

71. Machiels K, Joossens M, Sabino J, et al. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut*. 2013:1-9. doi:10.1136/gutjnl-2013-304833.

72. Gałęcka M, Szachta P, Bartnicka A, Szuber LŁ. Faecalibacterium prausnitzii and Crohn's Disease – is there any Connection? *Polish J Microbiol*. 2013;62(1):91-95.

73. Trompette A, Gollwitzer ES, Yadava K, et al. Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat Med*. 2014;20(2):159-66. doi:10.1038/nm.3444.

74. Boesjes M, Brufau G. Metabolic Effects of Bile Acids in the Gut in Health and Disease. *Curr Med Chem*. 2014.

75. Zwicker BL, Agellon LB. Transport and biological activities of bile acids. *Int J Biochem Cell Biol*. 2013;45(7):1389-98. doi:10.1016/j.biocel.2013.04.012.

76. Bortolini O, Medici A, Poli S. Biotransformations on steroid nucleus of bile acids. *Steroids*. 1997;62:564-77. doi:10.1016/S0039-128X(97)00043-3.

77. Begley M, Hill C, Gahan CGM. Bile Salt Hydrolase Activity in Probiotics Bile Salt Hydrolase Activity in Probiotics. *Appl Environ Microbiol*. 2006;72(3):1729-38. doi:10.1128/AEM.72.3.1729.

78. Joyce SA, MacSharry J, Casey PG, et al. Regulation of host weight gain and lipid metabolism by bacterial bile acid modification in the gut. *PNAS*. 2014;111(20):7421-6. doi:10.1073/pnas.1323599111.

79. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486(7402):207-14. doi:10.1038/nature11234.

80. Duytschaever G, Huys G, Bekaert M, Boulanger L, De Boeck K, Vandamme P. Dysbiosis of Bifidobacteria and Clostridium cluster XIVa in the cystic fibrosis faecal microbiota. *J Cyst Fibros*. 2013;12(3):206-15. doi:10.1016/j.jcf.2012.10.003.

81. Wu N, Yang X, Zhang R, et al. Dysbiosis signature of fecal microbiota in colorectal cancer patients. *Microb Ecol*. 2013;66(2):462-70. doi:10.1007/s00248-013-0245-9.

82. Rajilić-Stojanović M, Biagi E, Heilig HGHJ, et al. Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology*. 2011;141(5):1792-801. doi:10.1053/j.gastro.2011.07.043.

83. Iebba V, Santangelo F, Totino V, et al. Higher prevalence and abundance of Bdellovibrio bacteriovorus in the human gut of healthy subjects. *PLoS One*. 2013;8(4):e61608. doi:10.1371/journal.pone.0061608.

84. Abusleme L, Hong B-Y, Dupuy AK, Strausbaugh LD, Diaz PI. Influence of DNA extraction on oral microbial profiles obtained via 16S rRNA gene sequencing. *J Oral Microbiol*. 2014;6:1-7. doi:10.3402/jom.v6.23990.

85. Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464(7285):59-65. doi:10.1038/nature08821.

86. Li K, Bihan M, Yooseph S, Methe BA. Analyses of the Microbial Diversity across the Human Microbiome. *PLoS One*. 2012;7(6):e32118. doi:10.1371/journal.pone.0032118.

87. Jalanka-Tuovinen J, Salonen A, Nikkilä J, et al. Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms. *PLoS One*. 2011;6(7):e23035. doi:10.1371/journal.pone.0023035.

88. Caporaso JG, Lauber CL, Costello EK, et al. Moving pictures of the human microbiome. *Genome Biol*. 2011;12(5):R50. doi:10.1186/gb-2011-12-5-r50.

89. Scott KP, Duncan SH, Louis P, Flint HJ. Nutritional influences on the gut microbiota and the consequences for gastrointestinal health. *Biochem Soc Trans*. 2011;39(4):1073-8. doi:10.1042/BST0391073.

90. Van Leuwenhoek A. An Extract of a Letter from Mr . Anth . Van Leuwenhoek , concerning Animalcules Found on the Teeth ; Of the Scaleyness of the Skin. *Philos Trans*. 1693;17:646-649.

91. Nagpal R, Yadav H, Marotta F. Gut microbiota : the next-gen frontier in preventive and therapeutic medicine ? *Front Med*. 2014. doi:10.3389/fmed.2014.00015.

92. Walters WA, Knight R. Technology and techniques for microbial ecology via DNA sequencing. *Ann Am Thorac Soc*. 2014;11(Suppl 1):S16-20. doi:10.1513/AnnalsATS.201306-160MG.

93. Cénit MC, Matzaraki V, Tigchelaar EF, Zhernakova A. Rapidly expanding knowledge on the role of the gut microbiome in health and disease. *Biochim Biophys Acta*. 2014. doi:10.1016/j.bbadis.2014.05.023.

94. Sekirov I, Russell SL, Antunes LCM, Finlay BB. Gut microbiota in health and disease. *Physiol Rev*. 2010;90(3):859-904. doi:10.1152/physrev.00045.2009.

95. Fraher MH, O'Toole PW, Quigley EMM. Techniques used to characterize the gut microbiota: a guide for the clinician. *Nat Rev Gastroenterol Hepatol*. 2012;9(6):312.

96. Lagier J-C, Million M, Hugon P, Armougom F, Raoult D. Human gut microbiota: repertoire and variations. *Front Cell Infect Microbiol*. 2012;2(November):136. doi:10.3389/fcimb.2012.00136.

97. Olle B. Medicines from microbiota. *Nat Biotechnol*. 2013;31(4):309-15. doi:10.1038/nbt.2548.

98. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012;489(7415):220-30. doi:10.1038/nature11550.

99. Allen-Vercoe E. Bringing the gut microbiota into focus through microbial culture: recent progress and future perspective. *Curr Opin Microbiol*. 2013;16(5):625-9. doi:10.1016/j.mib.2013.09.008.

100. Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal microbial flora. *Science*. 2005;308(June):1635-8. doi:10.1126/science.1110591.

101. Hayashi H, Sakamoto M, Benno Y. Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. *Microbiol Immunol*. 2002;46(8):535-48. doi:10.1111/j.1348-0421.2002.tb02731.x.

102. Suau A, Bonnet R, Sutren M, et al. Direct Analysis of Genes Encoding 16S rRNA from Complex Communities Reveals Many Novel Molecular Species within the Human Gut Direct Analysis of Genes Encoding 16S rRNA from Complex Communities Reveals Many Novel Molecular Species within the Human Gut. *Appl Environ Microbiol*. 1999;65(11):4799-4807.

103. Siqueira JF, Rôças IN. As-yet-uncultivated oral bacteria: breadth and association with oral and extra-oral diseases. *J Oral Microbiol*. 2013;5(21077):1-14. doi:10.3402/jom.v5i0.21077.

104. Vartoukian SR, Palmer RM, Wade WG. Strategies for culture of "unculturable" bacteria. *FEMS Microbiol Lett*. 2010;309(1):1-7. doi:10.1111/j.1574-6968.2010.02000.x.

105. Lagier J-C, Armougom F, Million M, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect*. 2012;18(12):1185-93. doi:10.1111/1469-0691.12023.

106. Pham TAN, Lawley TD. Emerging insights on intestinal dysbiosis during bacterial infections. *Curr Opin Microbiol*. 2014;17:67-74. doi:10.1016/j.mib.2013.12.002.

107. Goodman AL, Kallstrom G, Faith JJ, et al. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc Natl Acad Sci*. 2011;108(15):6252-57. doi:10.1073/pnas.1102938108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1102938108.

108. Greub G. Culturomics: a new approach to study the human microbiome. *Clin Microbiol Infect*. 2012;18(12):1157-9. doi:10.1111/1469-0691.12032.

109. Lagier J-C, Hugon P, Khelaifia S, Fournier P-E, La Scola B, Raoult D. The Rebirth of Culture in Microbiology through the Example of Culturomics To Study Human Gut Microbiota. *Clin Microbiol Rev*. 2015;28(1):237-264. doi:10.1128/CMR.00014-14.

110. Dubourg G, Lagier JC, Robert C, et al. Culturomics and pyrosequencing evidence of the reduction in gut microbiota diversity in patients with broad-spectrum antibiotics. *Int J Antimicrob Agents*. 2014;44:117-124. doi:10.1016/j.ijantimicag.2014.04.020.

111. Pfleiderer A, Lagier JC, Armougom F, Robert C, Vialettes B, Raoult D. Culturomics identified 11 new bacterial species from a single anorexia nervosa stool sample. *Eur J Clin Microbiol Infect Dis*. 2013;32:1471-1481. doi:10.1007/s10096-013-1900-2.

112. Peterson DA, Frank DN, Pace NR, Gordon JI. Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. *Cell Host Microbe*. 2008;3(6):417-27. doi:10.1016/j.chom.2008.05.001.

113. Patel JB. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Mol diagnosis*. 2001;6(4):313-21. doi:10.1054/modi.2001.29158.

114. Cox MJ, Cookson WOCM, Moffatt MF. Sequencing the human microbiome in health and disease. *Hum Mol Genet*. 2013;22(1):R88-94. doi:10.1093/hmg/ddt398.

115. Langille MGI, Zaneveld J, Caporaso JG, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol*. 2013;31(9):814-21. doi:10.1038/nbt.2676.

116. Woo PCY, Lau SKP, Teng JLL, Yuen K-Y. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin Microbiol Infect*. 2008;14(10):908-34. doi:10.1111/j.1469-0691.2008.02070.x.

117. Reigstad CS, Kashyap PC. Beyond phylotyping: understanding the impact of gut microbiota on host biology. *Neurogastroenterol Motil*. 2013;25(5):358-72. doi:10.1111/nmo.12134.

118. Marchesi JR. Human distal gut microbiome. *Environ Microbiol*. 2011;13(12):3088-102. doi:10.1111/j.1462-2920.2011.02574.x.

119. Inglis G, Thomas M, Thomas D, Kalmokoff M, Brooks S, Selinger L. Molecular Methods to Measure Intestinal Bacteria : A Review. *J AOAC Int*. 2012;95(1):5-24. doi:10.5740/jaoacint.SGE.

120. Willner D, Daly J, Whiley D, Grimwood K, Wainwright CE, Hugenholtz P. Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PLoS One*. 2012;7(4):e34605. doi:10.1371/journal.pone.0034605.

121. Hwang C, Ling F, Andersen GL, LeChevallier MW, Liu W-T. Evaluation of Methods for the Extraction of DNA from Drinking Water Distribution System Biofilms. *Microbes Environ*. 2012;27(1):9-18. doi:10.1264/jsme2.ME11132.

122. Momozawa Y, Deffontaine V, Louis E, Medrano JF. Characterization of bacteria in biopsies of colon and stools by high throughput sequencing of the V2 region of bacterial 16S rRNA gene in human. *PLoS One*. 2011;6(2):e16952. doi:10.1371/journal.pone.0016952.

123. Wu GD, Lewis JD, Hoffmann C, et al. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiol*. 2010;10:206. doi:10.1186/1471-2180-10-206.

124. Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods*. 2007;69(2):330-9. doi:10.1016/j.mimet.2007.02.005.

125. Arrieta M-C, Stiemsma LT, Amenyogbe N, Brown EM, Finlay B. The Intestinal Microbiome in Early Life: Health and Disease. *Front Immunol*. 2014;5(September):1-18. doi:10.3389/fimmu.2014.00427.

126. Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J Biosci Bioeng*. 2003;96(4):317-23. doi:10.1016/S1389-1723(03)90130-7.

127. Sjögren YM, Jenmalm MC, Böttcher MF, Björkstén B, Sverremark-Ekström E. Altered early infant gut microbiota in children developing allergy up to 5 years of age. *Clin Exp allergy*. 2009;39(4):518-26. doi:10.1111/j.1365-2222.2008.03156.x.

128. Sjöberg F, Nowrouzian F, Rangel I, et al. Comparison between terminal-restriction fragment length polymorphism (T-RFLP) and quantitative culture for analysis of infants' gut microbiota. *J Microbiol Methods*. 2013;94(1):37-46. doi:10.1016/j.mimet.2013.04.002.

129. Collado MC, Calabuig M, Sanz Y. Differences between the Fecal Microbiota of Coeliac Infants and Healthy Controls. *Curr Issues Intest Microbiol*. 2007;8(1):9-14.

130. Paliy O, Kenche H, Abernathy F, Michail S. High-throughput quantitative analysis of the human intestinal microbiota with a phylogenetic microarray. *Appl Environ Microbiol*. 2009;75(11):3572-9. doi:10.1128/AEM.02764-08.

131. Siqueira JF, Fouad AF, Rôças IN. Pyrosequencing as a tool for better understanding of human microbiomes. *J Oral Microbiol*. 2012;4:1-15. doi:10.3402/jom.v4i0.10743.

132. Dey N, Soergel DAW, Repo S, Brenner SE. Association of gut microbiota with post-operative clinical course in Crohn's disease. *BMC Gastroenterol*. 2013;13(1):131. doi:10.1186/1471-230X-13-131.

133. Milani C, Hevia A, Foroni E, et al. Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. *PLoS One*. 2013;8(7):e68739. doi:10.1371/journal.pone.0068739.

134. Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J. Metagenomic pyrosequencing and microbial identification. *Clin Chem*. 2009;55(5):856-66. doi:10.1373/clinchem.2008.107565.

135. Sanger F, Nicklen S, Coulson R. DNA sequencing with chain-terminating. *Proc Natl Acad Sci*. 1977;74(12):5463-5467.

136. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*. 2008;26(10):1135-1145. doi:10.1038/nbt1486.

137. Strausberg RL, Levy S, Rogers Y-H. Emerging DNA sequencing technologies for human genomic medicine. *Drug Discov Today*. 2008;13(13-14):569-77. doi:10.1016/j.drudis.2008.03.025.

138. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005;437(7057):376-80. doi:10.1038/nature03959.

139. Glenn TC. Field guide to next-generation DNA sequencers. *Mol Ecol Resour*. 2011;11(5):759-69. doi:10.1111/j.1755-0998.2011.03024.x.

140. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *J Appl Genet*. 2011;52(4):413-35. doi:10.1007/s13353-011-0057-x.

141. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet*. 2008;24(3):133-41. doi:10.1016/j.tig.2007.12.007.

142. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *PNAS*. 2003;100(15):8817-22. doi:10.1073/pnas.1133470100.

143. Ronaghi M. DNA Sequencing: A Sequencing Method Based on Real-Time Pyrophosphate. *Science*. 1998;281(5375):363-365. doi:10.1126/science.281.5375.363.

144. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem*. 1996;242(1):84-9. doi:10.1006/abio.1996.0432.

145. Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res*. 2006;34(3):e22. doi:10.1093/nar/gnj023.

146. Voelkerding K V, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem*. 2009;55(4):641-58. doi:10.1373/clinchem.2008.112789.

147. Naseribafrouei A, Hestad K, Avershina E, et al. Correlation between the human fecal microbiota and depression. *Neurogastroenterol Motil*. 2014. doi:10.1111/nmo.12378.

148. Remely M, Aumueller E, Merold C, et al. Effects of short chain fatty acid producing bacteria on epigenetic regulation of FFAR3 in type 2 diabetes and obesity. *Gene*. 2014;537(1):85-92. doi:10.1016/j.gene.2013.11.081.

149. Frémont M, Coomans D, Massart S, De Meirleir K. High-throughput 16S rRNA gene sequencing reveals alterations of intestinal microbiota in myalgic encephalomyelitis/chronic fatigue syndrome patients. *Anaerobe*. 2013;22:50-6. doi:10.1016/j.anaerobe.2013.06.002.

150. Hamilton MJ, Weingarden AR, Unno T, Khoruts A, Sadowsky MJ. High-throughput DNA sequence analysis reveals stable engraftment of gut microbiota following transplantation of previously frozen fecal bacteria. *Gut Microbes*. 2013;4(2):125-135. doi:10.4161/gmic.23571.

151. Kim S-W, Suda W, Kim S, et al. Robustness of gut microbiota of healthy adults in response to probiotic intervention revealed by high-throughput pyrosequencing. *DNA Res*. 2013;20(3):241-53. doi:10.1093/dnares/dst006.

152. Martínez I, Muller CE, Walter J. Long-term temporal analysis of the human fecal microbiota revealed a stable core of dominant bacterial species. *PLoS One*. 2013;8(7):e69621. doi:10.1371/journal.pone.0069621.

153. Ukhanova M, Culpepper T, Baer D, et al. Gut microbiota correlates with energy gain from dietary fibre and appears to be associated with acute and chronic intestinal diseases. *Clin Microbiol Infect*. 2012;18(Suppl 4):62-66. doi:10.1111/j.1469-0691.2012.03859.x.

154. Mitra S, Förster-Fromme K, Damms-Machado A, et al. Analysis of the intestinal microbiota using SOLiD 16S rRNA gene sequencing and SOLiD shotgun sequencing. *BMC Genomics*. 2013;14(Suppl 5):S16. doi:10.1186/1471-2164-14-S5-S16.

155. Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One*. 2012;7(2):e30087. doi:10.1371/journal.pone.0030087.

156. Claesson MJ, Wang Q, O'Sullivan O, et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res*. 2010;38(22):e200. doi:10.4161/gmic.1.4.12306.

157. Weinstock GM. Genomic approaches to studying the human microbiota. *Nature*. 2012;489(7415):250-6. doi:10.1038/nature11553.

158. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. Analysis, Optimization and Verification of Illumina-Generated 16S rRNA Gene Amplicon Surveys. *PLoS One*. 2014;9(4):e94249. doi:10.1371/journal.pone.0094249.

159. Ong SH, Kukkillaya VU, Wilm A, et al. Species Identification and Profiling of Complex Microbial Communities Using Shotgun Illumina Sequencing of 16S rRNA Amplicon Sequences. *PLoS One*. 2013;8(4):e60811. doi:10.1371/journal.pone.0060811.

160. Mizrahi-Man O, Davenport ER, Gilad Y. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS One*. 2013;8(1):e53608. doi:10.1371/journal.pone.0053608.

161. Claesson MJ, O'Toole PW. Evaluating the latest high-throughput molecular techniques for the exploration of microbial gut communities. *Gut Microbes*. 2010;1(4):277-278. doi:10.4161/gmic.1.4.12306.

162. Youssef N, Sheik CS, Krumholz LR, Najar FZ, Roe BA, Elshahed MS. Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Appl Environ Microbiol*. 2009;75(16):5227-36. doi:10.1128/AEM.00592-09.

163. Zhao L, Wang G, Siegel P, et al. Quantitative genetic background of the host influences gut microbiomes in chickens. *Sci Rep*. 2013;3:1163. doi:10.1038/srep01163.

164. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486(7402):222-7. doi:10.1038/nature11053.

165. Chiu C-M, Lin F-M, Chang T-H, et al. Clinical detection of human probiotics and human pathogenic bacteria by using a novel high-throughput platform based on next generation sequencing. *J Clin Bioinforma*. 2014;4(1):1. doi:10.1186/2043-9113-4-1.

166. Li Y, Liu X-Y, Ma M-M, et al. Changes in intestinal microflora in rats with acute respiratory distress syndrome. *World J Gastroenterol*. 2014;20(19):5849-58. doi:10.3748/wjg.v20.i19.5849.

167. Lewis ZT, Bokulich NA, Kalanetra KM, Ruiz-Moyano S, Underwood MA, Mills DA. Anaerobe Use of bifidobacterial specific terminal restriction fragment length polymorphisms to complement next generation sequence profiling of infant gut communities. *Anaerobe*. 2013;19:62-9. doi:10.1016/j.anaerobe.2012.12.005.

168. Gut IG. New sequencing technologies. *Clin Transl Oncol*. 2013;15(11):879-81. doi:10.1007/s12094-013-1073-6.

169. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. *Hum Mol Genet*. 2010;19(R2):R227-40. doi:10.1093/hmg/ddq416.

170. Gilboa T, Meller A. Optical sensing and analyte manipulation in solid-state nanopores. *Analyst*. 2015. doi:10.1039/C4AN02388A.

171. O'Donnell C., Wang H, Dunbar W. Error analysis of idealized nanopore sequencing. *Electrophoresis*. 2013;34(15):2137-44. doi:10.1002/elps.201300174.

172. Ma J, Prince A, Aagaard KM. Use of whole genome shotgun metagenomics: A practical guide for the microbiome-minded physician scientist. *Semin Reprod Med*. 2014;32:5-13. doi:10.1055/s-0033-1361817.

173. Hodkinson BP, Grice EA. Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research. *Adv Wound Care*. 2015;4(1):50-58. doi:10.1089/wound.2014.0542.

174. Goll J., Szpakowski S, Krampis K, Nelson K. Metagenomics and Microbiomes. In: Bishop OT, ed. *Bioinformatics and Data Analysis in Microbiology*. Norfolk, UK: Caister Academic Press; 2014:163-192.

175. Mah KW, Sangsupawanich P, Tunyapanit W, et al. Gut microbiota of children living in rural south Thailand and urban Singapore. *Allergol Int*. 2008;57(1):65-71. doi:10.2332/allergolint.O-07-501.

176. Fouhy F, Fitzgerald G, Food E, Authority S. Composition of the early intestinal microbiota: Knowledge , knowledge gaps and the use of high-throughput sequencing to address these gaps. *Gut Microbes*. 2012;3(3):203-220. doi:10.4161/gmic.20169.

177. Zhang T, Fang HHP. Applications of real-time polymerase chain reaction for quantification of microorganisms in environmental samples. *Appl Microbiol Biotechnol*. 2006;70(3):281-9. doi:10.1007/s00253-006-0333-6.

178. Muyzer G, De Waal E, Uitierlinden A. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol*. 1993;59(3):695-700.

179. Li F, Hullar MAJ, Lampe JW. Optimization of terminal restriction fragment polymorphism (TRFLP) analysis of human gut microbiota. *J Microbiol Methods*. 2007;68(2):303-11. doi:10.1016/j.mimet.2006.09.006.

180. Osborn AM, Moore ER, Timmis KN. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environ Microbiol*. 2000;2(1):39-50. doi:10.1046/j.1462-2920.2000.00081.x.

181. Palmer C, Bik EM, Eisen MB, et al. Rapid quantitative profiling of complex microbial populations. *Nucleic Acids Res*. 2006;34(1):e5. doi:10.1093/nar/gnj007.

182. Ratan A, Miller W, Guillory J, Stinson J, Seshagiri S, Schuster SC. Comparison of sequencing platforms for single nucleotide variant calls in a human sample. *PLoS One*. 2013;8(2):e55089. doi:10.1371/journal.pone.0055089.

183. Venkatesan BM, Bashir R. Nanopore sensors for nucleic acid analysis. *Nat Nanotechnol*. 2011;6(10):615-24. doi:10.1038/nnano.2011.129.

184. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. *Science*. 2005;307(5717):1915-20. doi:10.1126/science.1104816.

185. Reid G, Younes JA, Van Der Mei HC, Gloor GB, Knight R, Busscher HJ. Microbiota restoration : natural and supplemented recovery of human microbial communities. *Nat Rev Microbiol*. 2011;9(1):27-38. doi:10.1038/nrmicro2473.

186. Kosiewicz MM, Dryden GW, Chhabra A, Alard P. Relationship between gut microbiota and development of T cell associated disease. *FEBS Lett*. 2014. doi:10.1016/j.febslet.2014.03.019.

187. Gerritsen J, Smidt H, Rijkers GT, De Vos WM. Intestinal microbiota in human health and disease: the impact of probiotics. *Genes Nutr*. 2011;6(3):209-40. doi:10.1007/s12263-011-0229-7.

188. Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*. 2006;124(4):837-48. doi:10.1016/j.cell.2006.02.017.

189. The Human Microbiome Jumpstart Reference Strains Consortium. A Catalog of Reference Genomes from the Human Microbiome. *Science*. 2010;328(5981):994-999. doi:10.1126/science.1183605.A.

190. Turnbaugh PJ, Gordon JI. The core gut microbiome, energy balance and obesity. *J Physiol*. 2009;587(17):4153-58. doi:10.1113/jphysiol.2009.174136.

191. Durbán A, Abellán JJ, Jiménez-Hernández N, Latorre A, Moya A. Daily follow-up of bacterial communities in the human gut reveals stable composition and host-specific patterns of interaction. *FEMS Microbiol Ecol*. 2012;81(2):427-437. doi:10.1111/j.1574-6941.2012.01368.x.

192. Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. Rapid evolution of the human gut virome. 2013;110(30). doi:10.1073/pnas.1300833110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1300833110.

193. Minot S, Sinha R, Chen J, et al. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res*. 2011;21(10):1616-25. doi:10.1101/gr.122705.111.

194. O'Brien AD, Newland JW, Miller SF, Holmes RK, Williams H, Formal SB. Shiga-like toxin-converting phages from Escherichia coli strains that cause hemorrhagic colitis or infantile diarrhea. *Science*. 1984;226(4675):694-6. doi:10.1126/science.6387911.

195. Waldor MK, Mekalanos JJ. Lysogenic Conversion by a Filamentous Phage Encoding Cholera Toxin. *Science*. 1996;272(5270):1910-1914. doi:10.1126/science.272.5270.1910.

196. Reyes A, Haynes M, Hanson N, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature*. 2010;466(7304):334-8. doi:10.1038/nature09199.

197. Wang Y, Pfeiffer J. Microbiology: A backup for bacteria. *Nature*. 2014;516:42-43. doi:10.1038/nature13938.

198. Cadwell K. Expanding the Role of the Virome: Commensalism in the Gut. *J Virol*. 2015;89(4):1951-1953. doi:10.1128/JVI.02966-14.

199. Kernbauer E, Ding Y, Cadwell K. An enteric virus can replace the beneficial function of commensal bacteria. *Nature*. 2014;516(7529):94-98. doi:10.1038/nature13960.

200. Wang ZK, Yang YS, Stefka AT, Sun G, Peng LH. Review article: fungal microbiota and digestive diseases. *Aliment Pharmacol Ther*. 2014;39(8):751-66. doi:10.1111/apt.12665.

201. Salonen A, Salojärvi J, Lahti L, De Vos WM. The adult intestinal core microbiota is determined by analysis depth and health status. *Clin Microbiol Infect*. 2012;18(Suppl 4):16-20. doi:10.1111/j.1469-0691.2012.03855.x.

202. Sekelja M, Berget I, Næs T, Rudi K. Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach. *ISME J*. 2011;5(3):519-31. doi:10.1038/ismej.2010.129.

203. David LA, Materna AC, Friedman J, et al. Host lifestyle affects human microbiota on daily timescales. *Genome Biol*. 2014;15(7):R89. doi:10.1186/gb-2014-15-7-r89.

204. Vital M, Howe AC, Tiedje JM. Revealing the Bacterial Butyrate Synthesis Pathways by Analyzing (Meta) genomic Data. *MBio*. 2014. doi:10.1128/mBio.00889-14.

205. Tap J, Mondot S, Levenez F, et al. Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol*. 2009;11(10):2574-84. doi:10.1111/j.1462-2920.2009.01982.x.

206. Ze X, Le Mougen F, Duncan SH, Louis P, Flint HJ. Some are more equal than others. *Gut Microbes*. 2013;4(3):263-40. doi:10.1016/j.jtos.2014.01.003.

207. Siggers RH, Siggers J, Boye M, et al. Early Administration of Probiotics Alters Bacterial Colonization and Limits Diet-Induced Gut Dysfunction and Severity of Necrotizing Enterocolitis in Preterm Pigs. *J Nutr*. 2008;138:1437-1444.

208. Zhao Y, Wu J, Li J V, Zhou N, Tang H, Wang Y. Gut Microbiota Composition Modifies Fecal Metabolic Profiles in Mice. *J Proteome Res*. 2013;12(6):2987-99. doi:10.1021/pr400263n.

209. Wopereis H, Oozeer R, Knipping K, Belzer C, Knol J. The first thousand days - intestinal microbiology of early life: establishing a symbiosis. *Pediatr allergy Immunol*. 2014. doi:10.1111/pai.12232.

210. Aidy S El, Hooiveld G, Tremaroli V, Bäckhed F, Kleerebezem M. The gut microbiota and mucosal homeostasis: Colonized at birth or at adulthood , does it matter ? *Gut Microbes*. 2013;4(2):118-124. doi:10.4161/gmic.23362.

211. Abrahamsson TR, Jakobsson HE, Andersson AF, Björkstén B, Engstrand L, Jenmalm MC. Low gut microbiota diversity in early infancy precedes asthma at school age. *Clin Exp allergy*. 2013;44(6):842-850. doi:10.1111/cea.12253.

212. Ismail IH, Oppedisano F, Joseph SJ, et al. Reduced gut microbial diversity in early life is associated with later development of eczema but not atopy in high-risk infants. *Pediatr allergy Immunol*. 2012;23(7):4674-81. doi:10.1111/j.1399-3038.2012.01328.x.

213. Isolauri E. Development of healthy gut microbiota early in life. *J Paediatr Child Health*. 2012;48 Suppl 3:1-6. doi:10.1111/j.1440-1754.2012.02489.x.

214. Vael C, Vanheirstraeten L, Desager KN, Goossens H. Denaturing gradient gel electrophoresis of neonatal intestinal microbiota in relation to the development of asthma. *BMC Microbiol*. 2011;11(1):68. doi:10.1186/1471-2180-11-68.

215. Penders J, Thijs C, Van Den Brandt PA, et al. Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. *Gut*. 2007;56(5):661-667. doi:10.1136/gut.2006.100164.

216. Li M, Wang M, Donovan SM. Early development of the gut microbiome and immune-mediated childhood disorders. *Semin Reprod Med*. 2014;32(1):74-86. doi:10.1055/s-0033-1361825.

217. Collado M, Cernada M, Bäuerl C, Vento M, Pérez-Martínez G. Microbial ecology and host-microbiota interactions during early life stages. *Gut Microbes*. 2012;3(4):352-365. doi:10.4161/gmic.21215.

218. Scholtens PAMJ, Oozeer R, Martin R, Amor K Ben, Knol J. The early settlers: intestinal microbiology in early life. *Annu Rev Food Sci Technol*. 2012;3:425-47. doi:10.1146/annurev-food-022811-101120.

219. Mackie RI, Sghir A, Gaskins HR. Developmental microbial ecology of the neonatal gastrointestinal tract. *Am J Clin Nutr*. 1999;69(5):1035S-1045S.

220. Voreades N, Kozil A, Weir TL. Diet and the development of the human intestinal microbiome. *Front Microbiol*. 2014;5(September):494. doi:10.3389/fmicb.2014.00494.

221. Francino MP. Early development of the gut microbiota and immune health. *Pathog (Basel, Switzerland)*. 2014;3(3):769-90. doi:10.3390/pathogens3030769.

222. Gosalbes MJ, Llop S, Vallès Y, Moya A, Ballester F, Francino MP. Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clin Exp Allergy*. 2012;43(2):198-211. doi:10.1111/cea.12063.

223. Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. *Cell*. 2012;148(6):1258-70. doi:10.1016/j.cell.2012.01.035.

224. Adlerberth I, Wold AE. Establishment of the gut microbiota in Western infants. *Acta Paediatr*. 2009;98(2):229-38. doi:10.1111/j.1651-2227.2008.01060.x.

225. Rotimi VO, Duerden BI. The development of the bacterial flora in normal neonates. *J Med Microbiol*. 1981;14:51-62. doi:10.1099/00222615-14-1-51.

226. Benno Y, Sawada K, Mitsuoka T. The Intestinal Microflora of Infants : Composition of Fecal Flora in Breast-Fed and Bottle-Fed Infants. *Microbiol Immunol*. 1984;28(9):975-986. doi:10.1111/j.1348-0421.1984.tb00754.x.

227. Garrity GM, Bell JA, Lilburn TG. Taxonomic Outline of the Procaryotes. In: *Bergey's Manual of Systematic Bacteriology, 2nd edition, Release 5.0, Springer-Verlag, New York.*; 2004:1-399. doi:10.1007/bergeysoutline200405.

228. Yoshioka H, Iseki K, Fujita K. Development and Differences of Intestinal Flora in the Neonatal Period in Breast-Fed and Bottle-Fed Infants. *Pediatrics*. 1983;72(3):317-21.

229. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. Ruan Y, ed. *PLoS Biol*. 2007;5(7):e177. doi:10.1371/journal.pbio.0050177.

230. Lee J-H, O'Sullivan DJ. Genomic insights into bifidobacteria. *Microbiol Mol Biol Rev*. 2010;74(3):378-416. doi:10.1128/MMBR.00004-10.

231. Favier CF, Vaughan EE, De Vos WM, Akkermans ADL. Molecular Monitoring of Succession of Bacterial Communities in Human Neonates. *Appl Environ Microbiol*. 2002;68(1):219-26. doi:10.1128/AEM.68.1.219.

232. Harmsen HJM, Wildeboer–Veloo ACM, Raangs GC, et al. Analysis of Intestinal Flora Development in Breast-Fed and Formula-Fed Infants by Using Molecular Identification and Detection Methods. *J Pediatr Gastroenterol Nutr*. 2000;30(1):61-67.

233. Penders J, Vink C, Driessen C, London N, Thijs C, Stobberingh EE. Quantification of Bifidobacterium spp., Escherichia coli and Clostridium difficile in faecal samples of breast-fed and formula-fed infants by real-time PCR. *FEMS Microbiol Lett*. 2005;243(1):141-7. doi:10.1016/j.femsle.2004.11.052.

234. Turroni F, Peano C, Pass DA, et al. Diversity of bifidobacteria within the infant gut microbiota. *PLoS One*. 2012;7(5):e36957. doi:10.1371/journal.pone.0036957.

235. Sergeant MJ, Constantinidou C, Cogan T, Penn CW, Pallen MJ. High-throughput sequencing of 16S rRNA gene amplicons: effects of extraction procedure, primer length and annealing temperature. *PLoS One*. 2012;7(5):e38094. doi:10.1371/journal.pone.0038094.

236. Sim K, Cox MJ, Wopereis H, et al. Improved detection of bifidobacteria with optimised 16S rRNA-gene based pyrosequencing. *PLoS One*. 2012;7(3):e32543. doi:10.1371/journal.pone.0032543.

237. Roger LC, Costabile A, Holland DT, Hoyles L, McCartney AL. Examination of faecal Bifidobacterium populations in breast- and formula-fed infants during the first 18 months of life. *Microbiology*. 2010;156(Pt 11):3329-41. doi:10.1099/mic.0.043224-0.

238. Wharton BA, Balmer SE, Scott PH. Sorrento studies of diet and fecal flora in the newborn. *Pediatr Int*. 1994;36(5):579-84.

239. Jost T, Lacroix C, Braegger CP, Chassard C. New insights in gut microbiota establishment in healthy breast fed neonates. *PLoS One*. 2012;7(8):e44595. doi:10.1371/journal.pone.0044595.

240. Avershina E, Storrø O, Øien T, et al. Bifidobacterial succession and correlation networks in a large unselected cohort of mothers and their children. *Appl Environ Microbiol*. 2013;79(2):497-507. doi:10.1128/AEM.02359-12.

241. Koenig JE, Spor A, Scalfone N, et al. Succession of microbial consortia in the developing infant gut microbiome. *PNAS*. 2011;108(Suppl 1):4578-85. doi:10.1073/pnas.1000081107.

242. Funkhouser LJ, Bordenstein SR. Mom knows best: the universality of maternal microbial transmission. *PLoS Biol*. 2013;11(8):e1001631. doi:10.1371/journal.pbio.1001631.

243. West CE, Jenmalm MC, Prescott SL. The gut microbiota and its role in the development of allergic disease: a wider perspective. *Clin Exp Allergy*. 2014. doi:10.1111/cea.12332.

244. Jiménez E, Marín ML, Martín R, et al. Is meconium from healthy newborns actually sterile? *Res Microbiol*. 2008;159(3):187-93. doi:10.1016/j.resmic.2007.12.007.

245. Thompson-Chagoyán OC, Maldonado J, Gil A. Colonization and impact of disease and other factors on intestinal microbiota. *Dig Dis Sci*. 2007;52(9):2069-77. doi:10.1007/s10620-006-9285-z.

246. Penders J, Thijs C, Vink C, et al. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*. 2006;118(2):511-21. doi:10.1542/peds.2005-2824.

247. Biasucci G, Rubini M, Riboni S, Morelli L, Bessi E, Retetangos C. Mode of delivery affects the bacterial community in the newborn gut. *Early Hum Dev*. 2010;86(Suppl 1):S13-S15. doi:10.1016/j.earlhumdev.2010.01.004.

248. Dominguez-Bello MG, Costello EK, Contreras M, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *PNAS*. 2010;107(26):11971-5. doi:10.1073/pnas.1002601107.

249. Fanaro S, Chierici R, Guerrini P, Vigi V. Intestinal microflora in early infancy: composition and development. *Acta Paediatr*. 2003;91(441):48-55. doi:10.1111/j.1651-2227.2003.tb00646.x.

250. Roger LC, McCartney AL. Longitudinal investigation of the faecal microbiota of healthy full-term infants using fluorescence in situ hybridization and denaturing gradient gel electrophoresis. *Microbiology*. 2010;156:3317-28. doi:10.1099/mic.0.041913-0.

251. Jakobsson HE, Abrahamsson TR, Jenmalm MC, et al. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut*. 2014;63(4):559-66. doi:10.1136/gutjnl-2012-303249.

252. Penders J, Gerhold K, Stobberingh EE, et al. Establishment of the intestinal microbiota and its role for atopic dermatitis in early childhood. *J Allergy Clin Immunol*. 2013;132(3):601-607. doi:10.1016/j.jaci.2013.05.043.

253. Pandey PK, Verma P, Kumar H, Bavdekar A, Patole MS, Shouche YS. Comparative analysis of fecal microflora of healthy full-term Indian infants born with different methods of delivery (vaginal vs cesarean): Acinetobacter sp. prevalence in vaginally born infants. *J Biosci*. 2012;37(S1):989-998. doi:10.1007/s12038-012-9268-5.

254. Van Nimwegen FA, Penders J, Stobberingh EE, et al. Mode and place of delivery, gastrointestinal microbiota, and their influence on asthma and atopy. *J Allergy Clin Immunol*. 2011;128(5):948-955. e3. doi:10.1016/j.jaci.2011.07.027.

255. Thavagnanam S, Fleming J, Bromley A, Shields MD, Cardwell CR. A meta-analysis of the association between Caesarean section and childhood asthma. *Clin Exp Allergy*. 2008;38(4):629-33. doi:10.1111/j.1365-2222.2007.02780.x.

256. Tsuji H, Oozeer R, Matsuda K, et al. Molecular monitoring of the development of intetinal microbiota in Japanese infants. *Benef Microbes*. 2012;3(2):113-25. doi:10.3920/BM2011.0038.

257. Schwiertz A, Jacobi M, Frick J-S, Richter M, Rusch K, Köhler H. Microbiota in pediatric inflammatory bowel disease. *J Pediatr*. 2010;157(2):240-244.e1. doi:10.1016/j.jpeds.2010.02.046.

258. Kerckhoffs APM, Samsom M, Van der Rest ME, et al. Lower Bifidobacteria counts in both duodenal mucosa-associated and fecal microbiota in irritable bowel syndrome patients. *World J Gastroenterol*. 2009;15(23):2887. doi:10.3748/wjg.15.2887.

259. Kirjavainen P V, Arvola T, Salminen SJ, Isolauri E. Aberrant composition of gut microbiota of allergic infants: a target of bifidobacterial therapy at weaning? *Gut*. 2002;51:51-55.

260. O'Sullivan A, He X, McNiven EMS, Haggarty NW, Lönnerdal B, Slupsky CM. Early diet impacts infant rhesus gut microbiome, immunity, and metabolism. *J Proteome Res*. 2013;12(6):2833-45. doi:10.1021/pr4001702.

261. Carlisle EM, Poroyko V, Caplan MS, Alverdy J, Morowitz MJ, Liu D. Murine gut microbiota and transcriptome are diet dependent. *Ann Surg*. 2013;257(2):287-94. doi:10.1097/SLA.0b013e318262a6a6.

262. De Palma G, Capilla A, Nova E, et al. Influence of milk-feeding type and genetic risk of developing coeliac disease on intestinal microbiota of infants: the PROFICEL study. *PLoS One*. 2012;7(2):e30791. doi:10.1371/journal.pone.0030791.

263. Azad MB, Kozyrskyj AL. Perinatal programming of asthma: the role of gut microbiota. *Clin Dev Immunol*. 2012;2012:932072. doi:10.1155/2012/932072.

264. Azad MB, Konya T, Maughan H, et al. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *Can Med Assoc J*. 2013;185(5):385-94. doi:10.1503/cmaj.121189.

265. Konieczna P, Akdis CA, Quigley EMM, Shanahan F, O'Mahony L. Portrait of an immunoregulatory Bifidobacterium. *Gut Microbes*. 2012;3(3):261-6. doi:10.4161/gmic.20358.

266. Russell SL, Gold MJ, Hartmann M, et al. Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. *EMBO Rep*. 2012;13(5):440-7. doi:10.1038/embor.2012.32.

267. Wang Y, Hoenig JD, Malin KJ, et al. 16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis. *ISME J*. 2009;3(8):944-54. doi:10.1038/ismej.2009.37.

268. Tanaka S, Kobayashi T, Songjinda P, et al. Influence of antibiotic exposure in the early postnatal period on the development of intestinal microbiota. *FEMS Immunol Med Microbiol*. 2009;56(1):80-7. doi:10.1111/j.1574-695X.2009.00553.x.

269. Mshvildadze M, Neu J, Shuster J, Theriaque D, Li N, Mai V. Intestinal microbial ecology in premature infants assessed with non-culture-based techniques. *J Pediatr*. 2010;156(1):20-5. doi:10.1016/j.jpeds.2009.06.063.

270. Angelakis E, Merhej V, Raoult D. Related actions of probiotics and antibiotics on gut microbiota and weight modification. *Lancet Infect Dis*. 2013;13(10):889-99. doi:10.1016/S1473-3099(13)70179-8.

271. Von der Weid T, Ibnou-Zekri N, Pfeifer A. Novel probiotics for the management of allergic inflammation. *Dig liver Dis*. 2002;34(Suppl. 2):S25-8.

272. Feleszko W, Jaworska J, Rha R-D, et al. Probiotics and prebiotic galacto-oligosaccharides in the prevention of allergic diseases: a randomized, double-blind, placebo-controlled trial. *Clin Exp allergy*. 2007;37(4):498-505. doi:10.1111/j.1365-2222.2006.02629.x.

273. Jakaitis BM, Denning PW. Commensal and probiotic bacteria may prevent NEC by maturing intestinal host defenses. *Pathophysiology*. 2014;21(4):47-54. doi:10.1016/j.pathophys.2013.11.012.

274. Kukkonen K, Savilahti E, Haahtela T, et al. Probiotics and prebiotic galacto-oligosaccharides in the prevention of allergic diseases: a randomized, double-blind, placebo-controlled trial. *J Allergy Clin Immunol*. 2007;119(1):192-8. doi:10.1016/j.jaci.2006.09.009.

275. Rautava S, Collado MC, Salminen S, Isolauri E. Probiotics modulate host-microbe interaction in the placenta and fetal gut: a randomized, double-blind, placebo-controlled trial. *Neonatology*. 2012;102(3):178-84. doi:10.1159/000339182.

276. Satokari R, Grönroos T, Laitinen K, Salminen S, Isolauri E. Bifidobacterium and Lactobacillus DNA in the human placenta. *Lett Appl Microbiol*. 2009;48(1):8-12. doi:10.1111/j.1472-765X.2008.02475.x.

277. Ardissone AN, De La Cruz DM, Davis-Richardson AG, et al. Meconium microbiome analysis identifies bacteria correlated with premature birth. *PLoS One*. 2014;9(3):e90784. doi:10.1371/journal.pone.0090784.

278. Steel JH, Malatos S, Kennea N, et al. Bacteria and inflammatory cells in fetal membranes do not always cause preterm labor. *Pediatr Res*. 2005;57(3):404-11. doi:10.1203/01.PDR.0000153869.96337.90.

279. Jiménez E, Fernández L, Marín ML, et al. Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. *Curr Microbiol*. 2005;51(4):270-4. doi:10.1007/s00284-005-0020-3.

280. Penders J, Thijs C, Vink C, et al. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*. 2006;118(2):511-21. doi:10.1542/peds.2005-2824.

281. Matsumiya Y, Kato N, Watanabe K, Kato H. Molecular epidemiological study of vertical transmission of vaginal Lactobacillus species from mothers to newborn infants in Japanese, by arbitrarily primed polymerase chain reaction. *J Infect Chemother*. 2002;8(1):43-9.

282. Grönlund M, Lehtonen O, Eerola E, Kero P. Fecal microflora in healthy infants born by different methods of delivery: permanent changes in intestinal flora after cesarean delivery. *J Paediatr Gastroenterol Nutr*. 1999;28(1):19-25.

283. Mitsou EK, Kirtzalidou E, Oikonomou I, Liosis G, Kyriacou A. Fecal microflora of Greek healthy neonates. *Anaerobe*. 2008;14(2):94-101. doi:10.1016/j.anaerobe.2007.11.002.

284. Neu J, Rushing J. Cesarean versus Vaginal Delivery: Long term infant outcomes and the Hygiene Hypothesis. *Clin Perinatol*. 2012;38(2):321-331. doi:10.1016/j.clp.2011.03.008.

285. Nathan AM, De Bruyne J, Khalid F, Arumugam K. Caesarean section and asthma in Malaysian children : a case-control study. *Asian Pacific J allergy Immunol*. 2012;30(3):204-208.

286. Fernández L, Langa S, Martín V, et al. The human milk microbiota: Origin and potential roles in health and disease. *Pharmacol Res*. 2013;69(1):1-10. doi:10.1016/j.phrs.2012.09.001.

287. Martín V, Maldonado-Barragán A, Moles L, et al. Sharing of bacterial strains between breast milk and infant feces. *J Hum Lact*. 2012;28(1):36-44. doi:10.1177/0890334411424729.

288. Solís G, De Los Reyes-Gavilan CG, Fernández N, Margolles A, Gueimonde M. Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut. *Anaerobe*. 2010;16(3):307-10. doi:10.1016/j.anaerobe.2010.02.004.

289. Martín R, Jiménez E, Heilig H, 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*. 2009;75(4):965-9. doi:10.1128/AEM.02063-08.

290. Martin R, Langa S, Reviriego C, et al. Human milk is a source of lactic acid bacteria for the infant gut. *J Pediatr*. 2003;143(6):754-8.

291. Hunt KM, Foster JA, Forney LJ, et al. Characterization of the diversity and temporal stability of bacterial communities in human milk. *PLoS One*. 2011;6(6):e21313. doi:10.1371/journal.pone.0021313.

292. Fan W, Huo G, Li X, Yang L, Duan C. Impact of Diet in Shaping Gut Microbiota Revealed by a Comparative Study in Infants During the First Six Months of Life. *J Microbiol Biotechnol*. 2014;24(2):133-143.

293. Bezirtzoglou E, Tsiotsias A, Welling GW. Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe*. 2011;17(6):478-82. doi:10.1016/j.anaerobe.2011.03.009.

294. Díaz-Ropero MP, Martín R, Sierra S, et al. Two Lactobacillus strains, isolated from breast milk, differently modulate the immune response. *J Appl Microbiol*. 2007;102(2):337-43. doi:10.1111/j.1365-2672.2006.03102.x.

295. Olivares M, Díaz-Ropero MP, Martín R, Rodríguez JM, Xaus J. Antimicrobial potential of four Lactobacillus strains isolated from breast milk. *J Appl Microbiol*. 2006;101(1):72-9. doi:10.1111/j.1365-2672.2006.02981.x.

296. Martín R, Olivares M, Marín ML, Fernández L, Xaus J, Rodríguez JM. Probiotic potential of 3 Lactobacilli strains isolated from breast milk. *J Hum Lact*. 2005;21(1):8-17. doi:10.1177/0890334404272393.

297. Musilova S, Rada V, Vlkova E, Bunesova V. Beneficial effects of human milk oligosaccharides on gut microbiota. *Benef Microbes*. 2014;5(3):273-83. doi:10.3920/BM2013.0080.

298. German JB, Freeman SL, Lebrilla CB, Mills DA. Human Milk Oligosaccharides: Evolution, Structures and Bioselectivity as Substrates for Intestinal Bacteria. *Nestlé Nutr Work Ser Paediatr Program*. 2008;62:205-18. doi:10.1159/000146322.

299. Brunser O, Figueroa G, Gotteland M, et al. Effects of probiotic or prebiotic supplemented milk formulas on fecal microbiota composition of infants. *Asia Pac J Clin Nutr*. 2006;15(3):368-76.

300. Fallani M, Young D, Scott J, et al. Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. *J Pediatr Gastroenterol Nutr*. 2010;51(1):77-84. doi:10.1097/MPG.0b013e3181d1b11e.

301. Kleessen B, Bunke H, Tovar K, Noack J, Sawatzki G. Influence of two infant formulas and human milk on the development of the faecal flora in newborn infants. *Acta Paediatr*. 1995;84(12):1347-56. doi:10.1111/j.1651-2227.1995.tb13567.x.

302. Le Huërou-Luron I, Blat S, Boudry G. Breast- v. formula-feeding: impacts on the digestive tract and immediate and long-term health effects. *Nutr Res Rev*. 2010;23(1):23-36. doi:10.1017/S0954422410000065.

303. Kurokawa K, Itoh T, Kuwahara T, et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res*. 2007;14(4):169-81. doi:10.1093/dnares/dsm018.

304. Fallani M, Amarri S, Uusijarvi A, et al. Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology*. 2011;157:1385-92. doi:10.1099/mic.0.042143-0.

305. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem*. 2009;284(37):24673-7. doi:10.1074/jbc.R109.022848.

306. David LA, Maurice CF, Carmody RN, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014;505(7484):559-63. doi:10.1038/nature12820.

307. Stark PL, Lee A. The Microbiolial Ecology of The Large Bowel of Breast-fed and Formula-fed infants during the first year of life. *J Med Microbiol*. 1982;15(2):189-203.

308. Hamzelou J. Born dirty: Babies arrive with gutful of bacteria. *New Sci*. 2012;214(2860):6-7.

309. DiGiulio DB. Diversity of microbes in amniotic fluid. *Semin Fetal Neonatal Med*. 2012;17(1):2-11. doi:10.1016/j.siny.2011.10.001.

310. White BA, Creedon DJ, Nelson KE, Wilson BA. The vaginal microbiome in health and disease. *Trends Endocrinol Metab*. 2011;22(10):389-93. doi:10.1016/j.tem.2011.06.001.

311. DiGiulio DB, Romero R, Kusanovic JP, et al. Prevalence and diversity of microbes in the amniotic fluid, the fetal inflammatory response, and pregnancy outcome in women with preterm prelabor rupture of membranes. *Am J Reprod Immunol*. 2010;64(1):38-57. doi:10.1111/j.1600-0897.2010.00830.x.

312. Boggess KA, Madianos PN, Preisser JS, Moise KJ, Offenbacher S. Chronic maternal and fetal Porphyromonas gingivalis exposure during pregnancy in rabbits. *Am J Obstet Gynecol*. 2005;192(2):554-7. doi:10.1016/j.ajog.2004.09.001.

313. Gonzales-Marin C, Spratt DA, Allaker RP. Maternal oral origin of Fusobacterium nucleatum in adverse pregnancy outcomes as determined using the 16S-23S rRNA gene intergenic transcribed spacer region. *J Med Microbiol*. 2013;62:133-44. doi:10.1099/jmm.0.049452-0.

314. León R, Silva N, Ovalle A, et al. Detection of Porphyromonas gingivalis in the amniotic fluid in pregnant women with a diagnosis of threatened premature labor. *J Periodontol*. 2007;78(7):1249-55. doi:10.1902/jop.2007.060368.

315. Bearfield C, Davenport ES, Sivapathasundaram V, Allaker RP. Possible association between amniotic fluid micro-organism infection and microflora in the mouth. *BJOG*. 2002;109(5):527-33.

316. Rescigno M, Rotta G, Valzasina B, Ricciardi-Castagnoli P. Dendritic cells shuttle microbes across gut epithelial monolayers. *Immunobiology*. 2001;204(5):572-81. doi:10.1078/0171-2985-00094.

317. Makino H, Kushiro A, Ishikawa E, et al. Transmission of intestinal Bifidobacterium longum subsp. longum strains from mother to infant, determined by multilocus sequencing typing and amplified fragment length polymorphism. *Appl Environ Microbiol*. 2011;77(19):6788-93. doi:10.1128/AEM.05346-11.

318. Bjerke GA, Wilson R, Storrø O, Øyen T, Johnsen R, Rudi K. Mother-to-Child Transmission of and Multiple-Strain Colonization by Bacteroides fragilis in a Cohort of Mothers and Their Children. *Appl Environ Microbiol*. 2011;77(23):8318-24. doi:10.1128/AEM.05293-11.

319. Moles L, Gómez M, Heilig H, et al. Bacterial diversity in meconium of preterm neonates and evolution of their fecal microbiota during the first month of life. *PLoS One*. 2013;8(6):e66986. doi:10.1371/journal.pone.0066986.

320. Hu J, Nomura Y, Bashir A, et al. Diversified microbiota of meconium is affected by maternal diabetes status. *PLoS One*. 2013;8(11):e78257. doi:10.1371/journal.pone.0078257.

321. Rautava S, Luoto R, Salminen S, Isolauri E. Microbial contact during pregnancy, intestinal colonization and human disease. *Nat Rev Gastroenterol Hepatol*. 2012;9(10):565-76.

322. Madan JC, Salari RC, Saxena D, et al. Gut microbial colonisation in premature neonates predicts neonatal sepsis. *Arch Dis childhood Fetal neonatal Ed*. 2012;97(6):F456-62. doi:10.1136/archdischild-2011-301373.

323. Stewart CJ, Marrs ECL, Magorrian S, et al. The preterm gut microbiota: changes associated with necrotizing enterocolitis and infection. *Acta Paediatr*. 2012;101(11):1121-7. doi:10.1111/j.1651-2227.2012.02801.x.

CHAPTER 2

GENERATING HIGH QUALITY ILLUMINA MISEQ SEQUENCING DATA: EXPERIMENTAL AND COMPUTATIONAL APPROACHES

Summary

High-throughput sequencing has provided great insight into our GIT microbial inhabitants. However, despite the advances associated with sequence-based techniques, they are not without their own limitations. The aims of this chapter therefore were to assess the data generated using Illumina Miseq sequencing technology. We mainly aimed to determine whether we had sequenced a true representation of bacteria present from the faecal specimens under study and whether we introduced any biasing factors during our experimental and sequencing approaches.

Using the QIAsymphony® SP instrument, we extracted nucleic acid from faecal specimens sampled from apparently healthy mothers and their infants at birth as well as a subset of infants at 4-12 and 20-28 weeks of age. Following nucleic acid extraction, we sequenced the V4 region of the bacterial 16S rRNA gene using Illumina-Miseq sequencing technology. We then filtered raw sequenced reads based on their quality scores and other sequencing artifacts and clustered and classified them using a bio-informatics workflow. Quality assessment following the bio-informatics workflow included determining whether sufficient sequencing depth was obtained for the specimens under study; evaluation of contaminants from non-template controls; and determining sequencing reproducibility using technical repeats. We also determined whether low template concentrations used during library preparation had any effect on the reproducibility of our sequencing results.

Faecal specimens were collected from 90 mothers and 107 infants at birth; as well as 72 infants at 4 to 12 and 36 infants at 20 to 28 weeks of age. The quality scores from Illumina's sequencing software indicated that 82% of the bases sequenced had passed quality filters with a quality score greater than 30. Following the bio-informatics workflow, the ability to detect bacteria at genus-level was greatest for meconium specimens. This was followed by infant faecal specimens collected at later time points. Maternal faecal specimens showed the lowest ability to detect bacteria at genus-level. We observed potential contamination of 69 bacterial genera within our two non-template controls, which were corrected for across all sequencing reactions. Correction resulted in a loss of 10% of OTUs from faecal samples. Overall, the experimental reproducibility of our sequencing approach was estimated at $R^2 = 0.85$. We noted that as template concentrations used for library preparation increased, reproducibility measures also increased.

Great care needs to be taken when planning and executing microbiome 16S rRNA gene high-throughput sequencing. Our results clearly emphasise that low template concentrations, used during library preparation, may result in reduced sequencing reproducibility. We also show the importance of bio-informatics workflows for quality filtering; despite high quality scores obtained from the sequencing platform.

2.1 Introduction

As discussed in the literature review (Chapter 1), Illumina sequencing of the 16S rRNA gene has been successfully used to sequence complex bacterial profiles from the human GIT.[1–3] One of the reasons for this is that the GIT contains a highly diverse microbial community,[4,5] requiring large amounts of reads to fully describe hundreds to thousands of taxa inhabiting it.[6,7] In the event that sequencing is not performed with sufficient depth, researchers may miss out on the less abundant species, resulting in representational bias of the species inhabiting the GIT.[8,9] Despite these concerns around the depth of sequencing, other factors may also impact the community profiles of the GIT. These include, for example, sequencing artifacts,[10] sampling strategies,[11] DNA isolation protocols,[1213] DNA template concentration, [14] variation among technical replicates,[15] library contamination[16] and data adjustment prior to analysis.[17,18]

The sequencing quality needs to be high in order to perform high quality downstream analyses.[10] Although sequencing output is filtered by Illumina sequencing analysis software, several sequencing artifacts, such as poor quality reads, chimeras and reads with adapters, still remain in the dataset.[16] It is therefore advisable to perform additional quality control steps to filter high quality reads obtained from sequencing softwares to exclude for sequencing artifacts.[10] The number of specimens tested also affects the estimate of the community's diversity and species richness,[19] since the number of different types of organisms within any community tend to increase with the number of specimens tested.[11] Another important aspect of concern when studying GIT microbiota profiles is the protocol used to isolate microbial DNA from the faecal specimens.[20] Studies have shown that the use of low DNA template during library preparation may impact on the variability of the sequencing results[14] and may also result in cross-contamination during the generation of amplicons using PCR.[21,22] Cross-contamination may result in false-positive PCR products, therefore leading to a false representation of diversity within and between specimens, and affect data analyses. Furthermore, the data obtained from microbiome studies are multivariate count data which are statistically challenging to analyse due to over-dispersion and the number of zeros they contain.[23] Even though zero values may represent rare components or components truly absent from the community, these zero values may compromise analyses of the microbiome data.[23]

The specific objectives for this chapter were (i) to provide summary statistics of the data metrics obtained from the Illumina Miseq sequencer software CASAVA version 1.8.2 (which includes quality statistics and the number of reads sequenced); (ii) to provide data metrics obtained from the bio-informatics workflow performed (which includes the number of filtered and trimed reads as well as the number of OTUs classified); (iii) to assess if sequencing was performed with sufficient depth in

order to measure all genera present within faecal specimens; (iv) to determine whether any contamination occurred during library preparation and to correct for it, if necessary; (v) to assess whether DNA concentrations used during library preparation had any effect on the technical reproducibility of this study; (vi) to determine the intra-individual diversity for faecal specimens under study; and finally (vii) to determine if there were any correlations between DNA concentrations used for library preparation, the number of reads sequenced, the number of OTUs classified and intra-individual diversities measured.

2.2 Materials and methods

2.2.1 Study design and setting

We conducted a pilot study nested within the Drakenstein Child Health Study (DCHS), a birth cohort study, which aims to investigate the epidemiology, risk factors and aetiology of lower respiratory tract infections and their impact on child health in a low middle income country.[24]

The Drakenstein municipality (Figure 5) is a peri-urban sub-district within the Cape Winelands, Western Cape, South Africa, with an estimated population of around 200 000.[25] The public health sector in the Drakenstein municipality is comprised of 23 medical facilities, including one centralized hospital, Paarl hospital.[26] Paarl hospital offers obstetric care, which includes routine ultrasound for all pregnant women at 20-24 weeks of gestation, as well as all hospital-based paediatric care. The public health system, together with its close proximity to the University of Cape Town, made this sub-district a highly appropriate setting for this research.

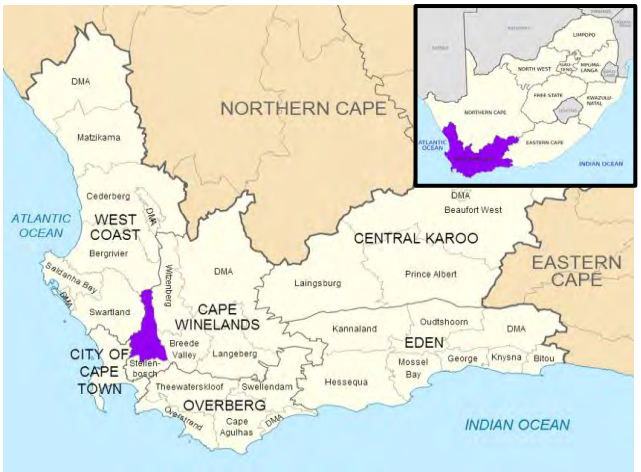


Figure 5. Geographical location (in purple) of the Drakenstein Child Health Study, Western Cape, South Africa
Source for country map: http://en.wikipedia.org/wiki/Administrative_divisions_of_South_Africa#/media/File:Map_of_South_Africa_with_English_labels.svg Source for province map: http://commons.wikimedia.org/wiki/File:Map_of_the_Western_Cape_with_municipalities_labelled_(2006).svg

2.2.2 Ethics statement

Both the DCHS and our pilot study received approval from the Faculty of Health Sciences, Human Research Ethics Committee (HREC) of the University of Cape Town, South Africa. The HREC reference numbers assigned to these studies were 401/2009 for the DCHS, and 742/2013 for this pilot study.

2.2.3 Study participant reqruitment

Enrolment of pregnant mothers occurred at 20 to 28 weeks of gestation upon routine antenatal ultrasound scanning at Paarl Hospital or primary health care facilities in two semi-rural areas, TC Newman and Mbekweni. In order for mothers to be enrolled, they had to provide informed consent and had to be residents of one of the two semi-rural areas mentioned above. Infant enrolment occurred at birth. Routine study follow-ups took place during public sector and vaccination clinic visits, scheduled at six, ten and 14 weeks, as well as six months of age.

2.2.4 Faecal specimen collection and selection

Collection of faecal specimens from mothers and their infants occurred at birth. All faecal specimen containers were labelled with pre-printed stickers. Caregivers collected maternal faecal specimens from bedpans using a sterile spatula attached to the lid of a sterile stool collection container. Collection of meconium specimens from newborn infants occurred prior to hospital discharge. In the event where hospital-based collections were not possible, the study staff encouraged mothers with freezers at home to sample the infant's first faecal discharge and to store specimens at -20 ˚C. In addition to meconium specimens, mothers having a freezer at home also collected longitudinal infant faecal specimens one day prior to monthly scheduled visits. Infant faecal specimens were decanted from the diaper into a sterile stool collection container using the sterile spatula attached to the container's lid. All faecal specimens immediately received a study lable, followed by storage at -20 ˚C. The next day, mothers transported these specimens to the clinics in ice boxes containing ice blocks, provided by the study. The process of home collections was not monitored in any way. Upon arrival at the clinics, specimens were stored at -20 ˚C, whereafter they were transferred to the Division of Medical Microbiology, University of Cape Town on the same day. The study made use of cooler boxes containing ice blocks to transport specimens to the laboratory and specimens were stored at -80 ˚C until further processing.

This pilot study aimed to include as many specimens as possible from mothers and their infants at birth, as well as a subset of longitudinal specimens from infants up to the age of seven months. Only a small proportion of the total births in the cohort had occurred when samples were selected for this study. Specimen selection from mothers and infants included in the DCHS database

required that specimens were collected at birth. Longitudinal specimens were included from infants in the event that a faecal specimen was collected from the infant at birth and that longitudinal collections occurred between 4 and 12 weeks or 20 and 28 weeks of age. When referring collectively to these specimens or participants in subsequent sections; we will be using the term "groups under study".

2.2.5 Nucleic acid extraction from faecal specimens

Approximately 50 mg of faecal specimen served as the starting material for nucleic acid extraction. We performed these extractions using the QIAsymphony DSP Virus/Pathogen Mini Kit® (Qiagen GmbH, Hilden, Germany), as previously described.[27] Fluorometrical DNA quantifications of all extractions took place using the Qubit® 2.0 Fluorometer (Invitrogen™, CA, USA) together with the Qubit™ dsDNA HS Assay Kit (Invitrogen™, CA, USA). The lower and upper detection limit of DNA concentrations using the Qubit™ dsDNA HS Assay Kit (Invitrogen™, CA, USA) were 0.05 ng/µl and 60 ng/µl, respectively. Specimens were excluded from DNA concentration summary statistics (Section 2.2.9.1) in the event that the DNA concentrations were outside the detection limits of the Qubit™ dsDNA HS Assay Kit (Invitrogen™, CA, USA).

2.2.6 16S ribosomal RNA gene library preparation

Following nucleic acid extractions, all 16S rRNA gene library preparation steps were carried out at the J Craig Venter Instutitute (JCVI), Maryland, United States of America (USA).

2.2.6.1 *Faecal specimens and controls*

Faecal specimens were collected from mothers and infants at birth, as well as a subset of infants at 4 to 12 weeks and 20 to 28 weeks of age. These specimens are referred to herein as the "groups under study". We also sequenced a set of controls, herein referred to as the "sequencing controls". "Sequencing controls" consisted of positive controls, which included genomic DNA from two microbial mock communities known as BEI controls HM-782D and HM-783D (BEI Resources, VA, USA) as well as genomic DNA extracted from *Escherichia coli*; non-template controls which consisted of PCR-grade water; and "technical repeats" which consisted of nucleic acid randomly selected from the sampling "groups under study".

2.2.6.2 *PCR amplification of 16S ribosomal bacterial DNA*

We performed two PCR reactions using previously published primers (515F and 806R) targeting the V4 hypervariable region of the 16S rRNA gene,[28] synthesised by Sigma Adrich (Sigma-

Aldrich®, MO, USA) . Nucleic acid extracts from faecal specimens of the study participants (described in Section 2.2.5) served as template in PCR amplification reactions.

The first PCR reaction contained 12.5 µl of 2X MyTaq™ HS Mix (Bioline, MA, USA), 2 µl of each primer (10µM initial concentration), 0.75 µl of dimethyl sulfoxide (Sigma-Aldrich®, MO, USA) and 4 µl of nucleic acid per specimen made up to a final volume of 25.25 µl using PCR-grade water (Thermo Fisher Scientific Inc., MA, USA). Positive controls, non-template controls and "technical repeats" (described in Section 2.2.6.1) were included at 4 µl each. Genomic DNA of the positive control *E. coli* was added at 15 ng/µl. We performed the PCR in the MJ Research PTC-225 Tetrad Peltier Thermal Cycler (MJ Research Inc., Quebec, Canada). Cycling conditions of the first PCR included a denaturation step at 95˚C for 3 min, an amplification step proceeding for 10 cycles at 95˚C for 30 sec, 50˚C for 30 sec and 72˚C for 1 sec; and a final extension step at 72˚C for 5 min.

The second PCR reaction made use of the same set of primers (515F and 806R), but both these primers contained sequencing adapters, priming regions as well as 12 to 15 random nucleotides. In addition, to facilitate in multiplexing of specimens, the reverse primer was barcoded with a 12-base Golay code.[28] The second PCR reaction contained 12.5 µl of 2X MyTaq™ HS Mix (Bioline, MA, USA), 4 µl of each primer (10µM initial concentration), 0.75 µl of dimethyl sulfoxide (Sigma-Aldrich®, MO, USA) and 4 µl of amplified DNA from the first PCR reaction, made up to a final volume of 25.25 µl using PCR-grade water (Thermo Fisher Scientific Inc., MA, USA). Cycling conditions were the same as for the first PCR reaction; however the amplification step was set at 30 cycles.

2.2.6.3 *Purification of PCR products*

We purified PCR products using the Agencourt® AMPure® XP PCR Purification kit (Beckman Coulter, CA, USA) as per manufacturer's protocol, with slight modifications. Briefly, we modified step 2 by adding 0.65 µl of Agencourt AMPure XP solution per PCR reaction volume. Modifications to step 3 included an additional homogenisation step using the Fisher Vortex Genie 2 (Fisher Scientific, NY, USA), followed by quick centrifugation using the GS-6R Centrifuge (Beckman Coulter, CA, USA). At step 7, we used Tris-EDTA (pH 8.0) (Amresco, OH, USA) as elution buffer, whereafter modifications to step 9 included the transfer of 35 µl of eluent into a new plate.

2.2.6.4 *Viewing of PCR products on agarose gels*

First, we mixed 5 µl of purified PCR products with 2 µl of 1X loading dye and loaded the mixture on a 1.5% agarose gel containing 0.1 µg/µl EtBr (Invitrogen™, CA, USA), submerged in 1X TAE buffer (Thermo Scientific, PA, USA). Then 5 µl of the molecular weight marker, Trackit™ 1 Kb

Plus DNA Ladder (Invitrogen™, CA, USA) was loaded on the 1.5% agarose gel. We performed electrophoresis for 90 min at 80 V using the Power Pac 300 (Bio-Rad Laboratories Inc., CA, USA) and viewed PCR products using the Typhoon 9410 Molecular Imager and Typhoon Scanner Control 5.0 software (GE Amersham Molecular Dynamics, PA, USA).

2.2.6.5 *Quantification of PCR products*

We quantified PCR products using the Quanti-iT™ dsDNA High-Sensitivity Assay Kit (Invitrogen™, CA, USA), as per manufacturer's instructions. In order to record DNA absorbance; we used the Infinite M1000 Pro® microplate reader (Tecan Group Ltd., Grödig, Austria) equipped with Tecan i-Control™ 1.7 software.

2.2.6.6 *Pooling of PCR products at equimolar ratio, purification and gel extraction*

In order to pool 100 ng of the respective PCR products, we calculated equimolar pooling volumes for each PCR product based on their initial DNA concentrations. Following pooling, the Nanodrop ND 1000 (Thermo Scientific, DE, USA), equipped with ND-1000 3.7.1 software, served as spectrophotometer for quantifying pooled PCR products.

Purification of pooled PCR products took place using Agencourt AMPure XP solution (Beckman Coulter, CA, USA), as per the manufacturer's instructions. Slight modifications to the protocol included aliquoting the pooled PCR products into volumes of 400 µl to which we added Agencourt AMPure XP solution in a 1:1 ratio. We eluted purified PCR products in 60 µl Tris-EDTA buffer (pH 8.0) (Amresco, OH, USA) and combined the purified PCR products in a single pool which was quantified using the Nanodrop ND 1000 spectrophotometer equipped with ND-1000 3.7.1 software.

Next, we mixed the purified pooled PCR products (6.6 µg) with 5 µl of 1X loading dye and loaded the mixture on a 1.5% agarose gel containing 0.1 µg/µl EtBr (Invitrogen™, CA, USA), submerged in 1X TAE buffer (Thermo Scientific, PA, USA) containing 1 µg/µl EtBr (Invitrogen™, CA, USA). Also loaded on the above-mentioned agarose gel was 7 µl of Trackit™ 1 Kb Plus DNA Ladder (Invitrogen™, CA, USA). We performed electrophoresis for 30 min at 35 V, followed by 45 min at 40 V and 3 hours at 70 V using the Power Pac 300 (Bio-Rad Laboratories Inc., CA, USA). This was followed by viewing the pooled PCR product with the Typhoon 9410 Molecular Imager and Typhoon Scanner Control 5.0 software (GE Amersham Molecular Dynamics, PA, USA). We then cut the PCR product from the agarose gel at the expected target size (between 400 and 500 bp) and gel purified it using the QIAquick Gel Extraction kit (QIAgen, MA, USA), as per manufacturer's instructions. Slight modifications for gel purification included centrifugation at 13 000 rpm during step 9, incubation at

37˚C for 5 min during step 10 and heating of elution buffer Tris-EDTA buffer (pH 8.0) (Amresco, OH, USA) to between 60 and 70 ˚C during step 13.

2.2.7 16S ribosomal RNA gene sequencing using Illumina MiSeq technology

Following library preparation steps, we performed 16S rRNA gene sequencing using Illumina MiSeq technology at JCVI, Maryland, USA.

2.2.7.1 *Quantification of pooled PCR products for sequencing purposes*

To quantify the pooled PCR products, we used the KAPA Library Quantification Kit K4835 (KAPA Biosystems, MA, USA) as per manufacturer's instructions, with slight modifications. These modifications included diluting DNA of pooled PCR products at 1:120, 1:1200 and 1:12000 during step 1. We performed the 1:120 dilution by adding 2 µl of PCR product to 238 µl of dilution buffer (Supplementary data), the 1:1200 dilution by adding 5 µl of the 1:120 dilution to 45 µl of dilution buffer (Supplementary data) and the 1:12000 dilution by adding 5 µl of the 1:1200 dilution to 45 µl of dilution buffer (Supplementary data). Each of these dilution steps required the use of the Vortex Genie 2 (Scientific Industries, NY, USA) before the next dilution was made. We made 1:1200 dilutions of our positive control Shrimp Aqu BIOFLOC3-PE-IL55-01. Each PCR reaction contained 6 µl of KAPA SYBR® FAST qPCR Master Mix Product (KAPA Biosystems, MA, USA) and 4 µl of either diluted PCR product, DNA quantification standards, positive control or non-template control. Performance of all PCR reactions took place in triplicate, using the ABI PRISM® 7900HT Sequence Detection System with SDS Plate Utility Software Version 2.1 (Applied Biosystems, CA, USA). The PCR cycling conditions included a denaturation step at 95˚C for 5 min and an amplification step proceeding for 35 cycles at 95˚C for 30 sec and 60˚C for 45 sec, as per manufacturer's instructions. We omitted any failed specimens before evaluating the quality of the PCR run by assessing the slope and R^2 values. An acceptable slope ranged between -3.58 to -3.10 and the accepted minimum for the R^2 value was 0.99.

2.2.7.2 *Integrity and sizes of PCR products for all dilutions*

To determine product sizes for all dilutions described Section 2.2.7.1, we used 0.8% E-Gel® pre-cast agarose gels (Invitrogen™, CA, USA). We loaded a final volume of 20 µl, made up of 16 µl of water, 1 µl of 6X TrackIt™ Cyan/Yellow Loading Buffer (Invitrogen™, CA, USA) and 3 µl of the respective pooled PCR product dilutions (described in Section 2.2.7.1). This was followed by loading a volume of 5 µl of TrackIt™ 1kb Plus DNA ladder (Invitrogen™, CA, USA) and performing electrophoresis for 30 min. Finally, we viewed the PCR products using the Bio-Rad Gel Doc™ EZ System (Bio-Rad, CA, USA).

2.2.7.3 *Library denaturation and sequencing*

We prepared sequencing reagents and the DNA library using the MiSeq Reagent Kit v3, 600 Cycles (Illumina, CA, USA) as per manufacturer's instructions.[29,30] We diluted the PCR products to a final concentration of 20 pM by using 10 mM Tris-HCl, pH 8.5 (Life Technologies™, CA, USA). In order to obtain a final concentration of 4 pM of denatured DNA product, we added 220 µl of the denatured DNA to 880 µl of hybridization buffer (HT1) (provided by the manufacturer). We diluted the denatured PhiX control to a final concentration of 4 pM, by adding 220 µl of the denatured PhiX control to 880 µl of HT1 buffer. The denatured PhiX control was spiked-in at 25%, by adding 250 µl of 4 pM denatured DNA to 750 µl of 4 pM denatured PhiX control. We loaded a total volume of 600 µl of the denatured DNA library onto the Illumina® MiSeq™ platform as per manufacturer's instructions.[31] CASAVA version 1.8.2 software (Illumina, CA, USA)[32] served to convert *.bcl files to compressed *.FASTQ files. This software also de-multiplexed barcoded specimens by assigning specimen names to each of the generated *.FASTQ files. In addition, the software also generated sequencing run quality metrics and read counts.

2.2.8 Bio-informatics analysis of the bacterial sequences: YAP workflow

The bioinformatics workflow, used to classify 16S rRNA sequences to bacterial taxonomies, was carried out at JCVI, Maryland, USA. The workflow performed consisted of a set of tools integrated into a computationally efficient workflow using a Python wrapper script called YAP[33] (Figure 6).

The first step in the above-mentioned workflow was to perform quality assessments of the *.FASTQ files using the read quality statistics embedded in these files. The workflow performed these quality assessments using the FASTX toolkit[34] via assessment of Phred quality scores (Q scores).[35] Following the quality assessment, trimming of each of the 300 base-pair long paired-end reads took place at a low threshold of Q = 3 using SolexaQA's Dynamic Trim function.[36] The workflow incorporated the software FLASH[37] to overlap the above-mentioned trimmed paired-end reads. The second step of the workflow consisted of a final trim using SolexaQA's Dynamic Trim function. The quality threshold for this trimming step was set at Q = 25. The third step mainly focussed on recording and removing any duplicate reads in order to minimize the data size to allevaite computational processing. During the fourth step, the workflow determined if all reads were in the 5' to 3' direction. By using the Nearest Alignment Space Termination (NAST) algorithm for multiple sequence alignments,[38] the workflow removed all reads not corresponding to the 16S rRNA reference gene when compared against the SILVA 16S rRNA database.[39,40] The fifth step in the YAP workflow consisted of removing chimeric reads by comparing each read to the SILVA database using the chimera detecting program UCHIME[41] integrated into the software mothur.[42] A second quality

check step followed by aligning reads to the 16S rRNA gene reference region of *Escherichia coli* (downloaded from the National Center for Biotechnology Information (NCBI)). The workflow then trimmed the resultant alignment in order to remove overly long reads as well as reads not aligning to the targeted V4 region of the 16S rRNA gene. The final step of the workflow aimed to cluster all the good quality reads into OTUs using a clustering algorithm implemented in the software, CD-HIT-EST within the CD-HIT Suite.[43] The workflow performed clustering at four sequence similarity levels of 90, 95, 97 and 99%. Prior to calculating the final OTU cluster size, the workflow accounted for recorded duplicate reads by adding them to the appropriate OTU clusters. Finally taxonomic information was assigned to OTUs using the naïve Bayesian classifier within mothur[44] as well as a normalized Ribosomal Database Project (RDP) training dataset.[44] For computational efficiency, only an exemplar sequence representing each of the OTU clusters received a taxonomic classification. The exemplar of any given cluster lies at the "centroid" of all the sequences in the cluster and is equidistant from them. Following taxonomic classification, the software computes the aggregate sequence counts for any given taxonomic level by adding together the cluster sizes of OTU clusters labelled with the taxonomic label of interest.

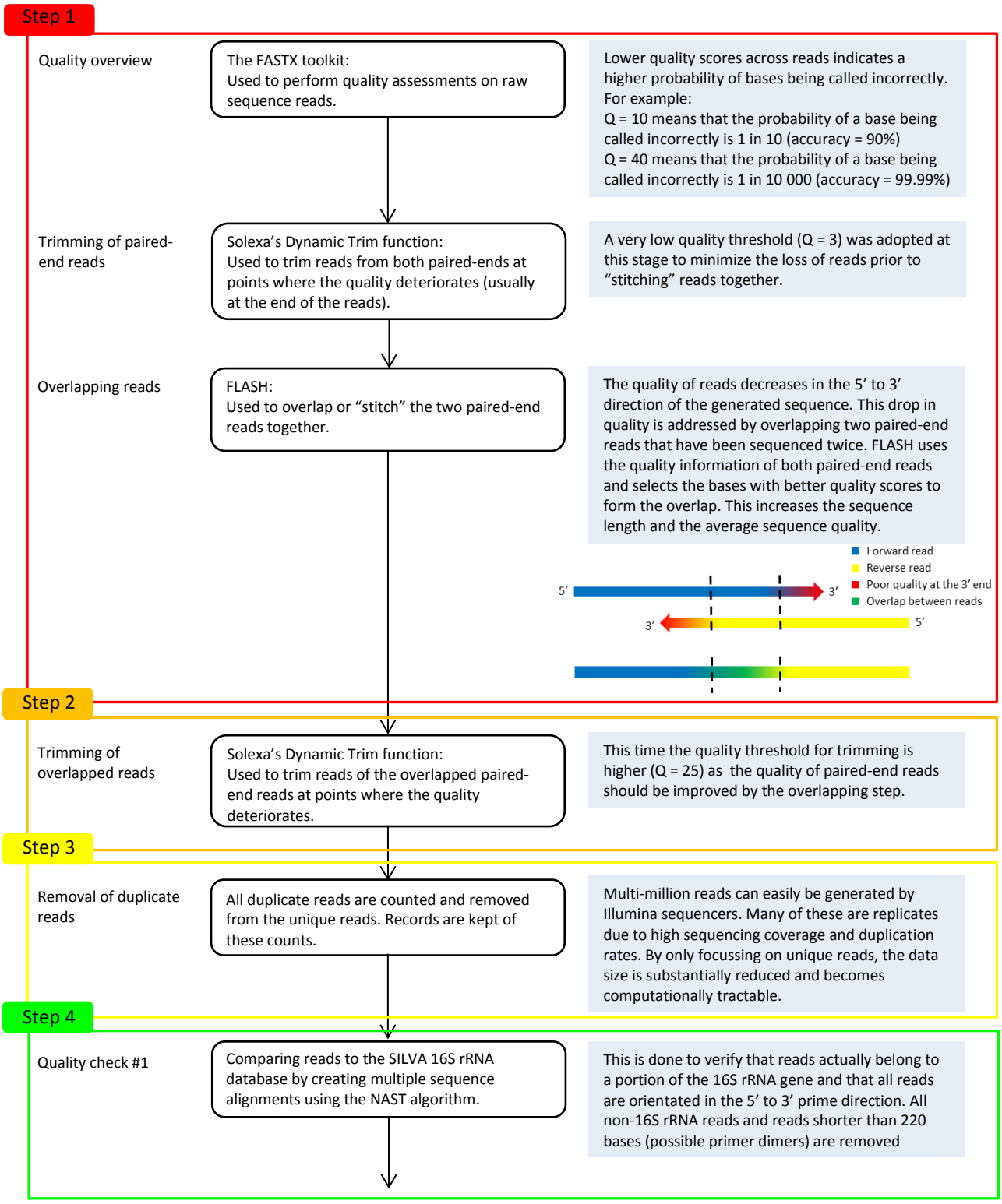


Figure 6. Schematic representation of the YAP workflow

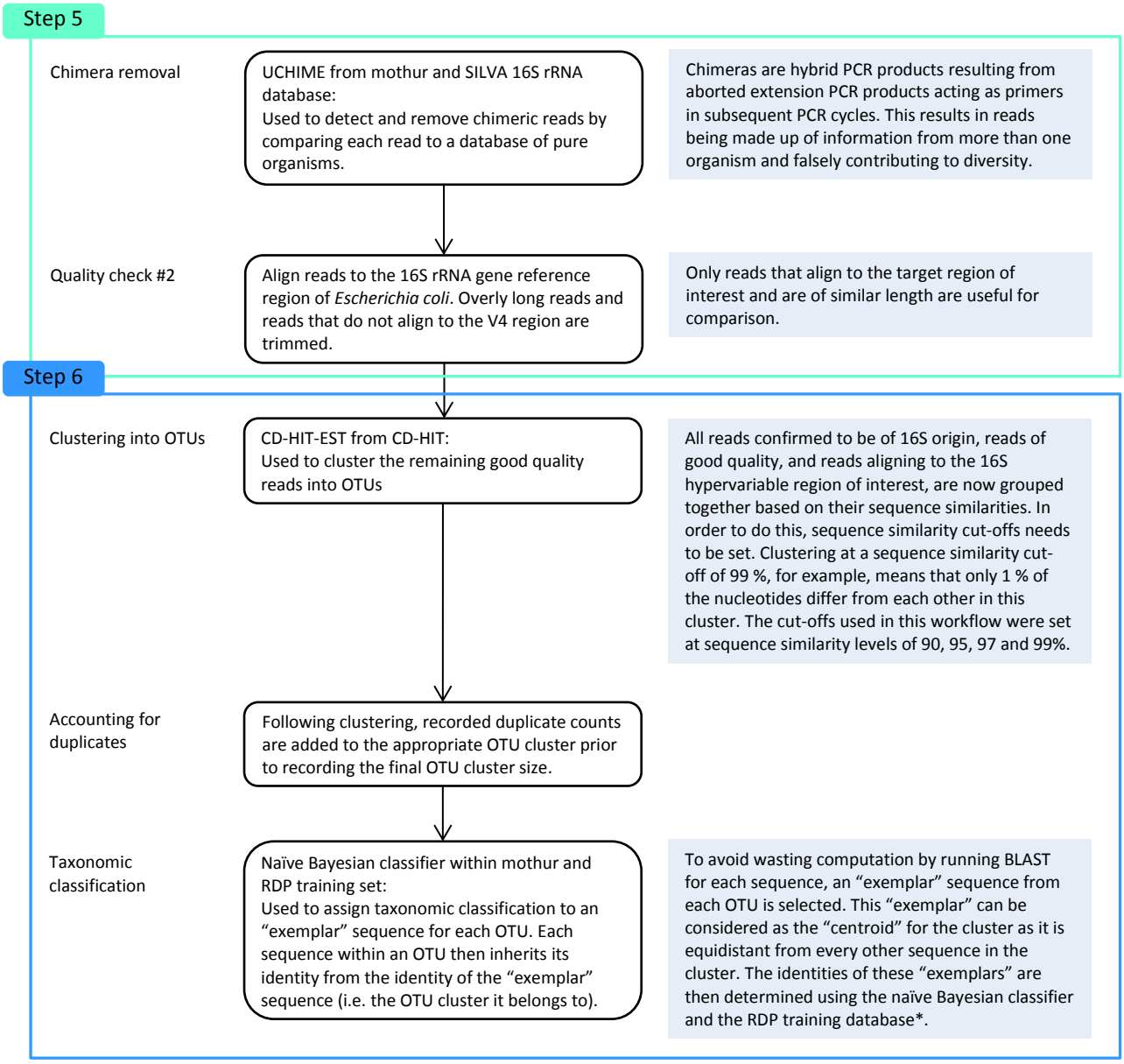


Figure 6. Schematic representation of the YAP workflow (continued)

2.2.9 Statistical analysis of sequencing data

We used R software version 3.1.1[45] together with RStudio software version 0.98.507[46] for all statistical analyses as well as graphical representations of the data. We used the output from the Illumina Miseq sequencing software CASAVA 1.8.2,[32] as well as the output from the YAP workflow[33] for quality control analyses. From the four sets of OTU tables contructed at 90, 95, 97 and 99% sequence similarity; the OTU table constructed at 97% sequence similarity (genus-level) was selected for quality control analysis.

2.2.9.1 *Summary statistics*

We used the [summ] and [tab1] functions of the R package *epicalc*[47] to compute summary statistics on DNA concentrations used for library preparation; data output from the Illumina Miseq sequencer; as well as data output form the YAP workflow. DNA concentration summary statistics only included specimens with DNA concentrations measurable using the Qubit™ dsDNA HS Assay Kit (Invitrogen™, CA, USA). Box and whisker plots were constructed for distribution analysis of the data. To compare non-parametric variables, we performed the Mann-Whitney-Wilcoxon and the Kruskall-Wallis tests using the functions [wilcox.test] and [kruskal.test] from the R package *epicalc*, as appropriate.[47] The Mann-Whitney-Wilcoxon test served for comparing two non-parametric variables, while the Kruskall-Wallis test served for multiple comparisons. By adding the argument "notch=TRUE" to the box and whisker plots constructed, we could graphically indicate which medians of the multiple comparisons are significantly different at the 5% level.[48]

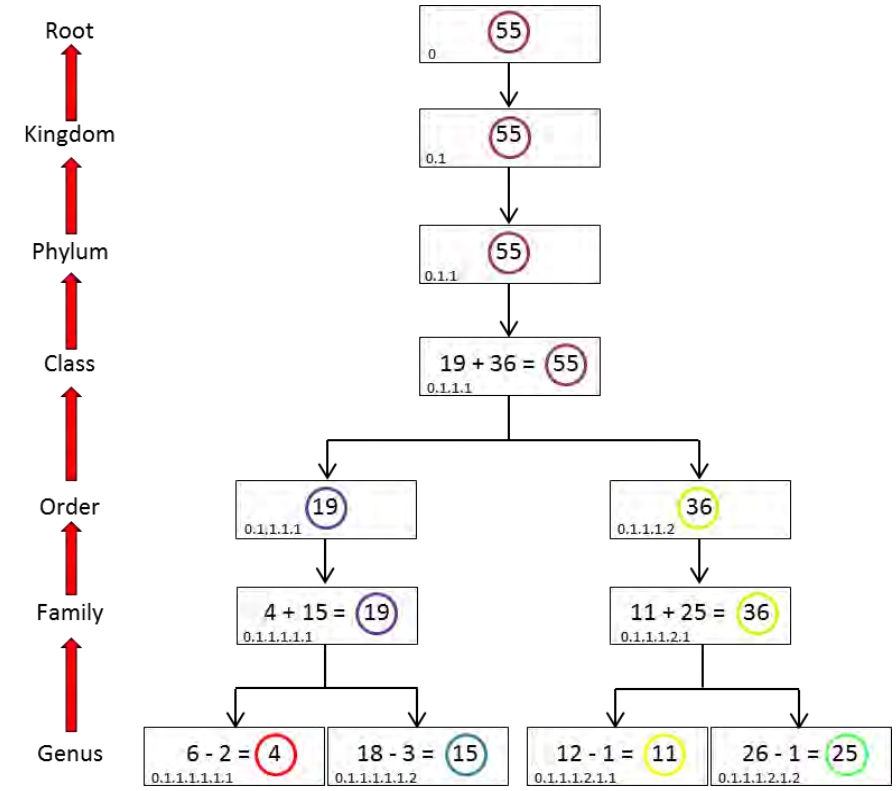
2.2.9.2 *Evaluation of sequencing depth for studying faecal bacterial communities*

We constructed accumulation curves[11] to determine if sufficient numbers of reads were obtained following the YAP workflow in order to adequately identify all bacterial genera from faecal specimens sampled from mothers and infants at birth, infants at 4 to 12 weeks of age and infants at 20 to 28 weeks of age. We evaluated the sequencing effort, by constructing accumulation curves of the number of genera (identified from OTUs classified at a similarity level of 97%) against the number of reads obtained following the YAP workflow using the function [rarefy][49] from the R package *vegan*.[50]

2.2.9.3 *Assessment and correction for potential contamination of the samples processed*

We determined whether contamination was evident in our sequencing run by analysing the two non-template controls. We also determined whether contamination occurred uniformly across all PCR plates used during library preparation. This was done by constructing barplots of the relative abundance as well as the actual numbers of potential contaminating OTUs at genus-level for the two non-template controls and all other sequencing reactions across all PCR plates. In addition, we searched the literature to determine what possible sources of contamination might exist for the putative contaminant sequences in our run. In correcting for contamination, we calculated the mean number of OTUs from the two non-template controls for each of the genera where contamination was evident. This mean was then subtracted from all OTUs representing these genera across all specimens, including the two non-template controls. The OTU table (at 97% sequence similarity), also contained OTUs at higher taxonomic classifications (namely family, order, class and phylum) which needed to be corrected. Consequently, we added the resultant numbers of OTUs at genus-

level to obtain the new number of OTUs at family-level. This was repeated for each of the higher taxonomic classifications until the root was reached (Figure 7).

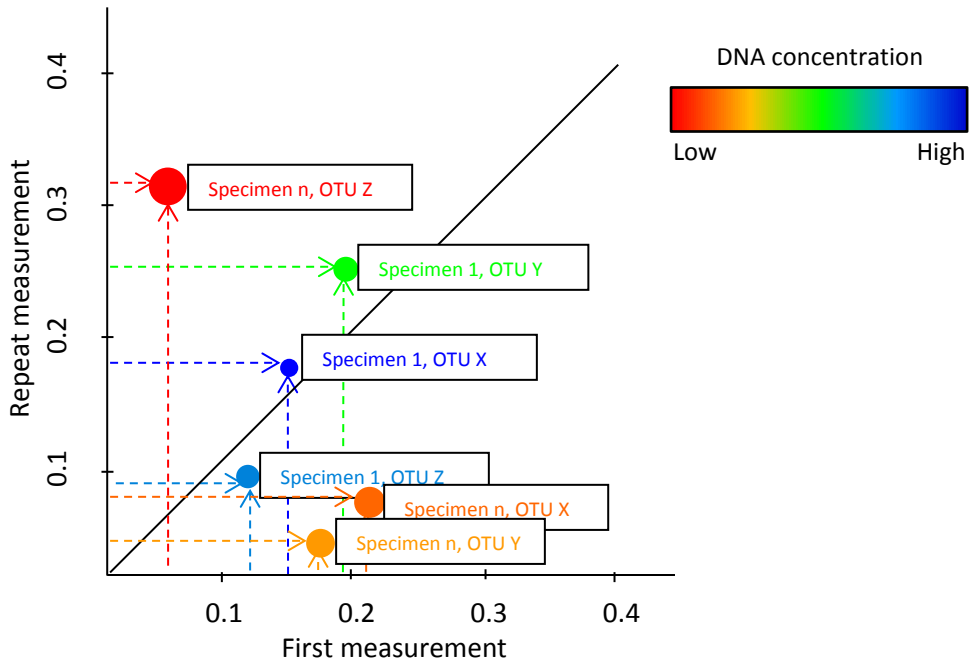


Taxonomic level	Rank ID of Taxon	Non-template control A	Non-template control B	Mean number of OTUs representing contamination	Specimen N prior to correction for contamination	Specimen N following correction for contamination
Root	0	8	6		62	55
Kingdom	0.1	8	6		62	55
Phylum	0.1.1	8	6		62	55
Class	0.1.1.1	8	6		62	55
Order	0.1.1.1.1	5	5		24	19
Order	0.1.1.1.2	3	1		38	36
Family	0.1.1.1.1.1	5	5		24	19
Family	0.1.1.1.2.1	3	1		38	36
Genus	0.1.1.1.1.1.1	1	3	2	6	4
Genus	0.1.1.1.1.1.2	4	2	3	18	15
Genus	0.1.1.1.2.1.1	1	1	1	12	11
Genus	0.1.1.1.2.1.2	2	0	1	26	25

Figure 7. Schematic representation of an OTU table during correction for potential contamination
Correction for potential contamination starts at the genus-level when the mean number of OTUs observed in the two non-template controls are calculated for each genus and subtracted from each faecal specimen. Following correction at genus-level, the resultant number of OTUs at genus-level within a family are added together to obtain the new number of OTUs at family-level. This process is repeated for all higher taxonomic levels until the root is reached. The calculations are carried out for all genera with OTUs present in the non-template controls and applied to all faecal specimens, including the two non-template controls.

2.2.9.4 *Analysis of variation between "technical repeats"*

To determine technical reproducibility of the sequencing process, the proportions of each OTU sequenced from 17 faecal specimens were compared to that of their "technical repeats". These specimens were randomly selected to be sequenced in duplicate. These specimens included three mother-infant pairs followed over time; two infants for whom we did not have maternal specimens to pair with; and a single mother for whom we did not have an infant specimen to pair with. We compared the proportion of each OTU present in the original specimen with those from their "technical repeats" using simple linear regression analysis.[51] The coefficient of determination (R^2) was calculated[51] for all 17 specimens used for repeat testing. This was also used to determine the R^2 at group level for infant specimens collected at birth, at 4-12 weeks, at 20-28 weeks of age and for maternal specimens collected at birth. Figure 8 schematically represents this strategy using two specimens, with their repeat measurements, and three OTUs as an example. In addition, we also determined the effect of DNA concentration on variation between "technical repeats". This was determined by simply assigning colours to the plotted proportion of variations, based on the DNA concentrations of the faecal specimens. DNA concentrations of < 0.05 ng/µl were coded as 0.025 ng/µl and DNA concentrations of > 60 ng/µl were coded as 65 ng/µl to enable data analysis using R software.[45]



OTU	Specimen 1		Specimen n	
	First measurement	Repeat measurement	First measurement	Repeat measurement
X	0.15	0.18	0.21	0.07
Y	0.19	0.25	0.17	0.03
Z	0.12	0.09	0.05	0.32

Figure 8. Schematic representation of reproducibility measurements correlated with DNA concentrations
The proportion of each OTU sequenced from the specimen (X axis) and its repeat measurement (Y axis) is plotted. The bigger the distance between the plot and the diagonal line, the bigger the variation measured for the respective OTU when comparing the specimen to its repeat. This variation is also highlighted by the size of each plot, with bigger sizes representing larger distances from the diagonal line. Colours are assigned based on the DNA concentrations of each specimen used for library preparation.

2.2.9.5 *Calculating intra-individual (alpha) diversity*

We calculated the Shannon diversity index (H')[52,53] (equation 3.1) to determine the alpha diversity amongst each of the "groups under study".

$$H' = -\sum_{i=1}^{s} p_i \ln p_i \quad [3.1]$$

Where i = species of interest,

p_i = proportion of species of interest (i) relative to the total number of species,

ln = natural logarithm,

Σ = sum of the resulting product across species, and

s = number of species recorded.

This diversity index measures both the richness and evenness of organisms within the specimen of interest. The greater the Shannon diversity index, the greater the diversity within the specimen. Figure 9 shows this in context of microbiome studies using OTUs to measure the diversity within, for example, nine specimens.

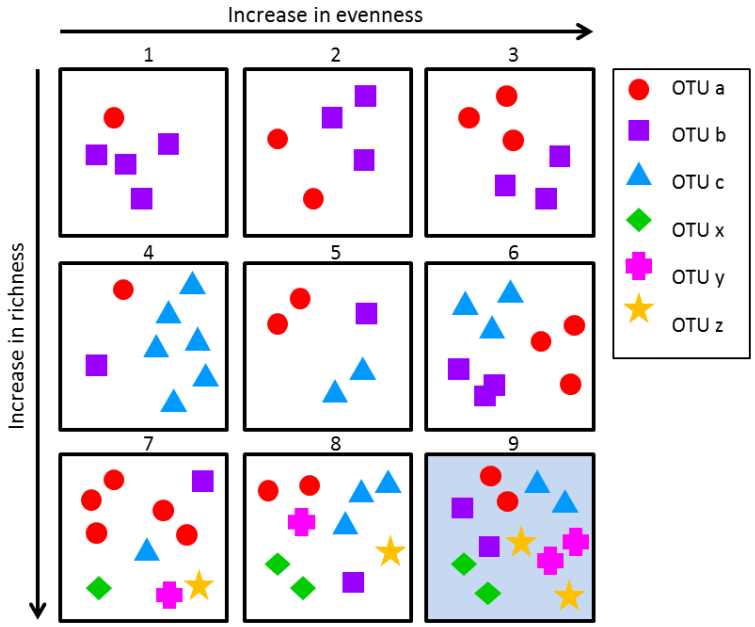


Figure 9. Schematic diagram showing species richness and evenness across nine different specimens
Each shape within the nine specimens represents an operational taxonomic unit (OTU). The Shannon index increases as both richness and evenness increase. The highest diversity will therefore be measured from specimen number 9.

2.2.9.6 *Correlations between DNA concentrations used for library preparation, the number of reads sequenced, OTUs classified and intra-individual diversity*

We used the [cor] function of the built-in R package *stats*[45] to investigate the relationships between i) the DNA concentrations used for library preparation; ii) the number of reads sequenced; iii) the number OTUs classified from the YAP workflow; iv) as well as the intra-individual diversity for

all “groups under study”. We also constructed pairwise scatterplots to visually determine the correlations between the variables i – iv listed above for all “groups under study”. These pairwise scatterplots indicate positive correlations between variables when scatter plots ascents from left to right and negative correlations when scatter plots descents from right to left. Strong correlations are indicated by plots that cluster as to resemble a straight line (slope). Correlation and pairwise scatterplot analyses were repeated separately for maternal faecal specimens collected at birth, infant meconium specimens collected at birth, and infant faecal specimens collected at 4 to 10 weeks and 20 to 28 weeks of age.

2.3 Results

2.3.1 Faecal specimens and controls sequenced using Illumina MiSeq technology

Faecal specimens from 90 mothers were sequenced. In infants, 107, 72 and 36 faecal specimen collected at birth, 4 to 12 weeks and 20 to 28 weeks of age were sequenced, respectively. All faecal specimens from both mothers and infants are referred to herein as the “groups under study” (n=305). In addition to the faecal specimens collected from the “groups under study”, we sequenced a set of controls, herein referred to as the “sequencing controls”. “Sequencing controls” consisted of two non-template controls, two BEI controls, *Escherichia coli* and 17 “technical repeats” randomly selected from the “groups under study”. The “groups under study”, together with the set of “sequencing controls”, are summarised in Figure 10 and referred to herein as “sequencing reactions” (n = 327).

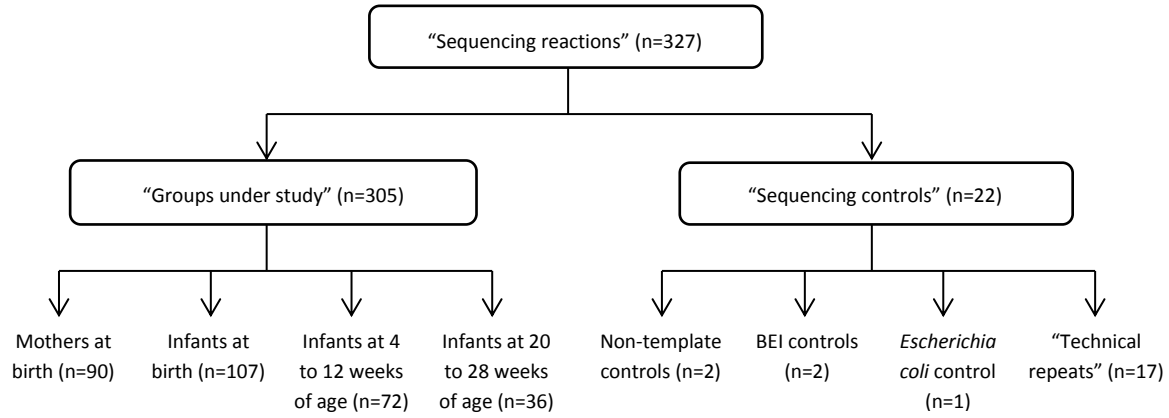


Figure 10. Summary of “sequencing reactions”
Faecal specimens sequenced from four “groups under study”. These included 90 mothers and 107 infants from which faecal specimens were collected at birth as well as 72 infants from which faecal specimens were collected at 4 to 12 weeks and 36 infants from which collection took place at 20 to 28 weeks of age. In addition, sequencing was also performed on 17 faecal specimens randomly collected from the “groups under study” to serve as “technical repeats”. Other “sequencing controls” included one *Escherichia coli* control, two BEI controls and two non-template controls.

2.3.2 DNA concentration from faecal specimens

We were not able to measure the DNA concentration in 19.6% (60/305) of the specimens tested. The majority (98.3%, 59/60) of these specimens had DNA concentrations too low to be detected by the Qubit™ dsDNA HS Assay Kit's lower detection limit (0.05 ng/µl). These specimens mainly consisted of infant meconium specimens (98.3%, 58/59), together with a single infant faecal specimen collected at 24 weeks. DNA concentration exceeding the Qubit™ dsDNA HS Assay Kit's upper detection limit (60 ng/µl) was observed in a single infant faecal specimen collected at 8 weeks of age. By excluding for all specimens outside the detection range of the Qubit™ dsDNA HS Assay Kit, we observed a statistically significant difference for the DNA concentrations extracted from the "groups under study" ($p < 0.0001$) (Figure 11). Infant meconium specimens collected at birth had the lowest DNA concentrations.

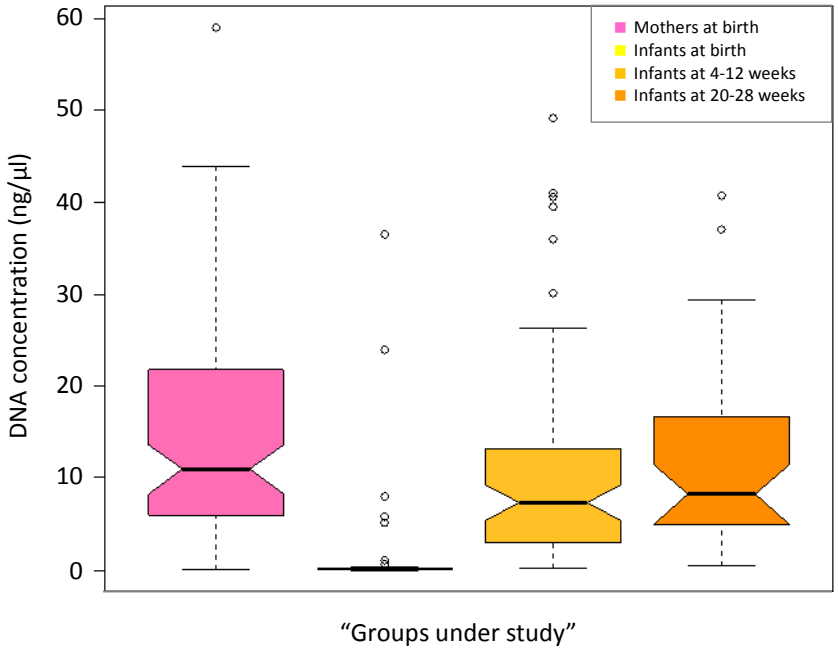


Figure 11. DNA concentrations extracted from faecal specimens for all "groups under study"
Specimens with undetectable DNA concentrations, using the Qubit™ dsDNA HS Assay Kit, were excluded from the box and whisker plots. DNA concentrationas are shown for faecal specimens collected from mothers at birth (n=90), meconium specimens collected from infants at birth (n=49), faecal specimens collected from infants at 4 to 12 weeks of age (n=71) and 20 to 28 weeks of age (n=35). The line within each box plot represents the median value. The median values of box plots are significantly different (at 5% significance level) in the event that their notches do not overlap. The box extends from the 25th to 75th percentile and whiskers show the minimum and maximum values.

2.3.3 Amplicon concentrations from PCR prior to pooling and sequencing

Figure 12 shows the amplicon concentrations from the Quanti-iT^{TM} dsDNA High-Sensitivity Assay Kit. We observed significant difference in amplicon concentrations when comparing all "groups under study" and all "sequencing controls" ($p < 0.0001$). The median amplicon concentration for all "groups under study" was 74.2 ng/µl (interquartile range (IQR), 34.4 - 102.4).

The infant meconium specimens produced significantly lower amplicon concentrations compared to the other three groups from the "groups under study" ($p < 0.0001$) (Figure 12A). All "sequencing reactions", including the two non-template controls, produced amplicons (Figure 12B). The median concentration of 6.0 ng/µl (IQR, 4.7 - 7.3) for the two non-template controls were significantly lower compared to those from the "groups under study" and the remainder of the "sequencing controls" ($p < 0.0001$). The median amplicon concentration of the "sequencing controls", excluding for the two non-template controls, was 85.3 ng/µl (IQR, 67.6 - 96.6).

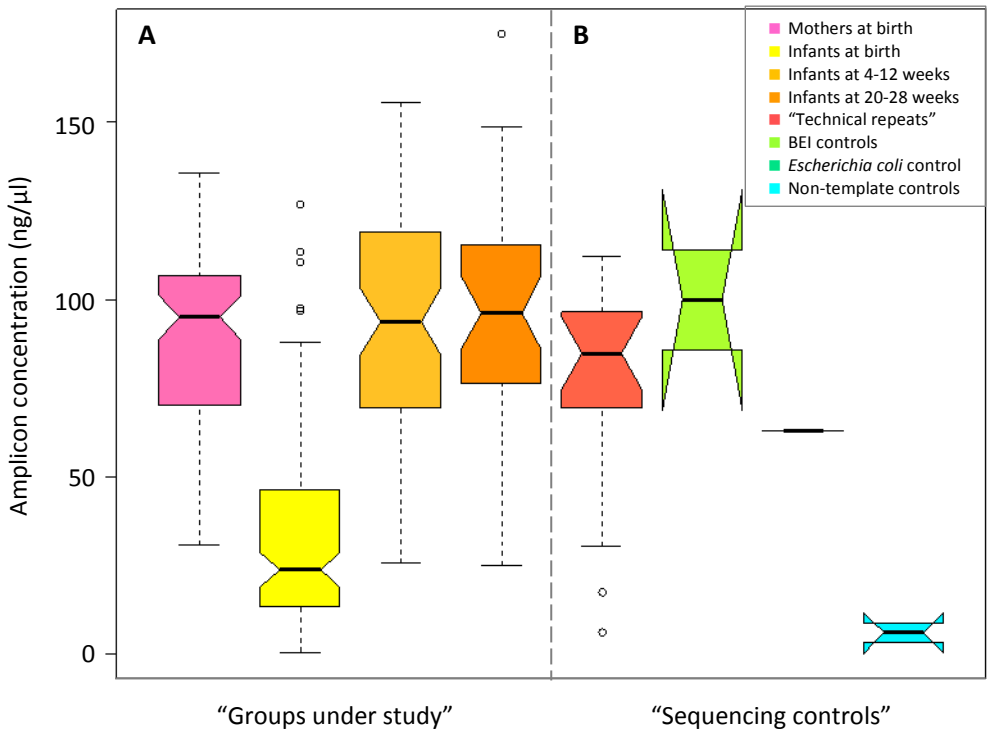


Figure 12. Amplicon concentrations from PCRs performed prior to pooling and sequencing steps
A) Amplicon concentrations are shown for faecal specimens collected from "groups under study", which included mothers at birth (n=90), meconium specimens collected from infants at birth (n=107), and faecal specimens collected from infants at 4 to 12 (n=72) and 20 to 28 weeks of age (n=36). **B)** Amplicon concentrations of our "sequencing controls" included "technical repeats" (n=17), BEI controls (n=2), an *E.coli* control (n=1) and non-template controls (n=2). The line within the box plot indicates the median values. The median values of box plots are significantly different (at 5% significance level) in the event that their notches do not overlap. The areas extending on either side of the notches (green and blue box plots) indicate that the confidence intervals exceeded the inter-quartile ranges. The box extends from the 25th to 75th percentile and whiskers show the minimum and maximum values.

2.3.4 Analysis of the sequencing output from Illumina platform generated by CASAVA 1.8.2 software

Quality metrics from CASAVA 1.8.2 software indicated that 94% (IQR, 93.0 - 94.5) of the sequencing reaction clusters, generated from "all sequencing reactions" (n = 327), passed Illumina sequencing filters. Of the clusters passing Illumina sequencing filters, 82% (IQR, 80.4 - 83.5) had bases with a Q score greater than 30. The median Q score of these bases was 33 (IQR, 32.3 - 33.3). We found a significant difference in the percentage of clusters that passed filtering ($p = 0.0003$) for the various "groups under study" (n=305). We also observed a significant difference in the percentage of clusters for which the bases had a Q score greater than or equal to 30 for the different "groups under study" ($p < 0.0001$). A significant difference was also observed between the "groups under study" based on their mean Q scores ($p < 0.0001$) (Figure 13). Meconium specimens sampled at birth yielded the lowest Q scores with nine specimens producing Q scores of less than 30 (Figure 13). Only the two non-template controls had significantly lower Q scores when compared to the "groups under study" ($p = 0.018$).

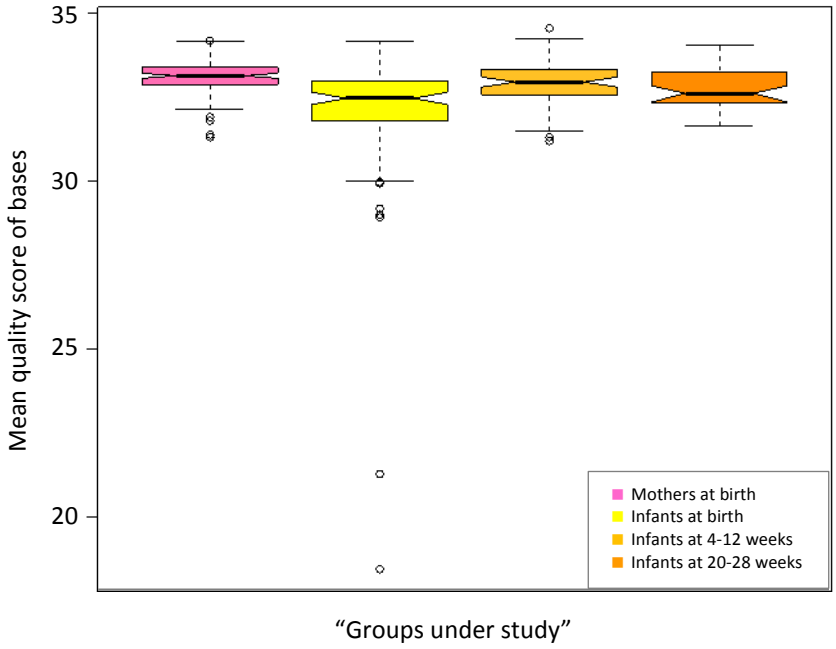


Figure 13. Quality scores of bases sequenced from faecal specimens for all "groups under study"
Quality scores are shown for faecal specimens collected from mothers at birth (n=90), meconium specimens collected from infants at birth (n=107), faecal specimens collected from infants at 4 to 12 weeks of age (n=72) and 20 to 28 weeks of age (n=36). The line within the box plot indicates the median values. The median values of box plots are significantly different (at 5% significance level) in the event that their notches do not overlap. The box extends from the 25th to 75th percentile and whiskers show the minimum and maximum values.

The total number of reads sequenced from "all sequencing reactions" (n = 327) was 17 722 962, with a median number of 43 098 (IQR, 27 706 – 72 353) reads per sequencing reaction. The total number of reads sequenced from all "groups under study" (n = 305) was 16 596 332 with a

median number of 43 198 (IQR, 27 764 – 73 256) reads per faecal specimen. We found a significant difference ($p < 0.0001$) in the median number of reads sequenced from each of the "groups under study" (Figure 14A). Infant meconium specimens sampled at birth produced the highest number of reads obtained from the sequencer (Figure 14A). The "sequencing controls" (n = 22) yielded a total of 1 126 630 reads. Of these, the "technical repeats" (n = 17) produced 950,706 reads. The two BEI controls yielded a total of 57 720 reads and the *E. coli* control yielded a total of 74 212. Of note, the two non-template controls produced a total of 43 992 reads.

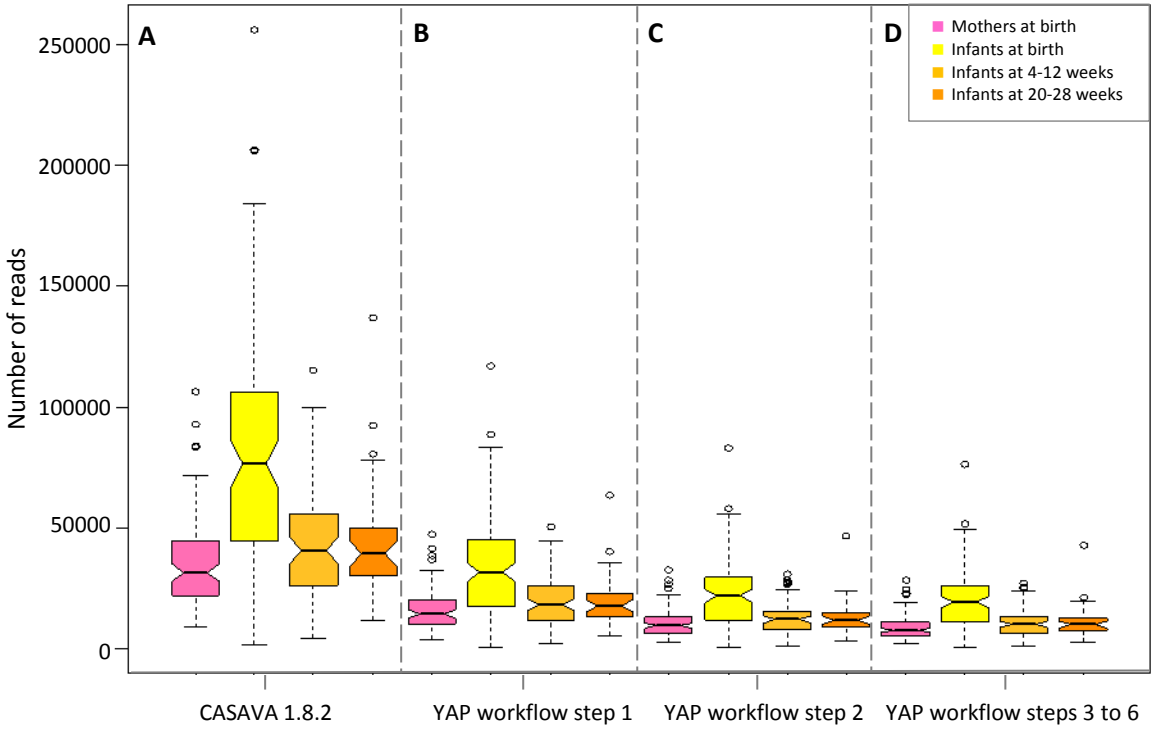


Figure 14. Reads obtained from CASAVA 1.8.2 software and from the YAP workflow for all "groups under study"
The number of reads are summarised for faecal specimens collected from mothers at birth (n=90), meconium specimens collected from infants at birth (n=107), faecal specimens collected from infants at 4 to 12 weeks of age (n=72) and 20 to 28 weeks of age (n=36). The line within the box plot indicates the median values. The median values of box plots are significantly different (at 5% significance level) in the event that their notches do not overlap. The box extends from the 25th to 75th percentile and whiskers show the minimum and maximum values. **A)** Reads obtained from CASAVA 1.8.2; **B)** Reads obtained following the first step of the YAP workflow; **C)** Reads obtained following the second step of the YAP workflow; and **D)** Reads obtained following steps 3 to 6 from the YAP workflow.

2.3.5 Sequencing output from the YAP workflow

Compared to the sequencing output from CASAVA 1.8.2 software (Figure 14A), we observed a significant reduction in the number of reads sequenced ($p < 0.0001$) as a result of the steps performed by the YAP workflow (Figure 14D). A significant reduction in the number of reads ($p < 0.0001$) was already evident following the first step of the YAP workflow (Figure 14B) which corresponded to trimming of paired end reads. Another significant reduction in the number of reads ($p < 0.0001$) was noted following the second step of the YAP workflow (Figure 14C), corresponding to trimming of overlapped reads. Removal of chimeras; reads not aligning to the V4 region of the

16S rRNA gene reference; and overly long reads, resulted in another significant drop in the number of reads (p = 0.0060)(Figure 14D). The total number of OTUs classified by the YAP workflow for all sequencing reactions (n = 327) was 121 570, with a median number of 268 (IQR, 161 - 477) OTUs per reaction. The OTUs classified for all "groups under study" (n = 305) was 115 536, with a median number of 276 (IQR, 164 - 494) OTUs per faecal specimen.

2.3.6 Determining whether the sequencing output from the YAP workflow was sufficient for detecting all bacterial genera present from the "groups under study"

The median number of reads observed for all "groups under study" (n = 305), as a result of the YAP workflow, was 10 730 (IQR, 6 660 - 18 068) (Figure 15). Infant meconium specimens sampled at birth are most concave-downward indicating that the majority of genera have been detected with the reads available (Figure 15). We also see a similar pattern for infants at 4-12 weeks and 20-28 weeks, however not as prominent as the meconium specimens (Figure 15). Maternal faecal specimens on the other hand have a more linear shape than any of the other specimens, indicating that more reads are required to more comprehensively identify all of the genera from these specimens (Figure 15).

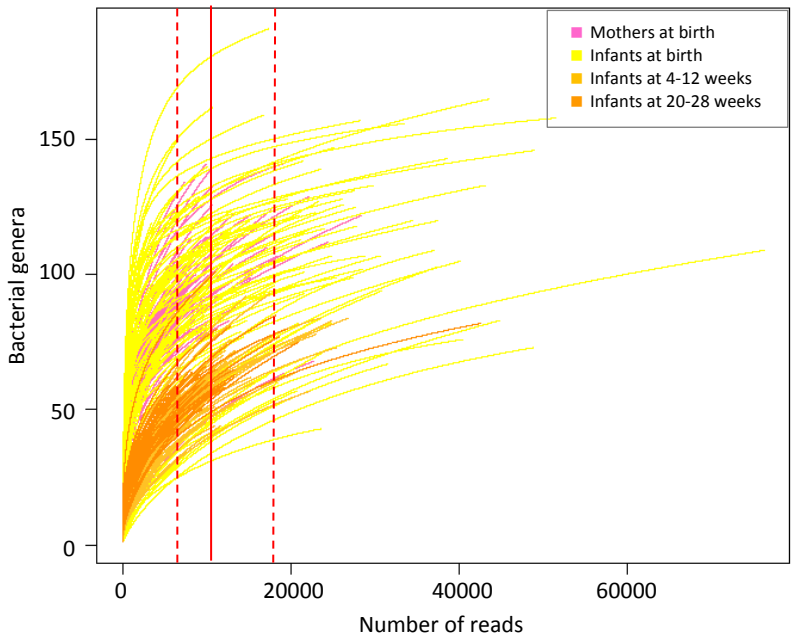


Figure 15. Accumulation curves comparing the number of reads from the YAP workflow to the number of bacterial genera identified
The median numbers of reads from all "groups under study" are shown by the red solid line with the inter-quartile ranges in red broken lines.

2.3.7 Potential contamination observed from non-template controls

Following the YAP workflow, 67 and 77 OTUs were classified respectively from the two non-template controls at genus-level. Together, these OTUs represented 69 genera from the two non-template controls, herein referred to as “contaminating genera”. The majority of these bacterial genera (64%) are in fact common colonizers of the human GIT, while 26% have also been identified from laboratory reagents such as DNA extraction kits and molecular biology reagents (Table 3).

Table 3. Potential "contaminating genera" from our study assessed according to the genera frequently detected from faecal specimens and laboratory reagents

Phylum	Class	Order	Family	Genus	Number of OTUs (n=144)	Faecal specimens from humans	Laboratory reagents	Comments
Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium	7	[54]	●[55]	Genus Corynebacterium is a typical skin colonizer, however within the order Actinomycetales it is one of the most frequently detected genera from the GIT.[56]
			Micrococcaceae	Arthrobacter	2	[57]	●[55]	
				Kocuria	1		●[55]	Genus Kocuria is a typical skin and oropharynx colonizer, however it has recently been reported in GIT samples.[56]
				Renibacterium	1			
				Unclassified	1			
		Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	6	[57–62]		Majority of species belonging to the genus Bifidobacterium have been recovered exclusively from GIT specimens.[56]
		Coriobacteriales	Coriobacteriaceae	Collinsella	1	[63]		The order Coriobacteriales is frequently detected from the GIT with Collinsella being the most dominant genus.[56]
Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas	3	[64,65]		Majority of GIT Bacteroidetes spp. belong to the families Porphyromonodaceae, Prevotellaceae, Rikenellacea and Bacteroidaceae.[56]
				Unclassified	1			
			Prevotellaceae	Prevotella	1	[58,61,62,66]		
			Rikenellaceae	Alistipes	1	[61,62,65]		
			Bacteroidaceae	Bacteroides	4	[54,58–62,65–67]		
	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	Sediminibacterium	1			Typically associated with ecosystems such as soil, however species belonging to the class Sphingobacteriales have occasionally been detected from GIT.[56]
Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Brevundimonas	3		●[55]	The class Alphaproteobacteria is one the five GIT inhabiting classes of the Proteobacteria.[56]
				Phenylobacterium	3			
		Rhizobiales	Bradyrhizobiaceae	Bradyrhizobium	2	[54]	●[55] ♦[68–70]	
				Unclassified	1			
			Methylobacteriaceae	Methylobacterium	3	[54]	●[55] ♦[68]	
			Rhizobiaceae	Rhizobium	1	[54]	●[55] ♦[68]	
		Rhodobacterales	Rhodobacteraceae	Unclassified	1			
		Rhodospirillales	Rhodospirillaceae	Rhodocista	1			
		Sphingomonadales	Sphingomonadaceae	Sphingomonas	1	[54,71]	●[55] ♦[68,69]	
		Unclassified	Unclassified	Unclassified	1			

Table 3. Potential "contaminating genera" from our study assessed according to the genera frequently detected from faecal specimens and laboratory reagents (continued)

Phylum	Class	Order	Family	Genus	Number of OTUs (n=144)	Faecal specimens from humans	Laboratory reagents	Comments
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Ralstonia	2	[54]	●[55] ♦[68,69] ■[72]	Bacterial members belonging to the class Betaproteobacteria are common and diverse GIT colonizers.[56]
			Comamonadaceae	Comamonas	3			
				Pelomonas	2	[54]	●[55] ♦[68]	
				Unclassified	1			
			Oxalobacteraceae	Massilia	1	[73]	●[55]	
				Undibacterium	1		●[55]	
				Unclassified	1			
			Sutterellaceae	Sutterella	1	[65,74]		
		Neisseriales	Neisseriaceae	Neisseria	2	[75]		
	Deltaproteobacteria	Myxococcales	Nannocystaceae	Nannocystis	1			
	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Escherichia/Shigella	2	[58]	●[55] ■[76]	The family Enterobacteriaceae is the most prevalent and diverse of all nine families of Proteobacteria detected from the GIT.[56]
				Unclassified	2			
		Aeromonadales	Aeromonadaceae	Aeromonas	1	[77]		Species of the genus Aeromonas are associated with gastroenteritis and diarrhea.[56]
			Succinivibrionaceae	Succinivibrio	1			Succinivibrio is one of three genera from the family Succinivibrionaceae detected in the GIT.[56]
		Alteromonadales	Shewanellaceae	Shewanella	1	[54]		
		Pasteurellales	Pasteurellaceae	Haemophilus	2	[58]		Members from the genus Haemophilus are elevated in irritable bowel syndrome.[56]
		Pseudomonadales	Moraxellaceae	Acinetobacter	1	[54,58,59]	●[55] ■[76]	Frequent detection of members from the family Moraxellaceae from the GIT. Acinetobacter species are commonly detected from infants.[56]
				Moraxella	6	[78]		
			Pseudomonadaceae	Pseudomonas	1	[54,58]	●[55] ♦[68,69] ■[72]	Genus Pseudomonas constitute of eight species which have been detected in the GIT.[56]
		Xanthomonadales	Xanthomonadaceae	Stenotrophomonas	2	[54]	●[55] ♦[68,69] ■[72,76]	Genus Stenotrophomonas is one of two genera isolated from the family Xantohomonadaceae in the GIT.[56]
				Xanthomonas	1	[79]	●[55] ■[72]	
		Unclassified	Unclassified	Unclassified	1			
	Unclassified	Unclassified	Unclassified	Unclassified	1			

Table 3. Potential "contaminating genera" from our study assessed according to the genera frequently detected from faecal specimens and laboratory reagents (continued)

Phylum	Class	Order	Family	Genus	Number of OTUs (n=144)	Faecal specimens from humans	Laboratory reagents	Comments
Tenericutes	Mollicutes	Mycoplasmatales	Mycoplasmataceae	Mycoplasma	2			The phylum Tenericutes has not yet been cultured, however molecular assays have detected these bacteria from the GIT.[56]
Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus	2	[54,58]		Genus Staphylococcus is a typical skin colonizer, however also detected from the GIT. Early colonizers of the GIT.[56]
		Lactobacillales	Carnobacteriaceae	Granulicatella	1	[80]		
			Enterococcaceae	Enterococcus	1	[58,60,61,81]		Genus Enterococcus is dominant in the upper section of the small intestine. They can be detected from the GIT as early as the first day of life.[56]
			Lactobacillaceae	Lactobacillus	4	[54,58,59,61,82]		Genus Lactobacilli are dominant in the upper GIT. They are ever-increasing in diversity as new species are still being reported from the GIT.[56]
			Leuconostocaceae	Weissella	1	[58]		Genus Weissella is highly abundant in the GIT and also colonize the GIT of some newborns.[56]
			Streptococcaceae	Streptococcus	8	[54,58,59,61,83]	●[55]	Genus Streptococcus is dominant in the upper section of the small intestine. They can be detected from the GIT as early as the first day of life.[56]
	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto	4	[84]		Thirty of the 72 Clostridium species detected from the GIT belong to the genus Clostridium sensu stricto [56].
			Lachnospiraceae	Anaerostipes	2	[62,85]		The family Lachnospiraceae is most diverse and abundant, with 24 genera of which the majority can be detected from the GIT. Members from this family are also amongst the first colonizers of the GIT.[56]
				Blautia	3	[54,60,67,86,87]		
				Clostridium XlVa	1	[84]		
				Dorea	1	[62,88]		
				Lachnospiracea incertae sedis	3	[89,90]		
				Roseburia	3	[62,65,85]		
				Uncassified	4			
			Peptococcaceae	Peptococcus	1	[84,91]		
			Peptostreptococcaceae	Clostridium XI	2	[92]		
			Ruminococcaceae	Clostridium IV	1	[56,61,84,92]		The family Ruminococcaceae is another important group of GIT inhabiting bacteria.[56]
				Faecalibacterium	5	[62,65,67,93]		
				Ruminococcus	5	[62,87]		
				Unclassified	3			
	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae	Erysipelotrichaceae incertae sedis	2			
	Unclassified	Unclassified	Unclassified	Unclassified	1			

● DNA extraction kits; ♦ Ultrapure industrial water systems; ■ Molecular biology reagents

Figure 16A indicates that the 69 “contaminating genera” from the two non-template controls (summarised in Table 3) corresponded to genera found at high relative abundances (between 60 and 80%) within the faecal specimens from our study (n = 305). All “contaminating genera”, except for the genera Nannocystis, Renibacterium and an unclassified genus from the family Bradyrhizobium, were identified from faecal specimens processed in all four PCR plates during library preparation. Overall, we observed a relatively uniform distribution of the “contaminating genera” at phylum-level amongst faecal specimens across all four of the PCR plates. DNA from faecal specimens amplified in PCR plates no. 2 and 3 had a higher relative abundance of the phylum Firmicutes, whereas DNA from faecal specimens amplified in plates no. 3 and 4 had a higher relative abundance of the phylum Proteobacteria. PCR plates no. 1 and 2 contained 68% (62/91) and 64% (61/95) of infant faecal specimens, while PCR plates no. 3 and 4 contained 79% (67/85) and 73% (38/52) infant faecal specimens.

Based on the actual number of OTUs classified from the two non-template controls, our findings indicated that the level of “contaminating genera” was very low in the non-template controls in comparison to the level of “contaminating genera” classified from faecal specimens (Figure 16B).

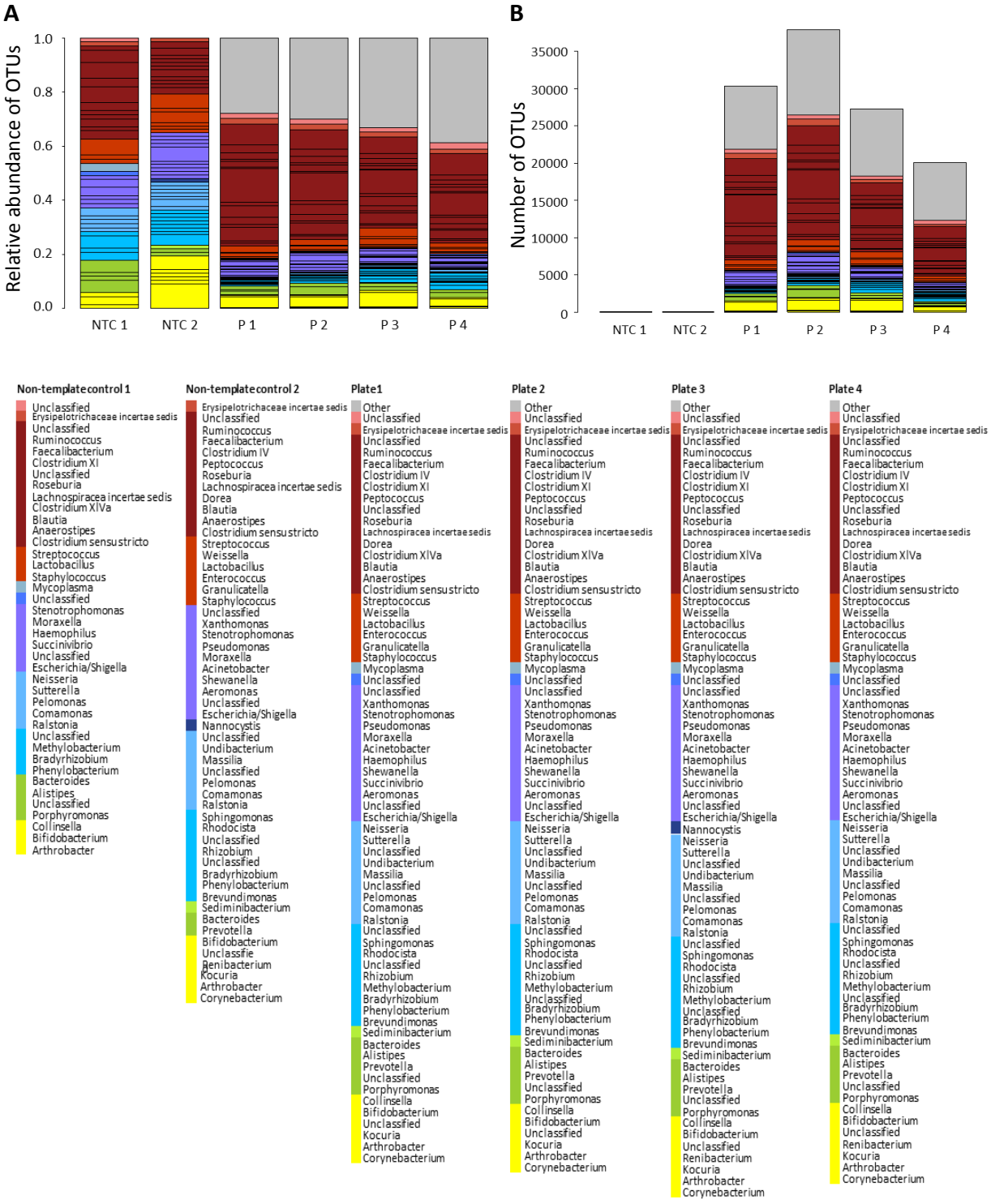


Figure 16. "Contaminating genera" observed from two non-template controls and faecal specimens studied
A) The relative abundance and **B)** number of OTUs representing "contaminating genera" from two non-template controls (NTC) and faecal specimens from the four PCR plates (P1 - P4) used during library preparation. Shades of red represent genera from the phylum Firmicutes, shades of blue from the phylum Proteobacteria, shades of green from the phylum Bacteroidetes and shades of yellow represent genera from the phylum Actinobacteria. "Other" represents genera not observed from the two non-template controls.

Following the process of correction for the contaminant sequences, a significant reduction (p = 0.011) in the total number of OTUs was evident with approximately 10% of the OTUs classified by the YAP workflow being removed (Figure 17). This resulted in a median number of 234 (IQR, 129 - 447) OTUs available per faecal specimen for consecutive data analyses.

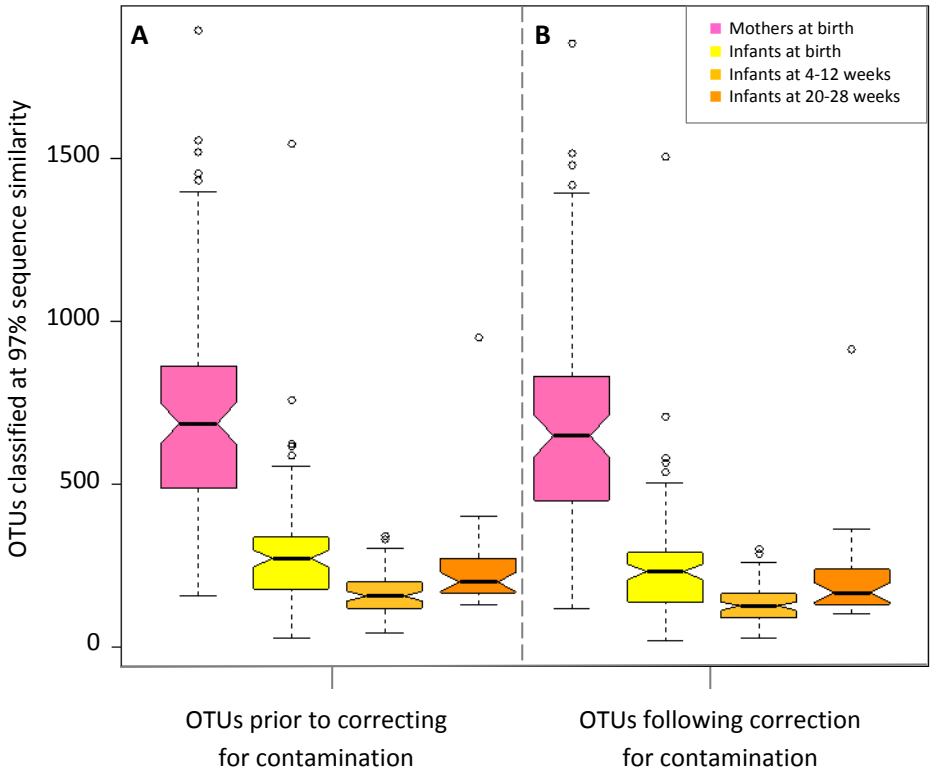


Figure 17. OTUs classified by the YAP workflow prior to and following correction for contamination
The number of OTUs are summarised for faecal specimens collected from mothers at birth (n=90), meconium specimens collected from infants at birth (n=107), faecal specimens collected from infants at 4 to 12 weeks of age (n=72) and 20 to 28 weeks of age (n=36).The line within the box plot indicates the median values. The line within each box plot represents the median value. The median values of box plots are significantly different (at 5% significance level) in the event that their notches do not overlap. The box extends from the 25th to 75th percentile and whiskers show the minimum and maximum values. **A)** OTUs classified by the YAP workflow; and **B)** OTUs resulting from correcting for contamination.

2.3.8 Variation between "technical repeats"

The experimental reproducibility was determined at R^2 = 0.85 based on the 17 "technical repeats". When assessing the variation between "technical repeats" for each of the "groups under study", we found that "technical repeats" for infant meconium specimens collected at birth were the least reproducible (R^2 = 0.61); followed by infant faecal specimens collected at 4-12 weeks of age (R^2 = 0.83). Faecal specimens from infants collected at 20-28 weeks of age showed higher reproducibility between "technical repeats" (R^2 = 0.92), while maternal faecal specimens collected at birth showed the highest reproducibility (R^2 = 0.98). Interestingly, we also observed an increase in the reproducibility of "technical repeats" as the DNA concentration, used during library preparation, increased (Figure 18).

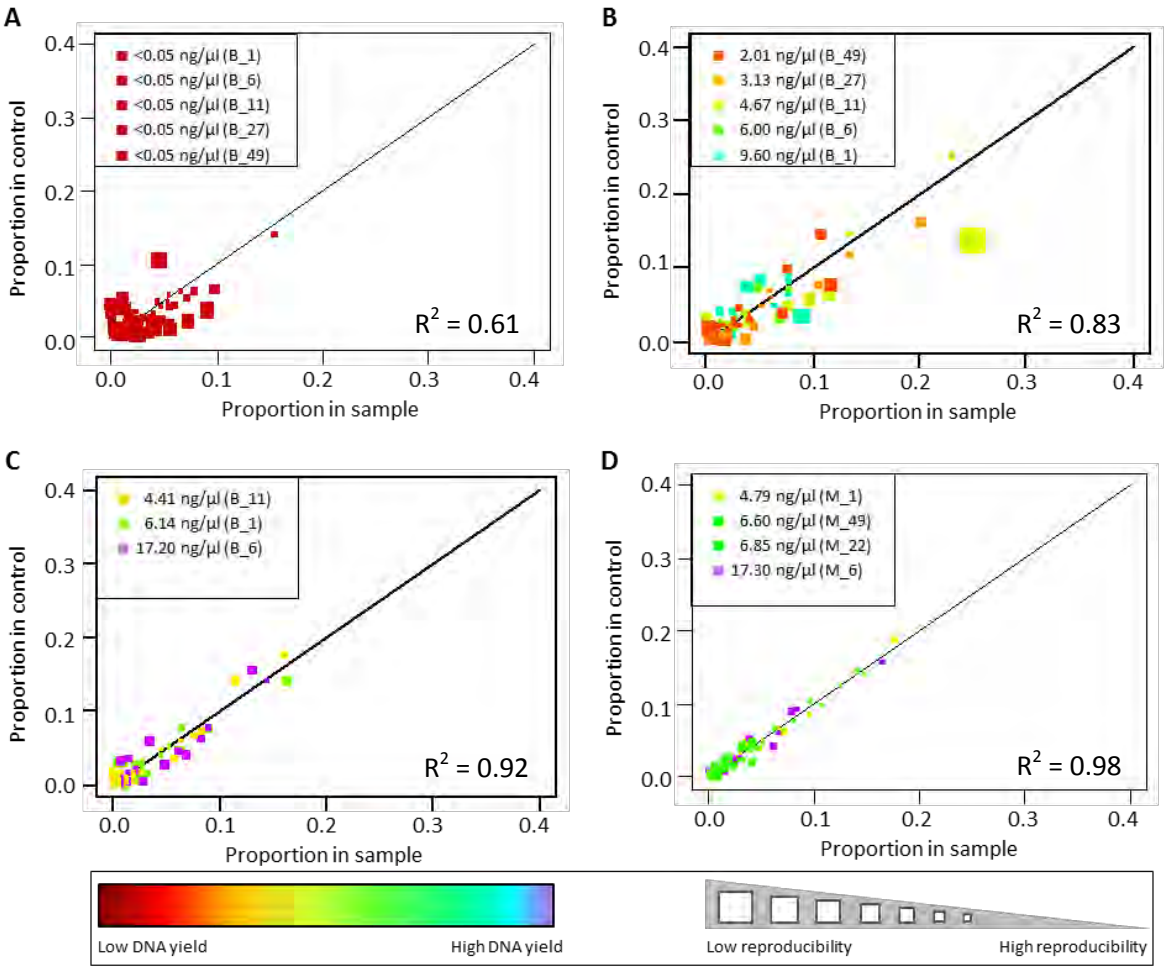


Figure 18. Reproducibility of "technical repeats"
Scatter plots show the proportion of each OTU detected from the original specimens (first measurement) and their "technical repeats" (second measurement). The bigger the distance between the plot and the diagonal line, the bigger the variation. This increase in variation is also shown by the increase in size of the scatter plots. In addition, the DNA concentrations obtained from each of the specimens are indicated using a range of colours. Colours range from red, representing low DNA concentrations, to purple, representing high DNA concentrations. The mother-infant pair ID is shown in brackets for each of the specimens and their repeats. **A)** Meconium specimens collected from infants at birth; **B)** Faecal specimens collected from infants between 4 and 12 weeks of age; **C)** Faecal specimens collected from infants between 20 and 28 weeks of age; and **D)** Faecal specimens collected from mothers at birth.

2.3.9 Intra-individual bacterial diversity of faecal specimens under study

The median Shannon diversity index for all "groups under study" (n = 305) was 3.4 (IQR, 3.2 - 3.8). The lowest diversity index measured for a faecal specimen was 2.3 and the highest was 4.7. There was a significant difference ($p < 0.0001$) between the various "groups under study"; with infants at birth having the highest diversity measures (Figure 19).

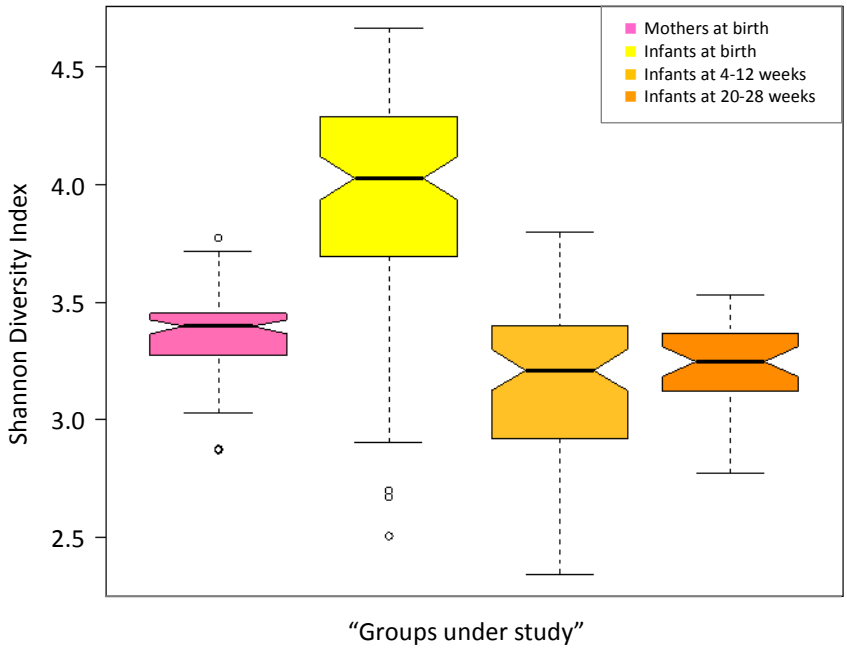


Figure 19. Intra-individual bacterial diversity indices for "groups under study"
The Shannon diversity indices are summarised for faecal specimens collected from mothers at birth (n=90), meconium specimens collected from infants at birth (n=107), faecal specimens collected from infants at 4 to 12 weeks of age (n=72) and 20 to 28 weeks of age (n=36). The line within each box plot represents the median value. The median values of box plots are significantly different (at 5% significance level) in the event that their notches do not overlap. The box extends from the 25th to 75th percentile and whiskers show the minimum and maximum values.

2.3.10 Correlations between DNA concentrations, reads sequenced, OTUs classified and intra-individual bacterial diversity

Overall, no strong correlations were seen between DNA concentrations, the number of reads sequenced, OTUs classified and intra-individual bacterial diversity indices for all "groups under study" (n = 305) (Table 4). Figure 20 visually highlights these correlations for the "groups under study". We observed a very strong positive correlation between the number of reads sequenced and the number of OTUs classified for maternal faecal specimens. In addition, infants sampled at 4-12 weeks also showed a strong positive correlation between the number of reads sequenced and the number of OTUs classified. Infants sampled at birth and 20-28 weeks of life showed a weak positive correlation between the number of reads sequenced and the number of OTUs classified. Another strong positive correlation was observed between the number of reads sequenced and the intra-individual bacterial diversity measured from infant faecal specimens collected at 4-12 weeks of age. This positive correlation (moderate to strong) was found in all other sampling groups, except for maternal faecal specimens. Finally, we also observed a very strong positive correlation between the number of OTUs classified and the bacterial diversity within infant faecal specimens collected at 4-12 weeks of age. A strong positive correlation was also observed between the number of OTUs classified and the bacterial diversity for infants sampled at birth.

Table 4. Correlations between DNA concentration, reads sequenced, OTUs classified and intra-individual bacterial diversity

Correlations		"Groups under study" (n = 305)	Mothers at birth (n = 90)	Infants at birth (n = 107)	Infants at 4-12 weeks (n = 72)	Infants at 20-28 weeks (n = 36)
DNA concentration	Reads sequenced	-0.28	-0.09	0.09	-0.19	-0.18
DNA concentration	OTUs classified	0.29	0.09	0.35	-0.04	0.61
DNA concentration	Diversity	-0.42	-0.36	-0.21	-0.20	-0.11
Reads sequenced	OTUs classified	0.05	**0.71**	0.27	0.69	0.23
Reads sequenced	Diversity	0.55	-0.14	0.41	0.69	0.38
OTUs classified	Diversity	0.02	-0.37	0.40	**0.78**	0.11

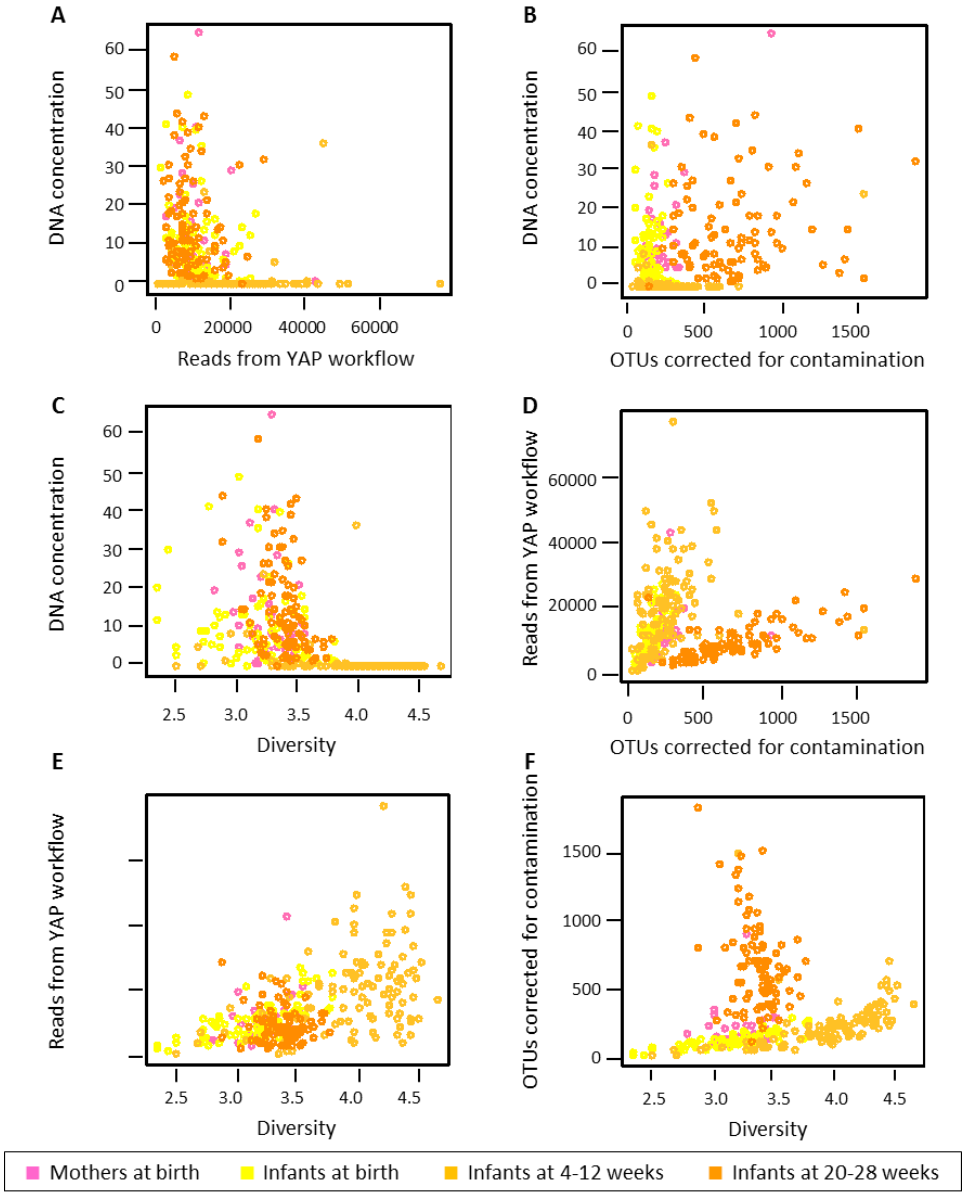


Figure 20. Correlations between DNA concentrations, reads, OTUs and intra-individual bacterial diversity
Correlations are summarised for faecal specimens collected from mothers at birth (n=90), meconium specimens collected from infants at birth (n=107), faecal specimens collected from infants at 4 to 12 weeks of age (n=72) and 20 to 28 weeks of age (n=36). Each of the sampling groups is indicated by a specific colour shown in the legend. Correlations between **A)** DNA concentrations and reads from the YAP workflow; **B)** DNA concentrations and OTUs corrected for contamination; **C)** DNA concentration and intra-individual diversity; **D)** reads from the YAP workflow and OTUs corrected for contamination; **E)** reads from the YAP workflow and intra-individual diversity; **F)** OTUs corrected for contamination and intra-individual diversity.

2.4 Discussion

High-throughput sequencing of the 16S rRNA gene, and in particular Illumina sequencing technology, [1–3] has been successfully used to sequence complex bacterial profiles from the human GIT. However, computational analysis of reads generated by high-throughput sequencing may be affected by a number of factors, such as the quality of the library prepared; the sequencing chemistry; as well as the bioinformatics workflow used to eliminate adapter sequences, chimeras, short reads and any sequencing artifacts.[16] This discussion primarily focuses on i) describing the sequencing depth obtained during from our faecal specimens in relation to previous reports; ii) adressing the effect of quality control steps performed in addition to analysis conducted by CASAVA 1.8.2 software; iii) identifying the possible source for potential contamination observed in our sequencing data and discussing the importance of correcting for contaminants; iv) and to highlight the importance of the use of optimal DNA yields as starting material during library preparation.

One of the important elements of studying highly diverse microbiomes (such as the GIT microbiome) using high-throughput sequencing, is to sequence deep enough to obtain a 'true' representation of the complex microbiomes.[89] Previously, studies targeting the V4, V6, and V2-V3 regions, using Illumina MiSeq technology, concluded that a sequencing depth of about 40 000 reads per specimen is sufficient for studying the GIT microbiome.[94,95] Our study successfully sequenced a median of 43 198 reads per faecal specimen, with the majority of reads ranging between approximately 28 000 to 73 000.

In addition to the sequencing depth; the quality of the sequening data is as important. High quality sequencing data is crucial for various downstream analyses since errors in base calling, poor quality reads, adaptor/primer contamination, and sequencing artifacts such as chimeras may bias results.[10] Overall, we generated high quality reads as indicated by more than 90% of the sequencing clusters passing sequencing filters. In addition, more than 80% of our bases sequenced had Q scores (or Phred scores) greater than 30, which are regarded as high quality scores according to Illumina Miseq technical team.[35] In support of this; the Illumina Miseq performance specificiations indicate that > 70% of bases should have a Q score greater than 30 when using the 2 x 300 bp Miseq Reagent Kit V3, as per our study.[96] Of note, however; the above-mentioned quality scores were generated using the Illumina sequencing software CASAVA 1.8.2 and it is known that machine-generated Q scores may over-estimate the accuracy of base calls.[97] Therefore, in addition to Illumina's sequencing software analysis; we included additional quality assessement steps by performing the YAP workflow. The YAP workflow consisted of a set of tools integrated into a computationally efficient workflow using a Python wrapper script to i) accurately filter erroneous reads by

overlapping paired-end reads;[97] to ii) remove sequencing artifacts such as chimeras;[10] and to iii) remove all reads not corresponding to the 16S rRNA reference gene which included overly long reads. Performing this quality control workflow, in addition to obtaining machine-assigned quality scores, was clearly of great importance since 25% of our reads sequenced were removed following the YAP workflow. These types of reductions in the number of reads sequenced are expected when utilising additional tools to filter high-quality reads produced by high-throughput platforms such as Illumina.[10] The reduced number of reads available for downstream data analysis led us to construct an accumulation curve to determine whether we still had a sufficient number of reads to accurately assess bacterial diversity present within the faecal specimens under study. Characteristically, accumulation curves initially increase steeply, whereafter they gradually level off as an increase in number of reads no longer contribute to new species being detected.[11] Although we did not reach the recommended sequencing depth of 40 000 reads per faecal specimen[94,95] (after applying the YAP workflow to our sequencing output); we did observe a plateau in the accumulation curves for a number of our specimens under study. To the best of our knowledge, no studies have reported a recommended sequencing depth for meconium specimens; however we observed a plateau at approximately 20 000 reads sequenced per meconium specimen. Meconium specimens also plataued most prominently in comparison to the faecal specimens collected at later stages in life as well as the maternal faecal specimens. These findings indicate that less reads may be required to obtain a comprehensive insight into the genera colonizing meconium specimens when compared to faecal specimens sampled at later stages in life. Interestingly, the study by Mshvildadze and colleagues only had an average of 888 sequences per meconium specimen (and only 18-46 OTUs classified at 97% sequence similarity) available to characterise bacteria from meconium specimens of healthy premature infants and infants suffering from necrotizing enterocolitis.[98]

Another potential artifact of high-throughput sequencing is the introduction of a considerable number of contaminating sources which may result in erroneous conclusions.[99] Several potential sources have been identified to contribute to sequence-based contamination from the time of specimen collection up until the final sequencing products are generated.[100] Some of these include, for example, specimen collection, DNA extraction, PCR amplification steps and the sequencing process itself. In the event where a marker gene is targeted, as performed by our study, reagent contaminants may also be amplified during the library preparation step.[100] Contamination was clearly evident from our sequencing run as we classified 67 and 77 OTUs at genus-level from our two non-template controls. However; determining the source of contamination for our sequencing data is not a straight forward process. Although studies have reported DNA extraction kits as possible sources for contamination,[55] we do not suspect that this was the case for our study. One of

the reasons for this being that previous studies have reported the QIAsymphony nucleic acid extraction platform to be free of cross-contamination.[101,102] Furthermore, we processed the contaminated non-template controls for the first time during library preparation in a different laboratory as to where the nucleic acid extractions took place. In addition to the potential role of DNA extraction kits in contamination; studies have also reported laboratory reagents such as PCR reagents and ultrapure water as potential sources for contamination.[55] Despite the fact that we did not have a baseline established to determine the effect of reagents used during library preparation as sources for contamination; we did, however, observe that approximately 10% of the "contaminating genera" in our run have previously been reported from laboratory reagents (Table 3). We suspect that the most influential factor for contamination in our study was the process of PCR amplification.[20,76,103] Since all "sequencing reactions" in our study underwent repeated amplification of the same target sequence, and were possibly exposed to aerosolized amplicons with high copy numbers of the target genes;[104] we suspect that contamination most likely resulted from cross-contamination during library preparation.[21] It has been reported that high copy numbers of the target sequences (evident for faecal specimens)[105–108] may result in cross-contamination during amplicon generation.[21] Moreover, studies have shown that the use of low template concentrations during amplification (evident for a number of our specimens) may result in contaminants being amplified instead of the desired template.[109,16,110,22,111] We also observed DNA from our two non-template controls after the process of amplification (as indicated in Section 2.3.3) which supports our suspicion of possible cross-contamination during library preparation. Furthermore, we also found that the majority of genera contaminating our non-template controls were in actual fact genera commonly found within the human GIT (Table 3). Another potential source for contamination is the use of contaminated laboratory equipment. This has been previously reported, for example, as carryover between consequtive MiSeq runs.[112] In our study, however, it is highly unlikely for the latter to have occurred since we observed amplification of the 16S rDNA fragment in the two non-template controls prior to sequencing.

Salter and colleagues[55] have provided a number of precautionary steps to minimize the impact of contaminants in high-througput sequencings studies. Some of these include minimising the risk of sample collection contamination; inclusion of technical controls (which may include storage, extraction kit, or PCR reagents) during each step of the laboratory workflow; treatment of reagents used for DNA extraction; using increased amounts of DNA yields as starting material during library preparation; treatment of reagents used for the process of PCR; quantification of negative controls and specimens of interest throughout the entire laboratory workflow in order to assess when contamination as it arises; and computationally controlling for negative controls when

analysing sequencing data. Since we were not in the position to repeat any experimental procedures, we decided to correct for the contaminants observed in our sequencing experiment by computationally removing OTUs representing "contaminating genera" from our faecal specimens. Following correction for contamination we noticed a 10% reduction in the number of OTUs classified from our faecal specimens. However, we do believe that it would be unlikely that this reduction would have any significant impact on downstream data analysis. The reason for this being that the OTUs present in our non-template controls were very low in numbers compared to their numbers in faecal specimens. In addition, the fact that these "contaminating genera" were highly abundant in our faecal specimens shows that the contaminants in our non-template controls represented genera naturally inhabiting the faecal specimens under study (supported by findings from the literature summarised in Table 3). In reality, exogenous microbial DNA representing contamination is to some extent a problem for every microbiome dataset;[100] however, the steps used to control or remove these contaminants during sequencing or computational processes is what adds to the credibility of high-throughput sequencing results.

Also of concern when conducting 16S rRNA-based high-throughput sequencing studies is low template concentrations which may not only be highly susceptible to external contaminants,[21] but may also result in less reproducible results.[14] This is of great concern when performing diversity measurments.[15] Since the poorest sequencing reproducibilities observed in our study were associated with meconium specimens which had the lowest DNA yields used during library preparation; our findings are in support of those from Kennedy et al. (2014) reporting a relationship between low DNA concentrations and low reproducibility.[14] Interestingly, not only did the infant meconium specimens included in this study associate with less reproducible results, but these specimens also produced the lowest amplicon concentrations and quality scores compared to any of the other "groups under study". This clearly highlight that low template concentrations incorporated during library preparation have a number of downstream effects on sequencing data. The low template concentrations extracted from meconium specimens in our study may be explained by the actual nature of the specimen rather than it resulting from the nucleic acid extraction process itself. The reason for this being that the faecal specimens included in this study provided significantly higher concentrations compared to the meconium specimens. In addition; we also used a nucleic acid extraction assay which have been reported to provide high quantities of DNA from faecal specimens.[27] Despite these findings from the meconium specimens under study; we did in actual fact observe an increase in reproducibility as DNA concentrations increased. This again is in support of the findings by Kennedy and colleagues.[14] We are, however, not in the position to compare our reproducibility measures with the literature, since we are not aware of any studies reporting on

reproducibility after conducting Illumina Miseq sequencing on faecal specimens. In addition, high-throughput sequencing data may vary substantially across studies due to factors such as the number of repeats analysed,[15] sampling methods and the type of specimen analysed,[15] as well as a number of machine artifacts, such as sequencing errors.[113]

An interesting finding from this study was that the meconium specimens, from which the lowest DNA concentrations were extracted, also produced the highest number of reads and intra-individual bacterial diversity, but the lowest numbers of OTUs. These dynamics are difficult to explain since we could not find any strong correlations between DNA yields, the number of reads sequenced, intra-individual bacterial diversities, nor the number of OTUs classified when assessing all "groups under study" (n = 305) (Table 4). We did, however, observe a weak positive correlation between the number of reads sequenced and the number of OTUs classified for infants sampled at birth; while maternal faecal specimens showed a very strong positive correlation. These interesting correlations, together with the summary statistics presented in Figures 14 and 17, indicate that infant meconium specimens produced more reads and less OTUs compared to maternal faecal specimens. A possible explanation for this may be that meconium specimens simply had more of the same reads (duplication of reads) sequenced which resulted in less OTUs being classified. These observations, however, does not necessarily suggest that meconium specimens are less diverse due to less OTUs being classified. This may be explained by the definition of the Shannon diversity index, used to calculate the bacterial diversity within a specimen (intra-individual diversity). Since the Shannon diversity index is based on both the number of OTUs (richness) and their distribution (evenness) within a specimen; intra-individual diversity does not necessarily suggest that high numbers of OTUs were classified from a specimen. This was clearly highlighted by the moderate negative correlation found when correlating the high number of OTUs classified from all maternal faecal specimens with their low intra-individual diversity indices. In addition, only faecal specimens sampled from infants at 4-12 weeks of age showed a very strong positive correlation when correlating their numbers of OTUs to their Shannon diversity indices.

Since DNA yields are commonly included when evaluating the DNA extraction methods for microbial community analyses;[27,114–116] we decided to also determine the effect of DNA yields used during library preparation on the microbial diversity obtained from sequencing data. Interestingly, we found only weak negative correlations when determining the effect of DNA yield on bacterial diversity measures. These weak negative correlations, together with previous reports that maximal DNA yields are not a guarantee for maximal microbial diversity,[117,118] emphasises that extraction protocols should be optimized for extracting the best representatives of the microbial community instead of aiming for high DNA yields only.[118]

In contrast to the weak correlations between DNA yields and bacterial diversity indices; our study shows that the lower the DNA yields used during library preparation, the lower the reproducibility of the sequencing run. What's more is that the reproducibility of a sequencing run has been reported to also affect diversity measures.[15] Therefore, taking into account the low DNA yields, the low sequencing reproducibility, and the very high intra-individual diversity indices obtained for our meconium specimens; we are under the impression that diversity measures obtained from meconium specimens may have been skewed. In support of this, are the reports from previous studies indicating that intra-individual diversity is lowest during infancy and increases with age.[88,119] In support of our opinion of a skewed intra-individual diversity for meconium and in agreement with previous studies[88,119] we did observe an increasing trend in intra-individual diversity from our infant faecal specimens sampled from 4-12 to 20-28 weeks of age. Maternal faecal specimens also had higher intra-individual diversities compared to infant faecal specimens sampled at later stages in life. These findings suggest the importance of optimal DNA yields for library preparation, as it may not only affect sequencing reproducibility but may also subsequently influence bacterial diversity estimations.[14]

2.5 Conclusion

We emphasize that microbiome studies, aiming to characterize bacterial communities from faecal specimens, first and foremostly need to ensure that optimal DNA extraction methods are used. We found that low DNA yields not only result in less reproducible results, but may in turn also impact on bacterial diversity measurements. Another critical step in high-throughput sequencing studies is to incorporate computational workflows in order to trim and select for good quality reads. This was clearly highlighted by the significant reduction in the number of reads resulting from quality control measures other than those performed by the sequencing platform's software. Finally, findings from this study and previous reports reinforced the need for including non-template controls during the DNA extraction process, library preparation and sequencing steps. This will allow researchers to effectively trace the source of contamination and to repeat experimental procedures or effectively correct for contamination during data analysis based on the source of contamination.

References

1. Naseribafrouei A, Hestad K, Avershina E, et al. Correlation between the human fecal microbiota and depression. *Neurogastroenterol Motil*. 2014. doi:10.1111/nmo.12378.

2. Dey N, Soergel DAW, Repo S, Brenner SE. Association of gut microbiota with post-operative clinical course in Crohn's disease. *BMC Gastroenterol*. 2013;13(1):131. doi:10.1186/1471-230X-13-131.

3. Hamilton MJ, Weingarden AR, Unno T, Khoruts A, Sadowsky MJ. High-throughput DNA sequence analysis reveals stable engraftment of gut microbiota following transplantation of previously frozen fecal bacteria. *Gut Microbes*. 2013;4(2):125-135. doi:10.4161/gmic.23571.

4. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012;489(7415):220-30. doi:10.1038/nature11550.

5. Umu OCO, Oostindjer M, Pope PB, et al. Potential applications of gut microbiota to control human physiology. *Antonie Van Leeuwenhoek*. 2013;104(5):609-18. doi:10.1007/s10482-013-0008-0.

6. Claesson MJ, O'Toole PW. Evaluating the latest high-throughput molecular techniques for the exploration of microbial gut communities. *Gut Microbes*. 2010;1(4):277-278. doi:10.4161/gmic.1.4.12306.

7. Maccaferri S, Biagi E, Brigidi P. Metagenomics: Key to human gut microbiota. *Dig Dis*. 2011;29:525-30. doi:10.1159/000332966.

8. Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464(7285):59-65. doi:10.1038/nature08821.

9. Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl Environ Microbiol*. 2011;77(11):3846-52. doi:10.1128/AEM.02772-10.

10. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*. 2012;7(2):e30619. doi:10.1371/journal.pone.0030619.

11. Hughes J, Hellmann J, Ricketts T, Bohannan B. Counting the Uncountable: Statistical Approaches to Estimating Microbial Diversity. *Appl Environ Microbiol*. 2001;67(10):4399-4406. doi:10.1128/AEM.67.10.4399.

12. Peng X, Yu K-Q, Deng G-H, et al. Comparison of direct boiling method with commercial kits for extracting fecal microbiome DNA by Illumina sequencing of 16S rRNA tags. *J Microbiol Methods*. 2013;95(3):455-62. doi:10.1016/j.mimet.2013.07.015.

13. Kennedy NA, Walker AW, Berry SH, et al. The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS One*. 2014;9(2):e88982. doi:10.1371/journal.pone.0088982.

14. Kennedy K, Hall MW, Lynch MDJ, Moreno-Hagelsieb G, Neufeld JD. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Appl Environ Microbiol*. 2014;80(18):5717-5722. doi:10.1128/AEM.01451-14.

15. Zhou J, Wu L, Deng Y, et al. Reproducibility and quantitation of amplicon sequencing-based detection. *ISME J*. 2011;5(8):1303-13. doi:10.1038/ismej.2011.11.

16. Kircher M, Heyn P, Kelso J. Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics*. 2011;12(1):382. doi:10.1186/1471-2164-12-382.

17. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol*. 2014;10(4):e1003531. doi:10.1371/journal.pcbi.1003531.

18. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*. 2014;2:15. doi:10.1186/2049-2618-2-15.

19. Rodriguez-Ramos T, Dornelas M, Maranon E, Cermeno P. Conventional sampling methods severely underestimate phytoplankton species richness. *J Plankton Res*. 2014;36(2):334-343. doi:10.1093/plankt/fbt115.

20. Inglis G, Thomas M, Thomas D, Kalmokoff M, Brooks S, Selinger L. Molecular Methods to Measure Intestinal Bacteria : A Review. *J AOAC Int*. 2012;95(1):5-24. doi:10.5740/jaoacint.SGE.

21. Aslanzadeh J. Preventing PCR Amplification Carryover Contamination in a Clinical Laboratory. *Ann Clin Lab Sci*. 2004;34(4):389-396.

22. Pennisi E. Contamination plagues some microbiome studies. *Sci Mag*. 2014;346(6211):801. doi:10.1126/science.346.6211.801.

23. Xia F, Chen J, Fung WK, Li H. A logistic normal multinomial regression model for microbiome compositional data analysis. *Biometrics*. 2013;69(4):1053-63. doi:10.1111/biom.12079.

24. Zar HJ, Barnett W, Myer L, Stein DJ, Nicol MP. Investigating the early-life determinants of illness in Africa: the Drakenstein Child Health Study. *Thorax*. 2014;0:1-3. doi:10.1136/thoraxjnl-2014-206242.

25. English R, Boland/Overberg Regional Office, Worcester. *Boland/Overberg Region. Annual Health Status Report 2007/2008*.; 2009:1-93.

26. Drakenstein Municipality. *Drakenstein Municipality Integrated Development Plan 2007-2012. Paarl, South Africa.*; 2007:1-93.

27. Claassen S, Du Toit E, Kaba M, Moodley C, Zar HJ, Nicol MP. A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples. *J Microbiol Methods*. 2013;94(2):103-10. doi:10.1016/j.mimet.2013.05.008.

28. Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS*. 2011;108(Supplement_1):4516-22. doi:10.1073/pnas.1000080107.

29. Illumina Proprietary. MiSeq ® Reagent Kit v3 Reagent Preparation Guide. 2013;(October 2013):1-14.

30. Illumina Proprietary. Preparing Libraries for Sequencing on the MiSeq ®. 2013;(August 2013):1-14.

31. Illumina Proprietary. MiSeq ® System User Guide. 2014;(January):1-94.

32. Illumina Proprietary. CASAVA v1.8.2 User Guide. 2011:1-181.

33. Szpakowski S. YAP: A Computationally Efficient Workflow for Taxonomic Analyses of Bacterial 16S and Fungal ITS Sequences. 2013.

34. Gordon J, Hannon GJ. FASTX-toolkit. 2010.

35. Illumina Proprietary. Quality Scores for Next-Generation Sequencing. 2011:1-2.

36. Cox MP, Peterson DA, Biggs PJ. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*. 2010;11:485. doi:10.1186/1471-2105-11-485.

37. Magoč T, Salzberg SL. FLASH : Fast Length Adjustment of Short Reads to Improve Genome Assemblies. *Bioinformatics*. 2011;27(21):2957-83. doi:10.1093/bioinformatics/btr507.

38. DeSantis TZ, Hugenholtz P, Keller K, et al. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res*. 2006;34(Web Server issue):W394-9. doi:10.1093/nar/gkl244.

39. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(Database issue):D590-6. doi:10.1093/nar/gks1219.

40. Pruesse E, Quast C, Knittel K, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res*. 2007;35(21):7188-96. doi:10.1093/nar/gkm864.

41. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. 2011;27(16):2194-200. doi:10.1093/bioinformatics/btr381.

42. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75(23):7537-41. doi:10.1128/AEM.01541-09.

43. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 2010;26(5):680-2. doi:10.1093/bioinformatics/btq003.

44. Cole JR, Wang Q, Cardenas E, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res*. 2009;37(Database issue):D141-5. doi:10.1093/nar/gkn879.

45. R Development Core Team. R: A language and environment for statistical computing. 2008.

46. RStudio. RStudio: Integrated development environment for R. 2012.

47. Chongsuvivatwong V. epicalc: Epidemiological calculator. 2012.

48. Chambers J, Cleveland W, Kleiner B, Tukey P. *Graphical methods for data analysis*. Boston, Massachusetts, USA: Wadsworth and Brooks/Cole; 1983:62.

49. Hurlbert SH. The Nonconcept of Species Diversity : A Critique and Alternative Parameters. *Ecology*. 1971;52(4):577-586.

50. Oksanen J, Blanchet F., Kindt R, et al. vegan: Community Ecology Package. 2013.

51. Draper N, Smith H. *Applied Regression Analysis*. Second edi. Wiley: New York; 1981.

52. Shannon CE. A Mathematical Theory of Communication. *Bell Syst Tech J*. 1948;27:379-423.

53. Morgan XC, Huttenhower C. Chapter 12: Human microbiome analysis. *PLoS Comput Biol*. 2012;8(12):e1002808. doi:10.1371/journal.pcbi.1002808.

54. Madan JC, Salari RC, Saxena D, et al. Gut microbial colonisation in premature neonates predicts neonatal sepsis. *Arch Dis childhood Fetal neonatal Ed*. 2012;97(6):F456-62. doi:10.1136/archdischild-2011-301373.

55. Salter SJ, Cox MJ, Turek EM, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12(1):87. doi:10.1186/s12915-014-0087-z.

56. Rajilić-Stojanović M, De Vos WM. The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol Rev*. 2014;38:996-1047. doi:10.1111/1574-6976.12075.

57. Hoyles L, Clear JA, McCartney AL. Use of denaturing gradient gel electrophoresis to detect Actinobacteria associated with the human faecal microbiota. *Anaerobe*. 2013;22:90-6. doi:10.1016/j.anaerobe.2013.06.001.

58. Gosalbes MJ, Llop S, Vallès Y, Moya A, Ballester F, Francino MP. Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clin Exp Allergy*. 2012;43(2):198-211. doi:10.1111/cea.12063.

59. Fan W, Huo G, Li X, et al. Diversity of the intestinal microbiota in different patterns of feeding infants by Illumina high-throughput sequencing. *World J Microbiol Biotechnol*. 2013;29(12):2365-72. doi:10.1007/s11274-013-1404-3.

60. Azad MB, Konya T, Maughan H, et al. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *Can Med Assoc J*. 2013;185(5):385-94. doi:10.1503/cmaj.121189.

61. Ringel-Kulka T, Cheng J, Ringel Y, et al. Intestinal microbiota in healthy U.S. young children and adults--a high throughput microarray analysis. *PLoS One*. 2013;8(5):e64315. doi:10.1371/journal.pone.0064315.

62. Tap J, Mondot S, Levenez F, et al. Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol*. 2009;11(10):2574-84. doi:10.1111/j.1462-2920.2009.01982.x.

63. Padmanabhan R, Dubourg G, Nguyen T-T, et al. Non-contiguous finished genome sequence and description of Collinsella massiliensis sp. nov. *Stand Genomic Sci*. 2014;9(3):1144-58. doi:10.4056/sigs.5399696.

64. Balamurugan R, Janardhan HP, George S, Chittaranjan SP, Ramakrishna BS. Bacterial succession in the colon during childhood and adolescence : molecular studies in a southern Indian village. *Am J Clin Nutr*. 2008;88:1643-1647. doi:10.3945/ajcn.2008.26511.1.

65. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. *Science*. 2009;326(5960):1694-7. doi:10.1126/science.1177486.

66. De Filippo C, Cavalieri D, Di Paola M, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A*. 2010;107(33):14691-6. doi:10.1073/pnas.1005963107.

67. Stearns JC, Lynch MDJ, Senadheera DB, et al. Bacterial biogeography of the human digestive tract. *Sci Rep*. 2011;1:170. doi:10.1038/srep00170.

68. Kéki Z, Grébner K, Bohus V, Márialigeti K, Tóth EM. Application of special oligotrophic media for cultivation of bacterial communities originated from ultrapure water. *Acta Microbiol Immunol Hung*. 2013;60(3):345-57. doi:10.1556/AMicr.60.2013.3.9.

69. Kulakov LA, Mcalister MB, Ogden KL, Larkin MJ, Hanlon JFO. Analysis of Bacteria Contaminating Ultrapure Water in Industrial Systems Analysis of Bacteria Contaminating Ultrapure Water in Industrial Systems. *Appl Environ Microbiol*. 2002;68(4):1548-1555. doi:10.1128/AEM.68.4.1548.

70. McAlister MB, Kulakov LA, O'Hanlon JF, Larkin MJ, Ogden KL. Survival and nutritional requirements of three bacteria isolated from ultrapure water. *J Ind Microbiol Biotechnol*. 2002;29(2):75-82. doi:10.1038/sj.jim.7000273.

71. Stewart CJ, Nelson A, Scribbins D, et al. Bacterial and fungal viability in the preterm gut: NEC and sepsis. *Arch Dis Child Fetal Neonatal Ed*. 2013;98(4):F298-303. doi:10.1136/archdischild-2012-302119.

72. Grahn N, Olofsson M, Ellnebo-Svedlund K, Monstein H-J, Jonasson J. Identification of mixed bacterial DNA contamination in broad-range PCR amplification of 16S rDNA V1 and V3 variable regions by pyrosequencing of cloned amplicons. *FEMS Microbiol Lett*. 2003;219(1):87-91. doi:10.1016/S0378-1097(02)01190-4.

73. Albenberg L, Esipova T V, Judge CP, et al. Correlation Between Intraluminal Oxygen Gradient and Radial Partitioning of Intestinal Microbiota. *Gastroenterology*. 2014;147(5):1055-1063.e8. doi:10.1053/j.gastro.2014.07.020.

74. Sakon H, Nagai F, Morotomi M, Tanaka R. Sutterella parvirubra sp. nov. and Megamonas funiformis sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol*. 2008;58(4):970-5. doi:10.1099/ijs.0.65456-0.

75. Goedert JJ, Hua X, Yu G, Shi J. Diversity and Composition of the Adult Fecal Microbiome Associated with History of Cesarean Birth or Appendectomy: Analysis of the American Gut Project. *EBioMedicine*. 2014:0-5. doi:10.1016/j.ebiom.2014.11.004.

76. Tanner MA, Goebel BM, Dojka MA, Pace NR. Specific Ribosomal DNA Sequences from Diverse Environmental Settings Correlate with Experimental Contaminants. *Appl Environ Microbiol*. 1998;64(8):3110.

77. Mulamattathil SG, Bezuidenhout C, Mbewe M, Ateba CN. Isolation of Environmental Bacteria from Surface and Drinking Water in Mafikeng , South Africa , and Characterization Using Their Antibiotic Resistance Profiles. *J Pathog*. 2014;2014:1-11. doi:10.1155/2014/371208.

78. Segata N, Haake SK, Mannon P, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol*. 2012;13(6):R42. doi:10.1186/gb-2012-13-6-r42.

79. Gupta SS, Mohammed MH, Ghosh TS, Kanungo S, Nair GB, Mande SS. Metagenome of the gut of a malnourished child. *Gut Pathog*. 2011;3(1):7. doi:10.1186/1757-4749-3-7.

80. De Vos P, Garrity G., Jones D, et al. *Bergey's Manual of Systematic Bacteriology. Volume 3: The Firmicutes*. Second. (Whitman W., Parte A., eds.). Springer; 2009.

81. Fisher K, Phillips C. The ecology, epidemiology and virulence of Enterococcus. *Microbiology*. 2009;155(Pt 6):1749-57. doi:10.1099/mic.0.026385-0.

82. Claesson MJ, Van Sinderen D, O'Toole PW. The genus Lactobacillus-a genomic basis for understanding its diversity. *FEMS Microbiol Lett*. 2007;269(1):22-8. doi:10.1111/j.1574-6968.2006.00596.x.

83. Hardie JM, Whiley RA. Oral Streptococci. In: *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community-Release 3.9*.; 2006:76-107.

84. Lopetuso LR, Scaldaferri F, Petito V, Gasbarrini A. Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog*. 2013;5(1):23. doi:10.1186/1757-4749-5-23.

85. Duncan SH, Louis P, Flint HJ. Lactate-Utilizing Bacteria, Isolated from Human Feces, That Produce Butyrate as a Major Fermentation Product. *Appl Environ Microbiol*. 2004;70(10):5810. doi:10.1128/AEM.70.10.5810.

86. Eren AM, Sogin ML, Morrison HG, et al. A single genus in the gut microbiome reflects host preference and specificity. *ISME J*. 2014:1-11. doi:10.1038/ismej.2014.97.

87. Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome. *Nature*. 2013;473(7346):174-180. doi:10.1038/nature09944.Enterotypes.

88. Vallès Y, Artacho A, Pascual-García A, et al. Microbial succession in the gut: directional trends of taxonomic and functional change in a birth cohort of Spanish infants. *PLoS Genet*. 2014;10(6):e1004406. doi:10.1371/journal.pgen.1004406.

89. Eggesbø M, Moen B, Peddada S, et al. Development of gut microbiota in infants not exposed to medical interventions. *APMIS*. 2010;119(1):17-35. doi:10.1111/j.1600-0463.2010.02688.x.

90. Ling Z, Li Z, Liu X, et al. Altered fecal microbiota composition associated with food allergy in infants. *Appl Environ Microbiol*. 2014;80(8):2546-54. doi:10.1128/AEM.00003-14.

91. Rekha R, Rizvi MA, Jaishree P. Designing and validation of genus-specific primers for human gut flora study. *Electron J Biotechnol*. 2006;9(5). doi:10.2225/vol9-issue5-fulltext-2.

92. Rajilić-Stojanović M, Smidt H, De Vos WM. Diversity of the human gastrointestinal tract microbiota revisited. *Environ Microbiol*. 2007;9(9):2125-36. doi:10.1111/j.1462-2920.2007.01369.x.

93. Lopez-Siles M, Khan TM, Duncan SH, Harmsen HJM, Garcia-Gil LJ, Flint HJ. Cultured representatives of two major phylogroups of human colonic Faecalibacterium prausnitzii can utilize pectin, uronic acids, and host-derived substrates for growth. *Appl Environ Microbiol*. 2012;78(2):420-8. doi:10.1128/AEM.06858-11.

94. Claesson MJ, Wang Q, O'Sullivan O, et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res*. 2010;38(22):e200. doi:10.4161/gmic.1.4.12306.

95. Tseng HE, Hullar MAJ, Li F, et al. A microbial profiling method for the human microbiota using high-throughput sequencing. *Metagenomics*. 2013;2(1):1-15.

96. Illumina Proprietary. MiSeq Specifications. Available at: http://www.illumina.com/systems/miseq/performance_specifications.html.

97. Eren AM, Vineis JH, Morrison HG, Sogin ML. A filtering method to generate high quality short reads using illumina paired-end technology. *PLoS One*. 2013;8(6):e66643. doi:10.1371/journal.pone.0066643.

98. Mshvildadze M, Neu J, Shuster J, Theriaque D, Li N, Mai V. Intestinal microbial ecology in premature infants assessed with non-culture-based techniques. *J Pediatr*. 2010;156(1):20-5. doi:10.1016/j.jpeds.2009.06.063.

99. Yong E. Contaminomics : Why Some Microbiome Studies May Be Wrong. *Natl Geogr*. 2014:1-9.

100. Weiss S, Amir A, Hyde ER, Metcalf JL, Song SJ, Knight R. Tracking down the sources of experimental contamination in microbiome studies. *Genome Biol*. 2014;15(564):1-3. doi:10.1186/s13059-014-0564-2.

101. Castella V, Kummer D, Gehrig C, Hall D. DNA extraction using the QIAsymphony: Evaluation of PCR inhibitor removal. *Forensic Sci Int Genet Suppl Ser*. 2011;3(1):e69-e70. doi:10.1016/j.fsigss.2011.08.034.

102. Scherer M, Weierstall T, Wegener U, et al. A novel platform for the modular integration of forensic assay setup and medium- to high-throughput purification of nucleic acids. *Forensic Sci Int Genet Suppl Ser*. 2009;2(1):89-90. doi:10.1016/j.fsigss.2009.08.120.

103. Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J Biosci Bioeng*. 2003;96(4):317-23. doi:10.1016/S1389-1723(03)90130-7.

104. Persing DH. Polymerase Chain Reaction: Trenches to Benches. *J Clin Microbiol*. 1991;29(7):1281-1285.

105. Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*. 2006;124(4):837-48. doi:10.1016/j.cell.2006.02.017.

106. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. *Science*. 2005;307(5717):1915-20. doi:10.1126/science.1104816.

107. Li K, Bihan M, Yooseph S, Methe BA. Analyses of the Microbial Diversity across the Human Microbiome. *PLoS One*. 2012;7(6):e32118. doi:10.1371/journal.pone.0032118.

108. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486(7402):207-14. doi:10.1038/nature11234.

109. Willner D, Daly J, Whiley D, Grimwood K, Wainwright CE, Hugenholtz P. Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PLoS One*. 2012;7(4):e34605. doi:10.1371/journal.pone.0034605.

110. Lazarevic V, Gaïa N, Emonet S, et al. Challenges in the culture-independent analysis of oral and respiratory samples from intubated patients. *Front Cell Infect Microbiol*. 2014;4(May):65. doi:10.3389/fcimb.2014.00065.

111. Lusk RW. Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS One*. 2014;9(10):e110808. doi:10.1101/002279.

112. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. Analysis, Optimization and Verification of Illumina-Generated 16S rRNA Gene Amplicon Surveys. *PLoS One*. 2014;9(4):e94249. doi:10.1371/journal.pone.0094249.

113. DePristo MA, Banks E, Poplin RE, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491-498. doi:10.1038/ng.806.

114. Ariefdjohan MW, Savaiano DA, Nakatsu CH. Comparison of DNA extraction kits for PCR-DGGE analysis of human intestinal microbial communities from fecal specimens. *Nutr J*. 2010;9:23. doi:10.1186/1475-2891-9-23.

115. Mirsepasi H, Persson S, Struve C, Andersen LOB, Petersen AM, Krogfelt KA. Microbial diversity in fecal samples depends on DNA extraction method: easyMag DNA extraction compared to QIAamp DNA stool mini kit extraction. *BMC Res Notes*. 2014;7:50. doi:10.1186/1756-0500-7-50.

116. Yu Z, Morrison M. Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques*. 2004;36:808-812.

117. Salonen A, Nikkilä J, Jalanka-Tuovinen J, 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 Methods*. 2010;81(2):127-34. doi:10.1016/j.mimet.2010.02.007.

118. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. Evaluation of methods for the extraction and purification of DNA from the human microbiome. Gilbert JA, ed. *PLoS One*. 2012;7(3):e33865. doi:10.1371/journal.pone.0033865.

119. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486(7402):222-7. doi:10.1038/nature11053.

Supplementary data

A) Preparation of Tris-HCl, pH 8.0 + 0.05% Tween-20 Solution

1. Add 500 µL of the 1 M Tris-HCl, pH 8.0 buffer solution to 49.5 ml of Molecular Biology grade water.
2. Measure out 49.25mL of 1X TE
3. Using a genomic pipette tip, measure out 25 µL of Tween-20.
4. Mix thoroughly.
5. Label bottle with the name of the solution, date prepared, and preparer's initials.
6. Store at room temperature.

CHAPTER 3

DYNAMICS OF THE STOOL MICROBIOTA FROM BIRTH TO SEVEN MONTHS OF AGE

Summary

Bacterial colonization of the GIT in utero has recently become a topic of great interest. Meconium microbiota have been associated with health disorders (neonatal sepsis, mucus congestion and premature birth) and have also been suggested as potential contributors to GIT bacterial profiles later in life. To date, few studies, using high-throughput sequencing, have included meconium specimens in longitudinal studies of early life faecal bacterial profiles of healthy term infants. The aims of this chapter were i) to molecularly characterize the maternal and infant faecal microbiota up until seven months of life, ii) to assess whether maternal faecal microbiota is a possible source of the infant's GIT bacteria, and iii) to study the dynamics of infant faecal bacterial profiles over time.

We applied high-throughput sequencing of the 16S rRNA gene to study the microbiota from meconium specimens and maternal faecal specimens collected at birth; as well as a subset of infant faecal specimens up until seven months of age. Diversity measures used included Shannon diversity and Bray Curtis dissimilarity indices. We constructed lambda scaled log ratio biplots of specimens to determine differences between them based on significantly different operational taxonomic units (OTUs). We used generalized linear and mixed models to determine the effect of external factors on microbial profiles; and to determine significant differences in OTUs from mothers and their infants and from infants over time.

Bacterial profiles were obtained from 90 maternal faecal specimens; 107 meconium; and 72 and 36 infant faecal specimens sampled at 4-12 and 20-28 weeks, respectively. Meconium specimens harboured the most diverse bacterial profiles, and (based on relative abundances of OTUs) harboured mainly the phylum Proteobacteria (60%). The phylum Firmicutes was most abundant at 4-12 weeks (49%) and 20-28 weeks (64%) of life. The phylum Actinobacteria was at its highest at 4-12 weeks of age (26%), followed by 20-28 weeks (9%). Maternal faecal specimens were mainly colonized by the phylum Firmicutes (79%). Infant faecal bacterial profiles became more adult-like over time; however no individual mother-infant pairs clustered at any timepoint studied. The most influential factor for changes in bacterial profiles was the increase in infant age.

Altogether, our results indicate that meconium specimens are diversely colonized by bacteria and are influenced by factors such as maternal education and gender. An interesting finding in our study was the association between maternal HIV status and bacterial diversities from infant faecal specimens sampled at 4-12 weeks of age. In general, meconium and faecal bacterial profiles are distinct from maternal faecal bacterial profiles, but the infant's microbiota do become more adult-like over time.

3.1 Introduction

The human GIT microbiota have been described as a 'plastic' entity with a number of environmental factors influencing its composition.[1] Early life bacterial colonization has conventionally been reported to start at birth;[2] and factors such as mode of delivery,[3–6] feeding practices,[6–8] weaning,[9–11] geography[12,13] as well as the use of antibiotics[14] have been considered as key contributors to these profiles. However, recent studies suggest that bacterial colonization occurs prior to the process of birth by identifying microbes from the previously considered "sterile" womb[15–19] as well as the newborn's meconium.[20–22] These findings raised questions on the origin of the bacteria found in the womb and meconium, and have also contributed to delineate different hypotheses. These hypotheses suggest either translocation of cervical and vaginal bacteria to amniotic fluid[23–25] or dissemination of maternal (oral or GIT lumen) bacteria into the maternal bloodstream followed by transplacental invasion[17,23,26–32] (described in Section 1.4.3 of Chapter 1).

Since our study focusses on characterising bacterial profiles from infant meconium and faecal specimens together with maternal faecal specimens; we found the hypothesis suggesting transfer from the maternal GIT lumen to the infant's GIT most interesting.[20,23] Despite the supporting data from the murine model study showing that bacterial strains orally administered to pregnant mice are indeed detected from the fetal meconium specimens;[20] studies comparing human infant meconium bacterial profiles to those from parent faecal bacterial profiles have provided conflicting results.[21,33] Despite these inconsistent findings, studies have not yet directly addressed this hypothesis by including a range of specimens such as meconium, amniotic fluid, umbilical cord blood, placental and maternal faecal specimens on large scale. In addition, factors influencing the meconium microbial profiles are still understudied.[33]

Bacterial colonization patterns early in life is a key determinant of microbial profiles[34] as well as disease states in both early[2,35–37] and in later stages of life.[38–40] Of note, specific meconium microbiota profiles have been linked to early life disease states such as neonatal sepsis,[41] infant mucus congestion,[21] and premature birth.[15] Meconium microbiota may also contribute to the development of GIT microbial profiles later in life. This concept is supported by the work of Gosalbes and colleagues[21] which showed similarities between meconium and faecal specimens collected during the first few months of life. Other studies have also reported on the important role of early life colonizers in preparing the environment for successive bacterial colonization.[42,43]

To date, only a few studies have included meconium specimens when longitudinally studying changes in early life faecal bacterial profiles of healthy term infants.[21,44] Therefore, the use of larger sample sizes may be very useful to determine what a healthy meconium microbiota looks like; to determine its roles in influencing successive faecal bacterial profiles; and to determine which

prenatal factors may influence the meconium microbiota. This chapter therefore aimed, using a larger sample size, to (i) characterise the maternal and infant faecal microbiota; (ii) to assess whether the maternal GIT microbiota is a possible source of the infant's GIT inhabitants; iii) to investigate how infant bacterial profiles change during the first seven months of life; and iv) to determine factors that may influence the maternal and infant faecal bacterial communities in our study population.

3.2 Materials and methods

3.2.1 Ethics statement

Both the DCHS and our pilot study received approval from the Faculty of Health Sciences, HREC of the University of Cape Town, South Africa. The HREC reference numbers assigned for these studies were 401/2009 (DCHS) and 742/2013 (this study).

3.2.2 Study population and sample selection

This pilot study was nested within the DCHS,[45] as described in detail in Chapter 2 (Sections 2.2.1 to 2.2.3). The aim of this pilot study was to include as many specimens as possible from mothers and their infants at birth. In addition, the study included a subset of longitudinal specimens from infants up to the age of seven months. The main criteria when selecting maternal and infant specimens from the DCHS database was that specimens were sampled at birth. The second criterion was that faecal specimens collected at later timepoints had to be from infants from which samples were collected at birth, and collections had to have taken place within the timeframes 4 and 12 weeks or 20 and 28 weeks of age. The term "groups under study" will be used when referring collectively to infant and maternal specimens in subsequent sections.

3.2.3 Laboratory processing of faecal specimens and16S ribosomal DNA sequencing

A detailed description of the laboratory processing of faecal specimens and the approach to Illumina Miseq sequencing is summarised in Chapter 2 (Sections 2.2.4 to 2.2.7). Briefly, nucleic acid extractions took place from 50 mg of faecal specimen using the QIAsymphony DSP Virus/Pathogen Mini Kit® (Qiagen GmbH, Hilden, Germany) (Section 2.2.5). These extracts served as template during library preparation for 16S ribosomal DNA sequencing (Section 2.2.6). The V4 hypervariable region of the 16S rRNA gene was targeted using previously published primers 515F and 806R.[46] Quantification of purified pooled PCR products took place using the KAPA Library Quantification Kit K4835 (KAPA

Biosystems, MA, USA) as per manufacturer's instructions, with slight modifications. We prepared sequencing reagents and the DNA library using the MiSeq Reagent Kit v3, 600 Cycles (Illumina, CA, USA) as per manufacturer's instructions (Section 2.2.7).[47,48]

3.2.4 Bio-informatics workflow and data manipulation

The bioinformatics workflow used in this study to classify 16S rRNA sequences to bacterial taxonomies in this study is referred to as YAP[49] and is described in detail in Section 2.2.8 of Chapter 2. In addition to the YAP workflow, we also corrected for potential contamination, which is described in Section 2.2.9.3 of Chapter 2.

3.2.5 Data analysis and statistics

We used R software version 3.1.1[50] together with RStudio version 0.98.507[51] for all data analyses. The OTU table constructed at 97% sequence similarity (genus-level), and corrected for contamination (described in Section 2.2.9.3 of Chapter 2), was selected for data analysis.

3.2.5.1 *Demographic and clinical characteristics of participants*

In order to describe participant characteristics we used the [summ] and [tab1] functions of the R package *epicalc.*[52]

3.2.5.2 *Calculating inter-individual (beta) diversity*

In order to determine the dissimilarity between faecal specimens, we used the Bray Curtis dissimilarity index[53,54] (equation 4.1) due to its general suitability for ecological data.[55,56] In order to calculate the above-mentioned, we used the [vegdist] function from the R package *vegan.*[57]

$$d_{BC} = \frac{\sum_{k=1}^{P} |x_{k1} - x_{k2}|}{\sum_{k=1}^{P} (x_{k1} + x_{k2})} \qquad [4.1]$$

Where, x_{k1} = abundance of species k in sampling unit 1

x_{k2} = abundance of species k in sampling unit 2, and

$\boldsymbol{P}$ is the total number of species recorded across both units.

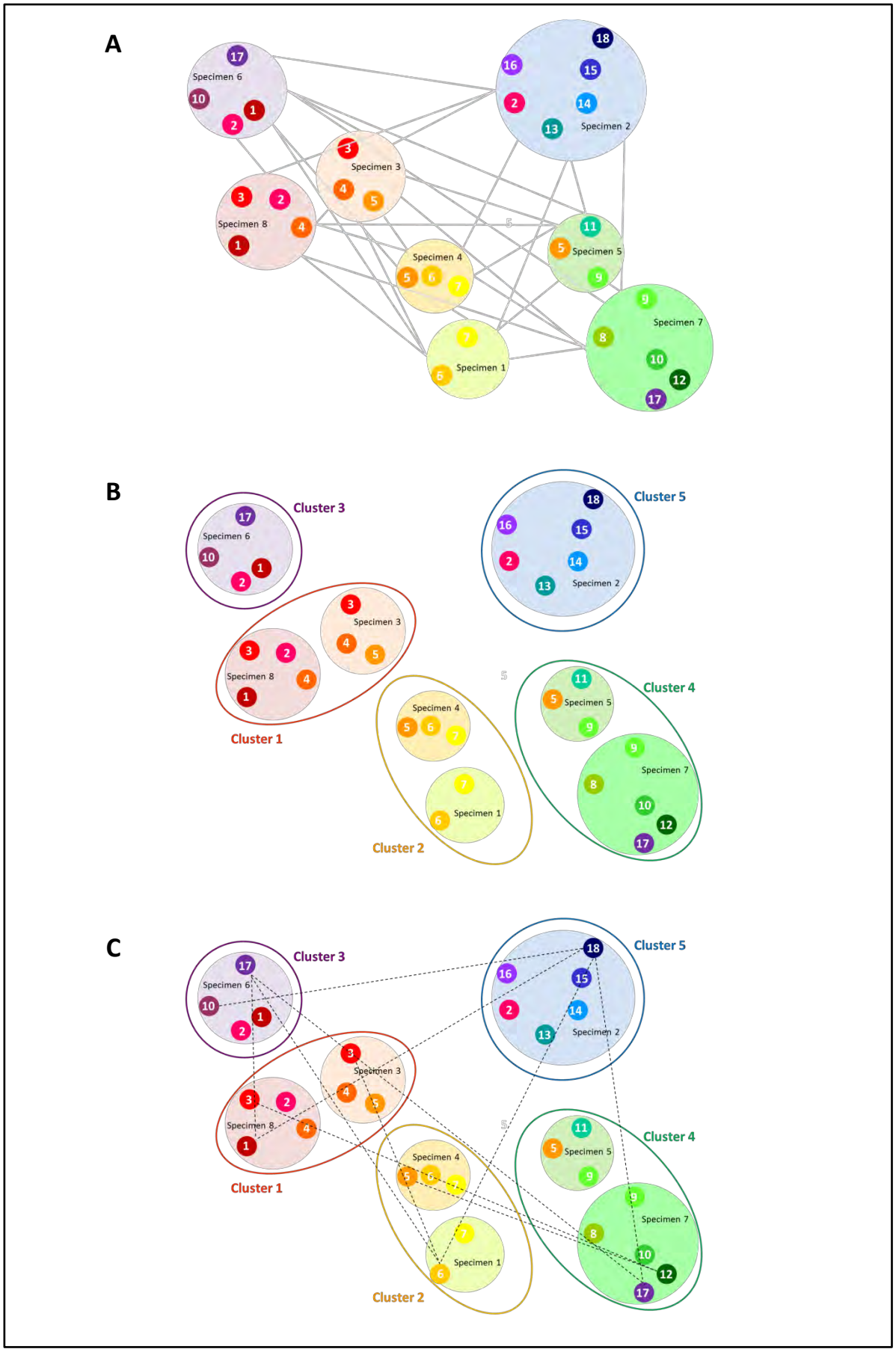
In this study, the Bray Curtis dissimilarity index measured the distance between faecal specimens based on their OTU composition, as explained in Table 5. Measures ranged from 0 to 1, where dissimilarity measures closer to zero indicated that communities are more similar in their compositions.

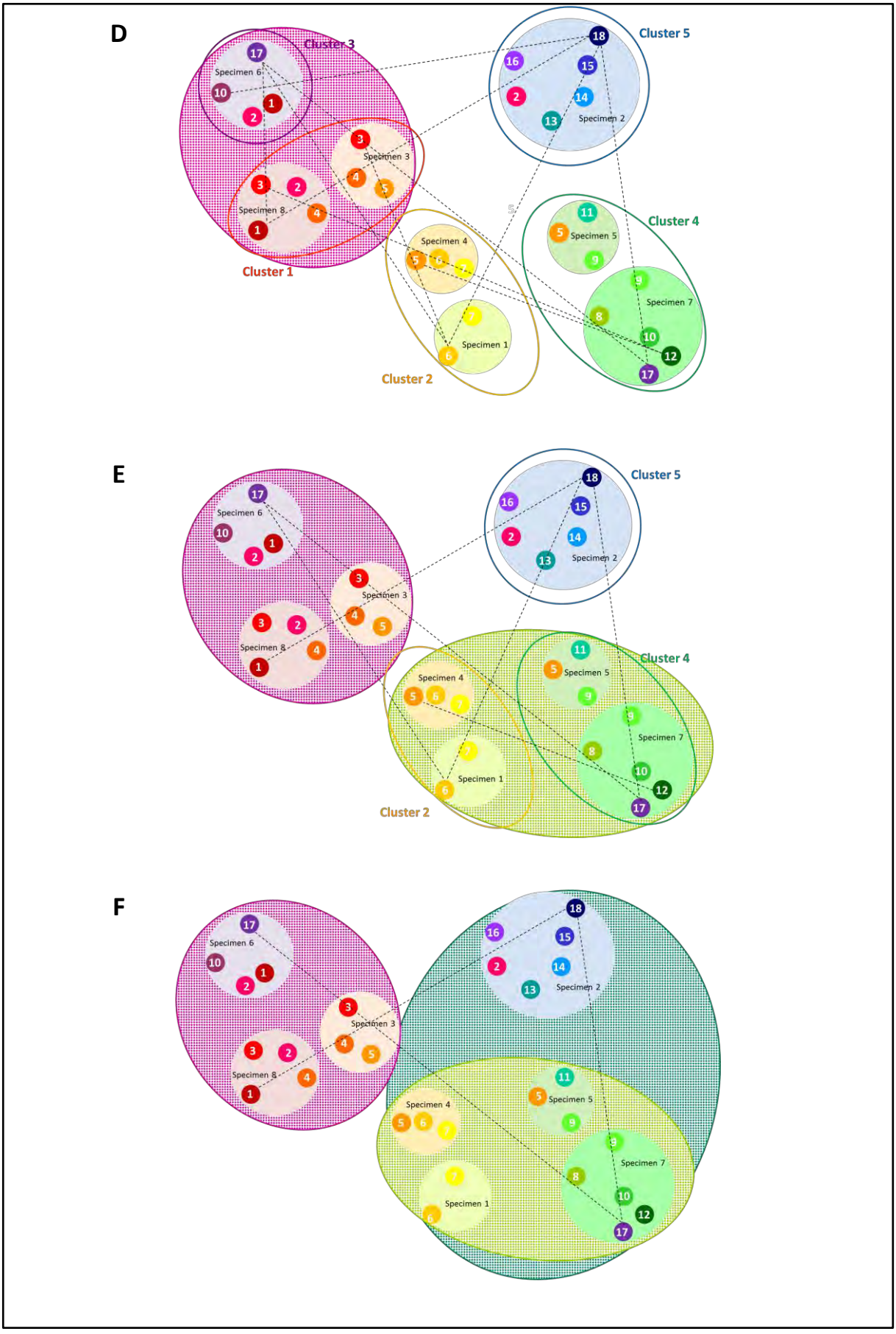
Table 5. An example of the Bray Curtis dissimilarity index calculated for two specimens based on their OTU composition

OTU composition	Relative abundance of each OTU from specimen 1 (x_1)	Relative abundance of each OTU from specimen N (x_N)	Absolute difference between abundances of each OTU from specimen 1 and N $\lvert x_1 - x_N \rvert$	Sum of abundances of each OTU from specimen 1 and N $(x_1 + x_N)$	Bray-Curtis dissimilarity index
OTU A	0.250	0.000	0.250	0.250	$d_{BC} = \frac{\lvert x_1 - x_N \rvert}{(x_1 + x_N)}$
OTU B	0.025	0.000	0.025	0.025	
OTU C	0.375	0.500	0.125	0.875	
OTU X	0.225	0.300	0.075	0.525	$d_{BC} = \frac{0.550}{2.000}$
OTU Z	0.125	0.200	0.075	0.325	
SUM	**1.000**	**1.000**	**0.550**	**2.000**	

3.2.5.3 *Clustering of faecal specimens*

We used the [hclust] function from the built-in R package *stats*[50] to perform agglomerative clustering using Complete Linkage (also referred to as furthest neighbour clustering).[55] We made use of the Bray-Curtis dissimilarity index (calculated in Section 3.2.5.2) to calculate the matrix used during the agglomerative clustering. Clustering was performed on all OTUs with a relative abundance > 0.5%. The OTUs with a relative abundance of < 0.5% (referred to as "other") are graphically represented according to the clustering of the OTUs with a relative abundance > 0.5%. This clustering method is explained schematically in Figure 21.


A
Specimen 6
Specimen 2
Specimen 3
Specimen 8
Specimen 4
Specimen 5
Specimen 7
Specimen 1
B
Cluster 3
Cluster 5
Cluster 1
Cluster 4
Cluster 2
C
Cluster 3
Cluster 5
Cluster 1
Cluster 4
Cluster 2


D
Cluster 3
Specimen 6
Specimen 3
Specimen 8
Cluster 1
Cluster 5
Specimen 2
Cluster 4
Specimen 5
Specimen 7
Specimen 4
Specimen 1
Cluster 2
E
Cluster 5
Specimen 2
Specimen 6
Specimen 3
Specimen 8
Cluster 4
Specimen 5
Specimen 7
Specimen 4
Specimen 1
Cluster 2
F
Specimen 6
Specimen 3
Specimen 8
Specimen 2
Specimen 5
Specimen 4
Specimen 7
Specimen 1

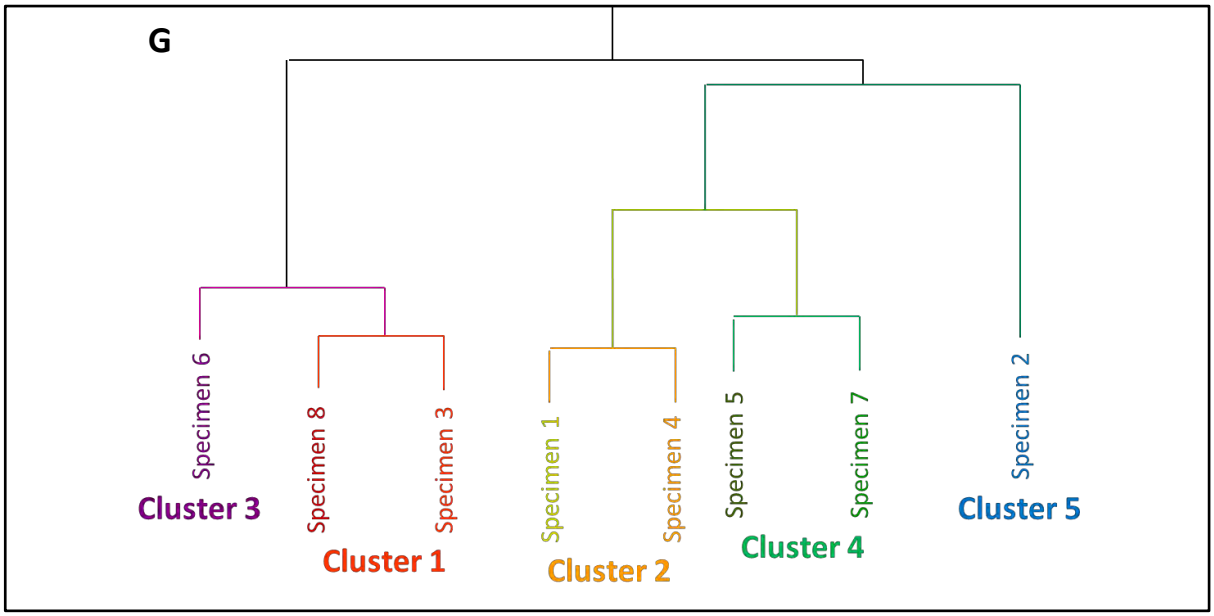


Figure 21. Clustering of faecal specimens based on their OTU compositions
Each small filled circle which has a number allocated represents an OTU. Each of the bigger filled circles represents a specimen containing OTUs. **A)** The Bray Curtis dissimilarity index calculates the distance between specimens, based on their OTU composition. **B)** The dissimilarity matrix is used to assign specimens to clusters based on the distances between specimens. **C)** Complete linkage then uses the Bray Curtis dissimilarity index to compute distances between the furthest OTUs from each cluster. **D)** The first cluster is formed between the two original clusters with the smallest dissimilarity. **E)** The second cluster is formed between the next set of clusters with the smallest dissimilarity. **F)** The procedure continues until all clusters are linked and a **G)** dendogram can be drawn.

3.2.5.4 *Multidimensional scaling of clusters*

In the previous section, dendograms are drawn to avoid crossing of lines and do not provide information on the similarity between adjacent clusters. In order to determine how similar clusters were, we performed multidimensional scaling (MDS)[58] on each of the clusters from the dendograms. Clusters were represented in MDS plots as pie charts summarising the proportions of OTUs within each of the clusters. We used the [smacofSym] function from the R package *smacof*[59] to perform MDS, using the Bray-Curtis dissimilarity matrix as "input data". The aim of MDS is to represent the clusters such that the ordinary Euclidean distances between clusters in the plot match as closely as possible the Bray-Curtis distances between the cluster means.

3.2.5.5 *Constructing biplots using sequencing data*

Data output resulting from high throughput sequencing is in the form of count data.[60] Due to the variation in the total number of reads sequenced across different specimens within a single run,[61] count data needs to be transformed to compositional data by calculating the relative abundance of each OTU per specimen.[61,62]

Since we were dealing with compositional data, we constructed log-ratio biplots,[63] in which only OTUs identified as statistically different (at the 5% significance level) from the "groups under

study" were used. We also made use of lambda scaling in our log-ratio biplots to ensure evenness in the "total spread" of each of the data sets (OTUs and specimens).[64] Lambda scaling adjusted for plots where one set of data points are plotted on top of each other; and data points from the other set are plotted further apart.

The use of log-ratio methodology, was made possible by first transforming count data to compositional data.[61,62] We then adjusted the data in a Bayesian context[65] to remove zeros.[66] Aitchinson and Greenacre (2002), shows that constructing a log ratio biplot is essentially equivalent to performing the central log ratio transformation.[62,67] The first step was to replace zero-values with non-zero values. The second step was to correct for the added non-zero values across all specimens within the respective taxa. This ensured that the compositional total after adjustment did not change. The transformation of count data to compositional data and the adjustment for zeros are explained using one specimen with three taxa (X, Y, Z), outlined in Tables 6 and 7. The examplar table, Table 6, presents OTUs in the form of count data prior to adjustment for zeros, while the exemplar table, Table 7, presents OTUs in the form of compositional data following adjustment for zeros.

Table 6. Example of a "count data" OTU table prior to adjustment for zeros

Taxon	Specimen 1	Specimen 2	Specimen 3	Specimen N
X	4	0	2	5
Y	5	10	0	0
Z	0	12	4	5
Total	**9**	**22**	**6**	**10**

An example of adjustment for zeros (using Specimen 1 as example):

Step 1: What is the proportion of each taxon in Specimen 1:

$$\frac{4}{9}+\frac{5}{9}+\frac{0}{9}=1$$

Step 2: Calculate a non-zero value to replace the zero:

$$\text{Fraction of zeros present in Specimen 1} = \frac{1}{3} = 0.333$$

$$\text{Non zero value for Specimen 1} = \frac{\text{Fraction of zeros present in Specimen 1}}{\text{Total number of OTUs in Specimen 1 +1}} = \frac{0.333}{9+1} = 0.033 = \frac{1}{30}$$

Step 3: Replace the zero value with a non-zero value:

$$\frac{4}{9}+\frac{5}{9}+\frac{1}{30}>1$$

Step 4: Subtract the non-zero value added to Specimen 1 from the remaining taxa in Specimen 1:

Other Taxa: $\frac{4}{9}+\frac{5}{9}=1$ [i]

What needs to be subtracted from other Taxa: $1\times\left(\frac{1}{30}\right)$

New total for other Taxa: $1-1\times\left(\frac{1}{30}\right)=\frac{29}{30}$

In order to obtain the new total we need to have: $\left(something<\frac{4}{9}\right)+\left(something<\frac{5}{9}\right)=\frac{29}{30}$ [ii]

Step 5: Multiplying all components of equation [i] with $\frac{29}{30}$ will satisfy the form of equation [ii]:

$$\left\{\frac{4}{9}\times\frac{29}{30}\right\}+\left\{\frac{5}{9}\times\frac{29}{30}\right\}=1\times\frac{29}{30}=\frac{29}{30}$$

$$\{0.4296\}+\{0.537\}=\frac{29}{30}$$

We subtracted more from Taxon Y than from Taxon X since Taxon Y had a greater number of OTUs compared to Taxon X.

Start with: $0.4444+0.5556+0=1$

End with: $0.4296+0.537+0.0333\cong 1$

This ensures that the reduction of the added non-zero value takes place in a proportionate manner.

Table 7. Example of a "compositional data" OTU table resulting from adjustment for zeros

Taxon	Specimen 1	Specimen 2	Specimen 3	Specimen N
X	0.43	0.15	0.32	0.48
Y	0.54	0.45	0.05	0.03
Z	0.03	0.52	0.63	0.48
Total	**~1.00**	**~1.00**	**~1.00**	**~1.00**

3.2.5.6 *Fitting of generalized linear models*

We performed hypothesis testing at a 5% significance level while controlling for the false discovery rate described by Benjamini & Hochberg.[68] Generalized linear models (GLMs) tested the effect of risk factors on the composition and diversity of the maternal and infant faecal microbiota profiles, respectively. We tested GLMs on all OTUs sequenced; on OTUs with a relative abundance greater than 0.25%; and on OTUs with a relative abundance greater than 0.5%. Final models were based on OTUs with a relative abundance greater than 0.5%.

In general, the Poisson distribution model (equation 4.2) is used to model count data.[69,70] Modelling count data variance, using the Poisson distribution models, assumes that the variance in the data is equal to its mean ($var(Y_i) = E(Y_i)$). However, the variance in count data obtained from high-throughput sequencing exceeds its mean ($var(Y_i) < E(Y_i)$) and is referred to as being over-dispersed.[69,71] An alternative choice to Poisson modelling is the negative binomial distribution. This distribution provides a model ($var(Y_i) = \phi E(Y_i)$) which allows for over-dispersion, as $\phi > 1$ is a parameter that can be estimated. We implemented the negative binomial model in RStudio[51] through the family function quasipoisson. To account for compositional data (instead of count data)[62] in our model, we specified the offset as equal to "root OTU counts" during modelling.[72] The parameters of a GLM are estimated by the iterative weighted least squares method, implemented in R in the function [glm] in the built in package *stats.*[50]

$$\log(Y) = \beta 0 + \beta 1x1 + \ldots\ldots + \beta nxn + error \qquad [4.2]$$

$$\log(Y) = \mu + \log(offest) + \beta_1(birth\ weight) + \beta_2(birth\ length) + \alpha_{gender} + \alpha_{mom-edu} + \alpha_{area} + \beta_3(gestation\ age) + \alpha_{delivery} + \alpha_{momHIV} + \varepsilon_{ij}$$

Where, Y = OTU count for the taxon currently tested
$offest$ = root OTU count
μ = overall mean effect
$\beta_1, \beta_2, \beta_3$ the coefficient parameters to be estimated for the continuous predictor variables
$\alpha_{gender} = \begin{cases} \alpha_F \\ \alpha_M \end{cases}$ the parameters to be estimated for the gender effect with $\alpha_F + \alpha_M = 0$
$\alpha_{mom-edu} = \begin{cases} \alpha_{Gr1-7} \\ \alpha_{Gr8-11} \\ \alpha_{Gr12} \\ \alpha_{Tert} \end{cases}$ the parameters to be estimated for the mother's education effect with $\alpha_{Gr1-7} + \alpha_{Gr8-11} + \alpha_{Gr12} + \alpha_{Tert} = 0$
$\alpha_{area} = \begin{cases} \alpha_1 \\ \alpha_2 \end{cases}$ the parameters to be estimated for the area effect with $\alpha_1 + \alpha_2 = 0$
$\alpha_{delivery} = \begin{cases} \alpha_{vaginal} \\ \alpha_{ceasarean} \end{cases}$ the parameters to be estimated for the gender effect with $\alpha_{vaginal} + \alpha_{ceasarean} = 0$
$\alpha_{momHIV} = \begin{cases} \alpha_{pos} \\ \alpha_{neg} \end{cases}$ the parameters to be estimated for the gender effect with $\alpha_{pos} + \alpha_{neg} = 0$

3.2.5.7 *Fitting of generalized linear mixed models*

We performed hypothesis testing at a 5% significance level while controlling for the false discovery rate described by Benjamini and Hochberg.[68] Generalized linear mixed models (GLMMs) were used to determine the difference in bacterial profiles between mothers and infants; as wells as

changes in an infant's bacterial profile over time. It is shown in a paper by Cnaan and colleagues[73] that the model is formulated in a two-stage approach. The first stage models the mothers and their infants or infants followed over time and the second stage models the individuals at each time point. Cnaan and colleagues[73] explains that all observations on the same pair will have the same between-pair errors, but stage two allows for different within-pair errors. By fitting this model, the within-pair differences (differences between mother and infant or infants over time), can be tested without being confounded with between-pair differences.The models were fitted with the glmmPQL function in the R package *MASS*[74] which implements a penalised quasilikelihood approach to estimating the model parameters.

3.3 Results

3.3.1 Participant characteristics

Table 8 summarises the clinical and demographical characteristics of the study participants. The study included faecal specimens from 90 mothers. The median age of mothers providing faecal specimens was 25 years (IQR; 22.2 - 32.1). The two residential areas were represented by similar numbers of mothers, with 9% more mothers residing in the area TC Newman. Most of the mothers included in this study had some form of secondary-level education, Grade 8-12 (84.5%), and were HIV negative (80%). Maternal faecal specimens sampled at birth were matched to 90 meconium specimens sampled from infants at birth; 72 infant faecal specimens sampled at 4-12 weeks of age; and 36 infant faecal specimens sampled at 20-28 weeks of age. The median age for the 107 infants providing meconium specimens at birth was 0 weeks (IQR; 0 - 0). The 72 infants providing faecal specimens at 4-12 weeks had a median age of 7 weeks (IQR; 6.8 - 8.3) and the 36 infants providing faecal specimens at 20-28 weeks had a median age of 24 weeks (IQR; 23 - 26). At all time-points, infant faecal specimen collections were slightly higher from the residential area Mbekweni, when compared to the residential area TC Newman. We also observed that the majority of infants had mothers with an secondary education level of Grade 8-12 at any of the three time points studied. The HIV status of infants is referred to as "HIV-exposed" if their mothers were tested HIV-positive or "HIV-unexposed" if their mothers were tested HIV-negative. The majority of infants were HIV-unexposed, of which none tested positive for HIV infection. The median gestational age of infants participating in this study was 39 weeks (IQR; 38 - 40); the median birth weight was 3 kilogram (kg) (IQR; 2.7 - 3.3); and the median birth length was 51 centimeter (cm) (IQR; 48.5 - 53.0). At any of the timepoints under study, the majority of infants (80.5 - 83.3%) were delivered via vaginal delivery. We

also observed a relatively similar distribution for gender with 45 - 50% of infants being males across all timepoints. We found that breastfeeding at six to ten week of age was recorded for more than 70% of the infants at 4-12 (72.5%) and 20-28 weeks of age (71.4%). Almost 80% of the infants studied at 20-28 weeks received breastfeeding at 14 weeks of life and 66% of these infants received breastfeeding at six months of life. Formula feeding was recorded for eight infants during the first week of life; for 13 infants during the first month of life; for 17 infants during the first two months of life; and for 20 infants during the first four months of life. The introduction of solid food was recorded for 18 infants studied at 4-12 weeks of age; for which three infants received solid food at one month of age, nine infants received solid food at two months of age and six infants received solid food at three months of age. With regards to the subset of infants studied at 20-28 weeks, 30 infants had received solid foods at some point during the first six months of life. Two of the infants studied at 20-28 weeks received solid food at one month of age, five at two months, four at three months, two at four months, six at five months and 11 at six months of age.

Table 8. Demographic and clinical characteristics of maternal and infant participants

Participant characteristics	Participants (%)	Median (IQR)
Mothers (n = 90)		
Age at which specimens were collected (years)		25.2 (22.2 - 32.1)
Residential area		
TC Newman	49 (54.4)	
Mbekweni	41 (45.6)	
Maternal education		
Primary education (Grade 1 to 7)	8 (8.9)	
Secondary education (Grade 8 to 11)	43 (47.8)	
Secondary education (Grade 12)	33 (36.7)	
Tertiary education (enrolment/completion)	6 (6.7)	
Maternal HIV status at time of delivery		
HIV-positive	18 (20)	
Infants at birth (n = 107)		
Age at which specimens were collected (weeks)		0 (0 - 0)
Residential area		
TC Newman	49 (45.8)	
Mbekweni	58 (54.2)	
Maternal education		
Primary education (Grade 1 to 7)	11 (10.3)	
Secondary education (Grade 8 to 11)	49 (45.8)	
Secondary education (Grade 12)	39 (36.5)	
Tertiary education (enrolment/completion)	8 (7.5)	
Maternal HIV status at time of delivery		
HIV-positive	26 (24.3)	
Gestational age (weeks)		39 (38 - 40)
Birth weight (kg)		3 (2.7 - 3.3)
Birth length (cm)		50 (48 - 53)
Mode of delivery		
Vaginal delivery	87 (81.3)	
Gender		
Male	48 (44.9)	

Table 8. Demographic and clinical characteristics of maternal and infant participants (continued)

Participant characteristics	Participants (%)	Median (IQR)
Infants at 4-12 weeks (n = 72)		
Age at which specimens were collected (weeks)		7 (6.8 - 8.3) ϯ
Residential area		
TC Newman	30 (41.7)	
Mbekweni	42 (58.3)	
Maternal education		
Primary education (Grade 1 to 7)	8 (11.1)	
Secondary education (Grade 8 to 11)	31 (43.0)	
Secondary education (Grade 12)	29 (40.3)	
Tertiary education (enrolment/completion)	4 (5.6)	
Maternal HIV status at time of delivery		
HIV-positive	19 (26.4)	
Gestational age (weeks)		39 (38 - 40)
Birth weight (kg)		3.1 (2.7- 3.4)
Birth length (cm)		51 (49 - 53)
Mode of delivery		
Vaginal delivery	60 (83.3)	
Gender		
Male	33 (45.8)	
Breastfed at 6-10 weeks of age	50 (72.5) ǂ	
Formula feeding	16 (38.1) δ	
Introduction of solid food recorded from 1-3 months of age	18 (25)	
Infants at 20-28 weeks (n = 36)		
Age at which specimens were collected (weeks)		24 (23 - 26)
Residential area		
TC Newman	16 (44.4)	
Mbekweni	20 (55.6)	
Maternal education		
Primary education (Grade 1 to 7)	7 (19.4)	
Secondary education (Grade 8 to 11)	14 (38.9)	
Secondary education (Grade 12)	13 (36.1)	
Tertiary education (enrolment/completion)	2 (5.6)	
Maternal HIV status at time of delivery		
HIV-positive	10 (27.8)	
Gestational age (weeks)		39 (38.8 - 40)
Birth weight (kg)		3.0 (2.6 - 3.3)
Birth length (cm)		50.5 (49 - 53)
Mode of delivery		
Vaginal delivery	29 (80.5)	
Gender		
Male	18 (50.0)	
Breastfed at 6-10 weeks of age	25 (71.4) ϕ	
Breastfed at 14 weeks of age	23 (79.3) ψ	
Breastfed at 6 months of age	21 (65.6) ϴ	
Formula feeding	2 (10) ʓ	
Introduction of solid food recorded from 1-6 months of age	30 (83.3)	

IQR: interquartile range;
Number of observations recorded: ϯ 71/72; ǂ 69/72; δ 42/72; ϕ 35/36; ψ 29/36; ϴ 32/36; ʓ 20/36

3.3.2 Microbial profiles from all "groups under study"

Most of the OTUs sequenced from all "groups under study" (n = 305) represented taxa from the domain Bacteria (99.88%). The remaining 0.12% of OTUs sequenced represented taxa from the domain Archaea (0.11%) as well as unclassifiable taxa (0.01%). Because the domain Bacteria is our main interest, the remainder of this section will focus on describing bacterial taxa sequenced from all "groups under study".

We classified 51 phyla from all "groups under study" (n =305) of which 28 phyla had relative abundances > 0.5% and 23 phyla had relative abundances < 0.5% (herein referred to as "other"). The most abundant phyla included the Firmicutes (50.44%), Proteobacteria (28.13%), Actinobacteria (13.54%), Bacteroidetes (5.39%) and an unclassified (1.33%) phylum. The remaining 23 phyla, which had relative abundances < 0.5%, constituted 1.17% of all OTUs classified at phylum-level. At class-level, only 11 of the 55 bacterial classes obtained in this study had relative abundances > 0.5%. The most abundant bacterial classes were Clostridia (32.86%), followed by the class Actinobacteria (13.54%) and the class Gammaproteobacteria (12.75%). At order-level, 23 of the 97 orders successfully classified had relative abundances > 0.5%. Of these, the order Clostridiales (32.71%) was most abundant, followed by the order Lactobacillales (9.36%). At family-level, 39 of the 222 families had relative abundances > 0.5%. The most abundant families obtained in this study were the Lachnospiraceae (17.70%) and Ruminococcaceae (7.29%). At genus-level, 616 genera were obtained, of which 49 had relative abundances > 0.5%. The most abundant genera were an unclassified genus (6.95%) within the phylum Firmicutes, followed by the genus Bifidobacterium (6.61%). Relative abundances for all bacterial taxa from the faecal specimens under study are summarised in Supplementary table S1.

3.3.3 Comparison of microbial profiles from infant and maternal faecal specimens sampled at birth

The following section reports the relative abundance of bacterial OTUs classified from 107 meconium specimens and 90 maternal faecal specimens sampled at birth. We also provide data comparing 90 meconium specimens, to their matched maternal faecal specimens sampled at birth. Descriptions are provided at both phylum- and genus-level for bacteria with proportions greater than 0.5% and those with less than 0.5%.

3.3.3.1 *Microbiota profiles at the phylum-level*

The most abundant phylum identified from infant meconium specimens (n = 107) was the phylum Proteobacteria (59.67%), followed by the phyla Firmicutes (22.91%), Actinobacteria (9.56%), and Bacteroidetes (4.31%) (Table S2). Overall, the phyla at abundances < 0.5% represented 2.38% of

all bacteria within infant meconium specimens. Of these, the phylum Cyanobacteria Chloroplast (0.46%) was most abundant, followed by Deinococcus Thermus (0.31%), Verrucomicrobia (0.18%), Chloroflexi (0.17%) and Fusobacteria (0.16%) (Table S3). In maternal faecal specimens collected at delivery (n = 90), the phylum Firmicutes (79.39%) was most abundant, followed by the phylum Bacteroidetes (7.06%), Actinobacteria (5.77%), Proteobacteria (4.71%) and unclassified phyla (2.44%) (Figure 22C and Table S2). The phyla at abundances < 0.5% represented 0.62% of all bacteria within maternal faecal specimens. Of these, the most abundant phyla were Euryarchaeota (0.23%) and Verrucomicrobia (0.17%) (Table S3 and Figure 22D).

Based on the clustering analysis of OTUs with abundances > 0.5% from 90 mother-infant pairs, two primary clusters were identified at phylum-level, and tentatively named clusters M and B (Figure 22A). The primary cluster M consisted of three sub-clusters (sub-clusters no. 1, 2 and 3). Two of these sub-clusters (sub-clusters no. 1 and 2) consisted mainly of maternal faecal specimens, for which two infant meconium specimens (B_44 and B_106) clustered within sub-cluster no. 1. Sub-cluster no. 3 consisted of 11 infant meconium specimens. The primary cluster B of Figure 22A also consisted of three sub-clusters (sub-clusters no. 4, 5 and 6). All three these sub-clusters consisted of infant meconium specimens, with the exception of one maternal faecal specimen (M_45) clustering within sub-cluster no. 6. The meconium specimen (B_67) clustered by itself within sub-cluster no. 5 and contained only two phyla at abundances > 0.5%, of which the phylum Proteobacteria was most abundant (Figure 22C). Following the clustering analysis, we plotted the participant characteristics for each of the participants against the position of their specimen within the dendogram. Except for the sample type (maternal versus infant), we did not observe any pattern of participant characteristics related to any of the clusters from the clustering analysis (Figures 22A and B).

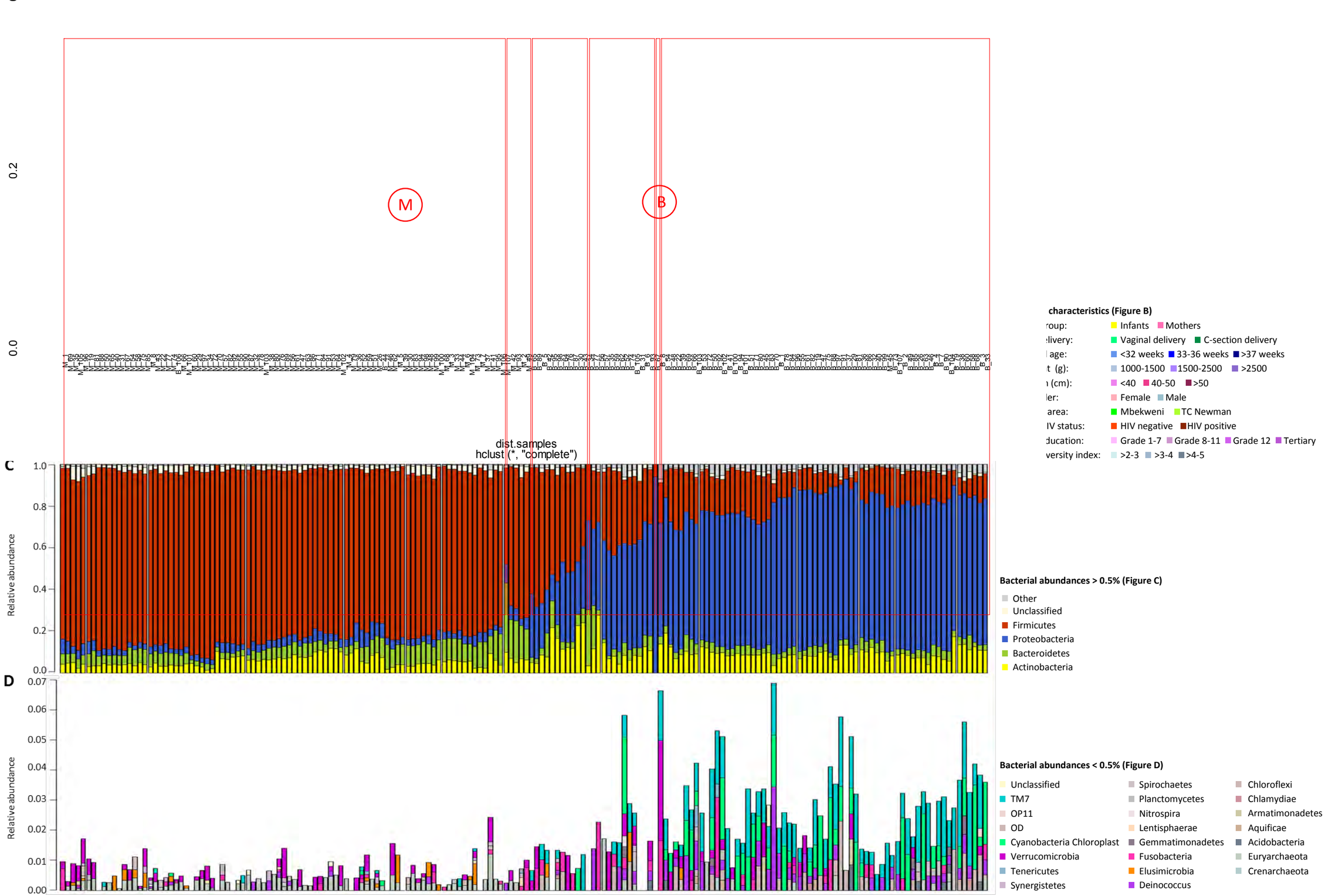


Figure 22. Phylum-level bacterial profiles of infant meconium and maternal faecal specimens sampled at birth

A) Dendogram of phylum-level bacterial profiles for infant meconium (n = 90) and maternal faecal (n = 90) specimens sampled at birth; **B)** Participant characteristics for each of the specimens under study; **C)** Relative abundance of bacteria with proportions > 0.5% at phylum-level; Relative abundance of bacteria with proportions < 0.5% at phylum-level.

3.3.3.2 *Microbiota profiles at the genus-level*

In infant meconium specimens (n = 107), the most abundant genus was Acinetobacter (6.28%), followed by the genera Aeromonas (3.56%), Shewanella (3.24%), an unclassified genus within the family Enterobacteriaceae (3.26%), and the genus Pseudomonas (2.65%) (Table S4). In this study, 14.95% of the genera with abundances > 0.5% were unclassifiable from meconium specimens. Genera with abundances of < 0.5% made up 35.96% of all bacteria within infant meconium specimens (Table S5). Unclassified genera within the families Lachnospiraceae (14.41%) and Ruminococcaceae (10.09%), followed by an unclassified genus within an unclassified family from the order Clostridiales (9.76%) were the most abundant genera from maternal faecal specimens (n = 90) (Table S4 and Figure 23C). These were followed by the genera Blautia (5.95%) and Lachnospiraceae incertae sedis (4.41%). In total, 41.38% of the genera with abundances > 0.5% represented unclassifiable genera from maternal faecal specimens. Genera with abundances < 0.5% are summarised in Table S5 and made up 17.45% of all bacteria within maternal faecal specimens.

Figure 23A summarises the two principle clusters (clusters M and B) evidenced at genus-level for 90 mother-infant pairs sampled at birth. Cluster M in Figure 23A only consisted of a single sub-cluster (sub-cluster no. 1) which contained the majority of maternal faecal specimens. It also contained three meconium specimens, all of which clustered with maternal faecal specimens at phylum-level (Figure 22A). Cluster B in Figure 23A contained sub-clusters no. 2 to 6, which mainly consisted of meconium specimens. Only sub-cluster no. 3 contained two maternal faecal specimens. The maternal faecal specimen M_45 (from sub-cluster no. 3) also clustered with meconium specimens at phylum-level (Figure 22A). The meconium specimen, B_67, which previously clustered by itself at phylum-level now clustered with other meconium specimens. Meconium specimen B_31 formed its own sub-cluster (sub-cluster no. 4) at genus-level. As observed for the phylum-level, plots of participant characteristics against the dendogram did not provide any interesting patterns for any of the clusters at genus-level except for the sample type (maternal versus infant) (Figures 23A and B).

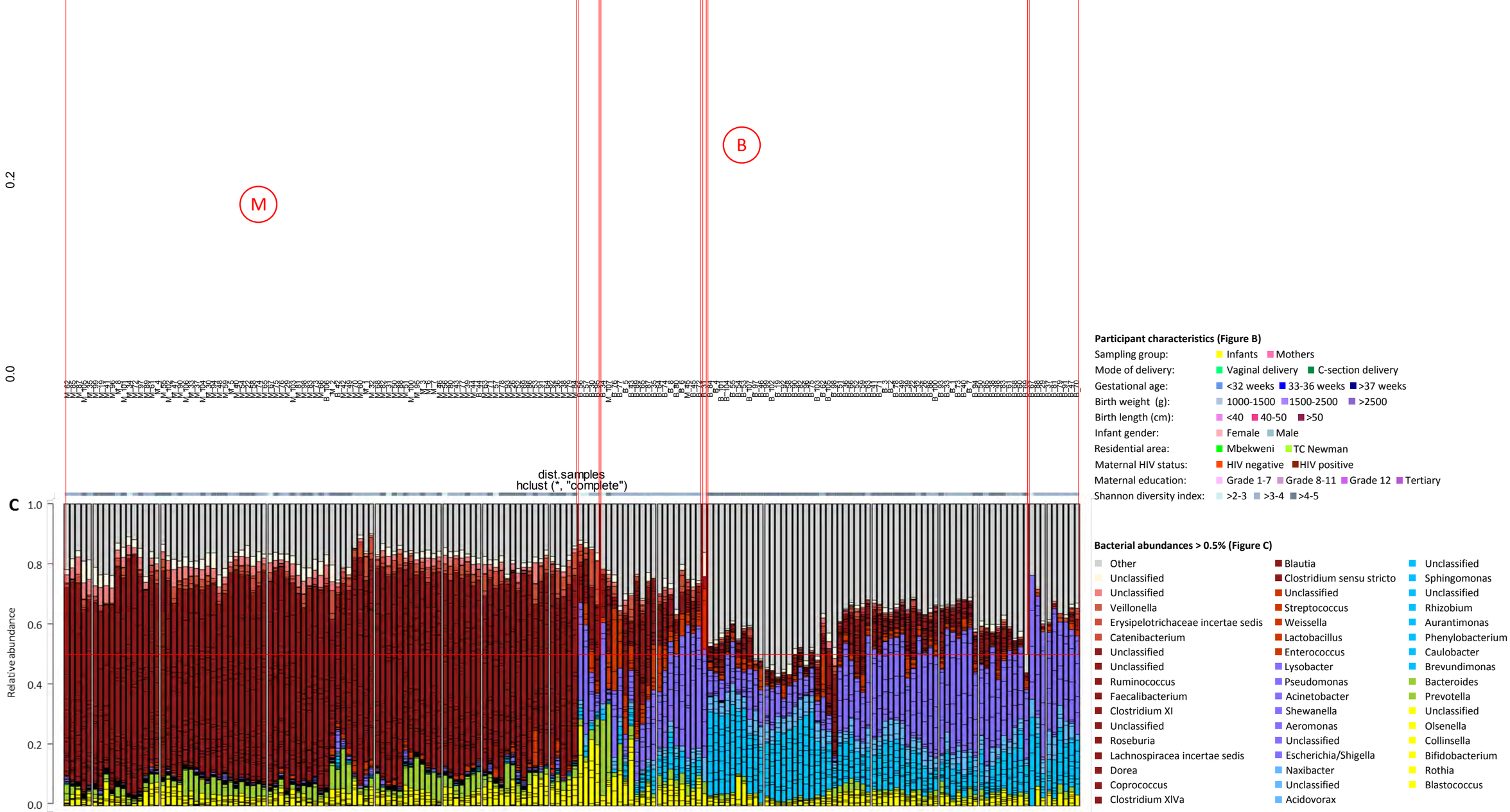


Figure 23. Genus-level bacterial profiles of infant meconium and maternal faecal specimens sampled at birth

A) Dendogram of genus-level bacterial profiles for infant meconium (n = 90) and maternal faecal (n = 90) specimens sampled at birth; **B)** Participant characteristics for each of the specimens under study; **C)** Relative abundance of bacteria with proportions > 0.5% at genus-level.

3.3.3.3 *Similarities between sub-clusters consisting of infant and maternal faecal specimens*

The MDS plots (Figures 24A and B) represent sub-clusters from the 90 mother-infant pairs (Figures 22 and 23) in the form of pie charts. These plots provide information on the similarity of sub-clusters (based on their bacterial composition), which are not provided by the clustering analysis. Meconium specimens (n = 90) clustered at greater distances over a bigger area compared to maternal faecal specimens (n = 90), at both phylum- and genus-level. Meconium specimens are more diverse and less stable, as indicated by the number of clusters, the distance between them as well as their composition.

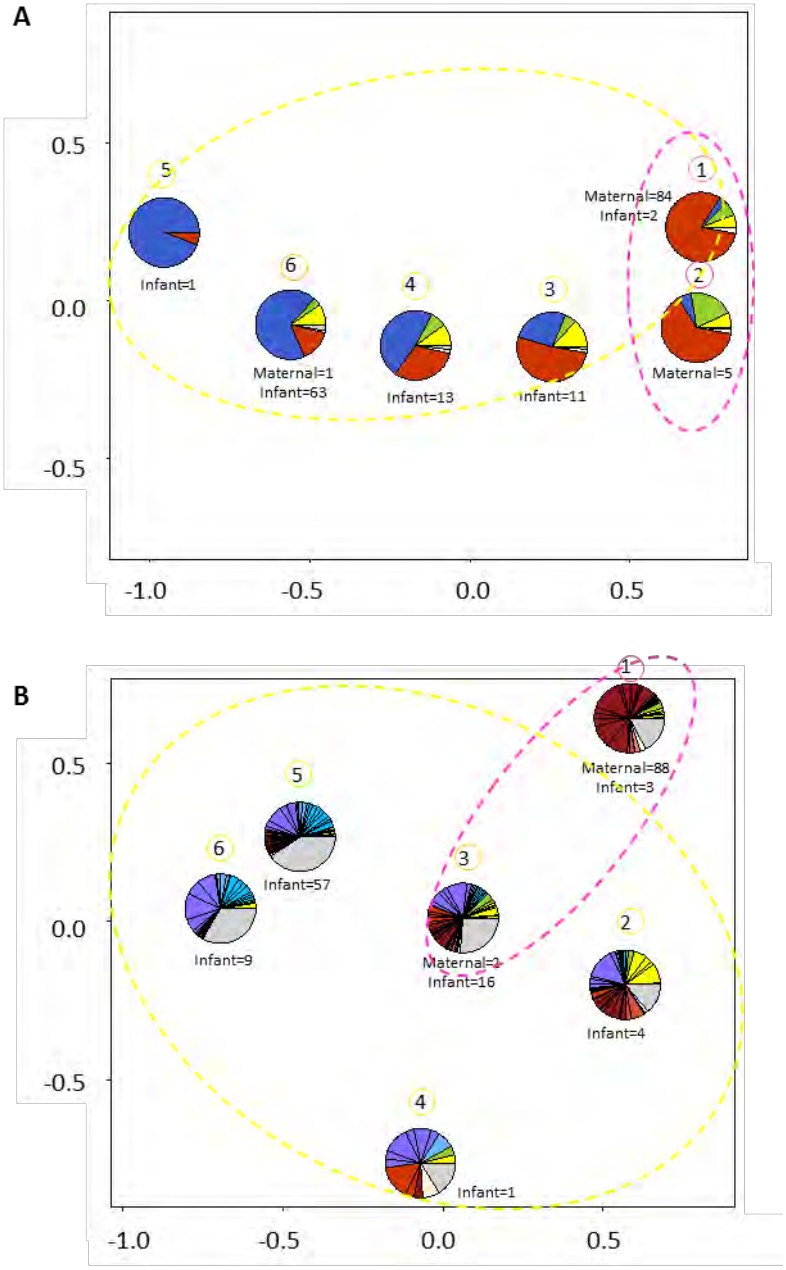


Figure 24. Multi-dimensional scaling (MDS) plots of infant meconium and maternal faecal specimens sampled at birth
A) MDS plots at phylum-level based on bacterial proportions > 0.5%; **B)** MDS plots at genus-level based on bacterial proportions > 0.5%.

3.3.3.4 *Comparing infant and maternal faecal specimens based on significant bacterial genera*

Log ratio biplots of genera with abundances > 0.5%, which were also statistically significant (at 5% significance) for infant (n = 107) and maternal (n = 90) faecal specimens, showed that infant and maternal specimens formed two distinct clusters (Figure 25). Maternal faecal specimens clustered tightly around genera within the phyla Firmicutes, Actinobacteria, Bacteroidetes and an unclassified phylum. Meconium specimens formed a less compact cluster, and clustered mainly around genera within the phylum Proteobacteria.

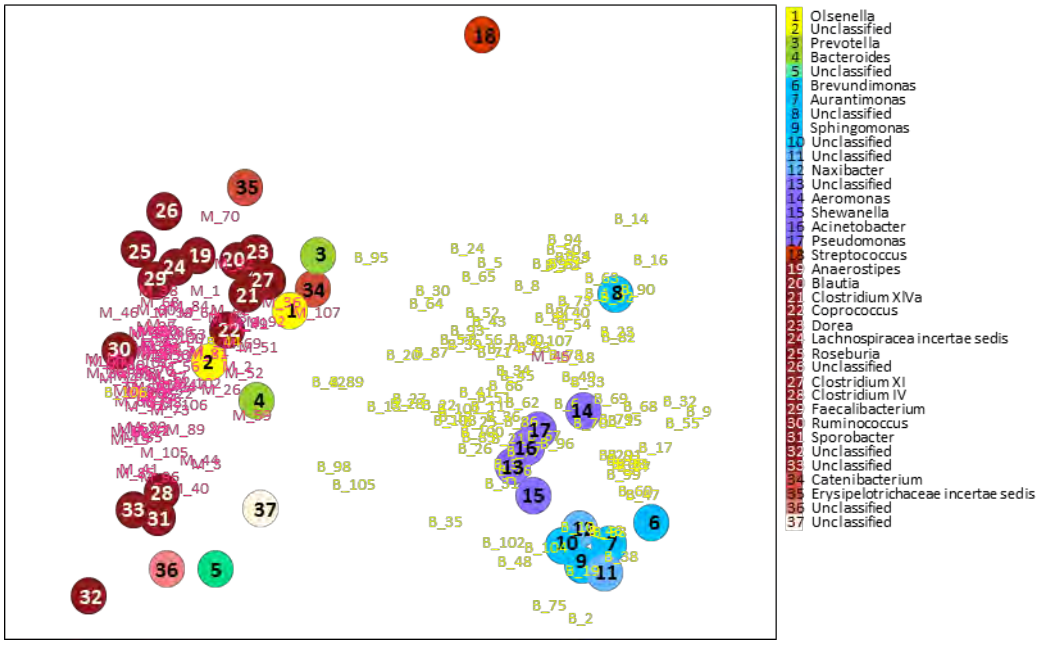


Figure 25. Log ratio biplot of infant meconium and maternal faecal specimens sampled at birth in relation to statistically significant genera

Meconium specimens are shown in yellow and maternal faecal specimens are shown in pink. Genera significantly associated with meconium and maternal faecal specimens range from 1 to 37 and their colours correspond to the class (and phyla) to which they belong.

3.3.3.5 *Significant differences in bacterial abundances from infant and maternal faecal specimens*

Table S6 summarises rate ratios for all bacterial OTUs which differed significantly (at 5% significance level) between infant (n = 107) and maternal (n = 90) faecal specimens. Meconium specimens sampled at birth had significantly higher proportions of the phyla Actinobacteria and Proteobacteria (rate ratios > 1.00). The maternal faecal specimens sampled at birth had significantly higher proportions of the phyla Firmicutes, Bacteroidetes as well as an unclassified phylum (rate ratios < 1.00). In total, 106 bacterial OTUs (ranging from phylum to genus-level) differed significantly in proportions when comparing meconium specimens with maternal faecal specimens sampled at birth.

3.3.3.6 *"Participant characteristics" influencing bacterial profiles in infant and maternal faecal specimens*

Table 9 summarises the association between participant characteristics (which included maternal education, residential area, maternal HIV status, infant's gestational age, gender, birth weight, birth length and mode of delivery) and meconium bacterial profiles (n = 107). None of the above-mentioned maternal characteristics had any significant association with maternal faecal (n = 90) bacterial taxa. In infants, only male gender and maternal education were significantly associated with meconium bacterial profiles. Gender was significantly associated with taxa from the class Clostridia. Females only had 60% of the proportions of the class Clostridia which were observed in males. This pattern was also observed down to genus-level for the families Lachnospiraceae and Ruminococcaceae (Table 9). The effect of maternal education on meconium bacterial profiles was clear for an unclassified family within the class Clostridia as well as an unclassified class within the phylum Firmicutes. We observed an increase in the proportions of both the unclassified family within the class Clostridia and the unclassified class within the phylum Firmicutes from meconium specimens as the maternal education levels increased from primary (Grades 1-7) to secondary (Grade 12) education level. A decrease in the bacterial proportions observed for the unclassified family within the class Clostridia and the unclassified class within the phylum Firmicutes was evident for infants with mothers having a tertiary eduction level. Finally, the intra- and inter-individual bacterial diversities of meconium specimens sampled at birth were not significantly associated with any of the participant characteristics tested (data not shown).

Table 9. Participant characteristics with significant effects on bacterial OTUs from meconium specimens (n = 107)

Taxon level	Bacterial OTUs	Gender (male: RR=1)		Maternal education (Grade 12: RR=1)			
				Grade 1-7	Grade 8-11	Tertiary education	
		Rate ratio	p-value	Rate ratio	Rate ratio	Rate ratio	p-value
Phylum	**Firmicutes**						
Class	Clostridia	0.6	0.0461	-	-	-	-
Order	Clostridiales	0.6	0.0461	-	-	-	-
Family	Lachnospiraceae	0.5	0.0079	-	-	-	-
Genus	Blautia	0.6	0.0369	-	-	-	-
Genus	Unclassified	0.3	0.0018	-	-	-	-
Family	Ruminococcaceae	0.4	0.0254	-	-	-	-
Genus	Faecalibacterium	0.4	0.0254	-	-	-	-
Genus	Ruminococcus	0.3	0.0183	-	-	-	-
Family	Unclassified	-	-	0.2	0.5	0.3	0.0400
Genus	Unclassified	-	-	0.2	0.5	0.3	0.0400
Class	Unclassified	-	-	0.2	0.5	0.1	0.0200
Order	Unclassified	-	-	0.2	0.5	0.1	0.0200
Family	Unclassified	-	-	0.2	0.5	0.1	0.0200
Genus	Unclassified	-	-	0.2	0.5	0.1	0.0200

OTU: Operational taxonomic unit

3.3.4 Comparison of microbial profiles from infant faecal specimens sampled at 4 to 12 weeks of age to those of maternal faecal specimens sampled at birth

The following section reports the relative abundance of bacterial OTUs classified from faecal specimens of 72 infants sampled at 4-12 weeks. We also provide data comparing 55 infant faecal specimens sampled at 4-12 weeks, to their matched maternal faecal specimens sampled at birth. Descriptions are provided at both phylum- and genus-level for bacteria with proportions greater than 0.5% and those with less than 0.5%.

3.3.4.1 *Microbiota profiles at the phylum-level*

The phylum Firmicutes (48.53%) was the most abundant phylum identified from infant faecal specimens collected at 4-12 weeks of age (n = 72); followed by the phylum Actinobacteria (26.27%), Proteobacteria (18.72%) and Bacteroidetes (5.42%). In total, 0.60% of the phyla with abundances > 0.5% were unclassifiable (Table S2). We observed a decrease in the proportions of phyla with abundances < 0.5% for infants at 4-12 weeks of age (0.45%) compared to infants sampled at birth (2.38%). Of the phyla with abundances < 0.5%; the phyla Verrucomicrobia (0.14%) and TM7 (0.14%) were most abundant, followed by the phylum Fusobacteria (0.06%) (Table S3). The 55 maternal faecal specimens (paired with the subset of infants studied at 4-12 weeks of age) had similar bacterial profiles at phylum-level when compared to the "main maternal cohort" of 90 mothers (Figures 22C-D and 26C-D).

The cluster analysis at phylum-level identified two principle clusters (clusters M and B) for the 55 mother-infant pairs based on OTUs with abundances > 0.5% (Figure 26A). Principle cluster M (sub-clusters no. 1 to 3) consisted of 36 infant faecal specimens sampled at 4-12 weeks of age, clustering with maternal faecal specimens collected at birth. Sub-cluster no. 1 mainly contained maternal faecal specimens, but also included two infant faecal specimens. Sub-clusters no. 2 and 3 consisted of 34 infant faecal specimens. Principle cluster B mainly grouped 19 infant faecal specimens into two sub-clusters (sub-clusters no. 4 and 6), while sub-cluster no. 5 contained a single maternal faecal specimen (M_45). This maternal faecal specimen, M_45, had high proportions of the phylum Proteobacteria and also clustered with infant meconium specimens studied at birth (Figures 22A and 23A). No interesting patterns, except for the sample type (maternal versus infant), were visible when plotting the participant characteristics against the order of the dendogram (Figures 26A and B).

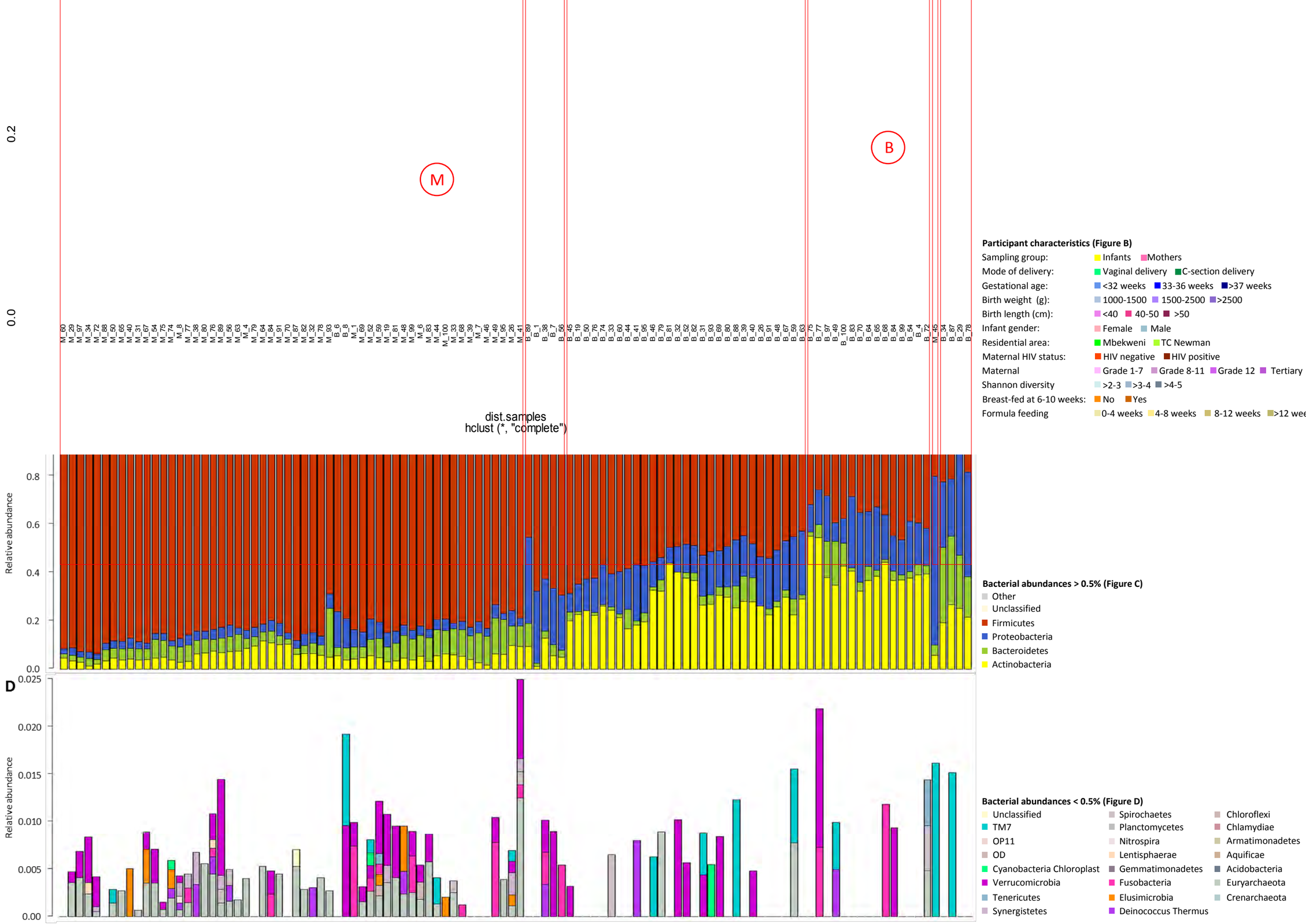


Figure 26. Phylum-level bacterial profiles of infant faecal specimens sampled at 4-12 weeks of age and maternal faecal specimens sampled at birth

A) Dendogram of phylum-level bacterial profiles for infant (n = 55) and maternal (n = 55) faecal specimens; **B)** Participant characteristics for each of the specimens under study; **C)** Relative abundance of bacteria with proportions > 0.5% at phylum-level; **D)** Relative abundance of bacteria with proportions < 0.5% at phylum-level.

3.3.4.2 *Microbiota profiles at the genus-level*

The genus Bifidobacterium (16.88%) was the most abundant in the infant faecal specimens (n = 72), followed by the genera Streptococcus (7.83%), Lactobacillus (5.73%), as well as unclassified genera from the families Enterobacteriaceae (3.84%) and Lachnospiraceae (3.24%). Overall, 15.25% of the phyla with abundances > 0.5% were unclassifiable (Table S4). Genera with abundances < 0.5% were reduced by half (17.61%) when compared to the proportions observed from meconium specimens (35.96%) (Table S5). The 55 maternal faecal specimens (paired with the subset of infants studied at 4-12 weeks of age) had similar bacterial profiles at the genus-level when compared to the "main maternal cohort" of 90 maternal faecal specimens (Figures 23C and 27C).

As at the phylum-level, two primary clusters (M and B) were observed for the 55 mother-infant pairs at genus-level when performing clustering analysis on OTUs with abundances > 0.5% (Figure 27 A). The main difference observed when compared to the phylum-level analysis is that at genus-level the principle cluster M mainly consisted of maternal faecal specimens (Figures 26A and 27A). In addition, primary cluster B (consisting of mainly infant faecal specimens) contained a single maternal faecal specimen (M_45) within sub-cluster no. 4 (Figure 27A). This maternal faecal specimen formed its own sub-cluster during phylum-level analysis at 4-12 weeks (Figure 26A). This maternal faecal specimen, M_45, also clustered with meconium specimens at both phylum- and genus-level. Once again the only pattern visible in relation to the clusters in the dendogram was the sample type analysed (maternal versus infant) (Figures 27A and B).

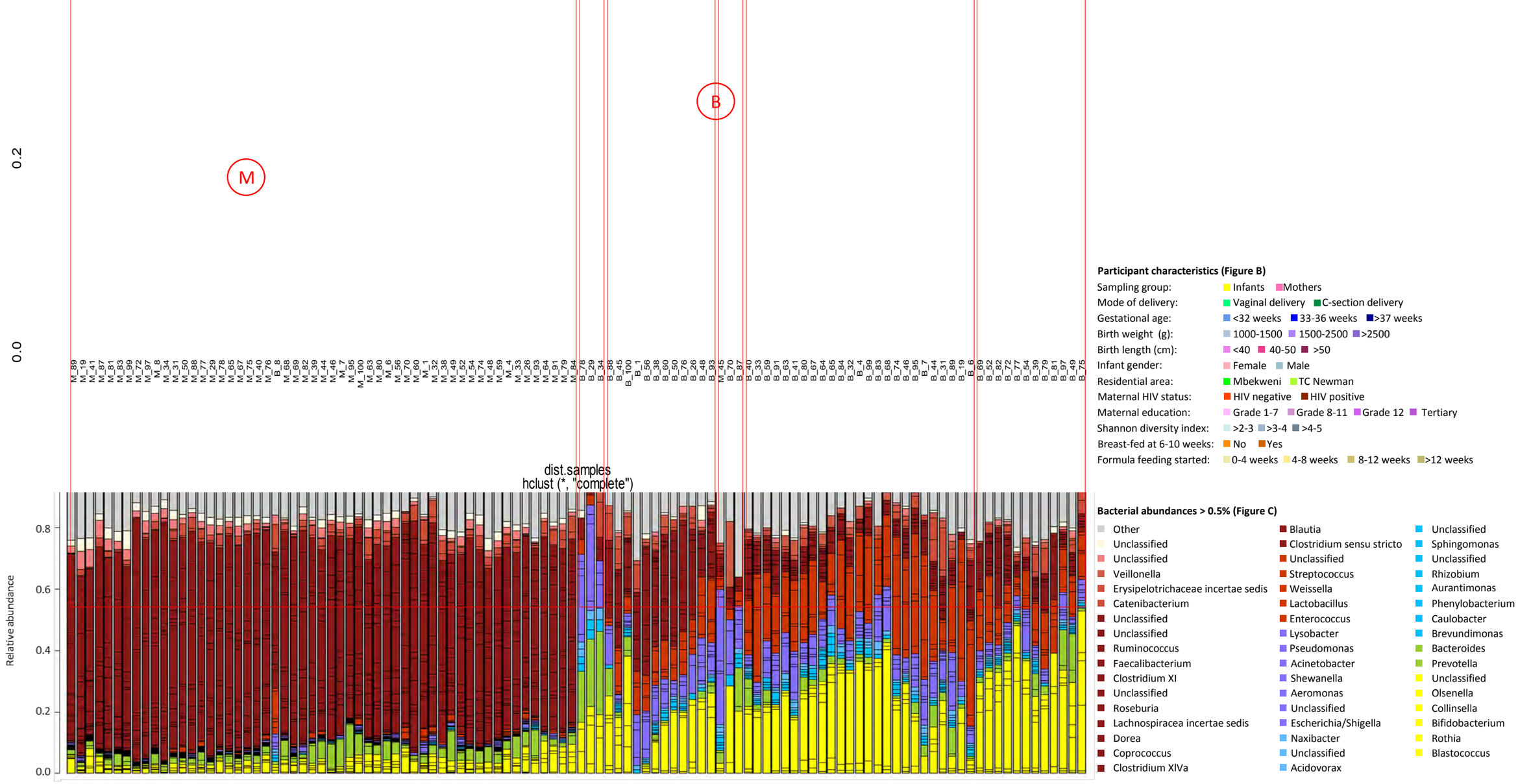


Figure 27. Genus-level bacterial profiles of infant faecal specimens sampled at 4-12 weeks of age and maternal faecal specimens sampled at birth
A) Dendogram of genus-level bacterial profiles for infant (n = 55) and maternal (n = 55) faecal specimens; **B)** Participant characteristics for each of the specimens under study; **C)** Relative abundance of bacteria with proportions > 0.5% at genus-level.

3.3.4.3 *Similarities between sub-clusters consisting of infant and maternal faecal specimen*

At both phylum- and genus-level, infant faecal specimens sampled at 4-12 weeks of age (n = 55) clustered at greater distances over bigger areas compared to their matched maternal faecal specimens (n = 55) (Figure 28). As observed for meconium specimens in this study (Figure 24), infant faecal specimens at 4-12 weeks are more diverse and less stable compared to maternal faecal specimens (Figure 28). This is clearly indicated by the number of sub-clusters, the area in which they are plotted, and their composition.

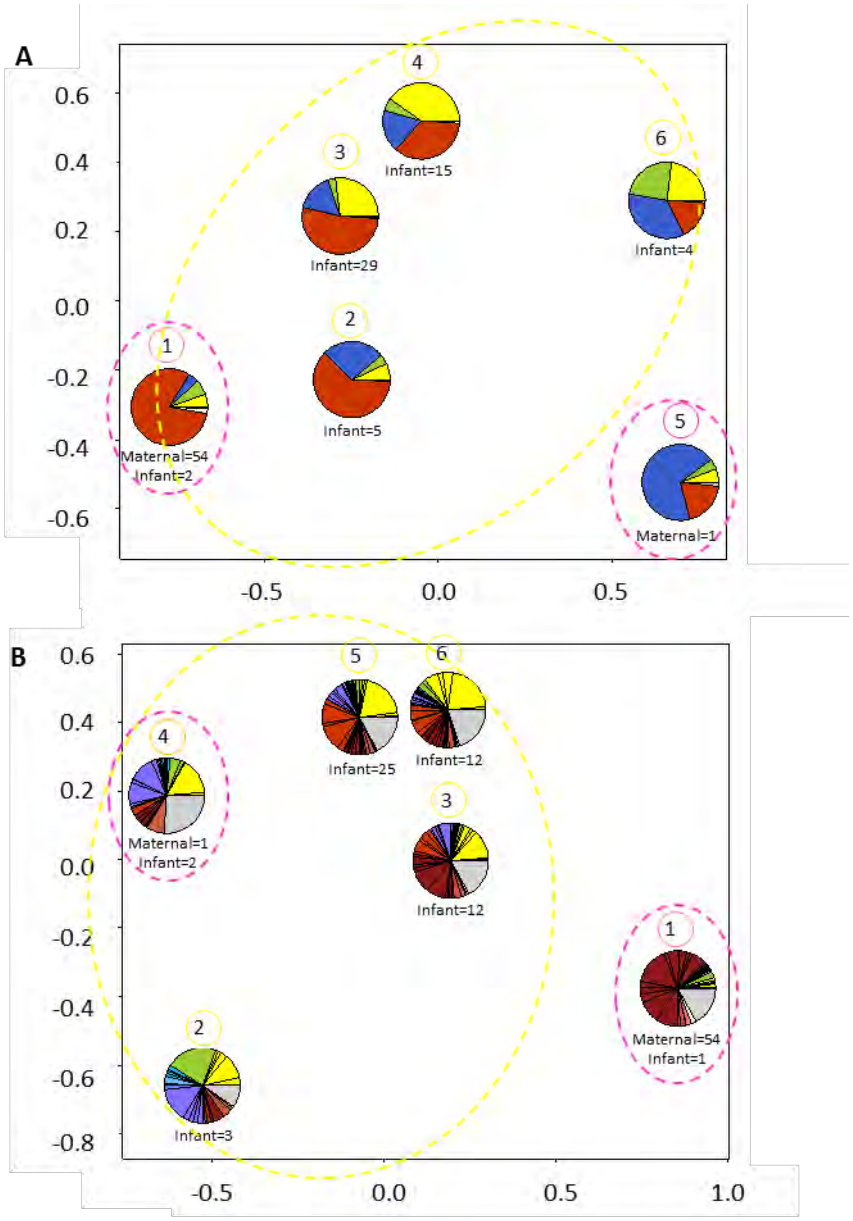


Figure 28. Multi-dimensional scaling (MDS) plots of infant faecal specimens sampled at 4-12 weeks of age and maternal faecal specimens sampled at birth
A) MDS plots at phylum-level based on bacterial proportions > 0.5%; **B)** MDS plots at genus-level based on bacterial proportions > 0.5%.

3.3.4.4 *Comparing infant and maternal faecal specimens based on significant bacterial genera*

Log ratio biplots of statistically significant genera (at 5% significance) showed that infant faecal specimens sampled at 4-12 weeks of age (n = 72) and maternal faecal specimens sampled at birth (n = 90) formed two distinct clusters (Figure 29). A notable difference for this biplot, compared to the one for specimens sampled at birth, is the reduced number of statistically different genera between maternal and infant faecal specimens. The number of genera that differed statistically between infants and mothers was reduced from 37 (Figure 25) to 28 (Figure 29). As with meconium specimens (Figure 25), infant faecal specimens sampled at 4-12 weeks of age formed a less compact cluster when compared to maternal faecal specimens and clustered around genera from the phylum Proteobacteria (Figure 29). In addition, these infant faecal specimens now also clustered more closely around genera from the phylum Actinobacteria.

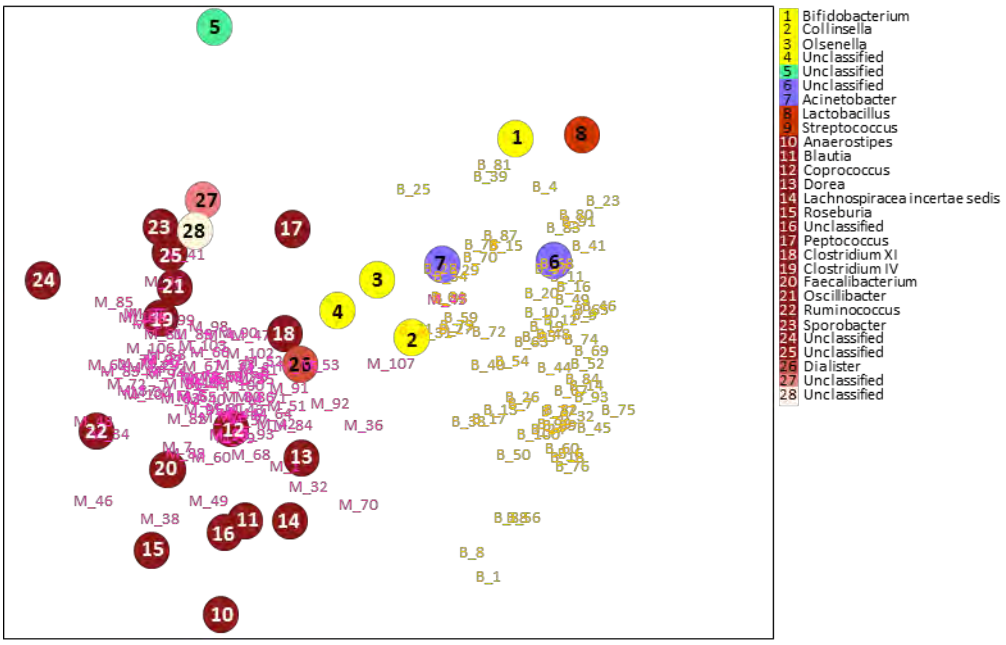


Figure 29. Log ratio biplot of infant faecal specimens sampled at 4-12 weeks of age and maternal faecal specimens sampled at birth in relation to statistically significant genera
Infant specimens are shown in light orange and maternal faecal specimens are shown in pink. Genera significantly associated with infant and maternal faecal specimens range from 1 to 28 and their colours correspond to the class (and phyla) to which they belong.

3.3.4.5 *Significant differences in bacterial abundances from infant and maternal faecal specimens*

Table S7 summarises the rate ratios for all bacterial OTUs which differed significantly (at 5% significance) between infant (n =72) and maternal faecal specimens (n = 90). Our results clearly showed that infant faecal specimens sampled at 4-12 weeks of age had statistically significantly higher proportions of the phyla Actinobacteria and Proteobacteria (rate ratios > 1.00). In total, 74

bacterial OTUs differed significantly in their proportions, as opposed to the 106 differing OTUs for mother-infant pairs sampled at birth.

3.3.4.6 *"Participant characteristics" influencing bacterial profiles in infant faecal specimens sampled at 4-12 weeks of age*

Table 10 summarises the association of participant characteristics (tested for meconium specimens with the addition of breastfeeding at 6-10 weeks and the effect of formula feeding) with bacterial profiles from infant faecal specimens collected at 4-12 weeks of age (n = 72). We found that gestational age was significantly associated with taxa from the order Coriobacteriales within the phylum Actinobacteria. A 20% reduction was seen for the order Coriobacteriales for every one week increase in gestational age. This pattern was evident down to the genus-level as shown for the genus Collinsella (Table 10). We also observed that HIV-unexposed infants only had 50% of the proportions of the genus Leuconostoc when compared to infants exposed to HIV. In addition, we observed a significant difference in the proportions of the phylum Actinobacteria for infants according to the residential areas. Infants residing in Mbekweni only had 20% of the proportions of the phylum Actinobacteria when compared to infants from TC Newman. This significant difference was clear down to the genus-level, as shown for the genus Bifidobacterium (Table 10). Conversely, at the phylum-level, infants from Mbekweni had 120% of the bacterial proportions for the phylum Firmicutes when compared to the bacterial proportions from infants residing in TC Newman. With regards to the feeding practices, infants breastfed at 6-10 weeks of age had 160% of the bacterial proportions belonging to the class Actinobacteria, compared to the proportions measured from infants not being breastfed. Moreover, breastfed infants also only had 50% of the bacterial proportions belonging to the class Clostridia, when compared to those not receiving breastfeeding.

Among the characteristics tested against the intra- and inter-individual bacterial diversities observed at 4-12 weeks of age, only maternal HIV status was significantly associated with both intra- ($p = 0.0073$) and inter- ($p = 0.0002$) individual bacterial diversity measures. HIV-unexposed infants only had around 80% of the intra-individual bacterial diversity found in HIV-exposed infants at 4-12 weeks of age (RR = 0.78). For inter-individual bacterial diversity, HIV-unexposed infants had even less diverse bacterial profiles with approximately 1% of the diversity found in infants exposed to HIV (RR = 0.01).

Table 10. Participant characteristics with significant effects on bacterial OTUs from infant faecal specimens at 4-12 weeks of life (n =72)

Taxon level	Taxa	Gestational age (Weekly increase in age†)		Maternal HIV status (HIV positive: RR=1.00)		Residential area (TC Newman: RR=1.00)		Breastfeeding (Not breastfed at 6-10 weeks: RR=1.00)	
		Rate ratio	p-value	Rate ratio	p-value	Rate ratio	p-value	Rate ratio	p-value
Phylum	**Actinobacteria**	**-**	**-**	**-**	**-**	**0.8**	**0.0183**	**1.6**	**0.0351**
Class	Actinobacteria	-	-	-	-	0.8	0.0183	1.6	0.0351
Order	Bifidobacteriales	-	-	-	-	0.8	0.0263	-	-
Family	Bifidobacteriaceae	-	-	-	-	0.8	0.0263	-	-
Genus	Bifidobacterium	-	-	-	-	0.8	0.0255	-	-
Order	Coriobacteriales	0.8	0.0309	-	-	-	-	-	-
Family	Coriobacteriaceae	0.8	0.0309	-	-	-	-	-	-
Genus	Collinsella	0.8	0.0309	-	-	-	-	-	-
Phylum	**Firmicutes**	**-**	**-**	**-**	**-**	**1.2**	**0.0340**	**0.8**	**0.0351**
Genus	Leuconostoc	-	-	0.5	0.0309	-	-	-	-
Class	Clostridia	-	-	-	-	-	-	0.5	0.0076
Order	Clostridiales	-	-	-	-	-	-	0.5	0.0076
Family	Clostridiaceae	-	-	-	-	-	-	0.3	0.0488
Genus	Clostridium sensu stricto	-	-	-	-	-	-	0.3	0.0464
Family	Lachnospiraceae	-	-	-	-	-	-	0.5	0.0256

† Rate ratios increase or decrease at a constant rate for every one week increase in gestational age

3.3.5 Comparison of microbial profiles from infant faecal specimens sampled at 20 to 28 weeks of age with those of maternal faecal sampled at birth

The following section reports the relative abundance of bacterial OTUs classified from faecal specimens of 36 infants sampled at 20-28 weeks. We also provide data comparing 28 infant faecal specimens sampled at 20-28 weeks, to their matched maternal faecal specimens sampled at birth. Descriptions are provided at both phylum- and genus-level for bacteria with proportions greater than 0.5% and those with less than 0.5%.

3.3.5.1 *Microbiota profiles at the phylum-level*

The most abundant phylum identified for infant faecal specimens collected at 20-28 weeks of age (n = 36) was the phylum Firmicutes (63.72%), followed by the phylum Actinobacteria (19.29%), Proteobacteria (11.76%), Bacteroidetes (4.38%). Of the phyla with abundances > 0.5%; 0.46% were unclassifiable (Table S2). Phyla with abundances < 0.5% were similar to those from infants at 4-12 weeks of age and made up 0.39% of all bacteria within infant faecal specimens. Of these, the phylum Verrucomicrobia (0.29%) was most abundant, followed by the phylum TM7 (0.0.03%) (Table S3). Similar phylum-level bacterial profiles were observed for the 28 maternal faecal specimens (used to match with infant faecal specimens), when compared to the bacterial profiles of the "main maternal cohort" of 90 mothers (Figures 22C-D and 30C-D). The only exception for this subset of 28 maternal faecal specimens was that the phylum Proteobacteria was slightly more abundant than the phyla Bacteroidetes and Actinobacteria.

With the exception of a single maternal faecal specimen (M_45), containing high proportions of the phylum Proteobacteria, we observed two primary clusters (clusters M and B) for the 28 mother-infant pair specimens based on their OTUs with abundances > 0.5% (Figure 30A). Primary cluster M contained maternal faecal specimens, together with seven infant faecal specimens. Four of these seven infant faecal specimens (B_1, B_6, B_8, B_40) also clustered with maternal faecal specimens at phylum-level during analysis of infant faecal specimens at 4-12 weeks of age (Figure 26A). The remainder of the infant faecal specimens sampled at 20-28 weeks of age, clustered in primary cluster B which consisted of sub-clusters no. 2 to 5. As shown for all prior analyses, the only participant characteristics which grouped with clusters from the dendogram was the sample types analysed (maternal versus infant) (Figures 30A and B).

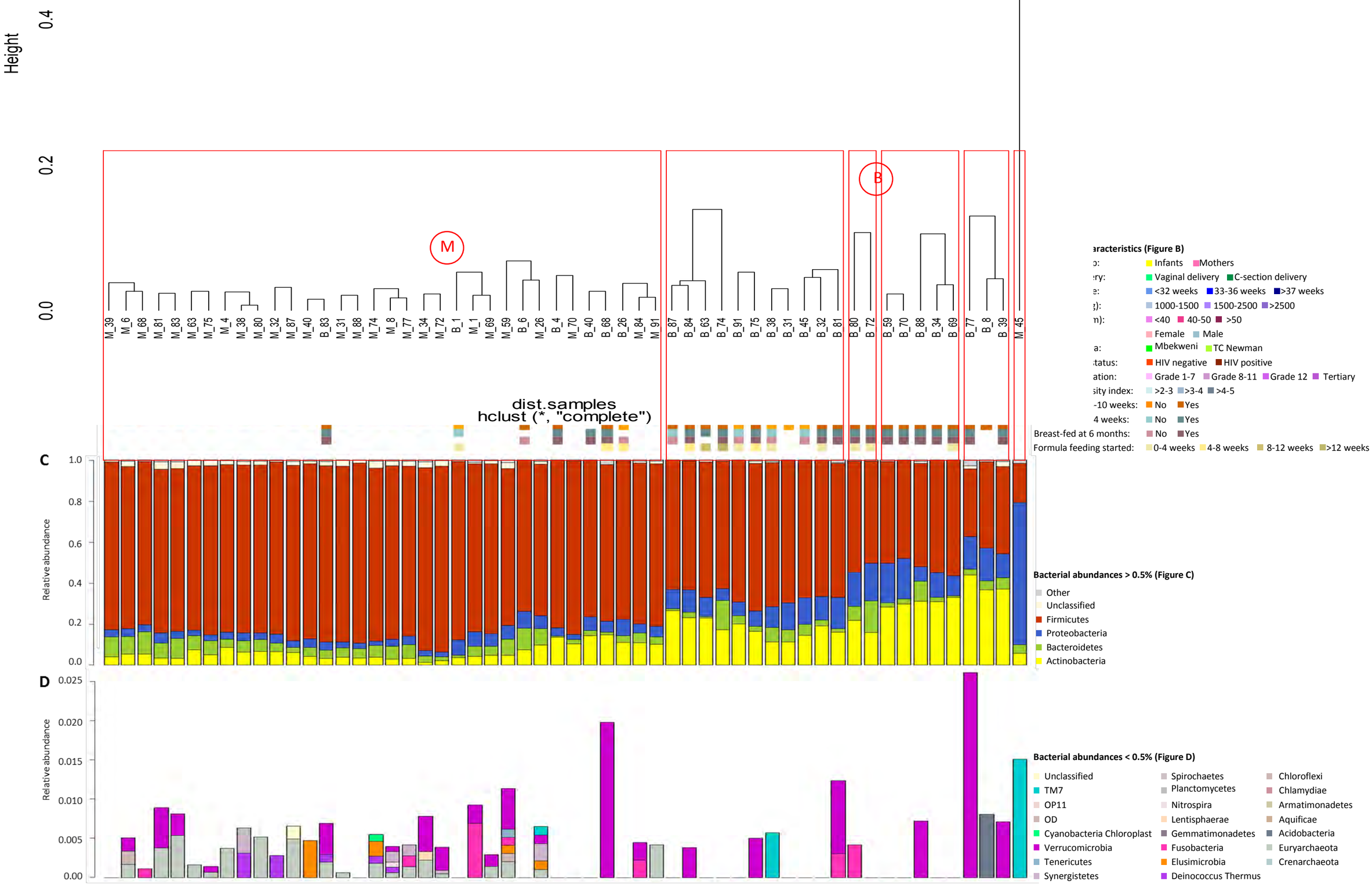


Figure 30. Phylum-level bacterial profiles of infant faecal specimens sampled at 20-28 weeks of age and maternal faecal specimens sampled at birth

A) Dendogram of phylum-level bacterial profiles for infant (n = 28) and maternal (n = 28) faecal specimens; **B)** Participant characteristics for each of the specimens under study; **C)** Relative abundance of bacteria with proportions > 0.5% at phylum-level; **D)** Relative abundance of bacteria with proportions < 0.5% at phylum-level.

3.3.5.2 *Microbiota profiles at the genus-level*

The genus Bifidobacterium (12.66%) was the most abundant in infant faecal specimens (n = 36), followed by an unclassified genus within the family Lachnospiraceaea (11.10%), as well as the genera Lachnospiraceae incertae sedis (6.67%), Streptococcus (5.06%) and Dorea (3.76%). Of the genera with abundances > 0.5%; 20.66% of the taxa were unclassifiable (Table S4). Genera with abundances < 0.5% were very similar to those found in faecal specimens sampled at 4-12 weeks of age and represented 17.65% of all bacteria within infant faecal specimens collected at 20-28 weeks of life (Table S5). The 28 maternal faecal specimens (paired with the subset of infants analysed at 20-28 weeks of age) had similar bacterial profiles at the genus-level compared to the "main maternal cohort" of 90 mothers (Figures 23C and 31C).

We observed two primary clusters (clusters M and B) of the 28 mother-infant pair specimens based on clustering of OTUs with abundances > 0.5% at genus-level (Figure 31A). Cluster M comprised of three sub-clusters which included 14/28 infant faecal specimens together with 35/28 maternal faecal specimens. Nine and six of these infant faecal specimens also clustered with maternal faecal specimens at phylum-level during analyses performed at 4-12 weeks of age (Figure 26A) and 20-28 weeks of age (Figure 30A), respectively. Cluster B (sub-clusters no. 4 to 6) consisted of the other 14 infant faecal specimens and a single maternal faecal specimen. Sub-cluster no. 6 contained the maternal faecal specimen (M_45) which clustered with infant faecal specimens for all clustering analyses performed in this study (Figures 22, 23, 26, 27, 30 and 31). With the exception of only the sample type analysed (maternal versus infant); no patterns were observed when plotting the participant characteristics against the dendogram constructed (Figures 31A and B).

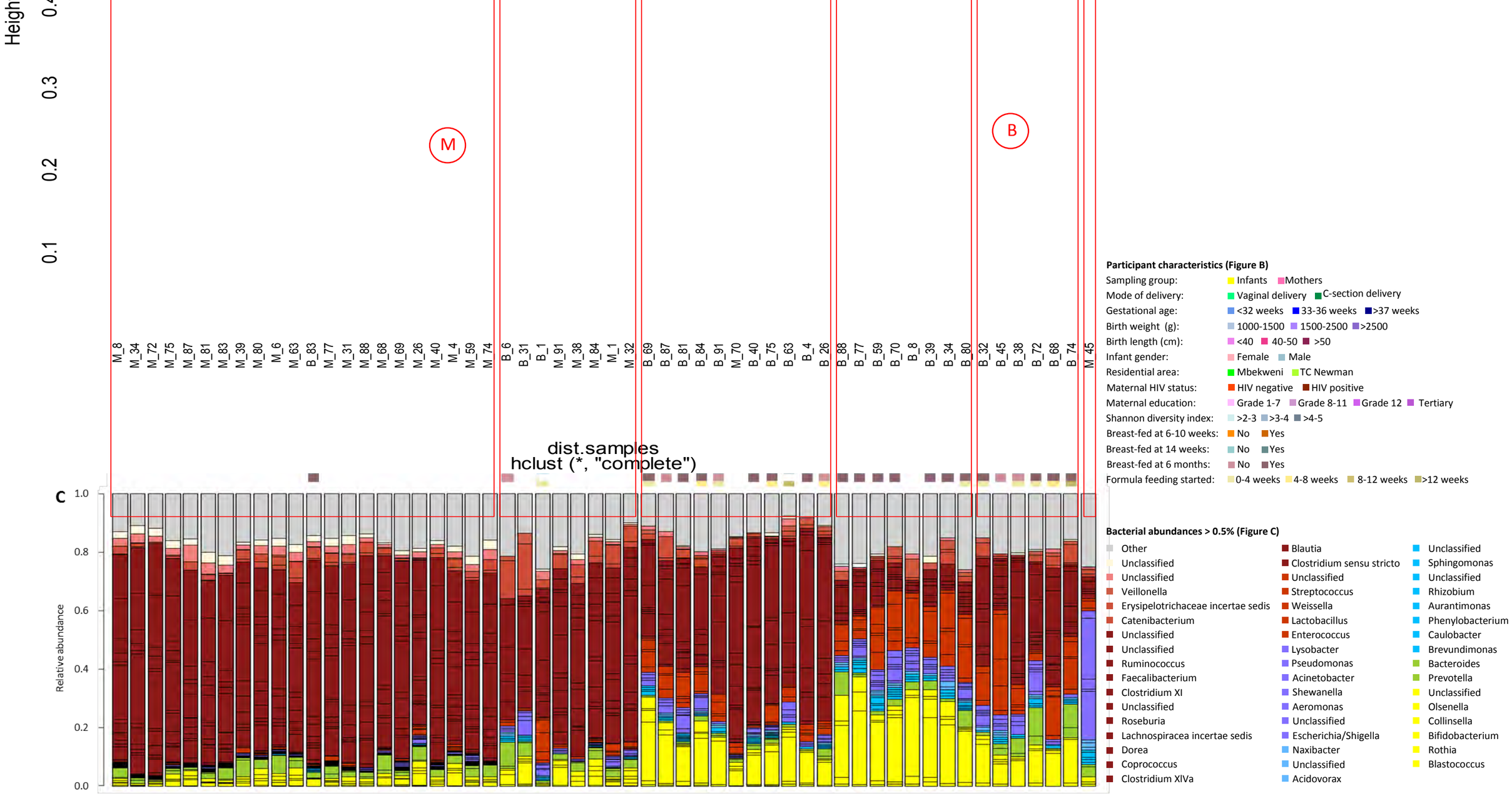


Figure 31. Genus-level bacterial profiles of infant faecal specimens sampled at 20-28 weeks of age and maternal faecal specimens sampled at birth
A) Dendogram of genus-level bacterial profiles for infant (n = 28) and maternal (n = 28) faecal specimens; **B)** Participant characteristics for each of the specimens under study; **C)** Relative abundance of bacteria with proportions > 0.5% at genus-level.

3.3.5.3 *Similarities between sub-clusters consisting of infant and maternal faecal specimen*

We observed that infant faecal specimens clustered at greater distances over a bigger area compared to maternal faecal specimens at both phylum- and genus-level when comparing 28 mother-infant pairs (Figure 28). As observed for meconium specimens and infant faecal specimens sampled at 4-12 weeks of age (Figures 24 and 28); infant faecal specimens collected at 20-28 weeks are more diverse and less stable (Figure 32). We did observe that the distances between infant faecal specimen sub-clusters sampled at 20-28 weeks are smaller compared to the sub-clusters observed for specimens collected earlier in life. At genus-level more sub-clusters consisted of both infant and maternal faecal specimens and distinction between the two types of faecal specimens is less clear.

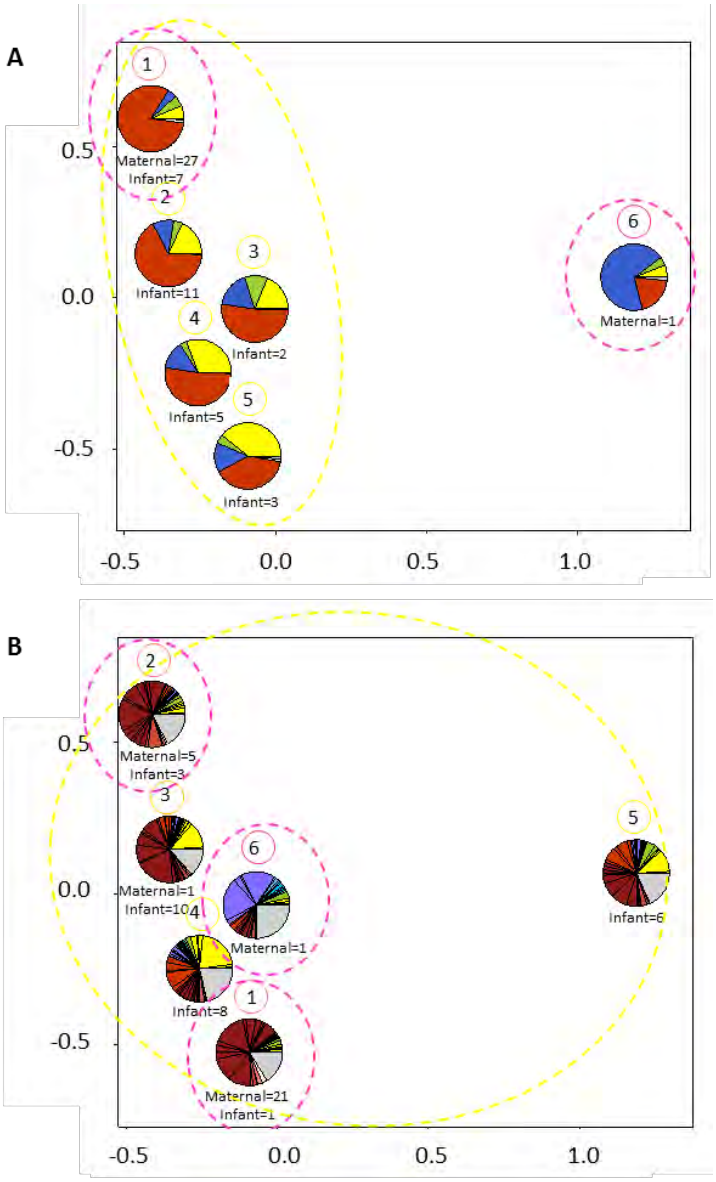


Figure 32. Multi-dimensional scaling (MDS) plots of infant faecal specimens sampled at 20-28 weeks of age and maternal faecal specimens sampled at birth

A) MDS plots at phylum-level based on bacterial proportions > 0.5%; **B)** MDS plots at genus-level based on bacterial proportions > 0.5%.

3.3.5.4 *Comparing infant and maternal faecal specimens based on significant bacterial genera*

Analysis at 20-28 weeks of age still showed two distinct clusters when constructing log-ratio biplots of statistically significant genera for infant (n = 36) and maternal (n = 90) faecal specimens. Of note, we observed a reduction in the number of genera which differed significantly between infant and maternal faecal specimens, compared to the analysis at 4-12 weeks of age (Figure 29). Infant faecal specimens sampled at 20-28 weeks of age formed a less compact cluster compared to maternal faecal specimens. This pattern was also observed for meconium and faecal specimens sampled earlier in life. In contrast to earlier specimens (and in agreement with maternal faecal specimens); infant faecal specimens sampled at 20-28 weeks of age clustered around genera from the phylum Firmicutes (Figure 33). Infant faecal specimens also clustered around genera from the phylum Actinobacteria (Figure 33), which was previously seen for specimens collected at 4-12 weeks of age (Figure 29).

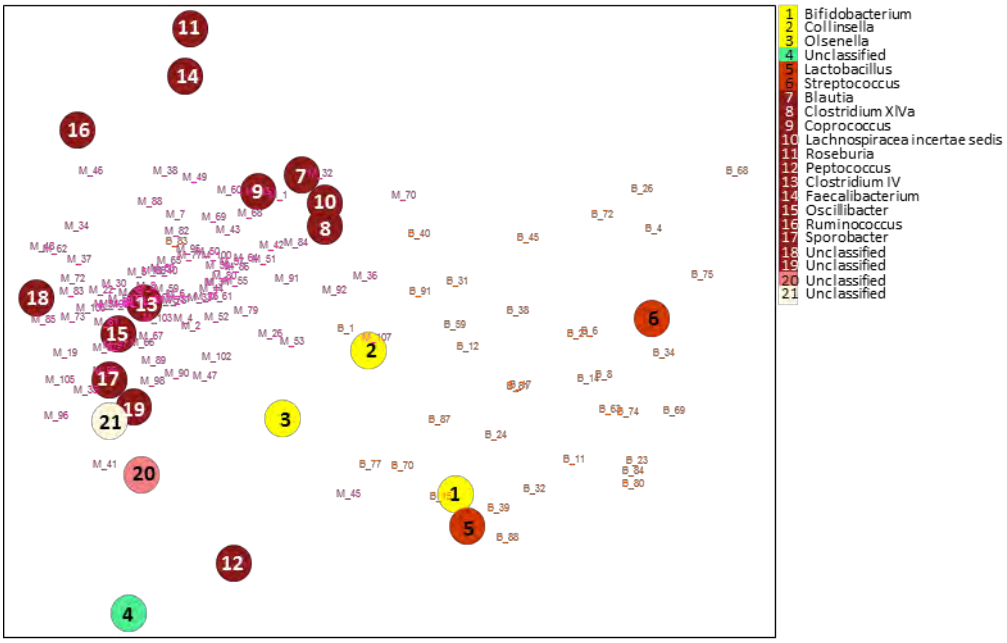


Figure 33. Log ratio biplot of infant faecal specimens sampled at 20-28 weeks of age and maternal faecal specimens sampled at birth in relation to statistically significant genera
Infant specimens are shown in light orange and maternal faecal specimens are shown in pink. Genera significantly associated with infant and maternal faecal specimens range from 1 to 21 and their colours correspond to the class (and phyla) to which they belong.

3.3.5.5 *Significant differences in bacterial abundances from infant and maternal faecal specimens*

Rate ratios for bacterial OTUs which differed significantly (at 5% significance) between infant (n = 36) and maternal (n = 90) faecal specimens are summarised in Table S8. As with infants studied at 4-12 weeks of age, infants studied at 20-28 weeks also had statistically significantly higher proportions of the phyla Actinobacteria and Proteobacteria (rate ratios > 1.00). In total, 60 bacterial

OTUs differed significantly in proportions as opposed to 74 statistically different OTUs for mother-infant pairs studied at 4-12 weeks of age.

3.3.5.6 *"Participant characteristics" influencing bacterial profiles in infant faecal specimens sampled at 20-28 weeks of age*

The association of participant characteristics (tested for infants at 4-12 weeks with the addition of breastfeeding at 14 weeks and six months of age) with bacterial profiles from infant faecal specimens collected at 20-28 weeks of age (n = 36) are summarised in Table 11. We found that for every one kilogram increase in birth weight, a significant increase of 20% was evident for the phylum Proteobacteria. Conversely; a significant decrease of 10% was observed for the phylum Firmicutes with every one kilogram increase in birth weight. Increased proportions of bacterial taxa were also seen for lower taxonomic levels from the classes Betaproteobacteria and Gammaproteobacteria; while decreased proportions for lower taxonomic levels from the class Clostridia were also evident. Females had 140% of the proportions of the family Leuconostocaceae observed in males. These high proportions were observed for females down to genus-level, shown by the high proportions of the genera Leuconostoc and Weissella when compared to males (Table 11). We also observed that females had almost four times greater proportions of OTUs representing the genus Lactococcus from the family Streptococcaceae when compared to males. Conversely, females only had had 30% of the proportions of the family Lactobacillaceae observed in males. Caesarean section delivered infants only had 20% of the proportions of the genus Lactococcus when compared to vaginal delivered infants. Infants born via caesarean section delivery had more than four times greater proportions of the genus Faecalibacterium compared to vaginal delivered infants. Infants receiving breastfeeding at 6-10 weeks of life had three times higher proportions of the order Bifidobacteriales at 20-28 weeks when compared to infants not receiving breastfeeding. This pattern was evident down to genus-level as shown for the genus Bifidobacterium (Table 11). Infants receiving breastfeeding only had 20% of the proportions of the genera Leuconostoc and Weissella; 30% of the proportions of the genus Lactococcus; and 40% of the proportions of the genus Clostridium XIVa at 20-28 weeks of life when compared to those not receiving breastfeeding.

None of the participant characteristics tested against intra- and inter-individual bacterial diversities observed at 20-28 weeks of age showed any significant associations (data not shown).

Table 11. Participant characteristics with significant effects on bacterial OTUs from infant faecal specimens at 20-28 weeks of life (n =36)

Taxon level	Bacterial OTUs	Birth weight (Per kilogram increase in weight†)		Gender(Male: RR=1.00)		Mode of delivery (Vaginal delivery: RR=1.00)		Breastfeeding (Not breastfed at 6-10 weeks: RR=1.00)	
		Rate ratio	p-value	Rate ratio	p-value	Rate ratio	p-value	Rate ratio	p-value
Phylum	**Actinobacteria**								
Order	Bifidobacteriales	-	-	-	-	-	-	2.5	0.0307
Family	Bifidobacteriaceae	-	-	-	-	-	-	2.5	0.0307
Genus	Bifidobacterium	-	-	-	-	-	-	2.5	0.0307
Phylum	**Proteobacteria**	**1.2**	**0.0085**	**-**	**-**	**-**	**-**		
Class	Betaproteobacteria	1.6	0.0485	-	-	-	-		
Order	Burkholderiales	1.6	0.0485	-	-	-	-		
Class	Gammaproteobacteria	1.1	0.0485	-	-	-	-		
Phylum	**Firmicutes**	**0.9**	**0.0087**	**-**	**-**	**-**	**-**	**-**	**-**
Family	Lactobacillaceae	-	-	0.3	0.0485	-	-	-	-
Family	Leuconostocaceae	-	-	1.4	0.0076	-	-	0.2	<0.0001
Genus	Leuconostoc	-	-	2.2	0.0223	-	-	0.2	0.0485
Genus	Weissella	-	-	1.2	0.0485	-	-	0.2	<0.0001
Family	Streptococcaceae								
Genus	Lactococcus	-	-	3.7	0.0019	0.2	0.0485	0.3	0.0485
Class	Clostridia								
Family	Lachnospiraceae								
Genus	Clostridium XIVa	-	-	-	-	-	-	0.4	0.0380
Family	Ruminococcaceae	0.8	<0.0001	-	-	-	-	-	-
Genus	Faecalibacterium	0.6	<0.0001	-	-	4.3	0.0485	-	-
Genus	Ruminococcus	0.1	<0.0001	-	-	-	-	-	-
Genus	Unclassified	0.5	<0.0001	-	-	-	-	-	-
Family	Unclassified	0.9	<0.0001	-	-	-	-	-	-
Genus	Unclassified	0.9	<0.0001	-	-	-	-	-	-

† Rate ratios increase or decrease at a constant rate for every one kilogram increase in infant birth weight

3.3.6 Dynamics of infant faecal bacterial profiles from birth throughout the first seven months of life

3.3.6.1 *Changes in infant faecal bacterial proportions over time*

At birth (Figure 34A), meconium specimens (n = 107) had a high abundance of the phylum Proteobacteria (60%), followed by Firmicutes (23%). As infants grew older, the phylum Proteobacteria decreased dramatically by approximately 40% around 4-12 weeks (n = 72) (Figure 34B) and a further 7% by 20-28 weeks of age (n = 36) (Figure 34C). The phylum Firmicutes almost doubled in abundance by 4-12 weeks (Figure 34 B) and increased even further by 15% at 20-28 weeks of age (Figure 34C). The phylum Actinobacteria was at its highest abundance (26%) around 4-12 weeks (Figure 34B), and 20-28 weeks (19%) (Figure 34C). The phylum Bacteroidetes remained relatively stable over time (Figures 34 A, B and C).

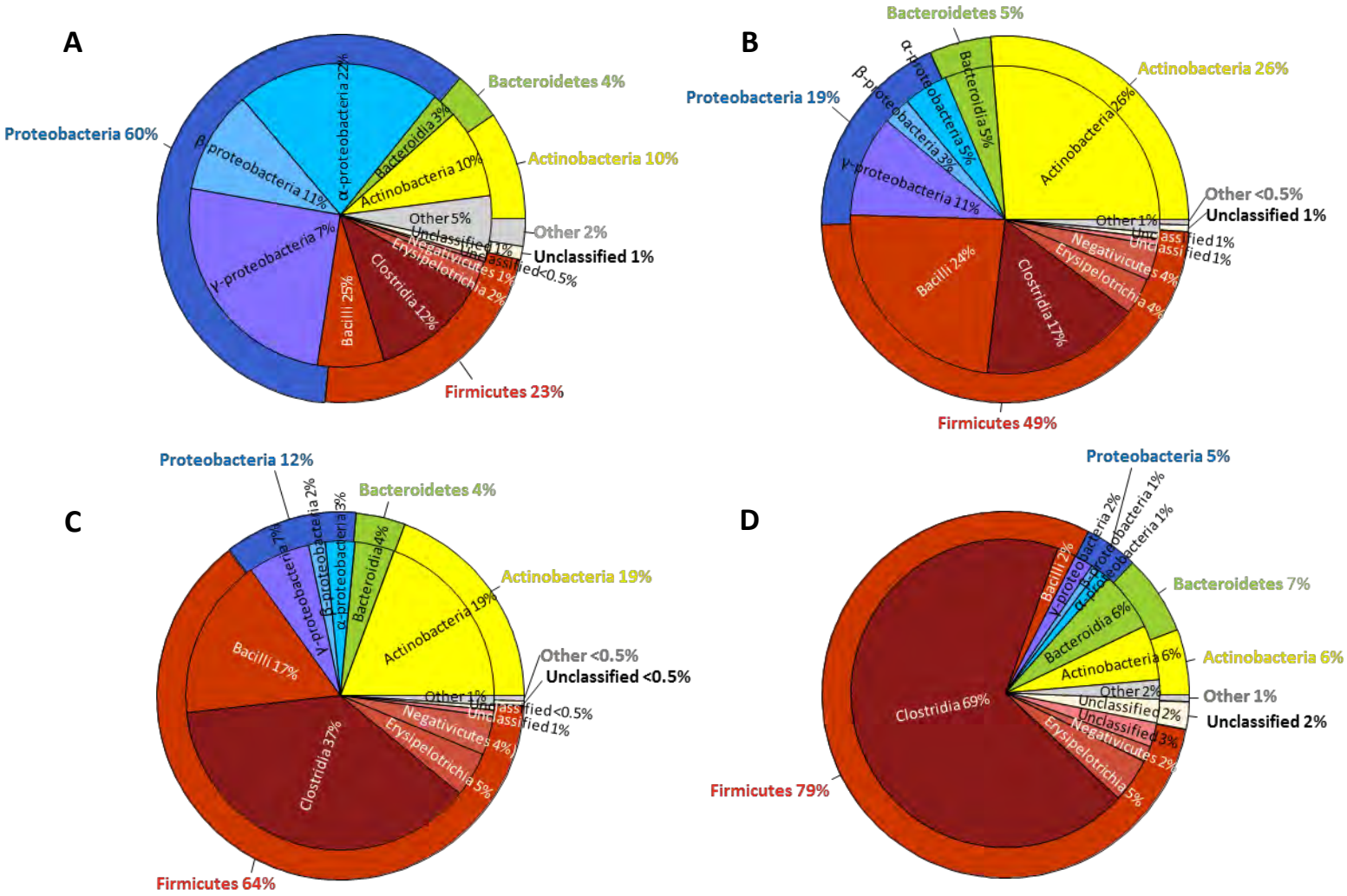


Figure 34. Pie charts summarising the changes in bacterial proportions for infant meconium and faecal specimens over time

The outer circles represent bacterial proportions at phylum-level while the inner circle represents class-level. **A)** Bacterial proportions from meconium specimens sampled at birth (n = 107); **B)** from infant faecal specimens sampled at 4-12 weeks of age (n = 72); **C)** from infant faecal specimens sampled at 20-28 weeks of age (n =36) ; and **D)** from maternal faecal specimens sampled at birth (n = 90).

3.3.6.2 *The significance of change over time on bacterial OTUs from infant faecal specimens*

When testing the effect of a weekly increase in infant age on the infant's faecal (n = 215) bacterial OTUs with proportions > 0.5%, we found significant changes in proportions of 87 OTUs (Table S9). At phylum-level, we observed a significant ($p < 0.0001$) decrease in the phylum Proteobacteria of approximately 10% for every one week increase in infant age. The phylum Firmicutes significantly ($p < 0.0001$) increased by 4% with every one week increase in infant age and the phylum Actinobacteria significantly ($p = 0.0274$) increased by 2% with every one week increase in infant age. A significant ($p = 0.0274$) reduction of 2% was observed for the unclassified phylum with every one week increase in infant age. No significant differences were observed for the phylum Bacteroidetes over time.

3.3.6.3 *Significant genera representing infant faecal specimens over time*

Log ratio biplots of statistically different genera for meconium and infant faecal specimens sampled shows a distinct profile for meconium (n = 90) when compared to infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of life (Figure 35). The meconium specimens (shown in yellow in Figure 35) formed a distinct cluster around genera from the phylum Proteobacteria. The other prominent cluster comprised of infant faecal specimens sampled at 4-12 and 20-28 weeks of age. The faecal specimens sampled at 4-12 weeks (Figure 35; light orange) mainly clustered around genera from the phylum Firmicutes and Actinobacteria, while infant faecal specimens sampled at 20-28 weeks (Figure 35; dark orange) mainly clustered around genera from the phylum Firmicutes.

.

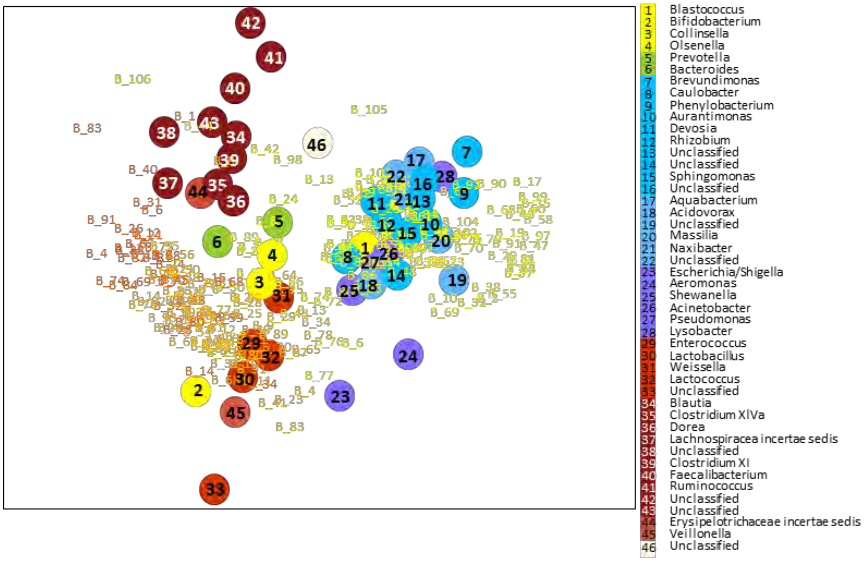


Figure 35. Faecal specimens sampled at birth, 4-12 and 20-28 weeks of age in relation to statistically significant genera
Infant meconium specimens are shown yellow, infant faecal specimens sampled at 4-12 weeks of age are shown in light orange and faecal specimens sampled at 20-28 weeks of age are shown in dark orange. Genera significantly associated with infant meconium and faecal specimens range from 1 to 46 and their colours correspond to the class (and phyla) to which they belong.

3.3.6.4 Relating longitudinal infant faecal bacterial profiles to maternal faecal bacterial profiles

Figure 34 showed the meconium and faecal bacterial profiles in comparison to the maternal faecal specimens. It is clear that infant bacterial profiles became more like the maternal profile as the infants grew older (Figure 34). At phylum-level, the phylum Proteobacteria was significantly ($p < 0.0001$) decreased from 60% at birth to 12% at 20-28 weeks of life. The proportion of Proteobacteria found within maternal faecal specimens sampled at birth was 5%. In contrast, the phylum Firmicutes signicantly ($p < 0.0001$) increased from 23% at birth to 64% at 20-28 weeks of life. The proportion of Firmicutes found within maternal faecal specimens sampled at birth was 79%. At class-level, the class Clostridia was the most abundant class classified from maternal faecal specimens (69%). This was also one of the classes which increased the most (from 12 to 37%) as the infants grew older. At genus-level, Table 12 shows a reduction in the number of genera present in meconium specimens when compared to other infant faecal specimens sampled at later timepoints.

Table 12. Number of genera from meconium specimens compared to other infant faecal specimens at later time points, as well as maternal faecal specimens

		Genera from infants at 4-12 weeks (n = 72)		Genera from infants at 20-28 weeks (n = 36)		Genera from maternal specimens (n = 90)	
		Absent	Present	Absent	Present	Absent	Present
Genera from infants at birth (n = 107)	Absent	261	27	265	23	258	30
	Present	131	197	145	183	173	155

In addition to Table 12, the log ratio biplot of the 616 genera identified from infant and maternal faecal specimens, shows three distinct clusters (Figure 36). Meconium specimens formed an overlap with the cluster consisting of other infant faecal specimens at 4-12 and 20-28 weeks of age. Maternal faecal specimens also formed a compact cluster detached from specimens sampled from infants (Figure 36). In addition, Figure 36 shows that the changes occurring in infant faecal bacterial profiles up until seven months of life are still of infant-like composition when compare to maternal faecal bacterial profiles.

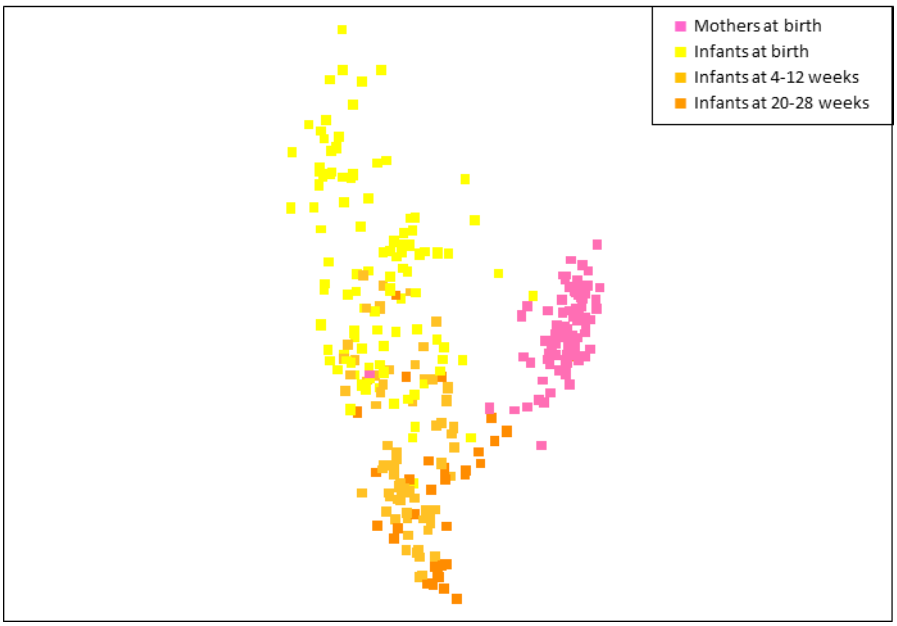


Figure 36. Log ratio biplot of infant meconium specimens sampled at birth (n = 107), infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age and maternal faecal specimens sampled at birth (n = 90)
Biplots represent faecal specimens in relation to the most abundant genera identified from them.

3.3.6.5 *"Participant characteristics" influencing infant faecal bacterial profiles during the first 28 weeks of life*

Table 13 summarises the association of participant characteristics with changes in infant bacterial profiles measured from birth up until 28 weeks of life (n = 215). In addition, we also determined if any of the participant characteristics had an effect on bacterial proportions as a function of a weekly increase in infant age (which is denoted by an asterisk in Table 13).

Mode of delivery was the only participant characteristic significantly associated with changes in bacterial profiles during the first 28 weeks of life as a function of a weekly increase in infant age. We observed changes in abundances of the families Burkholderiaceae and Ruminococcaceae for every one week increase in infant age. The family Burkolderiaceae decreased with every one week increase in age for both vaginal and caesarean sectional delivered infants, but the effect was greatest for vaginal delivered infants. In contrast to the family Burkholderiaceae, we observed a weekly increase in abundance for the family Ruminococcaceae from both vaginal and caesarean sectional delivered infants. This effect was again greatest for vaginally delivered infants. At genus-level, a per-week increase in abundance was evident for three genera (Faecalibacterium, Ruminococcus and an unclassified genus within the family Ruminococcaceae) from vaginally delivered infants. For caesarean sectional delivered infants, the genus Faecalibacterium also increased with every one week increase in age, but Ruminococcus and the unclassified genus decreased with every one week increase in age. Not all changes in bacterial profiles from vaginally

and caesarean sectionally delivered infants were associated with a “per-week increase” in infant age. We observed changes in the proportions of the genera Massilia and Escherichia/Shigella during the first 28 weeks of life, but these were not due to the effect of a weekly increase in infant age. We observed that infants born via vaginal delivery had 180% of proportions of the genus Escherichia/Shigella during the first 28 weeks of life compared to caesarean section delivered infants. On the other had, caesarean section delivered infants had 129% of the proportions of the genus Masillia observed for vaginal delivered infants during the first 28 weeks of life.

As seen for mode of delivery, significant changes in bacterial profiles during the first 28 weeks life were not due to a “per week increase” in infant. Participant characteristics which influenced bacterial profiles at some stage during the first 28 weeks of life, but not resulting from a “per week increase” in infant age were birth weight, gestational age, gender and maternal HIV status (Table 13). The two participant characteristics birth weight and gestational age were significantly associated with changes in the order Coriobacteriales, down to genus-level, during the first 28 weeks of life. The genus Collinsella increased by 76% with every one kilogram increase in birth weight during the first 28 weeks of life. The genera Collinsella and Olsenella were reduced by 13% and 12%, respectively, with every one week increase in gestational age. In this study, gender was significantly associated with changes in bacteria from the phylum Bacteroidetes. During the first 28 weeks of life lower proportions of the class Bacteroidia were observed for female infants when compared to males. The proportions of the class Bacteroidia (down to the genus Prevotella) observed for females were only around 68% of the proportions observed from males. Conversely, the class Flavobacteria, down to the family Flavobacteriaceae, was significantly more abundant in female infants when compared to males. Maternal HIV status was significantly associated with changes in the infant faecal bacterial genera Weissella and Lactococcus within the phylum Firmicutes. We observed that HIV-unexposed infants only had 37% and 45% of the proportions of the genera Weissella and Lactococcus, respectively, when compared to those exposed to HIV during the first 28 weeks of life.

Table 13. Participant characteristics with significant effects on bacterial OTUs from infant faecal specimens over time during the first 20-28 weeks of life (n = 215)

Taxon level	Bacterial OTUs	Birth weight (Weekly increase in kilograms†)		Gestational age (Weekly increase in age†)		Gender (Male: RR=1.00)		Vaginal delivery†	Caesarean section†		Maternal HIV status (HIV positive: RR=1.00)	
		Rate ratio	p-value	Rate ratio	p-value	Rate ratio	p-value	Rate ratio	Rate ratio	p-value	Rate ratio	p-value
Phylum	**Actinobacteria**											
Order	Coriobacteriales	-	-	0.88	0.0123	-	-	-	-	-	-	-
Family	Coriobacteriaceae	-	-	0.88	0.0123	-	-	-	-	-	-	-
Genus	Collinsella	1.76	0.0250	0.87	0.0133	-	-	-	-	-	-	-
Genus	Olsenella	-	-	0.88	0.0251	-	-	-	-	-	-	-
Phylum	**Bacteroidetes**											
Class	Bacteroidia	-	-	-	-	0.68	0.0279	-	-	-	-	-
Order	Bacteroidales	-	-	-	-	0.68	0.0279	-	-	-	-	-
Family	Prevotellaceae	-	-	-	-	0.66	0.0328	-	-	-	-	-
Genus	Prevotella	-	-	-	-	0.65	0.0328	-	-	-	-	-
Class	Flavobacteria	-	-	-	-	1.65	0.0466	-	-	-	-	-
Order	Flavobacteriales	-	-	-	-	1.65	0.0466	-	-	-	-	-
Family	Flavobacteriaceae	-	-	-	-	1.64	0.0495	-	-	-	-	-
Phylum	**Proteobacteria**											
Family	Burkholderiaceae	-	-	-	-	-	-	0.84*	0.95*	0.0033*	-	-
Genus	Massilia	-	-	-	-	-	-	0.77	1.29	0.0233	-	-
Genus	Escherichia/Shigella	-	-	-	-	-	-	1.80	0.55	0.0233	-	-
Phylum	**Firmicutes**											
Family	Leuconostocaceae	-	-	-	-	-	-	-	-	-	0.37	0.0001
Genus	Weissella	-	-	-	-	-	-	-	-	-	0.37	0.0002
Genus	Lactococcus	-	-	-	-	-	-	-	-	-	0.45	0.0155
Family	Ruminococcaceae	-	-	-	-	-	-	1.06*	1.02*	0.0014*	-	-
Genus	Faecalibacterium	-	-	-	-	-	-	1.05*	1.03*	0.0001*	-	-
Genus	Ruminococcus	-	-	-	-	-	-	1.39*	0.68*	0.0050*	-	-
Genus	Unclassified	-	-	-	-	-	-	1.09*	0.97*	0.0055*	-	-

* Characteristics associated with changes in infant faecal bacterial profiles during the first 28 week s of life also having an interaction with a weekly increase in infant age.
† Rate ratios increase or decrease at a constant rate over time based on the initial proportions of OTUs from faecal specimens collected at birth

3.3.6.6 *Changes in infant faecal bacterial diversity as a function of a weekly increase in infant age*

Our GLMs showed a significant decrease ($p < 0.0001$) in both the diversity within faecal specimens (intra-individual diversity) and the diversity between faecal specimens (inter-individual diversity) as an effect of a per-week increase in infant age (data not shown). The intra-individual bacterial diversities from infant faecal specimens decreased significantly by 3% with every one week increase in infant age. We also found that the intra-individual bacterial diversities in infant specimens differed significantly ($p < 0.0001$) from maternal faecal specimens (Figure 37A). Meconium specimens had much greater (80% higher) intra-individual bacterial diversity compared to maternal specimens, while infant faecal specimens at 4-12 and 20-28 weeks had less diverse (20% and 15% less) bacterial profiles compared to their mothers. The inter-individual bacterial diversities decreased by 62% with every one week increase in infant age (data not shown). When comparing inter-individual bacterial diversities of infant faecal specimens to maternal faecal specimens, we found that infant faecal bacterial diversities differed significantly from maternal faecal specimens (Figure 37B). This significant difference in bacterial diversities was observed at all timepoints studied (infant specimens collected at birth ($p < 0.0001$), at 4-12 weeks ($p = 0.027$) and at 20-28 weeks ($p = 0.028$)). Meconium specimens had 100% greater inter-individual bacterial diversity compared to maternal faecal specimens, while infant faecal specimens at 4-12 weeks and 20-28 weeks had much less diverse (99% and 80% less) bacterial profiles compared to mothers. Of the participant characteristics tested, only maternal HIV status was significantly associated with the two bacterial diversity measures.

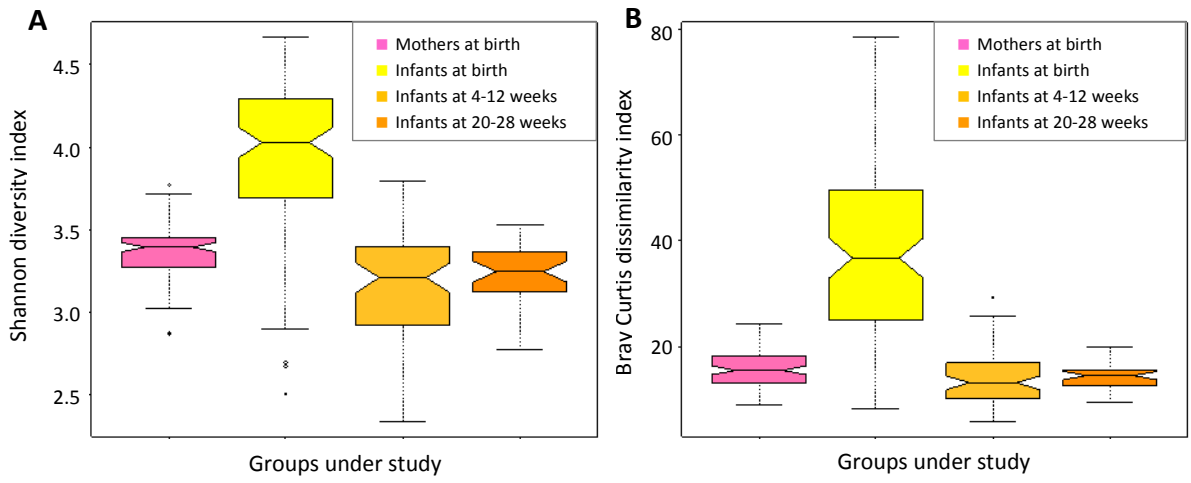


Figure 37. Intra- and inter-individual bacterial diversity measures of faecal specimens from the "groups under study"
A) Shannon diversity (intra-individual diversity) and **B)** Bray Curtis dissimilarity (inter-individual diversity) measures are shown for faecal specimens collected from mothers at birth (n=90), meconium specimens collected from infants (n=107), faecal specimens collected from infants at 4 to 12 (n=72) and at 20 to 28 weeks of age (n=36). The line within each box plot represents the median value. The median values of box plots are significantly different (at 5% significance level) in the event that their notches do not overlap. The box extends from the 25th to 75th percentile and whiskers show the minimum and maximum values.

3.4 Discussion

Meconium consists of amniotic fluid, metabolites such as pancreatic secretions and bile acids, intestinal epithelial cells, and mucus.[75] Remarkably, studies have recently also reported bacterial components from these specimens;[15,20–22,41] however the exact mechanisms of how the fetus acquires these bacterial components in-utero still remains unclear.[76]

Our study supports the notion that meconium is not sterile as we identified a number of genera previously reported from meconium specimens (Table S10).[15,20–22,41] The most abundant colonizers of meconium specimens were bacteria from the phylum Proteobacteria, followed by the phylum Firmicutes; which has also been reported from previous studies on meconium microbiota.[15,21,33] In contrast to this, two studies conducted on preterm infants[15,22] reported inverse bacterial proportions of these two phyla identified from infants born at gestational ages of less than 33 weeks. Interestingly, the study by Ardissone and colleagues[15] reported higher proportions of the phylum Proteobacteria in comparison to the phylum Firmicutes, as shown by our study, in infants born at gestational ages greater than 33 weeks. Of note, infants providing meconium specimens in our study had a median gestational age of 39 weeks (IQR, 38 – 40). In support of the high proportions of the phyla Proteobacteria and Firmicutes found in our study; studies have shown that select facultative anaerobes[42] belonging to these two phyla[2] are essential for successive GIT colonization by strict anaerobes around the first week of life.[42,77] One explanation for these colonization patterns is that facultative anaerobes prepare the environment for strict aerobes by consuming oxygen during growth and thereby reducing the positive redox potential of the infant's GIT.[42,77,78] Corroborating the above-mentioned "colonization dogma"; we observed a significant reduction in some facultative anaerobes (including Acinetobacter, Aeromonas, Shewanella, Pseudomonas and Aurantimonas) and a significant increase in strict anaerobes (including Bifidobacterium, Clostridium, Lactobacillus, Bacteroides and Veillonella) when studying infant faecal bacterial profiles over time. Despite the findings from our study, as well as the widely accepted "colonization dogma"; it is not set in stone that strictly anaerobic bacteria are completely absent very early on in life.[79] Tsuji et al. (2012) reported that 95% of the meconium specimens from 166 Japanese newborns were colonised by both facultative and strict anaerobes.[44] This is in agreement with what we observed in our study as we detected both facultative and strict anaerobic bacteria from meconium specimens sampled at birth.

Our study did not aim to address the mechanisms behind the meconium colonization process, as we did not analyse amniotic fluid, cord blood or placenta. Among several hypotheses; it has been suggested that bacteria may translocate from the maternal GIT lumen to placenta via

maternal systemic circulation.[23] Although we observed some overlap between bacterial profiles of infant and maternal samples, our study showed that meconium and maternal faecal bacterial profiles forms distinct clusters. This suggests that the in-utero translocation may have only a modest effect on the overall bacterial composition of the meconium at the time of birth. Our finding is in agreement with what has been reported in the study by Gosalbes and colleagues,[21] also showing separate clustering of meconium and maternal faecal specimens from two mother-infant pairs during early life using high-throughput sequencing. One possible reason for this may be that the anaerobic fetal/newborn GIT is not a suitable environment for adult-like microbiota which colonizes an aerobic GIT environment.[43,80–82] This is clearly emphasised by our study showing a significant increase in the number of strictly anaerobic Clostridia over time.[9] Furthermore, the murine model used to report on in-utero bacterial translocations of bacterial strains from the maternal to fetal GIT[20] studied facultative anaerobic Enterococcus species, which have been suggested as one of the early colonizers of the newborn's GIT. Considering that our meconium specimens were not sterile we do suggest the occurrence of in-utero colonization; however we also suspect that specific maternal GIT bacteria, such as anaerobes and facultative anaerobes belonging to the phyla Proteobacteria and Firmicutes,[2,42,77] may be better colonizers of the fetal GIT. Moreover, we did not observe direct clustering of bacterial profiles from any of the respective mother-infant pairs up until seven months of age, which supports previous reports of an adult-like microbiota only appearing between one to two and a half years of life.[83–85] Despite the fact that we did not detect direct clustering between respective mother-infant pairs; we did however note a changing trend in the infant faecal bacterial profiles over time. We noted that even though no mother-infant pair clustering was observed, infant faecal bacterial profiles became more adult-like during the first seven months of life. These observations are in line with those from Gosalbes et al. (2012) reporting closer clustering of infant faecal specimens with maternal faecal specimens as the infant's age increased.[21] We therefore suggest that faecal bacterial profiles from infants converge to profiles characteristic of the adult GIT during the first seven months of life and should be considered as an ongoing process starting from early infancy.

Our study and others[86–88] showed that changes in healthy infant faecal bacterial profiles during the first few months of life, are characterized by increased proportions of Actinobacteria and Firmicutes, with reduced proportions of the phylum Proteobacteria. These changes over time may be attributed to the introduction of milk feeding (breast or formula) during infancy.[79,89] The genus Bifidobacterium, from the phylum Actinobacteria, has been shown to increase in particular for breastfed infants.[6,7,12,87,89] In support of this, we found that the phylum Actinobacteria was most abundant around 4-12 weeks of age, and that the class Actinobacteria increased significantly with

the introduction of breastfeeding at 6-10 weeks of life. Moreover, the genus Bifidobacterium was most abundant within the phylum Actinobacteria from specimens sampled at 4-12 and 20-28 weeks (Figure S1), also reported by a Ringel-Kulka and co-workers.[90] In addition, our study also showed that the genus Bifidobacterium were the most abundant genus when compared to all other genera at 4-12 and 20-28 weeks of life (Table S4).[81,86,89,91–94] Interestingly, not only was breastfeeding associated with a significant increase in bacterial taxa from the class Actinobacteria at 4-12 and 20-28 weeks of age, but we also observed a significant reduction in bacterial taxa from the families Leuconostocaceae and Streptococcaceae, as well as the class Clostridia within the phylum Firmicutes for breastfed infants. This suggests that the introduction of formula feeding may contribute to the proliferation of bacteria within the families Leuconostocaceae and Streptococcaecae as well as the class Clostridia. In support of the latter, studies have shown that Bifidobacteria and Lactobacilli were dominant in breastfed infants, while formula fed infants contained higher proportions of Bacteroides, Staphylococci, Enterobacteria, Veillonella and Clostridia.[2,89]

In addition to breast- and formula feeding, weaning has also been associated with changes in infant faecal bacterial profiles, mainly contributing to shifts from infant-like bacterial profiles to adult-like profiles.[9,85,95] Bergstrom and colleagues[11] showed that solid foods dramatically impacted on infant GIT bacterial profiles between nine and 18 months of age. They reported an increase in Bacteroidetes species and a decrease in Bifidobacterium, Lactobacillus and Enterobacteriaceae.[11] In contrast, we did not find any significant association between the introduction of solid foods and any of our bacterial profiles; however our sample size was small. Of note, the Danish study by Bergstrom and colleagues[11] investigated a cohort of 330 infants, while we only studied faecal specimens of 30 infants who received solid food. The age at which solid foods were introduced to the 30 infants studied at 20-28 weeks of age, also varied across the first six months of life. More than half (17/30) of the infants only started to receive solid foods at five months of age of which seven provided samples after a short weaning period of less than two weeks. Bergstrom et al's study reported changes as from nine months of age, while our cohort only investigated infants up until seven months of age.[11]

Despite the fact that we did not observe significant changes in bacterial profiles associated with the introduction of solid food, which is in contrast with reports from previous studies, [9,84,85,95,96] we did observe that infant faecal bacterial profiles from our cohort are becoming more adult-like (Figure 34 and Table S9).[97] This is clearly highlighted by our study as we observed i) less distinct clustering of infant and maternal faecal specimens as infants grew older, ii) a reduction in significantly different genera between infant and maternal faecal specimens with increased age; iii) a significant increase in the phylum Firmicutes from our infant faecal specimens during the first

seven months of life; and iv) that the strictly anaerobic class Clostridia was the most abundant class from infant faecal specimens at 20-28 weeks of life. The latter being reported for adult faecal specimens from previous studies[9,11] as well as maternal faecal specimens in our cohort. Since our cohort was only studied up until seven months of age, these findings may be reflective of the effect of external factors, other than weaning,[9,84,89] on changes in the infant faecal bacterial profiles. In support of this, weaning seems to be associated with a significant increase in the proportions of Bacteroidetes.[84,96] Koenig and colleagues[84] found that the increase in Bacteroidetes only became apparent around six months of age; which too is in agreement with the findings from our study. Our results, and those from previous studies,[83–85] therefore suggest additional sampling points at later stages of infancy and early childhood in order to optimally address shifts towards adult-like bacterial profiles.

Adult faecal bacterial profiles are characterised by high proportions of the phylum Firmicutes; and when combined with the phylum Bacteroidetes, together consisted of approximately 90% of the bacteria found in the adult GIT.[98–105] In agreement with this, we found that the phyla Firmicutes and Bacteroidetes represented 86% of the bacteria identified from maternal faecal specimens under study. These were followed by the phylum Actinobacteria, Proteobacteria and other, which has also been reported by other studies.[86,102,103,105,106] This pattern of most abundant phyla was observed for all maternal faecal specimens; however the subset at 20-28 week of life had slightly higher abundances of the phylum Proteobacteria compared to the phyla Bacteroidetes and Actinobacteria. This may be explained by a sample-size effect since only 36 maternal samples were profiled of which one maternal faecal specimen, M_45, had a very high abundance of Proteobacteria. These findings highlight the importance of using large sample sizes when performing microbiota analyses in order to minimize the effect of outlier data.

In addition to feeding practices (i.e., breastfeeding, formula feeding and weaning), other external factors may act as potential “risk factors” for changes in both infant and maternal faecal bacterial profiles. Previous studies have reported an influence of maternal factors such as eczema and diabetes on infant meconium microbiota profiles.[21,33] We observed that the two maternal characteristics, maternal education and maternal HIV status, impacted infant faecal bacterial profiles early in life. An increase in maternal education levels was associated with higher proportions of an unclassified family within the class Clostridia, as well as an unclassified class within the phylum Firmicutes. Despite these increasing proportions, we surprisingly observed a reduction in these bacterial proportions from infants whose mothers had tertiary education levels. This does not seem biologically plausible since our findings up until the Grade 12 education level suggests that the higher the maternal education level, the higher the proportions of these unclassified bacteria.

Therefore, we suspect that our finding of an association with eduction levels may be ambiguous. However, our findings suggesting a potential role of maternal education in shaping meconium bacterial profiles are in keeping with those of Gosalbes et al. (2012) who reported two distinct bacterial profiles from infants whose mothers had either a low or a high level of education.[21]

When testing the effect of maternal HIV status on infant faecal bacterial profiles we found that HIV-exposed infants had increased bacterial diversities at 4-12 weeks of age and also had higher proportions of bacteria from the family Leuconostocaceae. Although the majority of children in this study were not exposed to HIV nor HIV-infected, our finding is in line with studies reporting increased bacterial diversities from faecal specimens of HIV positive participants.[107] Our observation, suggesting the role of maternal HIV status in altering infant faecal bacterial profiles, may be explained by a possible influence from breastfeeding. This is supported by the interesting findings from Gonzáles and colleagues[108] showing that the presence of HIV RNA in breast milk was associated with increased bacterial diversity from breastmilk specimens.[108]

Participant characteristics having a long-term effect on bacterial profiles from the time of birth up until the first 28 weeks of life were infant birth weight, gestational age, mode of delivery, gender and HIV status. Interestingly, we observed an increase in the proportions of the order Coriobacteriales with decreased gestational ages. Lower gestational ages have been shown to be associated with increased formula milk supplementation.[109–111] In addition, formula feeding has been associated with increased proportions of the order Coriobacteriales.[112] Therefore, we suggest that infants born to lower gestational ages may be exposed to increased supplementation with formula, which may in turn result in increased levels of Order Coriobacteriales. However; since we had a small number of infants being formula fed, and an even smaller number being born at low gestational ages we could not statistically determine whether low gestational ages were associated with increased formula feeding. Mode of delivery on the other hand was associated with changes in the genera Massilia and Escherichia/Shigella in this study. Although we could not find any literature reporting on the genus Massilia and its associations with mode of delivery, the higher abundances of Escherichia/Shigella observed for vaginal delivered infants are in agreement with findings from a previous study.[87] Besides measuring the effect of a number of participant characteristics on bacterial profiles throughout the first seven months of life; we also tested the effect of a per-week increase in infant age on these profiles. Interestingly, we only observed a significant association for the characteristic mode of delivery. Both vaginal and caesarean sectional delivered infants had a significant decrease in the proportions of bacteria within the phylum Proteobacteria, and conversely had a significant increase in bacteria within the phylum Firmicutes with every one week increase in infant age. However, the effects seen for vaginal delivered infants were much greater. This may be

supported by the findings of Grönlund et al.;[113] reporting that caesarean sectional delivered infants have delayed bacterial colonization rates.

3.5 Conclusion

Our results indicate colonization of meconium specimens by highly diverse bacterial profiles. We did not observe direct clustering of any of the respective mother-infant pairs when comparing their faecal bacterial profiles. Our study also emphasizes that a number of external factors, such as maternal education and HIV status, infant gestational age and birth weight, the infant's mode of delivery, the residential area in which the infant resides and whether the infant received breastfeeding or not, influence colonization of the GIT during early infancy. An interesting finding, not previously reported, was the association between HIV-exposed infants and increased bacterial diversities from infant faecal specimens. Furthermore, among the number of participant characteristics influencing bacterial profiles during early life, the most influential factor was the increase in infant age. Finally our study provides information on the shifts from an infant- to an adult-like microbiota profile during the first seven months of life.

References

1. Candela M, Biagi E, Maccaferri S, Turroni S, Brigidi P. Intestinal microbiota is a plastic factor responding to environmental changes. *Trends Microbiol*. 2012;20(8):385-91. doi:10.1016/j.tim.2012.05.003.

2. Arrieta M-C, Stiemsma LT, Amenyogbe N, Brown EM, Finlay B. The Intestinal Microbiome in Early Life: Health and Disease. *Front Immunol*. 2014;5(September):1-18. doi:10.3389/fimmu.2014.00427.

3. Dominguez-Bello MG, Costello EK, Contreras M, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *PNAS*. 2010;107(26):11971-5. doi:10.1073/pnas.1002601107.

4. Biasucci G, Rubini M, Riboni S, Morelli L, Bessi E, Retetangos C. Mode of delivery affects the bacterial community in the newborn gut. *Early Hum Dev*. 2010;86(Suppl 1):S13-S15. doi:10.1016/j.earlhumdev.2010.01.004.

5. Adlerberth I, Lindberg E, Aberg N, et al. Reduced enterobacterial and increased staphylococcal colonization of the infantile bowel: an effect of hygienic lifestyle? *Pediatr Res*. 2006;59(1):96-101. doi:10.1203/01.pdr.0000191137.12774.b2.

6. Fan W, Huo G, Li X, Yang L, Duan C. Impact of Diet in Shaping Gut Microbiota Revealed by a Comparative Study in Infants During the First Six Months of Life. *J Microbiol Biotechnol*. 2014;24(2):133-143.

7. Bezirtzoglou E, Tsiotsias A, Welling GW. Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe*. 2011;17(6):478-82. doi:10.1016/j.anaerobe.2011.03.009.

8. Kleessen B, Bunke H, Tovar K, Noack J, Sawatzki G. Influence of two infant formulas and human milk on the development of the faecal flora in newborn infants. *Acta Paediatr*. 1995;84(12):1347-56. doi:10.1111/j.1651-2227.1995.tb13567.x.

9. Fallani M, Amarri S, Uusijarvi A, et al. Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology*. 2011;157:1385-92. doi:10.1099/mic.0.042143-0.

10. Krebs NF, Sherlock LG, Westcott J, et al. Effects of different complementary feeding regimens on iron status and enteric microbiota in breastfed infants. *J Pediatr*. 2013;163(2):416-23. doi:10.1016/j.jpeds.2013.01.024.

11. Bergström A, Skov T, Bahl M, et al. Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of Danish infants. *Appl Environ Microbiol*. 2014;80(9):2889-900. doi:10.1128/AEM.00342-14.

12. Fallani M, Young D, Scott J, et al. Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. *J Pediatr Gastroenterol Nutr*. 2010;51(1):77-84. doi:10.1097/MPG.0b013e3181d1b11e.

13. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486(7402):222-7. doi:10.1038/nature11053.

14. Fouhy F, Guinane CM, Hussey S, et al. High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin. *Antimicrob Agents Chemother*. 2012;56(11):5811-20. doi:10.1128/AAC.00789-12.

15. Ardissone AN, De La Cruz DM, Davis-Richardson AG, et al. Meconium microbiome analysis identifies bacteria correlated with premature birth. *PLoS One*. 2014;9(3):e90784. doi:10.1371/journal.pone.0090784.

16. Rautava S, Collado MC, Salminen S, Isolauri E. Probiotics modulate host-microbe interaction in the placenta and fetal gut: a randomized, double-blind, placebo-controlled trial. *Neonatology*. 2012;102(3):178-84. doi:10.1159/000339182.

17. Satokari R, Grönroos T, Laitinen K, Salminen S, Isolauri E. Bifidobacterium and Lactobacillus DNA in the human placenta. *Lett Appl Microbiol*. 2009;48(1):8-12. doi:10.1111/j.1472-765X.2008.02475.x.

18. Steel JH, Malatos S, Kennea N, et al. Bacteria and inflammatory cells in fetal membranes do not always cause preterm labor. *Pediatr Res*. 2005;57(3):404-11. doi:10.1203/01.PDR.0000153869.96337.90.

19. Jiménez E, Fernández L, Marín ML, et al. Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. *Curr Microbiol*. 2005;51(4):270-4. doi:10.1007/s00284-005-0020-3.

20. Jiménez E, Marín ML, Martín R, et al. Is meconium from healthy newborns actually sterile? *Res Microbiol*. 2008;159(3):187-93. doi:10.1016/j.resmic.2007.12.007.

21. Gosalbes MJ, Llop S, Vallès Y, Moya A, Ballester F, Francino MP. Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clin Exp Allergy*. 2012;43(2):198-211. doi:10.1111/cea.12063.

22. Moles L, Gómez M, Heilig H, et al. Bacterial diversity in meconium of preterm neonates and evolution of their fecal microbiota during the first month of life. *PLoS One*. 2013;8(6):e66986. doi:10.1371/journal.pone.0066986.

23. DiGiulio DB. Diversity of microbes in amniotic fluid. *Semin Fetal Neonatal Med*. 2012;17(1):2-11. doi:10.1016/j.siny.2011.10.001.

24. White BA, Creedon DJ, Nelson KE, Wilson BA. The vaginal microbiome in health and disease. *Trends Endocrinol Metab*. 2011;22(10):389-93. doi:10.1016/j.tem.2011.06.001.

25. DiGiulio DB, Romero R, Kusanovic JP, et al. Prevalence and diversity of microbes in the amniotic fluid, the fetal inflammatory response, and pregnancy outcome in women with preterm prelabor rupture of membranes. *Am J Reprod Immunol*. 2010;64(1):38-57. doi:10.1111/j.1600-0897.2010.00830.x.

26. Rescigno M, Rotta G, Valzasina B, Ricciardi-Castagnoli P. Dendritic cells shuttle microbes across gut epithelial monolayers. *Immunobiology*. 2001;204(5):572-81. doi:10.1078/0171-2985-00094.

27. Makino H, Kushiro A, Ishikawa E, et al. Transmission of intestinal Bifidobacterium longum subsp. longum strains from mother to infant, determined by multilocus sequencing typing and amplified fragment length polymorphism. *Appl Environ Microbiol*. 2011;77(19):6788-93. doi:10.1128/AEM.05346-11.

28. Bjerke GA, Wilson R, Storrø O, Øyen T, Johnsen R, Rudi K. Mother-to-Child Transmission of and Multiple-Strain Colonization by Bacteroides fragilis in a Cohort of Mothers and Their Children. *Appl Environ Microbiol*. 2011;77(23):8318-24. doi:10.1128/AEM.05293-11.

29. Boggess KA, Madianos PN, Preisser JS, Moise KJ, Offenbacher S. Chronic maternal and fetal Porphyromonas gingivalis exposure during pregnancy in rabbits. *Am J Obstet Gynecol*. 2005;192(2):554-7. doi:10.1016/j.ajog.2004.09.001.

30. Gonzales-Marin C, Spratt DA, Allaker RP. Maternal oral origin of Fusobacterium nucleatum in adverse pregnancy outcomes as determined using the 16S-23S rRNA gene intergenic transcribed spacer region. *J Med Microbiol*. 2013;62:133-44. doi:10.1099/jmm.0.049452-0.

31. León R, Silva N, Ovalle A, et al. Detection of Porphyromonas gingivalis in the amniotic fluid in pregnant women with a diagnosis of threatened premature labor. *J Periodontol*. 2007;78(7):1249-55. doi:10.1902/jop.2007.060368.

32. Bearfield C, Davenport ES, Sivapathasundaram V, Allaker RP. Possible association between amniotic fluid micro-organism infection and microflora in the mouth. *BJOG*. 2002;109(5):527-33.

33. Hu J, Nomura Y, Bashir A, et al. Diversified microbiota of meconium is affected by maternal diabetes status. *PLoS One*. 2013;8(11):e78257. doi:10.1371/journal.pone.0078257.

34. Collado M, Cernada M, Bäuerl C, Vento M, Pérez-Martínez G. Microbial ecology and host-microbiota interactions during early life stages. *Gut Microbes*. 2012;3(4):352-365. doi:10.4161/gmic.21215.

35. Battersby AJ, Gibbons DL. The gut mucosal immune system in the neonatal period. *Pediatr allergy Immunol*. 2013;24(5):414-21. doi:10.1111/pai.12079.

36. Wopereis H, Oozeer R, Knipping K, Belzer C, Knol J. The first thousand days - intestinal microbiology of early life: establishing a symbiosis. *Pediatr allergy Immunol*. 2014. doi:10.1111/pai.12232.

37. Martin R, Nauta AJ, Ben Amor K, Knippels LMJ, Knol J, Garssen J. Early life: gut microbiota and immune development in infancy. *Benef Microbes*. 2010;1(4):367-82. doi:10.3920/BM2010.0027.

38. Hansen CHF, Nielsen DS, Kverka M, et al. Patterns of early gut colonization shape future immune responses of the host. *PLoS One*. 2012;7(3):e34043. doi:10.1371/journal.pone.0034043.

39. Kerr CA, Grice DM, Tran CD, et al. Early life events influence whole-of-life metabolic health via gut microflora and gut permeability. *Crit Rev Microbiol*. 2014. doi:10.3109/1040841X.2013.837863.

40. Clarke G, O'Mahony SM, Dinan TG, Cryan JF. Priming for health: gut microbiota acquired in early life regulates physiology, brain and behaviour. *Acta Paediatr*. 2014;103(8):812-9. doi:10.1111/apa.12674.

41. Madan JC, Salari RC, Saxena D, et al. Gut microbial colonisation in premature neonates predicts neonatal sepsis. *Arch Dis childhood Fetal neonatal Ed*. 2012;97(6):F456-62. doi:10.1136/archdischild-2011-301373.

42. Adlerberth I, Wold AE. Establishment of the gut microbiota in Western infants. *Acta Paediatr*. 2009;98(2):229-38. doi:10.1111/j.1651-2227.2008.01060.x.

43. Matamoros S, Gras-Leguen C, Le Vacon F, Potel G, De La Cochetiere M-F. Development of intestinal microbiota in infants and its impact on health. *Trends Microbiol*. 2013;21(4):167-73. doi:10.1016/j.tim.2012.12.001.

44. Tsuji H, Oozeer R, Matsuda K, et al. Molecular monitoring of the development of intetinal microbiota in Japanese infants. *Benef Microbes*. 2012;3(2):113-25. doi:10.3920/BM2011.0038.

45. Zar HJ, Barnett W, Myer L, Stein DJ, Nicol MP. Investigating the early-life determinants of illness in Africa: the Drakenstein Child Health Study. *Thorax*. 2014;0:1-3. doi:10.1136/thoraxjnl-2014-206242.

46. Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS*. 2011;108(Supplement_1):4516-22. doi:10.1073/pnas.1000080107.

47. Illumina Proprietary. MiSeq ® Reagent Kit v3 Reagent Preparation Guide. 2013;(October 2013):1-14.

48. Illumina Proprietary. Preparing Libraries for Sequencing on the MiSeq ®. 2013;(August 2013):1-14.

49. Szpakowski S. YAP: A Computationally Efficient Workflow for Taxonomic Analyses of Bacterial 16S and Fungal ITS Sequences. 2013.

50. R Development Core Team. R: A language and environment for statistical computing. 2008.

51. RStudio. RStudio: Integrated development environment for R. 2012.

52. Chongsuvivatwong V. epicalc: Epidemiological calculator. 2012.

53. Faith DP, Minchin PR, Belbin L. Compositional Dissimilarity as a Robust Measure of Ecological Distance. *Vegetatio*. 1987;69(1):57-68.

54. Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol Monogr*. 1957;27(4):325-349.

55. Clarke KR, Warwick RM. *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation*. Second edi. Plymouth, United Kingdom: PRIMER-E Ltd; 2001.

56. Morgan XC, Huttenhower C. Chapter 12: Human microbiome analysis. *PLoS Comput Biol*. 2012;8(12):e1002808. doi:10.1371/journal.pcbi.1002808.

57. Oksanen J, Blanchet F., Kindt R, et al. vegan: Community Ecology Package. 2013.

58. Cox T., Cox MA. *Multidimensional Scaling*. Second edi. Boca Raton: Chapman & Hall/CRC; 2001.

59. De Leeuw J, Mair P. Multidimensional scaling using majorization: The R package smacof. *J Stat Softw*. 2009;31(3):1-30.

60. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106. doi:10.1186/gb-2010-11-10-r106.

61. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol*. 2014;10(4):e1003531. doi:10.1371/journal.pcbi.1003531.

62. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*. 2014;2:15. doi:10.1186/2049-2618-2-15.

63. Greenacre M. Log-Ratio Biplots. In: *Biplots in Practice*. Fundación BBVA; 2010:69-78.

64. Gower J, Lubbe S, Le Roux N. *Understanding Biplots*. Chichester, United Kingdom: John Wiley & Sons Ltd.; 2011.

65. Martin-Fernandez J., Palarea-Albaladejo J, Olea R. Dealing with Zeros. In: Pawlowsky-Glahn V, Buccianti A, eds. *Compositional Data Analysis: Theory and Applications*. Chichester, United Kingdom: John Wiley & Sons Ltd.; 2011:43-58.

66. Aitchison J. The Statistical Analysis of Compositional Data. *J R Stat Soc*. 1982;44:139-160.

67. Aitchison J, Greenacre M. Biplots of compositional data. *J R Stat Soc Ser C (Applied Stat*. 2002;51(4):375-392.

68. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B*. 1995;57(1):289-300.

69. Plan EL. Modeling and simulation of count data. *CPT Pharmacometrics Syst Pharmacol*. 2014;3(August):e129. doi:10.1038/psp.2014.27.

70. McCullagh P, Nelder J. *Generalized linear models*. Second Edi. London: Chapman and Hall; 1989:199.

71. Aeberhard WH, Cantoni E, Heritier S. Robust inference in the negative binomial regression model with an application to falls data. *Biometrics*. 2014;(2001). doi:10.1111/biom.12212.

72. Dobson J. *An Introduction to Generalized Linear Models*. Second Edi. Boca Raton: Chapman and Hall/CRC; 2002.

73. Cnaan A, Laird NANM, Slasor P. Tutorial in Biostatistics. Using the General Linear Mixed Model to analyse unbalanced repeated measures and longitudinal data. *Stat Med*. 1997;16:2349-2380.

74. Venables W., Ripley B., Bates D., Hornik K, Gebhardt A, Firth D. Support Functions and Datasets for Venables and Ripley's MASS. 2014.

75. Kumagai M, Kimura A, Takei H, et al. Perinatal bile acid metabolism: bile acid analysis of meconium of preterm and full-term infants. *J Gastroenterol*. 2007;42(11):904-10. doi:10.1007/s00535-007-2108-y.

76. Funkhouser LJ, Bordenstein SR. Mom knows best: the universality of maternal microbial transmission. *PLoS Biol*. 2013;11(8):e1001631. doi:10.1371/journal.pbio.1001631.

77. Adlerberth I. Factors influencing the establishment of the intestinal microbiota in infancy. *Nestle Nutr Work Ser Pediatr Progr*. 2008;62:13-29. doi:10.1159/000146245.

78. Bezirtzoglou E. The intestinal microflora during the first weeks of life. *Anaerobe*. 1997;3(2-3):173-7. doi:10.1006/anae.1997.0102.

79. Jost T, Lacroix C, Braegger CP, Chassard C. New insights in gut microbiota establishment in healthy breast fed neonates. *PLoS One*. 2012;7(8):e44595. doi:10.1371/journal.pone.0044595.

80. Rotimi VO, Duerden BI. The development of the bacterial flora in normal neonates. *J Med Microbiol*. 1981;14:51-62. doi:10.1099/00222615-14-1-51.

81. Benno Y, Sawada K, Mitsuoka T. The Intestinal Microflora of Infants : Composition of Fecal Flora in Breast-Fed and Bottle-Fed Infants. *Microbiol Immunol*. 1984;28(9):975-986. doi:10.1111/j.1348-0421.1984.tb00754.x.

82. Garrity GM, Bell JA, Lilburn TG. Taxonomic Outline of the Procaryotes. In: *Bergey's Manual of Systematic Bacteriology, 2nd edition, Release 5.0, Springer-Verlag, New York.*; 2004:1-399. doi:10.1007/bergeysoutline200405.

83. Avershina E, Storrø O, Øien T, et al. Bifidobacterial succession and correlation networks in a large unselected cohort of mothers and their children. *Appl Environ Microbiol*. 2013;79(2):497-507. doi:10.1128/AEM.02359-12.

84. Koenig JE, Spor A, Scalfone N, et al. Succession of microbial consortia in the developing infant gut microbiome. *PNAS*. 2011;108(Suppl 1):4578-85. doi:10.1073/pnas.1000081107.

85. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. Ruan Y, ed. *PLoS Biol*. 2007;5(7):e177. doi:10.1371/journal.pbio.0050177.

86. Turroni F, Peano C, Pass DA, et al. Diversity of bifidobacteria within the infant gut microbiota. *PLoS One*. 2012;7(5):e36957. doi:10.1371/journal.pone.0036957.

87. Azad MB, Konya T, Maughan H, et al. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *Can Med Assoc J*. 2013;185(5):385-94. doi:10.1503/cmaj.121189.

88. Jakobsson HE, Abrahamsson TR, Jenmalm MC, et al. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut*. 2014;63(4):559-66. doi:10.1136/gutjnl-2012-303249.

89. Harmsen HJM, Wildeboer–Veloo ACM, Raangs GC, et al. Analysis of Intestinal Flora Development in Breast-Fed and Formula-Fed Infants by Using Molecular Identification and Detection Methods. *J Pediatr Gastroenterol Nutr*. 2000;30(1):61-67.

90. Ringel-Kulka T, Cheng J, Ringel Y, et al. Intestinal microbiota in healthy U.S. young children and adults--a high throughput microarray analysis. *PLoS One*. 2013;8(5):e64315. doi:10.1371/journal.pone.0064315.

91. Yoshioka H, Iseki K, Fujita K. Development and Differences of Intestinal Flora in the Neonatal Period in Breast-Fed and Bottle-Fed Infants. *Pediatrics*. 1983;72(3):317-21.

92. Favier CF, Vaughan EE, De Vos WM, Akkermans ADL. Molecular Monitoring of Succession of Bacterial Communities in Human Neonates. *Appl Environ Microbiol*. 2002;68(1):219-26. doi:10.1128/AEM.68.1.219.

93. Penders J, Vink C, Driessen C, London N, Thijs C, Stobberingh EE. Quantification of Bifidobacterium spp., Escherichia coli and Clostridium difficile in faecal samples of breast-fed and formula-fed infants by real-time PCR. *FEMS Microbiol Lett*. 2005;243(1):141-7. doi:10.1016/j.femsle.2004.11.052.

94. Solís G, De Los Reyes-Gavilan CG, Fernández N, Margolles A, Gueimonde M. Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut. *Anaerobe*. 2010;16(3):307-10. doi:10.1016/j.anaerobe.2010.02.004.

95. Kurokawa K, Itoh T, Kuwahara T, et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res*. 2007;14(4):169-81. doi:10.1093/dnares/dsm018.

96. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem*. 2009;284(37):24673-7. doi:10.1074/jbc.R109.022848.

97. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012;489(7415):220-30. doi:10.1038/nature11550.

98. Vital M, Howe AC, Tiedje JM. Revealing the Bacterial Butyrate Synthesis Pathways by Analyzing (Meta) genomic Data. *MBio*. 2014. doi:10.1128/mBio.00889-14.

99. Centanni M, Turroni S, Biagi E, et al. A novel combined approach based on HTF-Microbi.Array and qPCR for a reliable characterization of the Bifidobacterium-dominated gut microbiota of breast-fed infants. *FEMS Microbiol Lett*. 2013;343(2):121-6. doi:http://dx.doi.org/10.1111/1574-6968.12138.

100. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486(7402):207-14. doi:10.1038/nature11234.

101. Durbán A, Abellán JJ, Jiménez-Hernández N, Latorre A, Moya A. Daily follow-up of bacterial communities in the human gut reveals stable composition and host-specific patterns of interaction. *FEMS Microbiol Ecol*. 2012;81(2):427-437. doi:10.1111/j.1574-6941.2012.01368.x.

102. Jalanka-Tuovinen J, Salonen A, Nikkilä J, et al. Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms. *PLoS One*. 2011;6(7):e23035. doi:10.1371/journal.pone.0023035.

103. Tap J, Mondot S, Levenez F, et al. Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol*. 2009;11(10):2574-84. doi:10.1111/j.1462-2920.2009.01982.x.

104. Gill SR, Pop M, Deboy RT, et al. Metagenomic analysis of the human distal gut microbiome. *Science*. 2006;312(5778):1355-9. doi:10.1126/science.1124234.

105. Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal microbial flora. *Science*. 2005;308(June):1635-8. doi:10.1126/science.1110591.

106. Sekelja M, Berget I, Næs T, Rudi K. Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach. *ISME J*. 2011;5(3):519-31. doi:10.1038/ismej.2010.129.

107. Lozupone C, Li M, Campbell T, et al. Alterations in the gut microbiota associated with HIV-1 infection. *Cell Host Microbe*. 2013;14(3):329-39. doi:10.1016/j.chom.2013.08.006.

108. González R, Maldonado A, Martín V, et al. Breast milk and gut microbiota in African mothers and infants from an area of high HIV prevalence. *PLoS One*. 2013;8(11):e80299. doi:10.1371/journal.pone.0080299.

109. Altuntas N, Kocak M, Akkurt S, Razi HC, Kislal MF. LATCH Scores and Milk Intake in Preterm and Term Infants: A Prospective Comparative Study. *Breastfeed Med*. 2015;10(0):1-6. doi:10.1089/bfm.2014.0042.

110. Briere C-E, Lucas R, McGrath JM, Lussier M, Brownell E. Establishing Breastfeeding with the Late Preterm Infant in the NICU. *J Obstet Gynecol Neonatal Nurs*. 2015;44:102-113. doi:10.1111/1552-6909.12536.

111. Jang GJ, Lee SL, Kim HM. Breast Feeding Rates and Factors Influencing Breast Feeding Practice in Late Preterm Infants : Comparison with Preterm Born at Less than 34 Weeks of Gestational Age. *J Korean Acad Nurs*. 2012;42(2):181-189. doi:10.4040/jkan.2012.42.2.181.

112. Harmsen HJM, Wildeboer-Veloo a. CM, Grijpstra J, Knol J, Degener JE, Welling GW. Development of 16S rRNA-based probes for the Coriobacterium group and the Atopobium cluster and their application for enumeration of Coriobacteriaceae in human feces from volunteers of different age groups. *Appl Environ Microbiol*. 2000;66(10):4523-4527. doi:10.1128/AEM.66.10.4523-4527.2000.

113. Grönlund M, Lehtonen O, Eerola E, Kero P. Fecal microflora in healthy infants born by different methods of delivery: permanent changes in intestinal flora after cesarean delivery. *J Paediatr Gastroenterol Nutr*. 1999;28(1):19-25.

Supplementary data

A) Supplementary tables

Table S1. Bacterial taxa sequenced from all faecal and meconium specimens (n = 305) with relative abundances > 0.5% and < 0.5%

Phylum-level (> 0.5%)		Phylum-level (< 0.5%)							
Taxon	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**
Actinobacteria	13.54	Crenarchaeota	0.006	Chloroflexi	0.061	Nitrospira	0.007	Cyanobacteria Chloroplast	0.163
Bacteroidetes	5.39	Euryarchaeota	0.106	Deinococcus Thermus	0.125	Planctomycetes	0.008	OD1	0.001
Proteobacteria	28.13	Acidobacteria	0.017	Elusimicrobia	0.025	Spirochaetes	0.025	OP11	0.001
Firmicutes	50.44	Aquificae	0.003	Fusobacteria	0.083	Synergistetes	0.010	TM7	0.318
Unclassified	1.33	Armatimonadetes	0.007	Gemmatimonadetes	0.007	Tenericutes	0.006	Unclassified	0.005
Other	1.17	Chlamydiae	0.003	Lentisphaerae	0.003	Verrucomicrobia	0.180		
Class-level (> 0.5%)		**Class-level (< 0.5%)**							
Taxon	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**
Actinobacteria	13.54	Thermoprotei	0.006	Bacteroidetes incertae sedis class incertae sedis	0.001	Lentisphaeria	0.003	Verrucomicrobiae	0.163
Bacteroidia	4.48	Halobacteria	0.001	Sphingobacteria	0.332	Nitrospira	0.007	Unclassified	0.003
Alphaproteobacteria	9.53	Methanobacteria	0.104	Flavobacteria	0.313	Planctomycetacia	0.007	Chloroplast	0.149
Betaproteobacteria	5.02	Thermoplasmata	0.001	Unclassified	0.265	Unclassified	0.001	Cyanobacteria	0.014
Gammaproteobacteria	12.75	Acidobacteria Gp10	0.004	Chlamydiae	0.003	Deltaproteobacteria	0.404	OD1 class incertae sedis	0.001
Bacilli	10.65	Acidobacteria Gp16	0.001	Chloroflexi	0.000	Epsilonproteobacteria	0.075	OP11 class incertae sedis	0.001
Clostridia	32.86	Acidobacteria Gp3	0.000	Thermomicrobia	0.059	Unclassified	0.356	TM7 class incertae sedis	0.318
Erysipelotrichia	3.61	Acidobacteria Gp4	0.006	Unclassified	0.003	Spirochaetes	0.025	Unclassified	0.005
Negativicutes	2.18	Acidobacteria Gp6	0.005	Deinococci	0.125	Synergistia	0.010		
Unclassified	1.14	Acidobacteria Gp7	0.001	Elusimicrobia	0.025	Mollicutes	0.006		
Unclassified	1.33	Aquificae	0.003	Fusobacteria	0.083	Opitutae	0.009		
Other	2.92	Armatimonadetes gp5 class incertae sedis	0.007	Gemmatimonadetes	0.007	Subdivision5	0.004		

Table S1. Bacterial taxa sequenced from all faecal and meconium specimens (n = 305) with relative abundances > 0.5% and < 0.5% (continued)

Order-level (> 0.5%)		Order-level (< 0.5%)							
Taxon	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**
Actinomycetales	2.99	Fervidicoccales	0.006	Unclassified	0.003	Cardiobacteriales	0.005	TM7 order incertae sedis	0.318
Bifidobacteriales	6.82	Halobacteriales	0.001	Deinococcales	0.123	Chromatiales	0.175	Unclassified	0.005
Coriobacteriales	3.49	Methanobacteriales	0.104	Thermales	0.002	Gammaproteobacteria order incertae sedis	0.005		
Bacteroidales	4.48	Thermoplasmatales	0.001	Elusimicrobiales	0.025	Legionellales	0.016		
Caulobacterales	2.16	Acidobacteria Gp10 order incertae sedis	0.004	Fusobacteriales	0.083	Oceanospirillales	0.030		
Rhizobiales	3.74	Acidobacteria Gp16 order incertae sedis	0.001	Gemmatimonadales	0.007	Pasteurellales	0.217		
Rhodobacterales	1.02	Acidobacteria Gp3 order incertae sedis	0.000	Victivallales	0.002	Thiotrichales	0.002		
Rhodospirillales	0.65	Acidobacteria Gp4 order incertae sedis	0.006	Unclassified	0.000	Unclassified	0.238		
Sphingomonadales	1.36	Acidobacteria Gp6 order incertae sedis	0.005	Nitrospirales	0.007	Unclassified	0.356		
Unclassified	0.55	Acidobacteria Gp7 order incertae sedis	0.001	Planctomycetales	0.007	Spirochaetales	0.025		
Burkholderiales	4.77	Acidimicrobiales	0.007	Unclassified	0.001	Synergistales	0.010		
Enterobacteriales	3.31	Solirubrobacterales	0.080	Rickettsiales	0.044	Anaeroplasmatales	0.002		
Aeromonadales	1.68	Thermoleophilales	0.003	Hydrogenophilales	0.005	Mycoplasmatales	0.005		
Alteromonadales	1.60	Unclassified	0.147	Methylophilales	0.002	Opitutales	0.002		
Pseudomonadales	4.81	Aquificales	0.003	Neisseriales	0.188	Puniceicoccales	0.007		
Xanthomonadales	0.65	Armatimonadetes gp5 order incetae sedis	0.007	Nitrosomonadales	0.001	Subdivision5 order incertae sedis	0.004		
Bacillales	1.07	Bacteroidetes incertae sedis order incertae sedis	0.001	Rhodocyclales	0.027	Verrucomicrobiales	0.163		
Lactobacillales	9.36	Sphingobacteriales	0.332	Unclassified	0.029	Unclassified	0.003		
Clostridiales	32.71	Flavobacteriales	0.313	Bdellovibrionales	0.186	Chloroplast order incertae sedis	0.149		
Erysipelotrichales	3.61	Unclassified	0.265	Desulfovibrionales	0.200	Cyanobacteria order incertae sedis	0.014		
Selenomonadales	2.18	Chlamydiales	0.003	Myxococcales	0.011	Unclassified	0.220		
Unclassified	1.14	Herpetosiphonales	0.000	Unclassified	0.006	Unclassified	0.141		
Unclassified	1.33	Sphaerobacterales	0.058	Campylobacterales	0.075	OD1 order incertae sedis	0.001		
Other	4.53	Unclassified	0.001	Vibrionales	0.031	OP11 order incertae sedis	0.001		

Table S1. Bacterial taxa sequenced from all faecal and meconium specimens (n = 305) with relative abundances > 0.5% and < 0.5% (continued)

Family-level (> 0.5%)		Family-level (< 0.5%)							
Taxon	%	Taxon	%	Taxon	%	Taxon	%	Taxon	%
Geodermatophilaceae	0.61	Fervidicoccaceae	0.006	Bacteroidales incertae sedis	0.000	Unclassified	0.001	Bacillales Incertae Sedis XI	0.077
Micrococcaceae	0.82	Halobacteriaceae	0.001	Marinilabiaceae	0.057	Erythrobacteraceae	0.030	Bacillales Incertae Sedis XII	0.030
Nocardioidaceae	0.53	Methanobacteriaceae	0.104	Rikenellaceae	0.174	Unclassified	0.101	Anaeroplasmataceae	0.002
Bifidobacteriaceae	6.82	Unclassified	0.001	Unclassified	0.241	Alcaligenaceae	0.002	Mycoplasmataceae	0.005
Coriobacteriaceae	3.49	Acidobacteria Gp10 family incertae sedis	0.004	Cyclobacteriaceae	0.011	Burkholderiales incertae sedis	0.350	Opitutaceae	0.002
Porphyromonadaceae	0.64	Acidobacteria Gp16 family incertae sedis	0.001	Saprospiraceae	0.002	Sutterellaceae	0.175	Puniceicoccaceae	0.007
Prevotellaceae	2.32	Acidobacteria Gp3 family incertae sedis	0.000	Chitinophagaceae	0.132	Unclassified	0.432	Subdivision5 family incertae sedis	0.004
Bacteroidaceae	1.05	Acidobacteria Gp4 family incertae sedis	0.006	Cytophagaceae	0.107	Hydrogenophilaceae	0.005	Verrucomicrobiaceae	0.163
Caulobacteraceae	2.16	Acidobacteria Gp6 family incertae sedis	0.005	Sphingobacteriaceae	0.009	Methylophilaceae	0.002	Unclassified	0.003
Aurantimonadaceae	1.01	Acidobacteria Gp7 family incertae sedis	0.001	Unclassified	0.071	Neisseriaceae	0.188	Chloroplast	0.149
Hyphomicrobiaceae	0.60	Acidimicrobineae incertae sedis	0.001	Flavobacteriaceae	0.312	Nitrosomonadaceae	0.001	Family II	0.005
Rhizobiaceae	0.72	Acidimicrobiaceae	0.004	Unclassified	0.001	Rhodocyclaceae	0.027	Family IV	0.002
Unclassified	0.82	Unclassified	0.002	Unclassified	0.265	Unclassified	0.029	Family V	0.001
Rhodobacteraceae	1.02	Actinomycetaceae	0.300	Chlamydiaceae	0.000	Bacteriovoracaceae	0.114	Unclassified	0.006
Sphingomonadaceae	1.23	Beutenbergiaceae	0.001	Parachlamydiaceae	0.002	Bdellovibrionaceae	0.072	Alicyclobacillaceae	0.001
Unclassified	0.55	Bogoriellaceae	0.001	Herpetosiphonaceae	0.000	Desulfomicrobiaceae	0.002	Hahellaceae	0.001
Burkholderiaceae	0.52	Brevibacteriaceae	0.007	Sphaerobacteraceae	0.058	Desulfovibrionaceae	0.180	Halomonadaceae	0.028
Comamonadaceae	1.76	Cellulomonadaceae	0.017	Unclassified	0.001	Unclassified	0.019	Pasteurellaceae	0.217
Oxalobacteraceae	1.53	Corynebacteriaceae	0.104	Unclassified	0.003	Cystobacteraceae	0.002	Unclassified	0.001
Enterobacteriaceae	3.31	Demequinaceae	0.001	Deinococcaceae	0.121	Nannocystaceae	0.003	Bacillales incertae sedis	0.001
Aeromonadaceae	1.51	Dermabacteraceae	0.025	Trueperaceae	0.002	Phaselicystidaceae	0.004	Listeriaceae	0.002
Shewanellaceae	1.53	Dermacoccaceae	0.001	Thermaceae	0.002	Polyangiaceae	0.000	Paenibacillaceae 1	0.008
Moraxellaceae	3.27	Dermatophilaceae	0.007	Elusimicrobiaceae	0.025	Unclassified	0.002	Paenibacillaceae 2	0.003
Pseudomonadaceae	1.54	Dietziaceae	0.058	Fusobacteriaceae	0.060	Unclassified	0.006	Planococcaceae	0.254
Xanthomonadaceae	0.65	Intrasporangiaceae	0.044	Leptotrichiaceae	0.023	Campylobacteraceae	0.050	Sporolactobacillaceae	0.026
Enterococcaceae	1.27	Kineosporiaceae	0.003	Gemmatimonadaceae	0.007	Helicobacteraceae	0.025	Staphylococcaceae	0.413

Table S1. Bacterial taxa sequenced from all faecal and meconium specimens (n = 305) with relative abundances > 0.5% and < 0.5% (continued)

Family-level (> 0.5%)		Family-level (< 0.5%)							
Taxon	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**
Lactobacillaceae	2.29	Microbacteriaceae	0.107	Victivallaceae	0.002	Vibrionaceae	0.031	Unclassified	0.096
Leuconostocaceae	1.09	Micromonosporaceae	0.005	Unclassified	0.000	Succinivibrionaceae	0.161	Aerococcaceae	0.022
Streptococcaceae	3.58	Mycobacteriaceae	0.028	Nitrospiraceae	0.007	Unclassified	0.000	Carnobacteriaceae	0.153
Unclassified	0.95	Nocardiaceae	0.055	Planctomycetaceae	0.007	Alteromonadaceae	0.012	Unclassified	0.220
Clostridiaceae	1.21	Nocardiopsaceae	0.005	Unclassified	0.001	Unclassified	0.055	Clostridiaceae 2	0.058
Lachnospiraceae	17.7	Promicromonosporaceae	0.005	Hyphomonadaceae	0.001	Cardiobacteriaceae	0.005	Clostridiales Incertae Sedis XI	0.287
Peptostreptococcaceae	1.94	Propionibacteriaceae	0.055	Unclassified	0.005	Chromatiaceae	0.175	Clostridiales Incertae Sedis XIII	0.156
Ruminococcaceae	7.29	Pseudonocardiaceae	0.043	Bartonellaceae	0.002	Gammaproteobacteria family incertae sedis	0.005	Eubacteriaceae	0.084
Unclassified	3.70	Sporichthyaceae	0.001	Beijerinckiaceae	0.074	Coxiellaceae	0.005	Gracilibacteraceae	0.001
Erysipelotrichaceae	3.61	Streptomycetaceae	0.015	Bradyrhizobiaceae	0.216	Legionellaceae	0.011	Incertae Sedis XI	0.006
Veillonellaceae	2.02	Thermomonosporaceae	0.002	Brucellaceae	0.006	Alcanivoracaceae	0.001	Peptococcaceae 1	0.282
Unclassified	1.14	Unclassified	0.148	Methylobacteriaceae	0.179	Piscirickettsiaceae	0.002	Syntrophomonadaceae	0.002
Unclassified	1.33	Conexibacteraceae	0.070	Phyllobacteriaceae	0.004	Unclassified	0.238	Unclassified	0.141
Other	10.89	Solirubrobacteraceae	0.005	Rhodobiaceae	0.004	Unclassified	0.356	Acidaminococcaceae	0.151
		Unclassified	0.005	Xanthobacteraceae	0.113	Spirochaetaceae	0.025	Unclassified	0.006
		Thermoleophilaceae	0.003	Acetobacteraceae	0.174	Unclassified	0.000	OD1 family incertae sedis	0.001
		Unclassified	0.147	Rhodospirillaceae	0.469	Synergistaceae	0.010	OP11 family incertae sedis	0.001
		Aquificaceae	0.003	Unclassified	0.007	Bacillaceae 1	0.113	TM7 family incertae sedis	0.318
		Armatimonadetes gp5 family incetae sedis	0.007	Mitochondria	0.043	Bacillaceae 2	0.042	Unclassified	0.005
		Bacteroidetes incertae sedis family incertae sedis	0.001	Rickettsiaceae	0.000	Bacillales Incertae Sedis	0.005		

Table S1. Bacterial taxa sequenced from all faecal and meconium specimens (n = 305) with relative abundances > 0.5% and < 0.5% (continued)

Genus-level (> 0.5%)		Genus-level (< 0.5%)							
Taxon	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**
Blastococcus	0.54	Fervidicoccus	0.006	Sediminibacterium	0.062	Variovorax	0.003	Paenibacillus	0.007
Rothia	0.61	Unclassified	0.001	Unclassified	0.040	Duganella	0.004	Ammoniphilus	0.001
Bifidobacterium	6.61	Methanobrevibacter	0.074	Adhaeribacter	0.015	Herbaspirillum	0.018	Oxalophagus	0.002
Collinsella	0.89	Methanosphaera	0.026	Dyadobacter	0.031	Herminiimonas	0.001	Jeotgalibacillus	0.002
Olsenella	0.75	Methanothermobacter	0.002	Hymenobacter	0.004	Janthinobacterium	0.013	Lysinibacillus	0.001
Unclassified	1.32	Unclassified	0.002	Litoribacter	0.003	Massilia	0.325	Planomicrobium	0.051
Prevotella	2.08	Unclassified	0.001	Pontibacter	0.046	Oxalicibacterium	0.144	Ureibacillus	0.002
Bacteroides	1.05	Gp10	0.004	Unclassified	0.009	Oxalobacter	0.001	Unclassified	0.199
Brevundimonas	0.78	Gp16	0.001	Pedobacter	0.006	Undibacterium	0.085	Sporolactobacillaceae incertae sedis	0.026
Caulobacter	0.63	Gp3	0.000	Sphingobacterium	0.003	Unclassified	0.210	Jeotgalicoccus	0.024
Phenylobacterium	0.50	Gp4	0.006	Unclassified	0.071	Parasutterella	0.026	Macrococcus	0.000
Aurantimonas	1.00	Gp6	0.005	Capnocytophaga	0.003	Sutterella	0.149	Nosocomiicoccus	0.002
Rhizobium	0.72	Gp7	0.001	Chryseobacterium	0.008	Unclassified	0.000	Staphylococcus	0.298
Unclassified	0.82	Aciditerrimonas	0.001	Cloacibacterium	0.095	Unclassified	0.432	Unclassified	0.089
Sphingomonas	0.91	Ilumatobacter	0.002	Flavobacterium	0.189	Hydrogenophilus	0.001	Unclassified	0.096
Unclassified	0.55	Unclassified	0.002	Riemerella	0.008	Petrobacter	0.004	Aerococcus	0.005
Acidovorax	0.63	Unclassified	0.002	Unclassified	0.009	Unclassified	0.002	Facklamia	0.006
Unclassified	0.87	Actinobaculum	0.002	Unclassified	0.001	Iodobacter	0.000	Globicatella	0.008
Naxibacter	0.72	Actinomyces	0.257	Unclassified	0.265	Neisseria	0.062	Unclassified	0.003
Escherichia/Shigella	0.58	Arcanobacterium	0.003	Chlamydia	0.000	Vogesella	0.115	Atopostipes	0.020
Unclassified	2.44	Mobiluncus	0.003	Neochlamydia	0.001	Unclassified	0.011	Carnobacterium	0.003
Aeromonas	1.51	Trueperella	0.001	Parachlamydia	0.001	Unclassified	0.001	Desemzia	0.003
Shewanella	1.53	Varibaculum	0.034	Herpetosiphon	0.000	Azoarcus	0.006	Dolosigranulum	0.056
Acinetobacter	2.96	Unclassified	0.001	Sphaerobacter	0.058	Azospira	0.007	Granulicatella	0.042
Pseudomonas	1.30	Georgenia	0.001	Unclassified	0.001	Shinella	0.001	Isobaculum	0.002
Lysobacter	0.52	Brevibacterium	0.007	Unclassified	0.003	Zoogloea	0.001	Trichococcus	0.002

Table S1. Bacterial taxa sequenced from all faecal and meconium specimens (n = 305) with relative abundances > 0.5% and < 0.5% (continued)

Genus-level (> 0.5%)		Genus-level (< 0.5%)							
Taxon	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**
Enterococcus	1.13	Cellulomonas	0.009	Deinococcus	0.121	Unclassified	0.012	Unclassified	0.025
Lactobacillus	2.03	Unclassified	0.007	Truepera	0.002	Unclassified	0.029	Vagococcus	0.002
Weissella	0.80	Corynebacterium	0.104	Thermus	0.002	Bacteriovorax	0.011	Unclassified	0.140
Streptococcus	3.09	Demequina	0.001	Elusimicrobium	0.025	Peredibacter	0.103	Paralactobacillus	0.005
Unclassified	0.95	Brachybacterium	0.017	Cetobacterium	0.002	Bdellovibrio	0.070	Pediococcus	0.002
Clostridium sensu stricto	0.78	Dermabacter	0.007	Fusobacterium	0.059	Vampirovibrio	0.002	Unclassified	0.158
Blautia	3.01	Helcobacillus	0.001	Leptotrichia	0.011	Desulfomicrobium	0.002	Leuconostoc	0.250
Clostridium XlVa	0.86	Unclassified	0.000	Sneathia	0.007	Bilophila	0.049	Unclassified	0.041
Coprococcus	0.62	Unclassified	0.001	Streptobacillus	0.001	Desulfovibrio	0.128	Lactococcus	0.452
Dorea	2.17	Piscicoccus	0.007	Unclassified	0.004	Unclassified	0.002	Unclassified	0.041
Lachnospiracea incertae sedis	2.99	Dietzia	0.058	Gemmatimonas	0.007	Unclassified	0.019	Unclassified	0.220
Roseburia	0.52	Geodermatophilus	0.040	Victivallis	0.002	Cystobacter	0.001	Anaerosporobacter	0.009
Unclassified	6.95	Modestobacter	0.007	Unclassified	0.000	Unclassified	0.001	Proteiniclasticum	0.004
Clostridium XI	1.80	Unclassified	0.020	Nitrospira	0.007	Nannocystis	0.003	Sarcina	0.236
Faecalibacterium	1.17	Arsenicicoccus	0.000	Blastopirellula	0.002	Phaselicystis	0.004	Unclassified	0.180
Ruminococcus	1.19	Janibacter	0.018	Gemmata	0.001	Byssovorax	0.000	Natronincola	0.002
Unclassified	3.44	Ornithinimicrobium	0.018	Planctomyces	0.001	Unclassified	0.002	Unclassified	0.056
Unclassified	3.70	Unclassified	0.008	Singulisphaera	0.001	Unclassified	0.006	Anaerococcus	0.127
Catenibacterium	0.76	Kineococcus	0.000	Unclassified	0.002	Arcobacter	0.002	Finegoldia	0.069
Erysipelotrichaceae incertae sedis	1.90	Kineosporia	0.000	Unclassified	0.001	Campylobacter	0.049	Helcococcus	0.003
Veillonella	0.77	Quadrisphaera	0.001	Asticcacaulis	0.002	Helicobacter	0.025	Parvimonas	0.009
Unclassified	1.14	Unclassified	0.002	Unclassified	0.245	Buttiauxella	0.001	Peptoniphilus	0.066
Unclassified	1.33	Agrococcus	0.051	Hyphomonas	0.001	Citrobacter	0.073	Sedimentibacter	0.000
Other	24.00	Agromyces	0.001	Unclassified	0.005	Cosenzaea	0.000	Tissierella	0.002
		Frondihabitans	0.001	Unclassified	0.005	Cronobacter	0.005	Unclassified	0.011

Table S1. Bacterial taxa sequenced from all faecal and meconium specimens (n = 305) with relative abundances > 0.5% and < 0.5% (continued)

Genus-level (> 0.5%)		Genus-level (< 0.5%)							
Taxon	%	Taxon	%	Taxon	%	Taxon	%	Taxon	%
		Leucobacter	0.004	Bartonella	0.002	Enterobacter	0.001	Mogibacterium	0.087
		Microbacterium	0.004	Camelimonas	0.006	Klebsiella	0.013	Unclassified	0.069
		Rathayibacter	0.003	Chelatococcus	0.058	Morganella	0.009	Anaerofustis	0.004
		Unclassified	0.043	Unclassified	0.011	Pectobacterium	0.003	Eubacterium	0.063
		Arthrobacter	0.025	Afipia	0.001	Proteus	0.025	Pseudoramibacter	0.005
		Kocuria	0.089	Balneimonas	0.022	Providencia	0.003	Unclassified	0.013
		Micrococcus	0.059	Bosea	0.013	Raoultella	0.000	Gracilibacter	0.001
		Nesterenkonia	0.006	Bradyrhizobium	0.026	Salmonella	0.033	Anaerosphaera	0.001
		Renibacterium	0.001	Salinarimonas	0.127	Serratia	0.118	Murdochiella	0.005
		Yaniella	0.001	Unclassified	0.027	Photobacterium	0.003	Acetitomaculum	0.013
		Unclassified	0.029	Mycoplana	0.000	Vibrio	0.028	Anaerostipes	0.327
		Micromonospora	0.004	Unclassified	0.006	Succinatimonas	0.001	Catonella	0.003
		Unclassified	0.001	Devosia	0.344	Succinivibrio	0.161	Cellulosilyticum	0.003
		Mycobacterium	0.028	Gemmiger	0.246	Unclassified	0.000	Clostridium XlVb	0.085
		Gordonia	0.046	Hyphomicrobium	0.004	Alishewanella	0.009	Hespellia	0.009
		Millisia	0.000	Pedomicrobium	0.002	Haliea	0.001	Howardella	0.025
		Nocardia	0.002	Rhodoplanes	0.001	Marinobacter	0.003	Lachnobacterium	0.001
		Rhodococcus	0.004	Unclassified	0.001	Unclassified	0.055	Lactonifactor	0.014
		Williamsia	0.003	Methylobacterium	0.023	Cardiobacterium	0.005	Marvinbryantia	0.008
		Unclassified	0.000	Microvirga	0.127	Rheinheimera	0.175	Moryella	0.019
		Aeromicrobium	0.033	Unclassified	0.029	Methylonatrum	0.001	Oribacterium	0.064
		Kribbella	0.001	Aminobacter	0.001	Orbus	0.003	Parasporobacterium	0.000
		Marmoricola	0.018	Mesorhizobium	0.001	Thiohalorhabdus	0.001	Pseudobutyrivibrio	0.014
		Nocardioides	0.461	Unclassified	0.003	Aquicella	0.005	Robinsoniella	0.002
		Unclassified	0.013	Anderseniella	0.001	Legionella	0.011	Shuttleworthia	0.001
		Nocardiopsis	0.005	Parvibaculum	0.003	Alcanivorax	0.001	Syntrophococcus	0.002

Table S1. Bacterial taxa sequenced from all faecal and meconium specimens (n = 305) with relative abundances > 0.5% and < 0.5% (continued)

Genus-level (> 0.5%)		Genus-level (< 0.5%)							
Taxon	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**
		Cellulosimicrobium	0.001	Ancylobacter	0.004	Hahella	0.001	Peptococcus	0.279
		Isoptericola	0.002	Azorhizobium	0.003	Halomonas	0.028	Unclassified	0.003
		Promicromonospora	0.002	Xanthobacter	0.063	Actinobacillus	0.015	Peptostreptococcus	0.040
		Unclassified	0.000	Unclassified	0.043	Haemophilus	0.069	Sporacetigenium	0.000
		Friedmanniella	0.001	Amaricoccus	0.000	Unclassified	0.132	Unclassified	0.099
		Microlunatus	0.001	Catellibacterium	0.028	Alkanindiges	0.175	Acetanaerobacterium	0.004
		Propionibacterium	0.049	Haematobacter	0.000	Enhydrobacter	0.124	Acetivibrio	0.007
		Tessaracoccus	0.000	Pannonibacter	0.002	Moraxella	0.000	Anaerofilum	0.004
		Unclassified	0.004	Paracoccus	0.292	Perlucidibaca	0.003	Anaerotruncus	0.034
		Actinoalloteichus	0.001	Rhodobacter	0.055	Psychrobacter	0.006	Butyricicoccus	0.102
		Actinomycetospora	0.002	Rubellimicrobium	0.179	Unclassified	0.008	Clostridium III	0.030
		Prauserella	0.000	Unclassified	0.465	Azotobacter	0.002	Clostridium IV	0.469
		Pseudonocardia	0.008	Acetobacter	0.003	Cellvibrio	0.004	Fastidiosipila	0.001
		Saccharopolyspora	0.008	Craurococcus	0.005	Serpens	0.011	Flavonifractor	0.152
		Unclassified	0.024	Paracraurococcus	0.001	Unclassified	0.216	Hydrogenoanaerobacterium	0.003
		Sporichthya	0.001	Roseococcus	0.002	Unclassified	0.001	Oscillibacter	0.174
		Streptacidiphilus	0.002	Roseomonas	0.142	Methylophaga	0.002	Papillibacter	0.018
		Streptomyces	0.013	Rubritepida	0.002	Aquimonas	0.001	Pseudoflavonifractor	0.015
		Unclassified	0.002	Unclassified	0.017	Arenimonas	0.004	Saccharofermentans	0.001
		Unclassified	0.148	Azospirillum	0.145	Luteibacter	0.005	Sporobacter	0.239
		Alloscardovia	0.010	Caenispirillum	0.001	Luteimonas	0.002	Subdoligranulum	0.233
		Gardnerella	0.134	Desertibacter	0.020	Pseudoxanthomonas	0.026	Pelospora	0.000
		Parascardovia	0.000	Novispirillum	0.050	Silanimonas	0.004	Thermohydrogenium	0.002
		Scardovia	0.008	Rhodocista	0.156	Stenotrophomonas	0.004	Unclassified	0.000
		Unclassified	0.063	Skermanella	0.037	Thermomonas	0.001	Unclassified	0.141
		Asaccharobacter	0.014	Thalassospira	0.001	Xanthomonas	0.045	Allobaculum	0.000

Table S1. Bacterial taxa sequenced from all faecal and meconium specimens (n = 305) with relative abundances > 0.5% and < 0.5% (continued)

Genus-level (> 0.5%)		Genus-level (< 0.5%)							
Taxon	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**
		Atopobium	0.115	Unclassified	0.058	Unclassified	0.029	Bulleidia	0.022
		Cryptobacterium	0.000	Unclassified	0.007	Unclassified	0.238	Clostridium XVIII	0.338
		Eggerthella	0.142	Mitochondria genus incertae sedis	0.043	Unclassified	0.356	Coprobacillus	0.011
		Enterorhabdus	0.123	Rickettsia	0.000	Treponema	0.025	Erysipelothrix	0.004
		Gordonibacter	0.013	Unclassified	0.001	Unclassified	0.000	Holdemania	0.016
		Paraeggerthella	0.001	Altererythrobacter	0.004	Cloacibacillus	0.008	Solobacterium	0.022
		Slackia	0.124	Erythrobacter	0.001	Jonquetella	0.001	Turicibacter	0.279
		Conexibacter	0.070	Porphyrobacter	0.024	Pyramidobacter	0.000	Unclassified	0.266
		Solirubrobacter	0.005	Unclassified	0.001	Unclassified	0.001	Acidaminococcus	0.102
		Unclassified	0.005	Blastomonas	0.043	Asteroleplasma	0.002	Phascolarctobacterium	0.008
		Thermoleophilum	0.003	Novosphingobium	0.078	Mycoplasma	0.001	Succiniclasticum	0.002
		Unclassified	0.147	Sandaracinobacter	0.001	Ureaplasma	0.004	Unclassified	0.039
		Hydrogenobacter	0.003	Sphingobium	0.017	Unclassified	0.002	Allisonella	0.068
		Armatimonadetes gp5	0.007	Sphingopyxis	0.042	Cerasicoccus	0.005	Anaeroglobus	0.001
		Unclassified	0.001	Sphingosinicella	0.007	Unclassified	0.002	Anaerosinus	0.005
		Phocaeicola	0.000	Unclassified	0.130	5 genus incertae sedis	0.004	Centipeda	0.000
		Anaerophaga	0.026	Unclassified	0.101	Akkermansia	0.161	Dialister	0.341
		Unclassified	0.031	Alcaligenes	0.001	Prosthecobacter	0.002	Megamonas	0.197
		Barnesiella	0.048	Pusillimonas	0.001	Unclassified	0.003	Megasphaera	0.406
		Butyricimonas	0.062	Burkholderia	0.013	Chlorophyta	0.002	Mitsuokella	0.070
		Dysgonomonas	0.002	Chitinimonas	0.021	Streptophyta	0.145	Negativicoccus	0.024
		Odoribacter	0.029	Cupriavidus	0.177	Unclassified	0.003	Selenomonas	0.005
		Paludibacter	0.003	Limnobacter	0.051	GpIIa	0.005	Sporomusa	0.001
		Parabacteroides	0.376	Paucimonas	0.012	GpIV	0.002	Unclassified	0.127
		Porphyromonas	0.008	Polynucleobacter	0.006	GpV	0.001	Unclassified	0.006
		Tannerella	0.001	Ralstonia	0.143	Unclassified	0.006	OD1 genus incertae sedis	0.001

Table S1. Bacterial taxa sequenced from all faecal and meconium specimens (n = 305) with relative abundances > 0.5% and < 0.5% (continued)

Genus-level (> 0.5%)		Genus-level (< 0.5%)							
Taxon	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**	**Taxon**	**%**
		Unclassified	0.111	Unclassified	0.100	Tumebacillus	0.001	OP11 genus incertae sedis	0.001
		Hallella	0.005	Aquabacterium	0.333	Anoxybacillus	0.005	TM7 genus incertae sedis	0.318
		Paraprevotella	0.097	Inhella	0.004	Bacillus	0.078	Unclassified	0.005
		Xylanibacter	0.001	Sphaerotilus	0.002	Geobacillus	0.003		
		Unclassified	0.137	Tepidimonas	0.006	Unclassified	0.028		
		Alistipes	0.172	Unclassified	0.006	Alkalibacillus	0.001		
		Unclassified	0.002	Albidiferax	0.009	Marinococcus	0.001		
		Anaerorhabdus	0.002	Caldimonas	0.004	Terribacillus	0.002		
		Unclassified	0.241	Comamonas	0.006	Unclassified	0.039		
		Algoriphagus	0.011	Curvibacter	0.012	Thermicanus	0.005		
		Haliscomenobacter	0.002	Delftia	0.019	Gemella	0.077		
		XChitinophaga	0.001	Hydrogenophaga	0.202	Exiguobacterium	0.030		
		Ferruginibacter	0.001	Pelomonas	0.001	Caldalkalibacillus	0.001		
		Flavihumibacter	0.003	Polaromonas	0.005	Listeria	0.002		
		Flavisolibacter	0.007	Schlegelella	0.004	Unclassified	0.000		
		Parasegetibacter	0.017	Simplicispira	0.001	Cohnella	0.002		

Table S2. Phylum-level proportions of bacterial taxa with relative abundances > 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age

Phyla (> 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Actinobacteria	9.560	26.273	19.286	5.772
Bacteroidetes	4.310	5.425	4.384	7.064
Proteobacteria	59.671	18.721	11.759	4.714
Firmicutes	22.909	48.528	63.716	79.385
Unclassified	1.174	0.604	0.465	2.445
Other	2.376	0.450	0.391	0.621

Table S3. Phylum-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age

Phyla (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Crenarchaeota	0.020	0.000	0.000	0.000
Euryarchaeota	0.080	0.030	0.010	0.230
Acidobacteria	0.040	0.000	0.020	0.000
Aquificae	0.010	0.000	0.000	0.000
Armatimonadetes	0.020	0.000	0.000	0.000
Chlamydiae	0.010	0.000	0.000	0.000
Chloroflexi	0.170	0.000	0.000	0.000
Deinococcus Thermus	0.310	0.030	0.020	0.020
Elusimicrobia	0.020	0.010	0.000	0.050
Fusobacteria	0.160	0.060	0.020	0.040
Gemmatimonadetes	0.020	0.000	0.000	0.000
Lentisphaerae	0.000	0.000	0.000	0.010
Nitrospira	0.020	0.000	0.000	0.000
Planctomycetes	0.020	0.000	0.000	0.000
Spirochaetes	0.040	0.020	0.000	0.020
Synergistetes	0.010	0.000	0.000	0.030
Tenericutes	0.010	0.010	0.000	0.000
Verrucomicrobia	0.180	0.140	0.290	0.170
Cyanobacteria Chloroplast	0.450	0.010	0.000	0.000
OD1	0.000	0.000	0.000	0.000
OP11	0.000	0.000	0.000	0.000
TM7	0.770	0.140	0.030	0.030
Unclassified	0.010	0.000	0.000	0.010

Table S4. Genus-level proportions of bacterial taxa with relative abundances > 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age

Genera (> 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Blastococcus	1.010	0.480	0.360	0.100
Rothia	0.430	1.220	0.990	0.180
Bifidobacterium	1.710	16.880	12.660	1.790
Collinsella	0.490	1.700	1.200	0.570
Olsenella	0.510	1.210	0.440	0.800
Unclassified	0.640	2.270	1.320	1.350
Prevotella	1.280	2.540	1.500	2.890
Bacteroides	0.520	1.420	1.820	1.060
Brevundimonas	2.130	0.070	0.090	0.010
Caulobacter	1.060	0.700	0.490	0.110
Phenylobacterium	1.390	0.040	0.050	0.010
Aurantimonas	2.360	0.530	0.240	0.070
Rhizobium	1.510	0.530	0.340	0.070
Unclassified	1.940	0.340	0.300	0.090
Sphingomonas	2.030	0.560	0.300	0.100
Unclassified	1.340	0.140	0.200	0.070
Acidovorax	1.100	0.670	0.400	0.120
Unclassified	2.060	0.400	0.200	0.090
Naxibacter	1.530	0.520	0.350	0.080
Escherichia/Shigella	0.910	0.760	0.490	0.080
Unclassified	3.260	3.840	2.410	0.360
Aeromonas	3.560	0.760	0.400	0.130
Shewanella	3.240	1.060	0.660	0.210
Acinetobacter	6.280	1.980	0.980	0.580
Pseudomonas	2.650	1.040	0.560	0.210
Lysobacter	1.290	0.200	0.080	0.050
Enterococcus	1.010	2.160	1.740	0.200
Lactobacillus	0.720	5.730	3.280	0.460
Weissella	0.480	1.680	1.640	0.160
Streptococcus	1.420	7.830	5.060	0.500
Unclassified	0.360	2.480	1.540	0.180
Clostridium sensu stricto	0.850	0.600	0.890	0.790
Blautia	1.220	1.720	3.560	5.950
Clostridium XlVa	0.390	0.940	1.810	0.970
Coprococcus	0.350	0.360	0.500	1.210
Dorea	1.020	1.820	3.760	3.180
Lachnospiracea incertae sedis	0.650	2.850	6.670	4.410
Roseburia	0.140	0.020	0.210	1.500
Unclassified	1.770	3.240	11.100	14.410
Clostridium XI	1.370	1.190	2.390	2.590

Table S4. Genus-level proportions of bacterial taxa with relative abundances > 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age (continued)

Genera (> 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Faecalibacterium	0.350	0.400	0.470	3.040
Ruminococcus	0.340	0.030	0.210	3.510
Unclassified	0.880	0.220	0.870	10.090
Unclassified	1.130	1.010	1.520	9.760
Catenibacterium	0.630	0.800	0.700	0.900
Erysipelotrichaceae incertae sedis	0.900	2.200	2.840	2.460
Veillonella	0.280	1.940	1.540	0.120
Unclassified	0.400	0.690	0.720	2.530
Unclassified	1.170	0.600	0.460	2.440
Other	35.960	17.610	17.650	17.450

Table S5. Genus-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age

Genera (< 0.5%)	Relative abundance (%)				Genera (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth		Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Fervidicoccus	0.020	0.000	0.000	0.000	Brevibacterium	0.020	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Cellulomonas	0.030	0.000	0.000	0.000
Methanobrevibacter	0.070	0.030	0.000	0.140	Unclassified	0.020	0.000	0.000	0.000
Methanosphaera	0.010	0.000	0.000	0.080	Corynebacterium	0.280	0.000	0.030	0.000
Methanothermobacter	0.000	0.000	0.000	0.000	Demequina	0.000	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.010	Brachybacterium	0.050	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Dermabacter	0.010	0.000	0.040	0.000
Gp10	0.010	0.000	0.000	0.000	Helcobacillus	0.000	0.000	0.000	0.000
Gp16	0.000	0.000	0.000	0.000	Unclassified	0.000	0.000	0.000	0.000
Gp3	0.000	0.000	0.000	0.000	Unclassified	0.000	0.000	0.000	0.000
Gp4	0.010	0.000	0.020	0.000	Piscicoccus	0.010	0.000	0.020	0.000
Gp6	0.010	0.000	0.000	0.000	Dietzia	0.110	0.060	0.020	0.010
Gp7	0.000	0.000	0.000	0.000	Geodermatophilus	0.100	0.010	0.000	0.010
Aciditerrimonas	0.000	0.000	0.000	0.000	Modestobacter	0.020	0.000	0.000	0.000
Ilumatobacter	0.010	0.000	0.000	0.000	Unclassified	0.060	0.000	0.000	0.000
Unclassified	0.010	0.000	0.000	0.000	Arsenicicoccus	0.000	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Janibacter	0.040	0.010	0.000	0.000
Actinobaculum	0.000	0.010	0.000	0.000	Ornithinimicrobium	0.050	0.000	0.000	0.000
Actinomyces	0.100	0.480	0.470	0.180	Unclassified	0.020	0.000	0.000	0.000
Arcanobacterium	0.000	0.000	0.000	0.000	Kineococcus	0.000	0.000	0.000	0.000
Mobiluncus	0.000	0.000	0.000	0.010	Kineosporia	0.000	0.000	0.000	0.000
Trueperella	0.000	0.000	0.000	0.000	Quadrisphaera	0.000	0.000	0.000	0.000
Varibaculum	0.000	0.110	0.050	0.010	Unclassified	0.010	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Agrococcus	0.130	0.020	0.020	0.000
Georgenia	0.000	0.000	0.000	0.000	Agromyces	0.000	0.000	0.000	0.000

Table S5. Genus-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age (continued)

Genera (< 0.5%)	Relative abundance (%)				Genera (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth		Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Frondihabitans	0.000	0.000	0.000	0.000	Unclassified	0.040	0.000	0.000	0.000
Leucobacter	0.010	0.000	0.000	0.000	Nocardiopsis	0.010	0.010	0.000	0.000
Microbacterium	0.010	0.000	0.000	0.000	Cellulosimicrobium	0.000	0.000	0.000	0.000
Rathayibacter	0.010	0.000	0.000	0.000	Isoptericola	0.000	0.000	0.000	0.000
Unclassified	0.120	0.000	0.000	0.000	Promicromonospora	0.010	0.000	0.000	0.000
Arthrobacter	0.070	0.000	0.000	0.000	Unclassified	0.000	0.000	0.000	0.000
Kocuria	0.140	0.100	0.120	0.010	Friedmanniella	0.000	0.000	0.000	0.000
Micrococcus	0.130	0.030	0.000	0.020	Microlunatus	0.000	0.000	0.000	0.000
Nesterenkonia	0.020	0.000	0.000	0.000	Propionibacterium	0.130	0.010	0.020	0.000
Renibacterium	0.000	0.000	0.000	0.000	Tessaracoccus	0.000	0.000	0.000	0.000
Yaniella	0.000	0.000	0.000	0.000	Unclassified	0.010	0.000	0.000	0.000
Unclassified	0.060	0.030	0.010	0.000	Actinoalloteichus	0.000	0.000	0.000	0.000
Micromonospora	0.010	0.000	0.000	0.000	Actinomycetospora	0.010	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Prauserella	0.000	0.000	0.000	0.000
Mycobacterium	0.080	0.000	0.000	0.000	Pseudonocardia	0.020	0.000	0.000	0.000
Gordonia	0.120	0.010	0.000	0.000	Saccharopolyspora	0.020	0.010	0.000	0.000
Millisia	0.000	0.000	0.000	0.000	Unclassified	0.060	0.020	0.000	0.000
Nocardia	0.000	0.000	0.020	0.000	Sporichthya	0.000	0.000	0.000	0.000
Rhodococcus	0.010	0.000	0.000	0.000	Streptacidiphilus	0.010	0.000	0.000	0.000
Williamsia	0.000	0.010	0.000	0.000	Streptomyces	0.040	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Unclassified	0.000	0.000	0.000	0.000
Aeromicrobium	0.090	0.000	0.020	0.000	Unclassified	0.270	0.140	0.070	0.040
Kribbella	0.000	0.000	0.000	0.000	Alloscardovia	0.000	0.030	0.000	0.010
Marmoricola	0.040	0.010	0.000	0.000	Gardnerella	0.300	0.070	0.020	0.030
Nocardioides	1.180	0.120	0.060	0.040	Parascardovia	0.000	0.000	0.000	0.000

Table S5. Genus-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age (continued)

Genera (< 0.5%)	Relative abundance (%)				Genera (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth		Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Scardovia	0.000	0.020	0.000	0.010	Paludibacter	0.010	0.000	0.000	0.000
Unclassified	0.000	0.180	0.110	0.020	Parabacteroides	0.280	0.600	0.430	0.290
Asaccharobacter	0.000	0.000	0.030	0.030	Porphyromonas	0.020	0.000	0.000	0.000
Atopobium	0.050	0.290	0.180	0.020	Tannerella	0.000	0.000	0.000	0.000
Cryptobacterium	0.000	0.000	0.000	0.000	Unclassified	0.120	0.040	0.040	0.180
Eggerthella	0.040	0.240	0.570	0.020	Hallella	0.010	0.000	0.000	0.000
Enterorhabdus	0.070	0.080	0.120	0.210	Paraprevotella	0.040	0.000	0.020	0.270
Gordonibacter	0.020	0.010	0.020	0.010	Xylanibacter	0.000	0.000	0.000	0.000
Paraeggerthella	0.000	0.000	0.000	0.000	Unclassified	0.110	0.100	0.030	0.240
Slackia	0.030	0.160	0.170	0.180	Alistipes	0.090	0.100	0.190	0.320
Conexibacter	0.180	0.000	0.040	0.000	Unclassified	0.000	0.000	0.000	0.000
Solirubrobacter	0.010	0.000	0.000	0.000	Anaerorhabdus	0.000	0.000	0.000	0.010
Unclassified	0.010	0.000	0.000	0.000	Unclassified	0.090	0.150	0.100	0.540
Thermoleophilum	0.010	0.000	0.000	0.000	Algoriphagus	0.020	0.000	0.020	0.000
Unclassified	0.180	0.230	0.080	0.070	Haliscomenobacter	0.010	0.000	0.000	0.000
Hydrogenobacter	0.010	0.000	0.000	0.000	Chitinophaga	0.000	0.000	0.000	0.000
Armatimonadetes gp5	0.020	0.000	0.000	0.000	Ferruginibacter	0.000	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Flavihumibacter	0.010	0.000	0.000	0.000
Phocaeicola	0.000	0.000	0.000	0.000	Flavisolibacter	0.020	0.000	0.000	0.000
Anaerophaga	0.020	0.030	0.000	0.040	Parasegetibacter	0.050	0.000	0.000	0.000
Unclassified	0.020	0.010	0.000	0.070	Sediminibacterium	0.160	0.000	0.030	0.000
Barnesiella	0.010	0.020	0.010	0.140	Unclassified	0.090	0.020	0.020	0.000
Butyricimonas	0.010	0.020	0.040	0.160	Adhaeribacter	0.020	0.030	0.000	0.000
Dysgonomonas	0.000	0.010	0.000	0.000	Dyadobacter	0.080	0.000	0.030	0.000
Odoribacter	0.010	0.020	0.030	0.060	Hymenobacter	0.010	0.000	0.000	0.000

Table S5. Genus-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age (continued)

Genera (< 0.5%)	Relative abundance (%)				Genera (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth		Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Litoribacter	0.010	0.000	0.000	0.000	Cetobacterium	0.000	0.000	0.000	0.010
Pontibacter	0.120	0.020	0.000	0.000	Fusobacterium	0.110	0.040	0.020	0.030
Unclassified	0.030	0.000	0.000	0.000	Leptotrichia	0.030	0.010	0.000	0.000
Pedobacter	0.020	0.000	0.000	0.000	Sneathia	0.010	0.020	0.000	0.000
Sphingobacterium	0.010	0.000	0.000	0.000	Streptobacillus	0.000	0.000	0.000	0.000
Unclassified	0.100	0.020	0.000	0.100	Unclassified	0.010	0.000	0.000	0.000
Capnocytophaga	0.010	0.000	0.000	0.000	Gemmatimonas	0.020	0.000	0.000	0.000
Chryseobacterium	0.020	0.000	0.000	0.000	Victivallis	0.000	0.000	0.000	0.010
Cloacibacterium	0.220	0.050	0.030	0.010	Unclassified	0.000	0.000	0.000	0.000
Flavobacterium	0.450	0.100	0.030	0.020	Nitrospira	0.020	0.000	0.000	0.000
Riemerella	0.020	0.000	0.000	0.000	Blastopirellula	0.010	0.000	0.000	0.000
Unclassified	0.020	0.010	0.000	0.000	Gemmata	0.000	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Planctomyces	0.000	0.000	0.000	0.000
Unclassified	0.140	0.120	0.020	0.630	Singulisphaera	0.000	0.000	0.000	0.000
Chlamydia	0.000	0.000	0.000	0.000	Unclassified	0.010	0.000	0.000	0.000
Neochlamydia	0.000	0.000	0.000	0.000	Unclassified	0.000	0.000	0.000	0.000
Parachlamydia	0.000	0.000	0.000	0.000	Asticcacaulis	0.000	0.000	0.000	0.000
Herpetosiphon	0.000	0.000	0.000	0.000	Unclassified	0.630	0.070	0.040	0.010
Sphaerobacter	0.160	0.000	0.000	0.000	Hyphomonas	0.000	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Unclassified	0.010	0.000	0.000	0.000
Unclassified	0.010	0.000	0.000	0.000	Unclassified	0.010	0.000	0.000	0.000
Deinococcus	0.300	0.030	0.020	0.020	Bartonella	0.010	0.000	0.000	0.000
Truepera	0.000	0.000	0.000	0.000	Camelimonas	0.020	0.000	0.000	0.000
Thermus	0.010	0.000	0.000	0.000	Chelatococcus	0.140	0.020	0.000	0.020
Elusimicrobium	0.020	0.010	0.000	0.050	Unclassified	0.030	0.000	0.000	0.000

Table S5. Genus-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age (continued)

Genera (< 0.5%)	Relative abundance (%)				Genera (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth		Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Afipia	0.000	0.000	0.000	0.000	Unclassified	0.110	0.020	0.010	0.000
Balneimonas	0.050	0.010	0.000	0.000	Amaricoccus	0.000	0.000	0.000	0.000
Bosea	0.040	0.000	0.000	0.000	Catellibacterium	0.080	0.000	0.000	0.000
Bradyrhizobium	0.080	0.000	0.000	0.000	Haematobacter	0.000	0.000	0.000	0.000
Salinarimonas	0.270	0.100	0.040	0.010	Pannonibacter	0.010	0.000	0.000	0.000
Unclassified	0.080	0.000	0.000	0.000	Paracoccus	0.730	0.090	0.060	0.030
Mycoplana	0.000	0.000	0.000	0.000	Rhodobacter	0.150	0.000	0.000	0.010
Unclassified	0.010	0.000	0.000	0.000	Rubellimicrobium	0.470	0.040	0.000	0.010
Devosia	0.710	0.230	0.210	0.060	Unclassified	0.930	0.360	0.260	0.080
Gemmiger	0.160	0.200	0.190	0.410	Acetobacter	0.000	0.010	0.000	0.000
Hyphomicrobium	0.010	0.010	0.000	0.000	Craurococcus	0.010	0.000	0.000	0.000
Pedomicrobium	0.000	0.000	0.000	0.000	Paracraurococcus	0.000	0.000	0.000	0.000
Rhodoplanes	0.000	0.000	0.000	0.000	Roseococcus	0.010	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Roseomonas	0.330	0.070	0.050	0.020
Methylobacterium	0.070	0.000	0.000	0.000	Rubritepida	0.010	0.000	0.000	0.000
Microvirga	0.290	0.050	0.030	0.040	Unclassified	0.040	0.000	0.020	0.000
Unclassified	0.070	0.020	0.000	0.000	Azospirillum	0.360	0.070	0.020	0.000
Aminobacter	0.000	0.000	0.000	0.000	Caenispirillum	0.000	0.000	0.000	0.000
Mesorhizobium	0.000	0.000	0.000	0.000	Desertibacter	0.040	0.030	0.000	0.000
Unclassified	0.010	0.000	0.000	0.000	Novispirillum	0.140	0.000	0.000	0.000
Anderseniella	0.000	0.000	0.000	0.000	Rhodocista	0.410	0.040	0.010	0.010
Parvibaculum	0.010	0.000	0.000	0.000	Skermanella	0.070	0.030	0.020	0.000
Ancylobacter	0.000	0.010	0.000	0.000	Thalassospira	0.000	0.000	0.000	0.000
Azorhizobium	0.010	0.000	0.000	0.000	Unclassified	0.140	0.030	0.000	0.010
Xanthobacter	0.160	0.030	0.000	0.000	Unclassified	0.020	0.000	0.000	0.000

Table S5. Genus-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age (continued)

Genera (< 0.5%)	Relative abundance (%)				Genera (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth		Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Mitochondria genus incertae sedis	0.110	0.000	0.040	0.000	Unclassified	0.210	0.080	0.030	0.010
Rickettsia	0.000	0.000	0.000	0.000	Aquabacterium	0.780	0.170	0.090	0.030
Unclassified	0.000	0.000	0.000	0.000	Inhella	0.010	0.000	0.000	0.000
Altererythrobacter	0.010	0.000	0.000	0.000	Sphaerotilus	0.000	0.000	0.000	0.000
Erythrobacter	0.000	0.000	0.000	0.000	Tepidimonas	0.020	0.000	0.000	0.000
Porphyrobacter	0.070	0.000	0.000	0.000	Unclassified	0.010	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Albidiferax	0.020	0.000	0.000	0.000
Blastomonas	0.120	0.010	0.000	0.000	Caldimonas	0.010	0.000	0.000	0.000
Novosphingobium	0.190	0.030	0.020	0.010	Comamonas	0.020	0.000	0.000	0.000
Sandaracinobacter	0.000	0.000	0.000	0.000	Curvibacter	0.030	0.000	0.000	0.000
Sphingobium	0.050	0.000	0.000	0.000	Delftia	0.060	0.000	0.000	0.000
Sphingopyxis	0.110	0.010	0.000	0.000	Hydrogenophaga	0.480	0.120	0.030	0.010
Sphingosinicella	0.020	0.000	0.000	0.000	Pelomonas	0.000	0.000	0.000	0.000
Unclassified	0.320	0.040	0.020	0.020	Polaromonas	0.010	0.010	0.000	0.000
Unclassified	0.200	0.110	0.020	0.010	Schlegelella	0.010	0.000	0.000	0.000
Alcaligenes	0.000	0.000	0.000	0.000	Simplicispira	0.000	0.000	0.000	0.000
Pusillimonas	0.000	0.000	0.000	0.000	Variovorax	0.010	0.000	0.000	0.000
Burkholderia	0.040	0.000	0.000	0.000	Duganella	0.010	0.000	0.000	0.000
Chitinimonas	0.050	0.010	0.000	0.010	Herbaspirillum	0.050	0.000	0.000	0.000
Cupriavidus	0.420	0.050	0.060	0.030	Herminiimonas	0.000	0.000	0.000	0.000
Limnobacter	0.120	0.020	0.020	0.000	Janthinobacterium	0.040	0.000	0.000	0.000
Paucimonas	0.020	0.010	0.020	0.010	Massilia	0.770	0.170	0.030	0.040
Polynucleobacter	0.020	0.000	0.000	0.000	Oxalicibacterium	0.270	0.090	0.070	0.060
Ralstonia	0.380	0.030	0.020	0.000	Oxalobacter	0.000	0.000	0.000	0.000

Table S5. Genus-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age (continued)

Genera (< 0.5%)	Relative abundance (%)				Genera (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth		Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Undibacterium	0.220	0.020	0.000	0.000	Bilophila	0.020	0.050	0.020	0.090
Unclassified	0.510	0.040	0.050	0.050	Desulfovibrio	0.090	0.110	0.080	0.200
Parasutterella	0.020	0.010	0.050	0.030	Unclassified	0.000	0.000	0.000	0.010
Sutterella	0.110	0.190	0.240	0.130	Unclassified	0.010	0.000	0.000	0.050
Unclassified	0.000	0.000	0.000	0.000	Cystobacter	0.000	0.000	0.000	0.000
Unclassified	0.980	0.210	0.130	0.080	Unclassified	0.000	0.000	0.000	0.000
Hydrogenophilus	0.000	0.000	0.000	0.000	Nannocystis	0.000	0.010	0.000	0.000
Petrobacter	0.010	0.000	0.000	0.000	Phaselicystis	0.010	0.000	0.000	0.000
Unclassified	0.010	0.000	0.000	0.000	Byssovorax	0.000	0.000	0.000	0.000
Iodobacter	0.000	0.000	0.000	0.000	Unclassified	0.010	0.000	0.000	0.000
Vogesella	0.300	0.020	0.000	0.020	Unclassified	0.000	0.000	0.000	0.020
Unclassified	0.020	0.010	0.000	0.000	Arcobacter	0.000	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Campylobacter	0.050	0.080	0.060	0.010
Azoarcus	0.020	0.000	0.000	0.000	Helicobacter	0.040	0.010	0.070	0.000
Azospira	0.020	0.000	0.000	0.000	Buttiauxella	0.000	0.000	0.000	0.000
Shinella	0.000	0.000	0.000	0.000	Citrobacter	0.130	0.100	0.030	0.000
Zoogloea	0.000	0.000	0.000	0.000	Cosenzaea	0.000	0.000	0.000	0.000
Unclassified	0.030	0.000	0.000	0.000	Cronobacter	0.010	0.000	0.010	0.000
Unclassified	0.050	0.020	0.000	0.030	Enterobacter	0.000	0.000	0.000	0.000
Bacteriovorax	0.020	0.000	0.010	0.000	Klebsiella	0.030	0.000	0.010	0.000
Peredibacter	0.290	0.000	0.000	0.010	Morganella	0.000	0.000	0.070	0.000
Bdellovibrio	0.180	0.030	0.000	0.000	Pectobacterium	0.010	0.000	0.000	0.000
Vampirovibrio	0.000	0.000	0.000	0.010	Proteus	0.020	0.040	0.060	0.000
Desulfomicrobium	0.000	0.000	0.000	0.000	Providencia	0.010	0.000	0.000	0.000
Neisseria	0.180	0.000	0.000	0.000	Raoultella	0.000	0.000	0.000	0.000

Table S5. Genus-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age (continued)

Genera (< 0.5%)	Relative abundance (%)				Genera (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth		Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Salmonella	0.070	0.020	0.040	0.000	Unclassified	0.240	0.070	0.150	0.050
Serratia	0.230	0.100	0.100	0.010	Alkanindiges	0.400	0.090	0.040	0.020
Photobacterium	0.010	0.000	0.000	0.000	Enhydrobacter	0.290	0.050	0.040	0.020
Vibrio	0.080	0.000	0.000	0.000	Moraxella	0.000	0.000	0.000	0.000
Succinatimonas	0.000	0.000	0.000	0.000	Perlucidibaca	0.010	0.000	0.000	0.000
Succinivibrio	0.200	0.130	0.160	0.140	Psychrobacter	0.020	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Unclassified	0.010	0.020	0.000	0.000
Alishewanella	0.020	0.000	0.000	0.000	Azotobacter	0.010	0.000	0.000	0.000
Haliea	0.000	0.000	0.000	0.000	Cellvibrio	0.010	0.000	0.000	0.000
Marinobacter	0.010	0.000	0.000	0.000	Serpens	0.030	0.000	0.000	0.000
Unclassified	0.130	0.020	0.020	0.000	Unclassified	0.590	0.010	0.040	0.000
Cardiobacterium	0.020	0.000	0.000	0.000	Unclassified	0.000	0.000	0.000	0.000
Rheinheimera	0.440	0.070	0.020	0.010	Methylophaga	0.010	0.000	0.000	0.000
Methylonatrum	0.000	0.000	0.000	0.000	Aquimonas	0.000	0.000	0.000	0.000
Orbus	0.010	0.000	0.000	0.000	Arenimonas	0.000	0.010	0.000	0.000
Thiohalorhabdus	0.000	0.000	0.000	0.000	Luteibacter	0.010	0.000	0.010	0.000
Aquicella	0.010	0.000	0.000	0.000	Luteimonas	0.010	0.000	0.000	0.000
Legionella	0.030	0.000	0.000	0.000	Pseudoxanthomonas	0.050	0.010	0.030	0.000
Alcanivorax	0.000	0.000	0.000	0.000	Silanimonas	0.010	0.000	0.000	0.000
Hahella	0.000	0.000	0.000	0.000	Stenotrophomonas	0.010	0.000	0.000	0.000
Halomonas	0.080	0.000	0.000	0.000	Thermomonas	0.000	0.000	0.000	0.000
Actinobacillus	0.040	0.000	0.000	0.000	Xanthomonas	0.090	0.040	0.010	0.010
Haemophilus	0.190	0.000	0.000	0.000	Unclassified	0.070	0.010	0.010	0.010

Table S5. Genus-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age (continued)

Genera (< 0.5%)	Relative abundance (%)				Genera (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth		Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Unclassified	0.470	0.210	0.120	0.020	Tumebacillus	0.000	0.000	0.000	0.000
Unclassified	0.640	0.220	0.140	0.210	Anoxybacillus	0.010	0.000	0.000	0.000
Treponema	0.040	0.020	0.000	0.020	Bacillus	0.190	0.040	0.000	0.010
Unclassified	0.000	0.000	0.000	0.000	Geobacillus	0.000	0.010	0.000	0.000
Cloacibacillus	0.000	0.000	0.000	0.020	Unclassified	0.060	0.020	0.000	0.010
Jonquetella	0.000	0.000	0.000	0.000	Alkalibacillus	0.000	0.000	0.000	0.000
Pyramidobacter	0.000	0.000	0.000	0.000	Marinococcus	0.000	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.000	Terribacillus	0.000	0.000	0.000	0.000
Asteroleplasma	0.000	0.010	0.000	0.000	Unclassified	0.090	0.010	0.020	0.000
Mycoplasma	0.000	0.000	0.000	0.000	Thermicanus	0.010	0.000	0.000	0.000
Ureaplasma	0.010	0.000	0.000	0.000	Gemella	0.130	0.060	0.090	0.020
Unclassified	0.000	0.000	0.000	0.000	Exiguobacterium	0.090	0.000	0.000	0.000
Cerasicoccus	0.000	0.000	0.000	0.020	Caldalkalibacillus	0.000	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.010	Listeria	0.010	0.000	0.000	0.000
5 genus incertae sedis	0.010	0.000	0.000	0.000	Unclassified	0.000	0.000	0.000	0.000
Akkermansia	0.150	0.140	0.290	0.140	Cohnella	0.000	0.000	0.000	0.000
Prosthecobacter	0.010	0.000	0.000	0.000	Paenibacillus	0.020	0.000	0.000	0.000
Unclassified	0.010	0.000	0.000	0.000	Ammoniphilus	0.000	0.000	0.000	0.000
Chlorophyta	0.000	0.000	0.000	0.000	Oxalophagus	0.010	0.000	0.000	0.000
Streptophyta	0.400	0.010	0.000	0.000	Jeotgalibacillus	0.000	0.000	0.000	0.000
Unclassified	0.010	0.000	0.000	0.000	Lysinibacillus	0.000	0.000	0.000	0.000
GpIIa	0.010	0.000	0.000	0.000	Planomicrobium	0.150	0.000	0.000	0.000
GpIV	0.010	0.000	0.000	0.000	Ureibacillus	0.000	0.000	0.000	0.000
GpV	0.000	0.000	0.000	0.000	Unclassified	0.360	0.170	0.150	0.050
Unclassified	0.020	0.000	0.000	0.000	Sporolactobacillaceae incertae sedis	0.050	0.000	0.040	0.010

Table S5. Genus-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age (continued)

Genera (< 0.5%)	Relative abundance (%)				Genera (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth		Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Jeotgalicoccus	0.070	0.000	0.000	0.000	Lactococcus	0.200	0.990	1.070	0.070
Macrococcus	0.000	0.000	0.000	0.000	Unclassified	0.010	0.110	0.080	0.010
Nosocomiicoccus	0.000	0.000	0.010	0.000	Unclassified	0.220	0.390	0.320	0.040
Staphylococcus	0.530	0.330	0.220	0.020	Anaerosporobacter	0.000	0.000	0.000	0.030
Unclassified	0.160	0.110	0.060	0.000	Proteiniclasticum	0.010	0.000	0.020	0.000
Unclassified	0.230	0.040	0.000	0.010	Sarcina	0.230	0.270	0.160	0.250
Aerococcus	0.010	0.000	0.000	0.000	Unclassified	0.150	0.130	0.090	0.290
Facklamia	0.020	0.000	0.000	0.000	Natronincola	0.000	0.010	0.000	0.000
Globicatella	0.010	0.000	0.020	0.010	Unclassified	0.020	0.030	0.010	0.140
Unclassified	0.010	0.000	0.000	0.000	Anaerococcus	0.200	0.140	0.130	0.030
Atopostipes	0.060	0.000	0.000	0.000	Finegoldia	0.060	0.130	0.080	0.020
Carnobacterium	0.010	0.000	0.000	0.000	Helcococcus	0.010	0.000	0.000	0.000
Desemzia	0.010	0.000	0.000	0.000	Parvimonas	0.010	0.010	0.000	0.010
Dolosigranulum	0.100	0.050	0.060	0.010	Peptoniphilus	0.020	0.170	0.090	0.020
Granulicatella	0.030	0.050	0.100	0.020	Sedimentibacter	0.000	0.000	0.000	0.000
Isobaculum	0.000	0.000	0.000	0.000	Tissierella	0.000	0.000	0.000	0.000
Trichococcus	0.010	0.000	0.000	0.000	Unclassified	0.010	0.010	0.010	0.010
Unclassified	0.000	0.070	0.030	0.010	Mogibacterium	0.070	0.030	0.080	0.160
Vagococcus	0.010	0.000	0.000	0.000	Unclassified	0.020	0.020	0.000	0.200
Unclassified	0.080	0.320	0.260	0.010	Anaerofustis	0.000	0.000	0.000	0.010
Paralactobacillus	0.000	0.020	0.010	0.000	Eubacterium	0.010	0.100	0.070	0.090
Pediococcus	0.000	0.000	0.000	0.000	Pseudoramibacter	0.000	0.000	0.040	0.000
Unclassified	0.030	0.360	0.260	0.110	Unclassified	0.000	0.000	0.090	0.010
Leuconostoc	0.140	0.440	0.700	0.040	Gracilibacter	0.000	0.000	0.000	0.000
Unclassified	0.000	0.090	0.150	0.000	Anaerosphaera	0.000	0.000	0.000	0.000

Table S5. Genus-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age (continued)

Genera (< 0.5%)	Relative abundance (%)				Genera (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth		Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Murdochiella	0.000	0.010	0.000	0.010	Anaerofilum	0.000	0.000	0.000	0.010
Acetitomaculum	0.000	0.000	0.010	0.040	Anaerotruncus	0.010	0.020	0.020	0.080
Anaerostipes	0.080	0.020	0.700	0.720	Butyricicoccus	0.020	0.010	0.220	0.230
Catonella	0.010	0.000	0.000	0.000	Clostridium III	0.020	0.000	0.020	0.070
Cellulosilyticum	0.010	0.000	0.000	0.000	Clostridium IV	0.150	0.210	0.250	1.140
Clostridium XlVb	0.020	0.140	0.100	0.110	Fastidiosipila	0.000	0.000	0.000	0.000
Hespellia	0.010	0.000	0.000	0.020	Flavonifractor	0.050	0.210	0.490	0.100
Howardella	0.010	0.010	0.010	0.070	Hydrogenoanaerobacterium	0.000	0.000	0.000	0.010
Lachnobacterium	0.000	0.000	0.000	0.000	Oscillibacter	0.040	0.020	0.020	0.520
Lactonifactor	0.000	0.010	0.020	0.030	Papillibacter	0.000	0.000	0.000	0.060
Marvinbryantia	0.000	0.020	0.000	0.010	Pseudoflavonifractor	0.000	0.000	0.000	0.050
Moryella	0.000	0.030	0.060	0.010	Saccharofermentans	0.000	0.000	0.000	0.000
Oribacterium	0.070	0.030	0.040	0.100	Sporobacter	0.070	0.030	0.040	0.680
Parasporobacterium	0.000	0.000	0.000	0.000	Subdoligranulum	0.050	0.180	0.210	0.500
Pseudobutyrivibrio	0.010	0.010	0.010	0.030	Pelospora	0.000	0.000	0.000	0.000
Robinsoniella	0.000	0.000	0.000	0.010	Thermohydrogenium	0.010	0.000	0.000	0.000
Shuttleworthia	0.000	0.000	0.000	0.000	Unclassified	0.000	0.000	0.000	0.000
Syntrophococcus	0.000	0.000	0.020	0.000	Unclassified	0.050	0.040	0.010	0.380
Peptococcus	0.110	0.120	0.150	0.650	Allobaculum	0.000	0.000	0.000	0.000
Unclassified	0.000	0.000	0.000	0.010	Bulleidia	0.010	0.030	0.000	0.040
Peptostreptococcus	0.020	0.070	0.080	0.030	Clostridium XVIII	0.220	0.490	0.760	0.190
Sporacetigenium	0.000	0.000	0.000	0.000	Coprobacillus	0.000	0.000	0.050	0.020
Unclassified	0.090	0.020	0.080	0.180	Erysipelothrix	0.010	0.000	0.000	0.000
Acetanaerobacterium	0.000	0.000	0.000	0.010	Holdemania	0.000	0.000	0.000	0.050
Acetivibrio	0.000	0.000	0.000	0.020	Solobacterium	0.010	0.020	0.070	0.020

Table S5. Genus-level proportions of bacterial taxa with relative abundances < 0.5% sequenced from infant meconium specimens (n = 107) and maternal faecal specimens (n = 90) sampled at birth, infant faecal specimens sampled at 4-12 (n = 72) and 20-28 (n = 36) weeks of age (continued)

Genera (< 0.5%)	Relative abundance (%)				Genera (< 0.5%)	Relative abundance (%)			
	Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth		Infants at birth	Infants at 4 to 12 weeks	Infants at 20 to 28 weeks	Maternal at birth
Turicibacter	0.230	0.120	0.210	0.490	Megasphaera	0.220	0.620	0.950	0.240
Unclassified	0.160	0.250	0.130	0.450	Mitsuokella	0.050	0.040	0.050	0.130
Acidaminococcus	0.050	0.120	0.140	0.130	Negativicoccus	0.000	0.100	0.000	0.000
Phascolarctobacterium	0.000	0.010	0.000	0.010	Selenomonas	0.010	0.010	0.000	0.000
Succiniclasticum	0.000	0.000	0.000	0.000	Sporomusa	0.000	0.000	0.000	0.000
Unclassified	0.020	0.030	0.020	0.080	Unclassified	0.020	0.200	0.290	0.130
Allisonella	0.010	0.050	0.040	0.160	Unclassified	0.000	0.000	0.020	0.010
Anaeroglobus	0.000	0.000	0.000	0.000	OD1 genus incertae sedis	0.000	0.000	0.000	0.000
Anaerosinus	0.000	0.010	0.020	0.000	OP11 genus incertae sedis	0.000	0.000	0.000	0.000
Centipeda	0.000	0.000	0.000	0.000	TM7 genus incertae sedis	0.770	0.140	0.030	0.030
Dialister	0.170	0.360	0.280	0.550	Unclassified	0.010	0.000	0.000	0.010
Megamonas	0.140	0.180	0.630	0.100					

Table S6. Significant differences in proportions of bacterial OTUs from infant meconium specimens (n = 107) when compared to maternal faecal specimens sampled at birth (n = 90)

Taxon level	Bacterial OTUs	Rate ratio	p-value	Higher proportions from maternal or infant specimens
Phylum	**Actinobacteria**	**1.7**	<0.0001	**I**
Class	Actinobacteria	1.7	<0.0001	I
Order	Actinomycetales	9.8	<0.0001	I
Order	Coriobacteriales	0.6	<0.0001	M
Family	Coriobacteriaceae	0.6	<0.0001	M
Genus	Olsenella	0.6	0.0002	M
Genus	Unclassified	0.4	<0.0001	M
Phylum	**Bacteroidetes**	**0.6**	<0.0001	**M**
Class	Bacteroidia	0.4	<0.0001	M
Order	Bacteroidales	0.4	<0.0001	M
Family	Porphyromonadaceae	0.5	<0.0001	M
Family	Prevotellaceae	0.4	<0.0001	M
Genus	Prevotella	0.5	<0.0001	M
Family	Bacteroidaceae	0.3	<0.0001	M
Genus	Bacteroides	0.3	0.03753	M
Class	Unclassified	0.3	<0.0001	M
Order	Unclassified	0.3	<0.0001	M
Family	Unclassified	0.3	<0.0001	M
Genus	Unclassified	0.3	<0.0001	M
Phylum	**Proteobacteria**	**14.7**	<0.0001	**I**
Class	Alphaproteobacteria	19.3	<0.0001	I
Order	Caulobacterales	35.0	<0.0001	I
Family	Caulobacteraceae	35.0	<0.0001	I
Genus	Brevundimonas	170.0	<0.0001	I
Order	Rhizobiales	12.0	<0.0001	I
Family	Aurantimonadaceae	33.5	<0.0001	I
Genus	Aurantimonas	34.2	<0.0001	I
Family	Hyphomicrobiaceae	2.5	<0.0001	I
Family	Unclassified	26.9	<0.0001	I
Genus	Unclassified	26.9	<0.0001	I
Order	Rhodobacterales	25.5	<0.0001	I
Family	Rhodobacteraceae	25.5	<0.0001	I
Order	Sphingomonadales	25.8	<0.0001	I
Family	Sphingomonadaceae	25.1	<0.0001	I
Genus	Sphingomonas	21.2	<0.0001	I
Order	Unclassified	22.5	<0.0001	I
Family	Unclassified	22.5	<0.0001	I
Genus	Unclassified	22.5	<0.0001	I
Class	Betaproteobacteria	15.2	<0.0001	I
Order	Burkholderiales	15.2	<0.0001	I

Table S6. Significant differences in proportions of bacterial OTUs from infant meconium specimens (n = 107) when compared to maternal faecal specimens sampled at birth (n = 90) (continued)

Taxon level	Bacterial OTUs	Rate ratio	p-value	Higher proportions from maternal or infant specimens
Family	Comamonadaceae	19.8	<0.0001	I
Genus	Unclassified	26.2	<0.0001	I
Family	Oxalobacteraceae	15.2	<0.0001	I
Genus	Naxibacter	20.2	<0.0001	I
Class	Gammaproteobacteria	15.3	<0.0001	I
Order	Enterobacteriales	9.5	<0.0001	I
Family	Enterobacteriaceae	9.5	<0.0001	I
Genus	Unclassified	9.1	<0.0001	I
Order	Aeromonadales	12.6	<0.0001	I
Family	Aeromonadaceae	29.0	<0.0001	I
Genus	Aeromonas	29.0	<0.0001	I
Order	Alteromonadales	17.9	<0.0001	I
Family	Shewanellaceae	17.4	<0.0001	I
Genus	Shewanella	17.4	<0.0001	I
Order	Pseudomonadales	16.0	<0.0001	I
Family	Moraxellaceae	15.1	<0.0001	I
Genus	Acinetobacter	14.9	<0.0001	I
Family	Pseudomonadaceae	17.1	<0.0001	I
Genus	Pseudomonas	13.5	<0.0001	I
Order	Xanthomonadales	21.5	<0.0001	I
Family	Xanthomonadaceae	21.5	<0.0001	I
Phylum	**Firmicutes**	**0.3**	<0.0001	**M**
Class	Bacilli	3.7	<0.0001	I
Order	Bacillales	18.1	<0.0001	I
Order	Lactobacillales	2.5	<0.0001	I
Family	Lactobacillaceae	1.4	0.0179	I
Family	Streptococcaceae	2.6	<0.0001	I
Genus	Streptococcus	2.6	<0.0001	I
Class	Clostridia	0.2	<0.0001	M
Order	Clostridiales	0.2	<0.0001	M
Family	Lachnospiraceae	0.2	<0.0001	M
Genus	Anaerostipes	0.2	<0.0001	M
Genus	Blautia	0.3	<0.0001	M
Genus	Clostridium XlVa	0.4	<0.0001	M
Genus	Coprococcus	0.3	<0.0001	M
Genus	Dorea	0.4	<0.0001	M
Genus	Lachnospiracea incertae sedis	0.2	<0.0001	M
Genus	Roseburia	0.2	<0.0001	M
Genus	Unclassified	0.2	<0.0001	M

Table S6. Significant differences in proportions of bacterial OTUs from infant meconium specimens (n = 107) when compared to maternal faecal specimens sampled at birth (n = 90) (continued)

Taxon level	Bacterial OTUs	Rate ratio	p-value	Higher proportions from maternal or infant specimens
Family	Peptostreptococcaceae	0.6	0.0059	M
Genus	Clostridium XI	0.6	0.0047	M
Family	Ruminococcaceae	0.2	<0.0001	M
Genus	Clostridium IV	0.2	<0.0001	M
Genus	Faecalibacterium	0.2	<0.0001	M
Genus	Ruminococcus	0.2	<0.0001	M
Genus	Sporobacter	0.2	<0.0001	M
Genus	Unclassified	0.2	<0.0001	M
Family	Unclassified	0.2	<0.0001	M
Genus	Unclassified	0.2	<0.0001	M
Class	Erysipelotrichia	0.5	<0.0001	M
Order	Erysipelotrichales	0.5	<0.0001	M
Family	Erysipelotrichaceae	0.5	<0.0001	M
Genus	Catenibacterium	0.7	0.0007	M
Genus	Erysipelotrichaceae incertae sedis	0.4	<0.0001	M
Class	Negativicutes	0.6	<0.0001	M
Order	Selenomonadales	0.6	<0.0001	M
Family	Veillonellaceae	0.6	<0.0001	M
Class	Unclassified	0.2	<0.0001	M
Order	Unclassified	0.2	<0.0001	M
Family	Unclassified	0.2	<0.0001	M
Genus	Unclassified	0.2	<0.0001	M
Phylum	**Unclassified**	**0.5**	<0.0001	M
Class	Unclassified	0.5	<0.0001	M
Order	Unclassified	0.5	<0.0001	M
Family	Unclassified	0.5	<0.0001	M
Genus	Unclassified	0.5	<0.0001	M

OTU: Operational taxonomic unit; I: Infant; M: Mother

Table S7. Significant differences in proportions of bacterial OTUs from infant faecal specimens at 4 to 12 weeks of age (n = 72) when compared to maternal faecal specimens sampled at birth (n = 90)

Taxon level	Bacterial OTUs	Rate ratio	p-value	Higher proportions from maternal or infant specimens
Phylum	**Actinobacteria**	**4.9**	<0.0001	**I**
Class	Actinobacteria	4.9	<0.0001	I
Order	Actinomycetales	5.1	<0.0001	I
Order	Bifidobacteriales	9.8	<0.0001	I
Family	Bifidobacteriaceae	9.8	<0.0001	I
Genus	Bifidobacterium	10.0	<0.0001	I
Order	Coriobacteriales	2.3	<0.0001	I
Family	Coriobacteriaceae	2.3	<0.0001	I
Genus	Collinsella	3.6	<0.0001	I
Genus	Olsenella	1.9	<0.0001	I
Genus	Unclassified	1.9	<0.0001	I
Phylum	**Bacteroidetes**	**0.7**	**0.0281**	**M**
Class	Bacteroidia			
Order	Bacteroidales			
Family	Prevotellaceae	0.7	0.0333	M
Class	Unclassified	0.2	<0.0001	M
Order	Unclassified	0.2	<0.0001	M
Family	Unclassified	0.2	<0.0001	M
Genus	Unclassified	0.2	<0.0001	M
Phylum	**Proteobacteria**	**4.5**	<0.0001	**I**
Class	Alphaproteobacteria	4.1	<0.0001	I
Order	Rhizobiales	3.1	<0.0001	I
Class	Betaproteobacteria	3.6	<0.0001	I
Order	Burkholderiales	3.9	<0.0001	I
Class	Gammaproteobacteria	6.6	<0.0001	I
Order	Enterobacteriales	15.3	<0.0001	I
Family	Enterobacteriaceae	15.3	<0.0001	I
Genus	Unclassified	17.5	<0.0001	I
Order	Pseudomonadales	5.0	<0.0001	I
Family	Moraxellaceae	4.9	<0.0001	I
Genus	Acinetobacter	4.9	<0.0001	I
Phylum	**Firmicutes**	**0.6**	<0.0001	**M**
Class	Bacilli	14.4	<0.0001	I
Order	Lactobacillales	15.1	<0.0001	I
Family	Lactobacillaceae	12.3	<0.0001	I
Genus	Lactobacillus	14.4	<0.0001	I
Family	Streptococcaceae	18.5	<0.0001	I
Genus	Streptococcus	18.2	<0.0001	I

Table S7. Significant differences in proportions of bacterial OTUs from infant faecal specimens at 4 to 12 weeks of age (n = 72) when compared to maternal faecal specimens sampled at birth (n = 90) (continued)

Taxon level	Bacterial taxa	Rate ratio	p-value	Higher proportions from maternal or infant specimens
Class	Clostridia	0.3	<0.0001	M
Order	Clostridiales	0.3	<0.0001	M
Family	Lachnospiraceae	0.4	<0.0001	M
Genus	Anaerostipes	0.1	<0.0001	M
Genus	Blautia	0.3	<0.0001	M
Genus	Coprococcus	0.3	<0.0001	M
Genus	Dorea	0.5	<0.0001	M
Genus	Lachnospiracea incertae sedis	0.8	0.0480	M
Genus	Roseburia	0.0	<0.0001	M
Genus	Unclassified	0.3	<0.0001	M
Family	Peptococcaceae	0.2	<0.0001	M
Genus	Peptococcus	0.2	<0.0001	M
Family	Peptostreptococcaceae	0.5	<0.0001	M
Genus	Clostridium XI	0.5	<0.0001	M
Family	Ruminococcaceae	0.1	<0.0001	M
Genus	Clostridium IV	0.2	<0.0001	M
Genus	Faecalibacterium	0.1	<0.0001	M
Genus	Oscillibacter	0.1	<0.0001	M
Genus	Ruminococcus	0.0	<0.0001	M
Genus	Sporobacter	0.0	<0.0001	M
Genus	Unclassified	0.0	<0.0001	M
Family	Unclassified	0.1	<0.0001	M
Genus	Unclassified	0.1	<0.0001	M
Class	Negativicutes	2.2	<0.0001	I
Order	Selenomonadales	2.2	<0.0001	I
Family	Veillonellaceae	2.4	<0.0001	I
Genus	Dialister	0.6	0.0318	M
Class	Unclassified	0.3	<0.0001	M
Order	Unclassified	0.3	<0.0001	M
Family	Unclassified	0.3	<0.0001	M
Genus	Unclassified	0.3	<0.0001	M
Phylum	**Unclassified**	**0.3**	<0.0001	M
Class	Unclassified	0.3	<0.0001	M
Order	Unclassified	0.3	<0.0001	M
Family	Unclassified	0.3	<0.0001	M
Genus	Unclassified	0.3	<0.0001	M

OTU: Operational taxonomic unit; I: Infant; M: Mother

Table S8. Significant differences in proportions of bacterial OTUs from infant faecal specimens at 20 to 28 weeks of age (n = 36) when compared to maternal faecal specimens sampled at birth (n = 90)

Taxon level	Bacterial OTUs	Rate ratio	p-value	Higher proportions from maternal or infant specimens
Phylum	**Actinobacteria**	**3.2**	<0.0001	**I**
Class	Actinobacteria	3.2	<0.0001	I
Order	Actinomycetales	3.6	<0.0001	I
Order	Bifidobacteriales	7.0	<0.0001	I
Family	Bifidobacteriaceae	7.0	<0.0001	I
Genus	Bifidobacterium	7.2	<0.0001	I
Order	Coriobacteriales			I
Genus	Collinsella	2.0	0.0031	I
Genus	Olsenella	0.6	0.0239	I
Phylum	**Bacteroidetes**	**0.7**	**0.0183**	**M**
Class	Bacteroidia	0.7	0.0457	M
Order	Bacteroidales	0.7	0.0457	M
Family	Porphyromonadaceae	0.6	0.0037	M
Family	Prevotellaceae	0.6	0.0167	M
Phylum	**Proteobacteria**	**2.8**	<0.0001	**I**
Class	Alphaproteobacteria	2.6	<0.0001	I
Order	Rhizobiales	2.0	<0.0001	I
Class	Betaproteobacteria	2.3	<0.0001	I
Order	Burkholderiales	2.4	<0.0001	I
Class	Gammaproteobacteria	4.0	<0.0001	I
Order	Enterobacteriales	10.6	<0.0001	I
Family	Enterobacteriaceae	10.6	<0.0001	I
Order	Pseudomonadales	2.5	0.0172	I
Phylum	**Firmicutes**	**0.8**	<0.0001	**M**
Class	Bacilli	9.0	<0.0001	I
Order	Lactobacillales	9.5	<0.0001	I
Family	Lactobacillaceae	6.7	<0.0001	I
Genus	Lactobacillus	7.7	<0.0001	I
Family	Streptococcaceae	10.7	<0.0001	I
Genus	Streptococcus	10.1	<0.0001	I
Class	Clostridia	0.6	<0.0001	M
Order	Clostridiales	0.6	<0.0001	M
Genus	Blautia	0.7	0.0209	M
Genus	Clostridium_XlVa	2.0	0.0001	M
Genus	Coprococcus	0.5	0.0148	M
Genus	Lachnospiracea incertae sedis	1.6	0.0002	M
Genus	Roseburia	0.3	0.0044	M
Family	Peptococcaceae	0.2	0.0034	M
Genus	Peptococcus	0.2	0.0035	M

Table S8. Significant differences in proportions of bacterial OTUs from infant faecal specimens at 20 to 28 weeks of age (n = 36) when compared to maternal faecal specimens sampled at birth (n = 90) (continued)

Taxon level	Bacterial OTUs	Rate ratio	p-value	Higher proportions from maternal or infant specimens
Family	Ruminococcaceae	0.2	<0.0001	M
Genus	Clostridium IV	0.3	<0.0001	M
Genus	Faecalibacterium	0.3	<0.0001	M
Genus	Oscillibacter	0.1	<0.0001	M
Genus	Ruminococcus	0.2	<0.0001	M
Genus	Sporobacter	0.1	<0.0001	M
Genus	Unclassified	0.2	<0.0001	M
Family	Unclassified	0.2	<0.0001	M
Genus	Unclassified	0.2	<0.0001	M
Class	Negativicutes	2.2	<0.0001	I
Order	Selenomonadales	2.2	<0.0001	I
Family	Veillonellaceae	2.5	<0.0001	I
Class	Unclassified	0.3	<0.0001	M
Order	Unclassified	0.3	<0.0001	M
Family	Unclassified	0.3	<0.0001	M
Genus	Unclassified	0.3	<0.0001	M
Phylum	**Unclassified**	**0.3**	<0.0001	M
Class	Unclassified	0.3	<0.0001	M
Order	Unclassified	0.3	<0.0001	M
Family	Unclassified	0.3	<0.0001	M
Genus	Unclassified	0.3	<0.0001	M

OTU: Operational taxonomic unit; I: Infant; M: Mother

Table S9 Changes in bacterial proportions of infant faecal specimens from birth until 28 weeks of age (n = 215)

Taxon level	Bacterial OTUs	Increase/ decrease over time	Rate Ratio	p-value
Phylum	**Actinobacteria**	↑	**1.02**	**0.0274**
Class	Actinobacteria	↑	1.02	0.0274
Order	Actinomycetales	↓	0.95	<0.0001
Family	Geodermatophilaceae	↓	0.93	<0.0001
Genus	Blastococcus	↓	0.94	<0.0001
Order	Bifidobacteriales	↑	1.05	<0.0001
Family	Bifidobacteriaceae	↑	1.05	<0.0001
Genus	Bifidobacterium	↑	1.05	<0.0001
Phylum	**Bacteroidetes**			
Class	Bacteroidia			
Family	Bacteroidaceae	↑	1.07	0.0325
Genus	Bacteroides	↑	1.07	0.0325
Class	Sphingobacteria	↓	0.84	0.0124
Order	Sphingobacteriales	↓	0.84	0.0124
Class	Flavobacteria	↓	0.87	<0.0001
Order	Flavobacteriales	↓	0.87	<0.0001
Family	Flavobacteriaceae	↓	0.87	<0.0001
Phylum	**Proteobacteria**	↓	0.92	<0.0001
Class	Alphaproteobacteria	↓	0.90	<0.0001
Order	Caulobacterales	↓	0.88	<0.0001
Family	Caulobacteraceae	↓	0.88	<0.0001
Genus	Brevundimonas	↓	0.73	0.0369
Genus	Caulobacter	↓	0.96	<0.0001
Genus	Phenylobacterium	↓	0.75	0.0369
Order	Rhizobiales	↓	0.91	<0.0001
Family	Aurantimonadaceae	↓	0.88	<0.0001
Genus	Aurantimonas	↓	0.88	<0.0001
Family	Hyphomicrobiaceae	↓	0.95	<0.0001
Genus	Devosia	↓	0.89	<0.0001
Family	Rhizobiaceae	↓	0.93	<0.0001
Genus	Rhizobium	↓	0.93	<0.0001
Family	Unclassified	↓	0.89	<0.0001
Genus	Unclassified	↓	0.89	<0.0001
Order	Rhodobacterales	↓	0.90	<0.0001
Family	Rhodobacteraceae	↓	0.90	<0.0001
Genus	Unclassified	↓	0.94	<0.0001
Order	Rhodospirillales	↓	0.85	<0.0001
Family	Rhodospirillaceae	↓	0.84	<0.0001
Order	Sphingomonadales	↓	0.88	<0.0001
Family	Sphingomonadaceae	↓	0.89	<0.0001
Genus	Sphingomonas	↓	0.90	<0.0001

Table S9. Changes in bacterial proportions of infant faecal specimens from birth until 28 weeks of age (n = 215) (continued)

Taxon level	Bacterial OTUs	Increase/ decrease over time	Rate Ratio	p-value
Order	Unclassified	↓	0.89	<0.0001
Family	Unclassified	↓	0.87	<0.0001
Genus	Unclassified	↓	0.89	<0.0001
Class	Betaproteobacteria	↓	0.91	<0.0001
Order	Burkholderiales	↓	0.91	<0.0001
Family	Burkholderiaceae	↓	0.90	<0.0001
Family	Burkholderiales incertae sedis	↓	0.85	<0.0001
Genus	Aquabacterium	↓	0.86	<0.0001
Family	Comamonadaceae	↓	0.90	<0.0001
Genus	Acidovorax	↓	0.94	<0.0001
Genus	Unclassified	↓	0.86	<0.0001
Family	Oxalobacteraceae	↓	0.89	<0.0001
Genus	Massilia	↓	0.83	<0.0001
Genus	Naxibacter	↓	0.91	<0.0001
Family	Unclassified	↓	0.90	<0.0001
Genus	Unclassified	↓	0.90	<0.0001
Class	Deltaproteobacteria	↓	0.93	<0.0001
Class	Gammaproteobacteria	↓	0.93	<0.0001
Order	Aeromonadales	↓	0.89	<0.0001
Family	Aeromonadaceae	↓	0.87	<0.0001
Genus	Aeromonas	↓	0.87	<0.0001
Order	Alteromonadales	↓	0.92	<0.0001
Family	Shewanellaceae	↓	0.92	<0.0001
Genus	XShewanella	↓	0.92	<0.0001
Order	Pseudomonadales	↓	0.91	<0.0001
Family	Moraxellaceae	↓	0.91	<0.0001
Genus	Acinetobacter	↓	0.91	<0.0001
Family	Pseudomonadaceae	↓	0.91	<0.0001
Genus	Pseudomonas	↓	0.93	<0.0001
Order	Xanthomonadales	↓	0.88	<0.0001
Family	Xanthomonadaceae	↓	0.88	<0.0001
Genus	Lysobacter	↓	0.86	<0.0001
Phylum	**Firmicutes**	**↑**	**1.04**	<0.0001
Class	Bacilli	↑	1.02	0.0055
Order	Bacillales	↓	0.93	<0.0001
Family	Staphylococcaceae	↓	0.94	0.0110
Order	Lactobacillales	↑	1.03	0.0008
Family	Enterococcaceae			
Genus	Enterococcus	↑	1.03	0.0446

Table S9. Changes in bacterial proportions of infant faecal specimens from birth until 28 weeks of age (n = 215) (continued)

Taxon level	Bacterial OTUs	Increase/ decrease over time	Rate Ratio	p-value
Family	Lactobacillaceae	↑	1.05	<0.0001
Genus	Lactobacillus	↑	1.05	0.0002
Family	Leuconostocaceae	↑	1.04	0.0002
Genus	Weissella	↑	1.04	0.0002
Family	Unclassified	↑	1.04	0.0233
Genus	Unclassified	↑	1.04	0.023
Class	Clostridia	↑	1.05	<0.0001
Order	Clostridiales	↑	1.05	<0.0001
Family	Lachnospiraceae	↑	1.07	<0.0001
Genus	Blautia	↑	1.05	<0.0001
Genus	Clostridium XlVa	↑	1.06	<0.0001
Genus	Dorea	↑	1.06	<0.0001
Genus	Lachnospiracea incertae sedis	↑	1.09	<0.0001
Genus	Unclassified	↑	1.09	<0.0001
Family	Peptostreptococcaceae			
Genus	Clostridium XI	↑	1.03	0.0330
Family	Ruminococcaceae	↑	1.04	<0.0001
Genus	Faecalibacterium	↑	1.04	0.0031
Genus	Unclassified	↑	1.03	0.0391
Family	Unclassified	↑	1.03	0.0022
Genus	Unclassified	↑	1.03	0.0022
Class	Erysipelotrichia	↑	1.03	0.0031
Order	Erysipelotrichales	↑	1.03	0.0031
Family	Erysipelotrichaceae	↑	1.03	0.0031
Genus	Erysipelotrichaceae incertae sedis	↑	1.04	0.0041
Class	Negativicutes	↑	1.03	0.0108
Order	Selenomonadales	↑	1.03	0.0108
Family	Veillonellaceae	↑	1.03	0.0067
Genus	Veillonella	↑	1.04	0.0233
Phylum	**Unclassified**	↓	0.98	0.0274
Class	Unclassified	↓	0.98	0.0274
Order	Unclassified	↓	0.98	0.0274
Family	Unclassified	↓	0.98	0.0274
Genus	Unclassified	↓	0.98	0.0274

OTU: Operational taxonomic unit

Table S10. Genera identified from meconium specimens (n = 107) which have previously been reported from meconium specimens in the literature

Genus	Abundance (%) identified in this study	Reported in the literature				
		Jimenez (2008) [20]	Gosalbes et al. (2012) [21]	Madan et al. (2012) [41]	Moles et al. (2013) [22]	Ardissone et al. (2014) [15]
Bifidobacterium	6.61	✓	✓			✓
Streptococcus	3.09	✓	✓		✓	✓
Acinetobacter	2.96		✓			✓
Lactobacillus	2.13		✓		✓	
Prevotella	2.08		✓			
Clostridium			✓	✓		✓
Clostridium XI	1.80					
Clostridium XlVa	0.86					
Aeromonas	1.51		✓			
Pseudomonas	1.30		✓			
Enterococcus	1.13	✓	✓	✓	✓	✓
Bacteroides	1.05	✓	✓			✓
Weissella	0.80		✓			
Veillonella	0.77		✓	✓		✓
Acidovorax	0.63		✓			
Coprococcus	0.62					
Rothia	0.61					✓
Escherichia/Shigella	0.58	✓	✓			✓
Lysobacter	0.52					✓
Lactococcus	0.45		✓			
Parabacteroides	0.38	✓	✓			
Staphylococcus	0.30	✓	✓	✓	✓	✓
Paracoccus	0.29		✓			
Actinomyces	0.26		✓			
Leuconostoc	0.25	✓	✓			
Cupriavidus	0.18					✓
Ralstonia	0.14			✓		✓
Serratia	0.12		✓			
Bacillus	0.08					✓
Citrobacter	0.07		✓			
Haemophilus	0.07		✓			
Propionibacterium	0.05		✓			
Klebsiella	0.01	✓	✓	✓	✓	
Chryseobacterium	0.01		✓			
Comamonas	0.01		✓			
Stenotrophomonas	<0.01					✓
Arcobacter	<0.01		✓			
Enterobacter	<0.01	✓	✓			✓

B) Supplementary figures

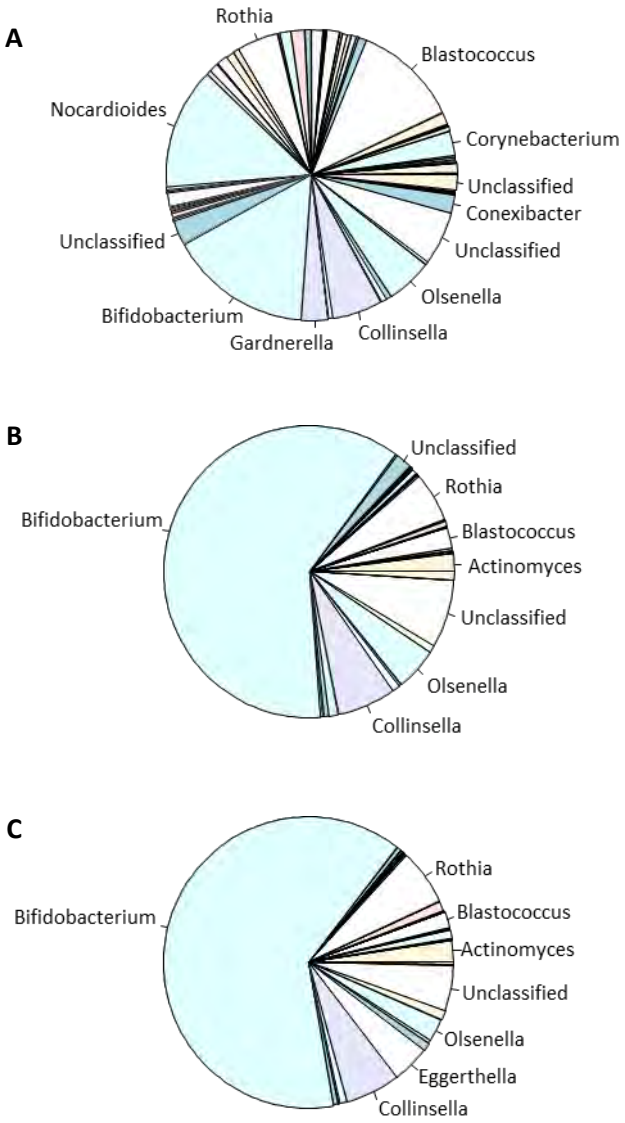


Figure S1. Bacterial proportions of the genera from the phylum Actinobacteria for infant meconium specimens sampled at birth as well as infant faecal specimens sampled at 4-12 and 20-28 week of age

Only the most abundant genera are labelled in the pie charts. Proportions of genera from the phylum Actinobacteria identified from **A)** infant meconium specimens sampled at birth (n = 107); **B)** infant faecal specimens sampled at 4-12 weeks of age (n = 72); and **C)** infant faecal specimens sampled at 20-28 weeks of age (n = 36).

GENERAL DISCUSSION AND CONCLUSION

Early life is regarded the most important period during which GIT microbiota are established.[1] Conventionally it has been reported that initial colonization occurs during the process of birth[2–9] and that these bacterial profiles are shaped by a number of external factors such as the mode of delivery, feeding practices and medication.[10] More recently, reports on in-utero bacterial colonization of the "sterile" fetus[11–17] have revolutionised our ideas around infantile GIT colonization.[18,19] Despite the increased interest in microbiota profiles from meconium specimens during the last decade; few studies have included meconium specimens when longitudinally profiling GIT microbiota from the healthy infants early in life.[11,12] To the best of our knowledge, this is the largest study to provide information on faecal bacterial dynamics from healthy term infants in which meconium specimens were inluded. The results generated by this study essentially provide a high-throughput overview of the bacterial profiles from a large number of meconium specimens. Results from this study also show how these bacterial profiles change throughout the first seven months of life. In addition, this study provides insights into the influencing factors involved in shaping early life faecal bacterial profiles.

Using Illumina Mi-Seq sequencing technology we were able to sequence 16 596 332 reads from 305 faecal specimens, for which the median number of reads per specimen was sufficient for GIT microbiota analysis.[20,21] However, when using any sequencing technology one needs to bear in mind that sequencing results are not error free and quality control measures are therefore essential in any microbiota study.[22,23] Following the use of an independent bio-informatics workflow (to accurately filter and remove erroneous reads and sequencing artifacts), we noted a reduction of 25% in the total number of reads sequenced. These types of reductions in the number of reads sequenced are expected when performing quality control steps (such as the removal of sequencing artifacts and trimming of low quality reads) on reads sequences by Illumina Miseq sequencing platforms.[23,24] A recent study by Sinclair and colleagues[24] reported on the reliability and biases of Illumina Miseq platforms by performing paired-end sequencing on the 16S rRNA gene extracted and amplified from sediment specimens. Sinclair and colleagues[24] reported a 50% reduction in the number of reads sequenced following quality control processes. In addition to the efficiency of the sequencing approach itself, experimental approaches towards library preparation prior to sequencing are just as important and also need to be carefully planned and executed. We observed that low template concentrations used for library preparation may have impacted on the reproducibility of our sequencing approach, as previously reported.[25] This in turn may also impact on diversity measures,[26]

since we observed very high intra-individual diversity indices from the meconium specimens compared to the other faecal specimens under study. Of note, in this study, meconium specimens had the lowest template concentrations and the lowest reproducibility measures. In addition to the downstream effects of low template concentrations, we also suspected contamination in our sequencing run after classifying OTUs from our non-template controls. These were corrected for during data analysis, and resulted in 10% of the OTUs classified being removed from our faecal specimens analysed. A number of potential contaminant sources, such as specimen collection, the nucleic acid extraction steps, PCR amplification steps and the sequencing process itself have previously been reported to affect sequencing results.[27,28] We were able to rule out potential contamination resulting from the nucleic acid extraction and sequencing procedures from our data set. We did, however, suspect that contamination occurred during the process of amplification;[29–31] either from the use of contaminated laboratory reagents[32] or cross-contamination between wells.[27,33–35] Even though we did not include controls to establish a baseline for reagent contamination, we did observe that approximately 10% of the "contaminating genera" identified fom our two non-template controls have been previously reported as laboratory reagent contaminants (Table 3). In addition, the effect of cross-contamination during the amplification process is also highly probable in our study due to the fact that i) the majority of our specimens provided low template concentrations for library preparation;[34–38] ii) we performed repeat amplification of a marker gene and had possible exposure to aerosilized amplicons;[27,39] and iii) that the majority of contaminating genera observed from our non-template controls were genera commonly characterised from the human GIT (Table 3). Correcting for contamination, however, was unlikely to have had a significant impact on downstream data analysis since very low levels of contaminants were present.

Despite the low template concentrations used for meconium specimens in this study, we successfully managed to classify bacteria from all meconium specimens analysed. Our findings on the bacterial profiles colonizing the meconium specimens were in agreement with previous studies.[11–13,40,41] We found that the most abundant phylum classified from the meconium cohort was the phylum Proteobacteria, followed by the phylum Firmicutes. Complete linkage clustering analysis clearly grouped infant and maternal faecal specimens into two primary clusters and none of the respective mother-infant pairs clustered at low clustering distances. In contrast to the findings from the specimens sampled at birth; we did observe a change in infant faecal bacterial profiles towards more adult-like profiles at later stages in life. Therefore, our findings suggest that maternal-fetal bacterial translocations may only have a modest effect on the overall bacterial composition identified from meconium, but its effect may become clearer at later stages of life. This may be

explained by reports of mainly the phyla Proteobacteria and Firmicutes colonizing meconium,[11,13,42,41] some of which have been identified as bacteria playing a role in priming the infant GIT for strict anaerobic bacterial colonization later in life.[43–45]

Although only few studies have characterised longitudinal bacterial profiles from meconium of healthy term infants to date,[11,46] external factors influencing these profiles are still understudied. In our healthy cohort, only maternal education level and infant gender were significantly associated with phyla from the family Firmicutes. In agreement with previous studies,[11,42,47] we also found that mode of delivery does not alter profiles of meconium microbiota and the effect of mode of delivery only becomes apparent at later stages of life. At 4-12 weeks of age, we found that both residential area and gestational age had a significant effect on specific bacterial taxa. The effect of residential area is of interest, as we observed changes in infant faecal bacterial profiles over a short period of time (4-12 weeks of age) from two ethnogeographic populations within the same sub-district in the Cape-Winelands. Our findings showed that infants residing in Mbekweni (comprising of a black African population) had lower proportions of the phylum Actinobacteria (down to the genus Bifidobacterium), but higher proportions of the phylum Firmicutes compared to infants residing in TC Newman (comprising of a mixed race population). The influence of geographical location on GIT microbiota has been reported by a number of studies and has been attributed to varying environmental pressures such as differences in genetic backgrounds, diets and cultural practices of different ethnogeographical populations and regional lifestyle.[10] To the best of our knowledge, our study is the first to associate geographical influence on a cohort residing within two residential areas from the same sub-district at less than three months of age. This highlights the importance of the community setting to which a young infant is exposed. Another interesting finding in this study is the observation that HIV-exposed infants had higher proportions of the genus Leuconostoc, as well as higher bacterial diversity indices. To date, the effect of maternal HIV status on faecal bacterial profiles of their offspring has not yet been reported. Of note, the factor most prominently influencing bacterial proportions within our study was an increase in infant age. This was clearly emphasised by the significant changes in proportions of 87 bacterial taxa from infant faecal specimens during the first 28 weeks of life. Although a number of studies have reported on changes in bacterial profiles over time, only few studies have reported on changes in bacterial profiles which included meconium specimens when analysing healthy infants.[11,46]

One of the limitations of our study was the possibility that not all meconium specimens may have been collected at hospital by the DCHS's clinical team. In the event that a first stool was not passed by the newborn during hospitalisation, the responsibility fell upon the participating mother to collect this specimen at home. This resulted in the data capturers having to rely on the

information provided by the mother and not by the DCHS clinical team for select specimens from the DCHS cohort. Despite this possible limitation, we did observe distinct bacterial profiles for the meconium specimens analysed in our study when compared to specimens collected at later stages in life. Other limitations to our study were low template concentrations obtained from meconium specimens and the suspicion of contamination during library preparation. Low template concentrations were clearly related to less reproducible results which may have resulted in "skewed" diversity measures from meconium specimens as shown by the very high intra-individual diversity measurements obtained from the first stool passed by infants.[41,47–52] On the other hand, the inter-individual diversity measures observed in our study were in agreement with previous findings.[48–51,53] Since we were not in the position to repeat any of the experimental steps towards sequencing (due to limited volumes of extracted nucleic acid exported to JCVI, USA) we decided to correct for contamination during data analysis. Our approach to computationally remove contaminating OTUs observed from all sequencing reactions performed (prior to conducting microbiome data analyses), is becoming apparent for next-generation sequencing datasets.[32,54,55] The final limitation from our study was the limited clinical data available for association studies. We were not able to determine, for example, the effect of medication (such as antibiotics) on the maternal and infant microbiota profiles. Neither were we in the position to determine the effect of infant HIV status on infantile faecal bacterial profiles, since none of the infants were HIV-infected.

In conclusion, our study provides a detailed description of the dynamics of infant faecal bacterial profiles during the first seven months of life, which includes meconium specimens. Although we could not directly address the hypothesis suggesting in-utero transfer of maternal GIT microbiota to the fetus; our study shows that infant and maternal faecal bacterial profiles are very different at birth up until seven months of life. In order to effectively address the pathways behind in-utero bacterial transfer from the maternal GIT to the fetus; studies need to consider analysing a range of specimens which may include maternal faecal specimens, cord blood, amniotic fluid, placenta as well as meconium. Since other hypotheses suggest in-utero translocation of maternal oral and vaginal bacteria,[56–59] studies should also attempt to analyse the microbiota from maternal oral and vaginal specimens in conjunction with uterine and faecal specimens. Even though a number of studies have reported on bacterial colonization of the uterine, meconium and maternal oral and faecal specimens using a range of techniques (some of which included high-throughput sequencing);[11–17,60] to the best of our knowledge no studies have analysed microbiota from more than two sample types from a single cohort. In addition, future studies also need to contemplate the use of techniques such as high-throughput culture or viability PCRs to provide an overview of viable microbiota profiles from these specimens.[60] Information on the dynamics of viable cells from

uterine, meconium and infant faecal specimens may contribute greatly to our insights on the mechanisms behind in-utero colonization and microbial succession over time.[60] Considering that we conducted a pilot study on the dynamics of infant faecal bacterial profiles during the first seven months of life; we plan to perform a more extensive study using larger sample sizes from the DCHS cohort which will include additional timepoints. This will allow us to optimally determine the progression of meconium faecal bacterial profiles to infant-like and finally adult-like profiles within a poor African community. Of note, our interesting findings on the higher bacterial diversities from HIV-exposed infants need to be extended and these infants should be followed up for longer periods of time to determine whether the effect of maternal HIV status is retained later in life.

Moreover, we suggest that studies should include a range of clinical characteristics to better understand the effect of prenatal factors such as medical interventions and maternal characteristics on meconium microbiota as well as subsequent microbiota profiles during infancy. Additionally, longitudinal studies need to also include meconium specimens when performing association studies with disease states developed later in life. Good examples of this have been highlighted by associations between specific meconium microbiota (reported to incite inflammatory reponses) and respiratory diseases and premature birth.[11,13] Of note, further investigation of meconium microbiota and their interaction with the host may provide innovative therapies for the prevention of subsequent health risks such as premature birth.

References

1. Wopereis H, Oozeer R, Knipping K, Belzer C, Knol J. The first thousand days - intestinal microbiology of early life: establishing a symbiosis. *Pediatr allergy Immunol*. 2014. doi:10.1111/pai.12232.

2. Scholtens PAMJ, Oozeer R, Martin R, Amor K Ben, Knol J. The early settlers: intestinal microbiology in early life. *Annu Rev Food Sci Technol*. 2012;3:425-47. doi:10.1146/annurev-food-022811-101120.

3. Mackie RI, Sghir A, Gaskins HR. Developmental microbial ecology of the neonatal gastrointestinal tract. *Am J Clin Nutr*. 1999;69(5):1035S-1045S.

4. Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. *Cell*. 2012;148(6):1258-70. doi:10.1016/j.cell.2012.01.035.

5. Thompson-Chagoyán OC, Maldonado J, Gil A. Colonization and impact of disease and other factors on intestinal microbiota. *Dig Dis Sci*. 2007;52(9):2069-77. doi:10.1007/s10620-006-9285-z.

6. Penders J, Thijs C, Vink C, et al. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*. 2006;118(2):511-21. doi:10.1542/peds.2005-2824.

7. Biasucci G, Rubini M, Riboni S, Morelli L, Bessi E, Retetangos C. Mode of delivery affects the bacterial community in the newborn gut. *Early Hum Dev*. 2010;86(Suppl 1):S13-S15. doi:10.1016/j.earlhumdev.2010.01.004.

8. Dominguez-Bello MG, Costello EK, Contreras M, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *PNAS*. 2010;107(26):11971-5. doi:10.1073/pnas.1002601107.

9. Fanaro S, Chierici R, Guerrini P, Vigi V. Intestinal microflora in early infancy: composition and development. *Acta Paediatr*. 2003;91(441):48-55. doi:10.1111/j.1651-2227.2003.tb00646.x.

10. Arrieta M-C, Stiemsma LT, Amenyogbe N, Brown EM, Finlay B. The Intestinal Microbiome in Early Life: Health and Disease. *Front Immunol*. 2014;5(September):1-18. doi:10.3389/fimmu.2014.00427.

11. Gosalbes MJ, Llop S, Vallès Y, Moya A, Ballester F, Francino MP. Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clin Exp Allergy*. 2012;43(2):198-211. doi:10.1111/cea.12063.

12. Jiménez E, Marín ML, Martín R, et al. Is meconium from healthy newborns actually sterile? *Res Microbiol*. 2008;159(3):187-93. doi:10.1016/j.resmic.2007.12.007.

13. Ardissone AN, De La Cruz DM, Davis-Richardson AG, et al. Meconium microbiome analysis identifies bacteria correlated with premature birth. *PLoS One*. 2014;9(3):e90784. doi:10.1371/journal.pone.0090784.

14. Rautava S, Collado MC, Salminen S, Isolauri E. Probiotics modulate host-microbe interaction in the placenta and fetal gut: a randomized, double-blind, placebo-controlled trial. *Neonatology*. 2012;102(3):178-84. doi:10.1159/000339182.

15. Satokari R, Grönroos T, Laitinen K, Salminen S, Isolauri E. Bifidobacterium and Lactobacillus DNA in the human placenta. *Lett Appl Microbiol*. 2009;48(1):8-12. doi:10.1111/j.1472-765X.2008.02475.x.

16. Steel JH, Malatos S, Kennea N, et al. Bacteria and inflammatory cells in fetal membranes do not always cause preterm labor. *Pediatr Res*. 2005;57(3):404-11. doi:10.1203/01.PDR.0000153869.96337.90.

17. Jiménez E, Fernández L, Marín ML, et al. Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. *Curr Microbiol*. 2005;51(4):270-4. doi:10.1007/s00284-005-0020-3.

18. Funkhouser LJ, Bordenstein SR. Mom knows best: the universality of maternal microbial transmission. *PLoS Biol*. 2013;11(8):e1001631. doi:10.1371/journal.pbio.1001631.

19. Francino MP. Early development of the gut microbiota and immune health. *Pathog (Basel, Switzerland)*. 2014;3(3):769-90. doi:10.3390/pathogens3030769.

20. Claesson MJ, Wang Q, O'Sullivan O, et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res*. 2010;38(22):e200. doi:10.4161/gmic.1.4.12306.

21. Tseng HE, Hullar MAJ, Li F, et al. A microbial profiling method for the human microbiota using high-throughput sequencing. *Metagenomics*. 2013;2(1):1-15.

22. Eren AM, Vineis JH, Morrison HG, Sogin ML. A filtering method to generate high quality short reads using illumina paired-end technology. *PLoS One*. 2013;8(6):e66643. doi:10.1371/journal.pone.0066643.

23. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*. 2012;7(2):e30619. doi:10.1371/journal.pone.0030619.

24. Sinclair L, Osman OA, Bertilsson S, Eiler A. Microbial Community Composition and Diversity via 16S rRNA Gene Amplicons: Evaluating the Illumina Platform. *PLoS One*. 2015;10:e0116955. doi:10.1371/journal.pone.0116955.

25. Kennedy NA, Walker AW, Berry SH, et al. The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS One*. 2014;9(2):e88982. doi:10.1371/journal.pone.0088982.

26. Zhou J, Wu L, Deng Y, et al. Reproducibility and quantitation of amplicon sequencing-based detection. *ISME J*. 2011;5(8):1303-13. doi:10.1038/ismej.2011.11.

27. Weiss S, Amir A, Hyde ER, Metcalf JL, Song SJ, Knight R. Tracking down the sources of experimental contamination in microbiome studies. *Genome Biol*. 2014;15(564):1-3. doi:10.1186/s13059-014-0564-2.

28. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. Analysis, Optimization and Verification of Illumina-Generated 16S rRNA Gene Amplicon Surveys. *PLoS One*. 2014;9(4):e94249. doi:10.1371/journal.pone.0094249.

29. Inglis G, Thomas M, Thomas D, Kalmokoff M, Brooks S, Selinger L. Molecular Methods to Measure Intestinal Bacteria : A Review. *J AOAC Int*. 2012;95(1):5-24. doi:10.5740/jaoacint.SGE.

30. Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J Biosci Bioeng*. 2003;96(4):317-23. doi:10.1016/S1389-1723(03)90130-7.

31. Tanner MA, Goebel BM, Dojka MA, Pace NR. Specific Ribosomal DNA Sequences from Diverse Environmental Settings Correlate with Experimental Contaminants. *Appl Environ Microbiol*. 1998;64(8):3110.

32. Salter SJ, Cox MJ, Turek EM, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12(1):87. doi:10.1186/s12915-014-0087-z.

33. Aslanzadeh J. Preventing PCR Amplification Carryover Contamination in a Clinical Laboratory. *Ann Clin Lab Sci*. 2004;34(4):389-396.

34. Willner D, Daly J, Whiley D, Grimwood K, Wainwright CE, Hugenholtz P. Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PLoS One*. 2012;7(4):e34605. doi:10.1371/journal.pone.0034605.

35. Kircher M, Heyn P, Kelso J. Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics*. 2011;12(1):382. doi:10.1186/1471-2164-12-382.

36. Lazarevic V, Gaïa N, Emonet S, et al. Challenges in the culture-independent analysis of oral and respiratory samples from intubated patients. *Front Cell Infect Microbiol*. 2014;4(May):65. doi:10.3389/fcimb.2014.00065.

37. Pennisi E. Contamination plagues some microbiome studies. *Sci Mag*. 2014;346(6211):801. doi:10.1126/science.346.6211.801.

38. Lusk RW. Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS One*. 2014;9(10):e110808. doi:10.1101/002279.

39. Persing DH. Polymerase Chain Reaction: Trenches to Benches. *J Clin Microbiol*. 1991;29(7):1281-1285.

40. Madan JC, Salari RC, Saxena D, et al. Gut microbial colonisation in premature neonates predicts neonatal sepsis. *Arch Dis childhood Fetal neonatal Ed*. 2012;97:F456-62. doi:10.1136/archdischild-2011-301373.

41. Moles L, Gómez M, Heilig H, et al. Bacterial diversity in meconium of preterm neonates and evolution of their fecal microbiota during the first month of life. *PLoS One*. 2013;8(6):e66986. doi:10.1371/journal.pone.0066986.

42. Hu J, Nomura Y, Bashir A, et al. Diversified microbiota of meconium is affected by maternal diabetes status. *PLoS One*. 2013;8(11):e78257. doi:10.1371/journal.pone.0078257.

43. Adlerberth I, Wold AE. Establishment of the gut microbiota in Western infants. *Acta Paediatr*. 2009;98(2):229-38. doi:10.1111/j.1651-2227.2008.01060.x.

44. Adlerberth I. Factors influencing the establishment of the intestinal microbiota in infancy. *Nestle Nutr Work Ser Pediatr Progr*. 2008;62:13-29. doi:10.1159/000146245.

45. Bezirtzoglou E. The intestinal microflora during the first weeks of life. *Anaerobe*. 1997;3(2-3):173-7. doi:10.1006/anae.1997.0102.

46. Tsuji H, Oozeer R, Matsuda K, et al. Molecular monitoring of the development of intetinal microbiota in Japanese infants. *Benef Microbes*. 2012;3(2):113-25. doi:10.3920/BM2011.0038.

47. Mshvildadze M, Neu J, Shuster J, Theriaque D, Li N, Mai V. Intestinal microbial ecology in premature infants assessed with non-culture-based techniques. *J Pediatr*. 2010;156(1):20-5. doi:10.1016/j.jpeds.2009.06.063.

48. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. Ruan Y, ed. *PLoS Biol*. 2007;5(7):e177. doi:10.1371/journal.pbio.0050177.

49. Koenig JE, Spor A, Scalfone N, et al. Succession of microbial consortia in the developing infant gut microbiome. *PNAS*. 2011;108(Suppl 1):4578-85. doi:10.1073/pnas.1000081107.

50. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486(7402):222-7. doi:10.1038/nature11053.

51. Avershina E, Storrø O, Øien T, Johnsen R, Pope P, Rudi K. Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *FEMS Microbiol Ecol*. 2013;87(1):280-90. doi:10.1111/1574-6941.12223.

52. Ringel-Kulka T, Cheng J, Ringel Y, et al. Intestinal microbiota in healthy U.S. young children and adults--a high throughput microarray analysis. *PLoS One*. 2013;8(5):e64315. doi:10.1371/journal.pone.0064315.

53. De Filippo C, Cavalieri D, Di Paola M, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A*. 2010;107(33):14691-6. doi:10.1073/pnas.1005963107.

54. Cibulskis K, McKenna A, Fennell T, Banks E, DePristo M, Getz G. ContEst: Estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics*. 2011;27(18):2601-2602. doi:10.1093/bioinformatics/btr446.

55. Laurence M, Hatzis C, Brash DE. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS One*. 2014;9(5):1-8. doi:10.1371/journal.pone.0097876.

56. DiGiulio DB. Diversity of microbes in amniotic fluid. *Semin Fetal Neonatal Med*. 2012;17(1):2-11. doi:10.1016/j.siny.2011.10.001.

57. White BA, Creedon DJ, Nelson KE, Wilson BA. The vaginal microbiome in health and disease. *Trends Endocrinol Metab*. 2011;22(10):389-93. doi:10.1016/j.tem.2011.06.001.

58. DiGiulio DB, Romero R, Kusanovic JP, et al. Prevalence and diversity of microbes in the amniotic fluid, the fetal inflammatory response, and pregnancy outcome in women with preterm prelabor rupture of membranes. *Am J Reprod Immunol*. 2010;64(1):38-57. doi:10.1111/j.1600-0897.2010.00830.x.

59. Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The placenta harbors a unique microbiome. *Sci Transl Med*. 2014;6(237):1-11. doi:10.1126/scitranslmed.3008599.

60. Payne MS, Bayatibojakhi S. Exploring Preterm Birth as a Polymicrobial Disease: An Overview of the Uterine Microbiome. *Front Immunol*. 2014;5(November):1-12. doi:10.3389/fimmu.2014.00595.