

Early-life Gut Microbiome and Resistome Trajectories and their Determinants in a Northern-Nigerian Birth Cohort

A thesis submitted for the degree of
Doctor of Philosophy in Interdisciplinary Bioscience
Michaelmas 2025



By

Chinenye Pascaline Akpulu

Interdisciplinary Bioscience, DTP

Department of Biology

Linacre College

CONTENTS

ABSTRACT	VIII
ACKNOWLEDGEMENTS	X
DECLARATION OF CONTRIBUTIONS TO THESIS	XII
CHAPTER 5,6: I AM ESPECIALLY GRATEFUL TO ADITYA LANKAPALLI AND LAURA DE NIES WHO AIDED WITH INITIATING THE BIOINFORMATIC ANALYSIS. ADITYA PLOTTED FIGURE 5.3	XII
PUBLICATIONS	XIII
LIST OF ABBREVIATIONS.....	XIV
CHAPTER 1: INTRODUCTION	1
1.1 GLOBAL AND CLINICAL CONTEXT OF ANTIMICROBIAL RESISTANCE	1
1.1.1 Contextual overview and global burden of antimicrobial resistance (AMR)	1
1.1.2 AMR Burden in Neonates and Infants in LMICs	3
1.1.3 Mechanisms and drivers of AMR relevant to early-life colonisation	8
1.2. THE DEVELOPING INFANT GUT MICROBIOME	10
1.2.1 Developmental stages and early colonisation	10
1.2.2 Functional roles in immunity and metabolism	11
1.2.3 Why early life is a critical window.....	13
1.3. MODULATORS OF THE MICROBIOME AND RESISTOME IN EARLY LIFE	16
1.3.1 Diet	19
1.3.2 Mode of delivery	19
1.3.3 Antibiotic Exposure.....	20
1.3.4 Environmental and Contextual Modulators	21
1.3.5 Geographic and Socioeconomic Context	21
1.4. THE EARLY LIFE RESISTOME	22
1.4.1 Definition and significance.....	22
1.4.2 Detection and carriage of ARGs in infants.....	23
1.4.3 Transmission routes and reservoirs.....	24
1.4.4 Clinical and public health implications	26
1.5 DETECTING ARGs IN THE MICROBIOME	27
1.5.1 Limitations of traditional detection methods	28
1.5.2 Metagenomic approaches to ARG detection	29
1.5.3 Strengths, limitations, and relevance to this study	30
1.5.4 Summary of knowledge gaps.....	31
1.6 STUDY RATIONALE AND OBJECTIVES	32

1.6.1 Rationale for the study	32
1.6.2 Study objectives	33
1.6.3 Thesis structure and chapter overview	34
CHAPTER 2: MATERIALS AND METHODS	36
2.1. STUDY DESIGN AND SETTING	36
2.2 ETHICAL APPROVALS	37
2.2.1 Institutional review and approvals	37
2.2.2 Informed consent procedures	37
2.2.3 Confidentiality and participant protection	38
2.2.4 Operational permissions	38
2.3. STUDY POPULATION	38
2.3.1 Inclusion and Exclusion Criteria	38
2.3.2 Recruitment Process	40
2.3.3 Participant Identification and Data Collection	40
2.3.4 Retention Strategies	40
2.4. SAMPLE SIZE CALCULATION	41
2.5 SAMPLE AND METADATA COLLECTION	42
2.5.1 Sample collection and storage	42
2.5.2 Exposure metadata collection	43
2.6 LABORATORY METHODS	48
2.6.1 DNA extraction and technical quality control	48
2.6.2 Library preparation and shotgun sequencing	49
2.6.3 Sequencing Output and Data Handling	50
2.7 QUALITY CONTROL PROCEDURES	52
2.7.1. Contamination controls: extraction and sequencing controls	52
2.7.2. Host DNA Removal	53
2.7.3. Sequencing depth normalisation	53
2.7.4 Data Traceability and Replicates	54
2.8 BIOINFORMATICS AND COMPUTATIONAL ANALYSIS	54
2.8.1. Preprocessing and quality control	55
2.8.2. Taxonomic profiling	58
2.8.3. ARG and MGE detection	58
2.8.4 Integration of taxonomic and resistome outputs	59
2.9. STATISTICAL ANALYSIS	60
2.9.1 Overview of data structure and general approach	60
2.9.2. Microbiome diversity and composition	60
2.9.3. Resistome summaries and composition	61

2.9.4. Exposure-Outcome association modelling and covariate adjustment	62
2.9.5. Sensitivity analyses	63
2.10 DATA MANAGEMENT AND REPRODUCIBILITY	64
2.10.1 Data Management	64
2.10.2 Reproducibility	64
2.10.3 Quality Assurance	65
CHAPTER 3: VALIDATING DNA EXTRACTION FOR NEONATAL, LOW-BIOMASS METAGENOMICS IN A RESOURCE-LIMITED SETTING	66
3.1 INTRODUCTION	66
Aim and objectives	71
3.2 METHODS	72
3.2.1. Ethics approval and study setting	72
3.2.2. Participants, sampling and design	72
3.2.3 Selection of DNA extraction kit	75
3.2.4. DNA extraction procedures	76
3.2.5. DNA quantification	76
3.2.6. Library preparation and sequencing.....	77
3.2.7. Bioinformatics processing.....	77
3.2.8. Statistical analysis and diversity.....	78
3.2.8. Data availability.....	79
3.3. RESULTS	79
3.3.1. Protocol-specific DNA yield	79
3.3.2. Effects of storage and time-to-extraction	80
3.3.3. ONT read-length distributions and assembly-relevant metrics	81
3.3.4. Community profiles signal	83
3.4. DISCUSSION	88
CHAPTER 4: COHORT PROFILE OF EARLY-LIFE GUT MICROBIOME AND RESISTOME LONGITUDINAL STUDY: NORTHERN NIGERIAN BIRTH COHORT	94
4.1 INTRODUCTION	94
4.1.1 Study aims	97
4.2 STUDY SITE CONTEXT	97
4.2.1 Kano city.....	97
4.2.2 Enrolment hub: Murtala Muhammed Specialist Hospital, Kano	100
4.3. STUDY IMPLEMENTATION.....	104
4.3.1 Operational objectives.....	104
4.3.2 Recruitment.....	104

4.3.3 Data collection.....	107
4.3.4 Biological sample types.....	107
4.4 COHORT PROFILE AND IMPLEMENTATION OUTCOMES.....	112
4.4.1 Attrition and cohort retention	112
4.4.2 Household characteristics	112
4.4.3. Maternal characteristics	113
4.4.4 Birth characteristics	118
4.4.5 Antibiotic use	119
4.4.6 Infant Feeding	121
4.4.7 Traditional and cultural practices.....	123
4.4.8 Biological samples	123
4.5 DISCUSSION	125
4.6. FIELDWORK EXPERIENCE: CHALLENGES, LESSONS LEARNT, AND THINGS I WISH I KNEW	131
CHAPTER 5: TEMPORAL DYNAMICS OF THE INFANT GUT MICROBIOME AND RESISTOME DURING EARLY LIFE.....	135
5.1 INTRODUCTION	135
5.2 AIMS AND OBJECTIVES	138
5.3 CHAPTER-SPECIFIC METHODS.....	139
5.3.1 ARG feature representation	139
5.3.2 Presence-weighted class and mechanism summaries.....	140
5.3.3 Clinically focused ARG definitions	141
5.3.4 Time-to-event and persistence.....	142
5.3.5 Co-presence and enrichment between resistance classes	142
5.3.6 Contig-linked taxonomic context for selected ARGs.....	143
5.4 RESULTS	143
5.4.0 Data readiness and QC outcomes (summary)	143
5.4.1 Early-life gut microbiome trajectory	144
5.4.2 Infant gut resistome trajectory: burden, diversity and antibiotic-class composition.....	150
5.4.3 Early acquisition of key ARGs and the “never-colonised” group	154
5.4.4 Dual β -lactam/aminoglycoside and co-carriage	158
5.4.5 Enrichment of resistance class co-presence.....	161
5.4.6 Host context and dynamics of <i>bla</i> _{CTX-M-15} and aminoglycoside resistance	165
5.4.6.1 Host context of ESBL and <i>bla</i> _{CTX-M-15}	166
5.5 DISCUSSION	178
5.5.1 Early microbiome assembly and resistome establishment.....	178
5.5.2 Multi-class resistance clustering and host-context dynamics.....	180
5.5.3 Rare but high-stakes resistance markers.....	182

5.5.4 Clinical significance: from gut colonisation to infection risk.....	183
5.6 CONCLUSION.....	184
CHAPTER 6: EARLY-LIFE MODULATORS OF THE INFANT GUT MICROBIOME AND RESISTOME: AN ASSOCIATION STUDY	185
6.1 INTRODUCTION	185
6.2 AIMS	188
6.3 CHAPTER-SPECIFIC METHODS.....	189
6.3.1 Early-life modulators	189
6.3.2 Outcomes.....	191
6.3.3 Statistical analysis	191
6.4 RESULTS	193
6.4.0 Results overview: distribution of key modulators and rationale for antibiotic windowing.	193
6.4.1 Mode of delivery: early microbiome and resistome difference	199
6.4.2 Infant antibiotics and acute-window changes in microbiome/resistome endpoints.....	207
6.4.3 Feeding pattern and clinically prioritised microbiome/resistome markers	214
6.4.4 Combined exposures and synergy	222
6.5 DISCUSSION	228
6.5.1. Delivery mode: early signatures with rapid convergence	229
6.5.2 Timing dependence and the neonatal window.....	231
6.5.3 Infant antibiotics: modifiers of an already-seeded resistome	231
6.5.4 Risk stacking: cumulative exposures and ESBL marker clustering.....	233
6.5.5 Synthesis and implications	234
CHAPTER 7: GENERAL DISCUSSION.....	236
7.1 GENERAL OVERVIEW	236
7.2 HOW EXPOSURE SHAPES EARLY-LIFE AMR ACQUISITION	241
7.2.1 Colonisation pressure in resource-limited settings	241
7.2.2 Trajectories of microbiome development	244
7.3 FROM MECHANISMS TO ACTION: IDENTIFYING PATHWAYS TO REDUCE AMR COLONISATION	248
7.3.1 Delivery Pathways and Early Care Contacts	249
7.3.2 Selection Pressure and Antibiotic Stewardship.....	250
7.3.3 Microbial Maturation, Feeding, and Hygiene Constraints	252
7.3.4 Why Marker-Specific Carriage Is More Actionable Than Community-Wide Metrics.....	254
7.4 STRENGTH AND LIMITATIONS OF THE STUDY	255
7.5 FUTURE DIRECTIONS	257
7.5.1 Immediate Priority: Characterising Maternal and Household Reservoirs	257
7.5.2 Resolve Genetic Context and Mobile Genetic Elements for Clinically Prioritised Markers	258

7.5.3 Longitudinal Stability and the Weaning Transition	258
7.5.4 Proposed Institutional Engagement and Protocol Review	259
7.6 A CONCEPTUAL FRAMEWORK TO IDENTIFY FUTURE INTERVENTIONAL STRATEGIES	261
REFERENCES.....	264
APPENDICES	307

Abstract

Antimicrobial resistance (AMR) remains a critical threat to infant survival, especially in low- and middle-income countries (LMICs). As gut colonisation by resistant pathogens often precedes clinical infection and sepsis, understanding early-life microbiome and resistome development is essential. This thesis characterises the longitudinal co-development of the gut microbiome and resistome in a northern Nigerian birth cohort (n=87) from birth to 90 days of life, investigating the impact of clinical and environmental modulators in a resource-limited setting.

Utilising shotgun metagenomic sequencing of longitudinal faecal samples, with high-intensity sampling during the first postnatal week, I employed Kaplan-Meier survival analysis and covariate-adjusted regression models to quantify the timing and determinants of antimicrobial resistance gene (ARG) acquisition.

Results revealed a rapid early acquisition of ARGs within the first week of life. Although 10% of infants carried β -lactam or aminoglycoside ARGs at birth, prevalence rose to 90% by day 7. Near-universal β -lactam or aminoglycoside ARG prevalence was established by the second week, preceding most recorded infant antibiotic exposure. Taxonomic composition was dominated early by *Escherichia* and *Klebsiella*, with delayed *Bifidobacterium* colonisation in caesarean-section infants (first appearance day 14 vs days 0-7 in vaginally delivered infants). Delivery mode explained 6.2-7.8% of microbiome and resistome variation in the first week; however, this was attenuated by the third month, suggesting faster convergence than that reported from high-income countries.

Contextual assessment of the cohort setting revealed routine postpartum antibiotic administration (contrary to global recommendations), constrained sanitation infrastructure,

and compact household contact networks. These findings demonstrate that in high-exposure settings, AMR acquisition is contracted in the immediate neonatal window and driven by a multifactorial combination of maternal, healthcare, and environmental factors rather than infant-level antibiotic use alone. This thesis proposes a four-stage framework to guide intervention design, prioritising infection prevention, antibiotic stewardship, and support for microbial maturation, applied to LMICs.

Acknowledgements

Looking back on this DPhil journey, it has been both challenging and incredibly rewarding. Attempting to express my heartfelt gratitude to everyone who has supported me feels like a monumental task, because every individual's contribution deserves more recognition than these pages can hold.

First and foremost, I extend my thanks to Almighty God for this amazing opportunity and for being my anchor through every challenge and victory. My life and experiences are a true testament to Your unwavering presence; through every twist and turn, You have consistently shown me that I am indeed “*Nwachinaemelu*”.

I am truly thankful for my wonderful supervisors: Prof. Timothy Walsh, Associate Prof. Mathew Stracy, Dr Kirsty Sands, and Dr Kathryn Thomson. Each of you has shaped my journey personally and academically, fostering my growth and sharing invaluable lessons and opportunities I will always cherish. I sincerely appreciate your guidance and the insightful feedback that inspired me to pursue excellence. A special thank you to Dr Kirsty Sands for your unwavering support, thoughtful insights, and immense patience.

My heartfelt thanks also go to my “backbones” both within and outside of the Walsh lab: Maria, Kate, and Shonnette. Thank you for powering through this DPhil with me and offering the encouragement and shared moments of reflection that made the journey easier. To Laura and Adi, thank you for your time, support and tutoring. To the entire Walsh lab, thank you for being a safe space.

I am endlessly thankful for my beloved family; your prayers, love, and support were immeasurable. To my parents, Engr and Mrs. Tony Uko-Akpulu, my siblings, Chinelo,

Nkee, Ogoo, Ekene: your love, pep talks, laughter, and constant calls meant more to me than words can capture. To my brother-in-law, Omalicha, thank you for keeping me in check and sane.

To my “Oxford family/hiding refuge”, the Jev-Achi family (Jev, Chioma, Mone, Jessy and Amy): thank you for being my home away from home. I am so glad I gained a family in you.

To the three rockstars and research assistants who supported my work in Nigeria – Hafsat, Hafsat-Sada, and Hassana: thank you for your sacrifices and tenacity. Your diligence made a meaningful difference in my data collection.

Lastly, to the people who made this work possible, the study participants: thank you for your resilience, courage, and contribution to the world. This thesis is dedicated to your strength and your aspirations for a better Nigeria.

Declaration of contributions to thesis

I declare that this thesis is substantively my own work with appropriate support and assistance from my supervisors and colleagues. My supervisors, Tim Walsh and Kirsty Sands, provided scientific guidance throughout all the stages of this work. Specific assistance I have received from others in relation to this thesis is outlined below. In general, I am grateful to Nigerian collaborators Professor Kenneth Iregbu (National Hospital Abuja, NHA) and Dr Aminu Aliyu (Aminu Kano Teaching Hospital, AKTH) for their general clinical microbiology guidance and advice.

Chapter 2: Laura de Nies designed the pipeline in Figure 2.3.

Chapter 3: The pilot study was done with the help of the Nigerian collaborators at the Murtala Muhammed Specialist Hospital, Kano (MMSH) site. Particularly, Rashida Yakubu and Aisha Mukkadas for helping with sample collection. Also thank the department of molecular microbiology laboratory at the Aminu Kano Teaching Hospital (AKTH) for providing ultra-low temperature freezer storage space for the experiment in this chapter. Aditya plotted figure 3.5

Chapter 4: All sampling was facilitated by collaborators from the study staff in Nigeria, particularly at Murtala Muhammed Specialist Hospital, Kano (MMSH), who helped enroll participants, collect samples and demographic data (questionnaire), and followed the participants at their various homes during the follow-up visits.

Chapter 5,6: I am especially grateful to Aditya Lankapalli and Laura de Nies who aided with initiating the bioinformatic analysis. Aditya plotted figure 5.3

Publications

Papers arising from work presented in this thesis

Akpulu C., Lankapali A. K., Toufiq R., Cook K., Portal E. AR, Yakubu Khalid R., Mukkadas A., Gambo S., Aminu A., Iregbu K., Thomson K., Walsh T., Sands K. (2025). Comparative evaluation of DNA extraction protocols for neonatal gut microbiota profiling in a resource-limited setting. *The Microbe*, 7: 100398.

Akpulu C., Yakubu R., Tukur F., *et al.* (2025). Beyond the samples: Nigerian research staff perspectives on a multicentre neonatal sepsis study's impact. *Nature Communications*, 16:6745.

Akpulu C., Maikudi H., Ahmed H. *et al.* Cohort Profile: Infant Gut Bacterial Study in Nigeria (INBUGS-NG). *BMJ Open* 2026;16: e111007.

Presentation at international conferences

Optimizing stool DNA extraction protocol to evaluate the development of the neonatal resistome in northern Nigeria 2023. *Poster presentation at the 4th International World of Microbiome Conference (WoM).*

Effect of antibiotic use on the gut microbiome of neonates: a northern Nigeria cohort study (2025). *Poster presentation at the 35th ESCMID Global.*

List of abbreviations

AKTH	Aminu Kano Teaching Hospital
AME	aminoglycoside-modifying enzymes
AMR	Antimicrobial resistance
AMRFinderPlus	Antimicrobial Resistance Gene Finder
ARG-ANNOT	Antibiotic Resistance Gene-ANNOtation
ANT	Aminoglycoside nucleotidyltransferase
APGAR	Appearance, Pulse, Grimace, Activity, Respiration
AR	Antibiotic Resistance
ARG	Antimicrobial resistance gene
ARO	Antibiotic Resistance Ontology
AST	Antimicrobial susceptibility testing
AWaRe	Access, Watch, Reserve
BARNARDS	Burden of Antibiotic Resistance in Neonates from Developing Societies
BF	Breastfeeding
<i>bla</i>	β -lactamase gene
CARD	Comprehensive Antibiotic Resistance Database
CI	Confidence Interval
CLR	Centered Log-Ratio
CRF	Clinical Report Form
CS	Caesarean Section
DALY	Disability-Adjusted Life Year
DNA	Deoxyribonucleic acid
EBF	Exclusive Breastfeeding
EEA	European Economic Area
EOS	Early onset Sepsis
ESBL	Extended-Spectrum Beta-Lactamase <i>Enterococcus faecium</i> , <i>Staphylococcus aureus</i> , <i>Klebsiella pneumoniae</i> , <i>Acinetobacter baumannii</i> , <i>Pseudomonas aeruginosa</i> ,
ESKAPE	<i>Enterobacter species</i>

EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
FGM	Female Genital Mutilation
GBD	Global Burden of Disease
GDP	Gross Domestic Product
HGT	Horizontal Gene Transfer
HIC	High-Income Country
HMO	Human Milk Oligosaccharide
HMW	High Molecular Weight
IAP	Intrapartum antibiotic prophylaxis
IHMS	International Human Microbiome Standards
IPC	Infection Prevention and Control
IQR	Interquartile range
IS	Inverse Simpson Index
LGA	Local government area
LMIC	Low- and Middle-Income Country
MDR	Multidrug-Resistant
MDRO	Multidrug-Resistant Organism
MGE	Mobile Genetic Elements
MMSH	Murtala Muhammed Specialist Hospital
NCBI	National Center for Biotechnology Information
NHREC	National Health Research Ethics Committee
NICU	Neonatal Intensive Care Unit
ONT	Oxford Nanopore Technologies
OTC	Over-the-Counter
OxTREC	Oxford Tropical Research Ethics Committee
PCR	Polymerase Chain Reaction
PERMANOVA	Permutational Multivariate Analysis of Variance
PHC	Primary health care
PROM	Prolonged Rupture of Membranes
QC	Quality control
RGI	Resistance Gene Identifier
RNA	Ribonucleic Acid

SCBU	Special Care Baby Unit
SCFA	Short Chain Fatty Acid
SDG	Sustainable Development Goals
SF	Solid food
SHV	sulphydryl variable
SPRI	Solid Phase Reversible Immobilization
SVD	Spontaneous Vaginal Delivery
SXT	Sulfamethoxazole-trimethoprim
UNICEF	United Nations Children's Fund
US	United States
VIM	Verona Integron Metallo-beta-lactamase
WASH	Water, Sanitation, and Hygiene
WHO	World Health Organization

Chapter 1: Introduction

1.1 Global and clinical context of antimicrobial resistance

1.1.1 Contextual overview and global burden of antimicrobial resistance (AMR)

Throughout history, bacterial infections have been a major cause of morbidity and mortality in humans. The discovery, development, and medical use of antibiotics marked the end of the pre-antibiotic era by providing effective treatments against bacterial infections. However, this advancement has been speedily accompanied by the growing challenge of antimicrobial resistance (AMR), defined as the ability of microorganisms such as bacteria, viruses, fungi, and parasites to survive exposure to antimicrobial agents that would ordinarily inhibit or kill them (World Health Organization, 2023). AMR undermines the effectiveness of standard treatments, allowing infections to persist and spread (World Health Organization, 2023). Although antibiotic resistance occurs through natural evolution and adaptation (D’Costa et al., 2011), it has been dangerously exacerbated by human activity, particularly the inappropriate use of antibiotics in healthcare, agriculture, and the environment. The clinical and economic consequences of AMR are critical and continue to rise. In 2019, an estimated 1.27 million deaths were directly attributable to bacterial AMR, with nearly 5 million deaths associated overall (Murray et al., 2022). This burden is unevenly distributed, with low- and middle-income countries (LMICs) disproportionately affected due to weaker health systems, limited diagnostic capacity, and inadequate regulatory oversight. More recent Global Burden of Disease (GBD) 2021 estimates (Naghavi et al., 2024) reported a slightly lower figure of 1.14 million attributable deaths in 2021; however, forecasts remain deeply concerning. The 2016 O’Neill Review predicted up to 10 million annual deaths by 2050 in the absence of effective action (O’Neill, 2016), a trajectory supported by the GBD 2021 study, which projects a steady rise in mortality throughout the

mid-century. These reports also present a striking regional breakdown (Figure 1.1), illustrating that South Asia and Sub-Saharan Africa are projected to bear the highest mortality burden by 2050. Global fatalities are forecasted to rise to approximately 1.91 million attributable and 8.22 million associated deaths per year due to bacterial AMR (Naghavi et al., 2024).

Beyond its clinical consequences, AMR poses a serious threat to the global economy and the achievement of the Sustainable Development Goals (SDGs), a set of 17 interlinked objectives adopted in 2015 to address poverty, health, inequality, and environmental sustainability (Aslam et al., 2024). Without effective intervention, AMR could reduce Gross Domestic Product (GDP) in LMICs by 5% or more by 2050 (Jonas et al., 2017). Under more alarming forecasts, regions such as Sub-Saharan Africa could face up to 4 million deaths annually, contributing to cumulative global economic losses of US\$100-210 trillion without substantial intervention (O'Neill, 2016). Economic assessments in high-income countries (HICs) also demonstrate substantial costs. In the United States, treating infections from just six high-priority drug-resistant pathogens costs the healthcare system approximately US\$4.6 billion annually (Centers for Disease Control and Prevention (CDC), 2023). In the European Union/European Economic Area (EU/EEA), the annual extra healthcare expenditure attributed to resistant pathogens is estimated at €6.6 billion, contributing to a combined economic burden of €11.7 billion when productivity losses are included (Organisation for Economic Co-operation and Development (OECD), 2023). If left unaddressed, AMR could push an estimated 24 million people into extreme poverty by 2030, particularly in LMICs (Jonas et al., 2017).

1.1.2 AMR Burden in Neonates and Infants in LMICs

While these global forecasts highlight the economic and developmental risk of AMR, the human cost is most evident in neonatal and infant infections. Neonatal and infant sepsis remains a leading cause of morbidity and mortality in LMICs, with the spread of AMR further compounding the challenge of effective management (Murray et al., 2022; Naghavi et al., 2024; Sands et al., 2021; Sands, Spiller, et al., 2022). Recent global estimates suggest that approximately 3 million cases of neonatal sepsis occur annually in LMICs, resulting in around 570,000 deaths (Fleischmann et al., 2021), accounting for the majority of global neonatal deaths (Naghavi et al., 2024). In sub-Saharan Africa alone, conservative estimates attribute over 8 million disability-adjusted life years (DALYs) annually to neonatal sepsis, with economic burden exceeding \$400 billion (Ranjeva et al., 2018).

Despite the magnitude of this burden, high-quality population-level data on the aetiology and antibiotic resistance (AR) profiles of neonatal and infant infections remain scarce (Fleischmann et al., 2021; Milton, Gillespie, et al., 2022; Russell et al., 2023). Most available data arise from tertiary hospitals, introducing biases that underestimate the broader community burden and mask regional differences in resistance profile. Large-scale studies conducted between 2011 and 2024, such as BARNARDS (Sands et al., 2021), BabyGERMS (Mashau et al., 2022), CHAMPS (Mahtab et al., 2024), DENIS (Investigators of the Delhi Neonatal Infection Study (DeNIS) collaboration, 2016), and NeoOBS (Russell et al., 2023), have provided valuable insights, consistently identifying Gram-negative bacteria, particularly *Klebsiella pneumoniae*, *Acinetobacter baumannii*, and *Escherichia coli*, as well as the Gram-positive *Staphylococcus aureus*, as leading aetiological agents in both early-onset (days 0-2) and late-onset (days 3-28) neonatal sepsis (Harrison et al., 2024). Data on infants beyond the neonatal period, in contrast, are far more limited and often less robust,

largely because surveillance systems and studies tend to aggregate outcomes for all infants under one year or under five years, and because most originate from tertiary hospitals rather than community settings (Dramowski et al., 2025; K. Iregbu et al., 2022). This lack of disaggregated, age-specific data is particularly problematic, since early colonisation patterns are known to shape infection risk well into infancy (Carvalho et al., 2022; Iqbal et al., 2024; Schwartz et al., 2023).

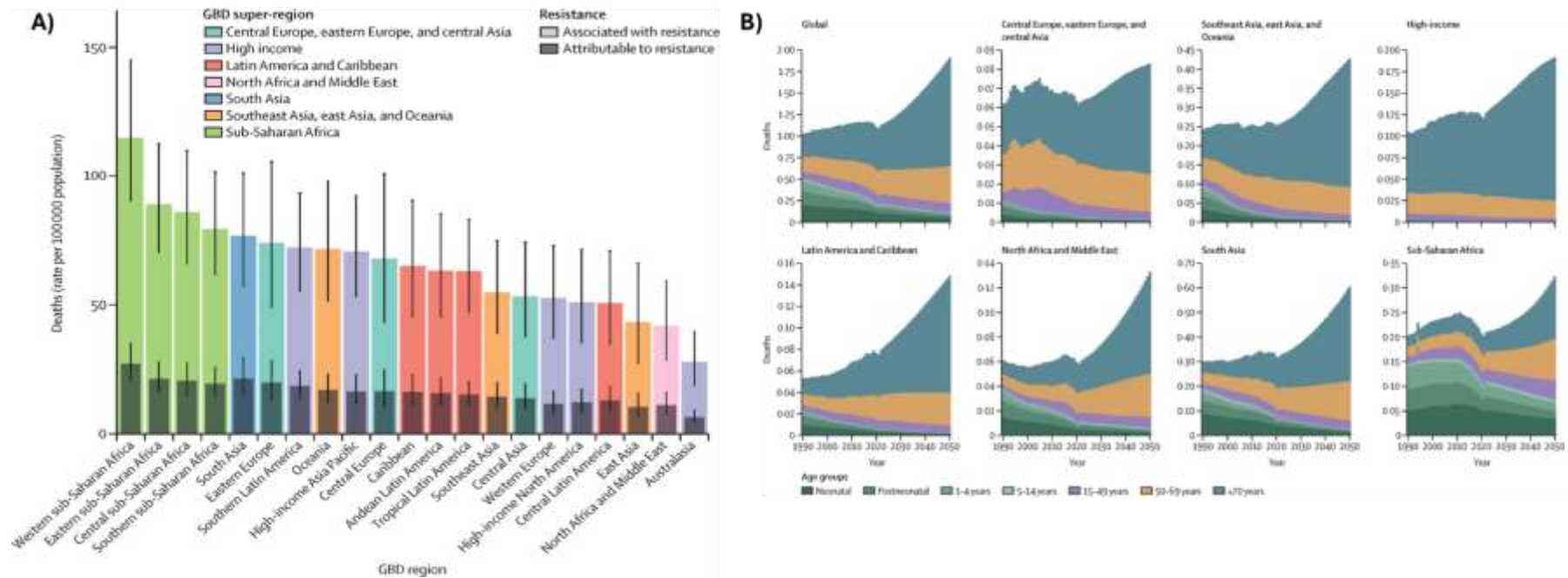


Figure 1.1 Global and regional burden of bacterial antimicrobial resistance (AMR)

(A) All-age death rates per 100,000 population attributable to and associated with bacterial AMR in 2019, by Global Burden of Disease (GBD) region. Sub-Saharan Africa and South Asia experienced the highest age-standardised mortality rates, highlighting the disproportionate burden in low- and middle-income regions (Murray et al., 2022)

(B) Observed and forecasted deaths attributable to bacterial AMR by GBD super-region and age group, 1990–2050. Projections suggest that, without intensified interventions, the AMR burden will continue to rise through mid-century, with substantial regional variation (Naghavi et al., 2024). Reproduced from Murray *et al.* (2022)(Murray et al., 2022), and GBD 2021 AMR Collaborators (2024) (Naghavi et al., 2024). Both articles are open access under a Creative Commons CC BY 4.0 license

Reported AR rates highlight the severity of the problem. Across LMICs, including multi-country studies and national reports from regions such as South Asia and sub-Saharan Africa, resistance to aminoglycosides ranges from 42%-69%, third-generation cephalosporins 59%-84%, and carbapenems 13%-82% (Investigators of the Delhi Neonatal Infection Study (DeNIS) collaboration, 2016; Mahtab et al., 2024; Mashau et al., 2022; Russell et al., 2023; Sands et al., 2021). These wide ranges reflect substantial regional variation, differing prescribing practices, and heterogeneous surveillance methods, which complicate cross-study comparisons. The prevalence of antimicrobial-resistant organisms (AROs) critically undermines the efficacy of WHO-recommended empiric neonatal regimens, likely contributing to persistently high sepsis-attributable mortality.

While much of the focus has been on clinical outcomes and therapeutic failure, a crucial but underexplored aspect of this challenge is the role of the developing gut microbiome. Acting as both a reservoir and potential amplifier of AMR, the infant gut highlights the need to examine colonisation dynamics in early life.

1.1.2.1 Microbiome and colonisation: the missing link in neonatal AMR dynamics

Colonisation with AROs is a critical precursor that often determines subsequent risk of invasive disease (Iqbal et al., 2024; Schwartz et al., 2023). The neonatal gut microbiome, still immature and ecologically unstable in the first weeks of life, not only serves as a potential reservoir for AR genes but also influences whether colonisation with AR strains is suppressed, transient, or persistent (Bargheet et al., 2025; Ravi et al., 2015). Microbiome composition, diversity, and functional capacity play a decisive role in shaping vulnerability to invasive infection (Pickard et al., 2017). Despite this, surveillance frameworks and burden estimates have focused predominantly on bloodstream infections, leaving minimal data on

colonisation patterns. This omission is significant, as colonisation may represent the missing link in understanding how AMR organisms establish, persist, and ultimately progress to invasive infections in early life.

Colonisation with AROs such as *K. pneumoniae*, *E. coli*, and *A. baumannii* has been documented within days of birth, particularly in hospital-born neonates exposed to intensive care units and empiric antibiotic therapies, with reported carriage rates of extended-spectrum β -lactamases (ESBL)-producing Enterobacteriaceae exceeding 40–70% in some LMIC neonatal wards (Labi et al., 2020; Sands et al., 2021). Transmission occurs through multiple pathways, including vertical transfer from the maternal microbiome, horizontal acquisition within healthcare facilities, and exposure to contaminated environments and/or healthcare workers (Busi et al., 2021; Jost et al., 2014; Labi et al., 2020).

In LMICs, where empiric antibiotic use is widespread and infection control measures are frequently inadequate, these pressures are magnified, accelerating the establishment and persistence of AR strains. Importantly, early-life studies have demonstrated that mobile genetic elements (MGEs), DNA segments that move within genome or between organisms to facilitate the spread of antimicrobial resistance genes (ARGs) (Partridge et al., 2018), can persist in the infant gut microbiome for years enabling colonising flora to act as a long-term reservoir for ARGs (Ravi et al., 2015).

Such colonisation may appear asymptomatic but carries significant risks. Carriage significantly increases the risk of subsequent bloodstream infection, facilitates horizontal gene transfer (HGT) - the exchange of genetic material, including ARG, between bacteria, within the microbial community - and contributes to recurrent hospital outbreaks (Bargheet et al., 2025; Iqbal et al., 2024; Labi et al., 2020). The neonatal microbiome also provides protection against colonisation by pathogenic bacteria, often referred to as colonisation

resistance (1.2.2). However, this protective effect is compromised by dysbiosis induced by factors including antibiotic exposure, caesarean delivery, or formula feeding tipping the balance towards persistent colonisation by potential pathogens and infection risk (Amenyogbe et al., 2017; Busi et al., 2021). In this context, the distinction between colonisation and invasive disease is not discrete but rather a spectrum shaped by microbial, host, and environmental factors.

Despite growing recognition of its importance, systematic data on AMR carriage in infants, particularly beyond the neonatal period, remain scarce, especially in LMIC contexts (1.1.2). This gap underscores the need to investigate the infant gut microbiome not only as a passive reservoir of resistant organisms but as an active driver of AMR dynamics in early life.

1.1.3 Mechanisms and drivers of AMR relevant to early-life colonisation

AMR in bacteria emerges through both intrinsic and acquired mechanisms, with the latter posing the greatest concern in early life. While intrinsic resistance derives from structural traits such as impermeable outer membranes or efflux pumps (Blair et al., 2014; Kumar & Schweizer, 2005), acquired resistance results from spontaneous mutations or HGT (Huddleston, 2014; Kumar & Schweizer, 2005), processes particularly relevant in the ecologically unstable infant gut.

Three principal mechanisms underpin acquired resistance: (i) enzymatic inactivation of antimicrobials, exemplified by ESBLs hydrolysing penicillins and cephalosporins; (ii) modification of antibiotic targets, such as alterations in penicillin-binding proteins and DNA gyrase that reduce β -lactam and fluoroquinolone susceptibility; and (iii) efflux pumps and permeability changes that decrease intracellular drug concentrations (Blair et al., 2014). The

prevalence and dynamics of these mechanisms in the neonatal microbiome remain poorly characterised.

HGT accelerates the dissemination of AR in early life. Conjugation, transformation, and transduction are classical routes, while emerging evidence highlights vesiduction, whereby extracellular vesicles transport resistance determinants, including integrons and plasmids, as a more recently discovered phenomenon (Huddleston, 2014; Soler & Forterre, 2020). The high microbial turnover and reduced colonisation resistance of the neonatal gut likely amplify the efficiency of these transfer events.

Beyond molecular mechanisms, the drivers of AMR in early life are strongly anthropological, shaped by human practices, healthcare structures, and environmental contexts. Maternal intrapartum prophylaxis, empiric neonatal antibiotic therapy, and prolonged neonatal intensive care unit (NICU) stays introduce intense antibiotic pressure, often before the gut microbiome has stabilised (Busi et al., 2021; World Health Organization, 2023). In LMICs, these pressures are magnified by limited diagnostic capacity, weak regulation of antibiotic sales, and suboptimal infection prevention and control (IPC) measures (Labi et al., 2020; O'Neill, 2016; Sands et al., 2021). Such conditions encourage both the acquisition of AR strains and the persistence of ARGs within commensal reservoirs. Thus, while the mechanisms of AMR are well described *in vitro* and in adult infections, the prevalence, expression, and clinical impact of ARGs in early-life colonisation remain insufficiently defined, a context in which the developing infant gut microbiome plays a central role.

1.2. The developing infant gut microbiome

1.2.1 Developmental stages and early colonisation

The establishment of the gut microbiome in early life is a dynamic process that plays a pivotal role in immune and metabolic development, with implications for both immediate and long-term health outcomes (Amenyogbe et al., 2017; Arrieta et al., 2014).

The foetal gut has traditionally been considered sterile, with colonisation thought to occur only during and after delivery. However, reports of microbial DNA in umbilical cords (Jiménez et al., 2005), amniotic fluid (Collado et al., 2016; DiGiulio et al., 2010), placenta tissue (Aagaard et al., 2014; Amarasekara et al., 2015; Collado et al., 2016; Onderdonk et al., 2008; Prince et al., 2016), and meconium (Gosalbes et al., 2013, 2016) have challenged this view. Whether these findings represent true colonisation or reflect low-level contamination remains debated, and their clinical significance for neonatal health is uncertain. If confirmed, prenatal microbial exposure may have implications for early establishment of the resistome, defined as the collection of all ARGs in microorganisms (Wright, 2007), and sometimes restricted to specific context (e.g. the neonatal resistome).

Postnatally, microbial colonisation begins immediately and is strongly influenced by the circumstances of birth and early exposures (Bäckhed et al., 2015; Dominguez-Bello et al., 2010). Vaginally delivered infants (SVD) typically acquire maternal vaginal and gut taxa, while those born by caesarean section (CS) are more often colonised by skin- and hospital-associated organisms (Dominguez-Bello et al., 2010; Palmer et al., 2007). Feeding practices further shape these trajectories: breastfeeding enriches *Bifidobacterium* and other taxa that promote gut barrier integrity (Laursen et al., 2021; Lokossou et al., 2022; Lyons et al., 2020), whereas formula feeding is associated with higher prevalence of Enterobacteriaceae and other facultative anaerobes (Ma et al., 2020; K. M. Pärnänen et al., 2022). These early

differences, particularly when compounded by antibiotic exposure, can persist for months and have been linked to increased risk of AMR Enterobacteriaceae carriage (Bergström et al., 2014; Busi et al., 2021; Roswall et al., 2021).

By the second year of life, the microbiome usually progresses through developmental, transitional, and stable phases, gradually approaching an adult-like configuration (Roswall et al., 2021; Stewart et al., 2018; Yatsunenko et al., 2012), although this trajectory may be delayed or altered by recurrent antibiotic exposure, poor sanitation, and high pathogen loads (Busi et al., 2021; K. M. Pärnänen et al., 2022; Russell et al., 2023). Despite these insights, the extent to which early colonisation trajectories determine the persistence and amplification of AMR determinants into later infancy remains unclear, representing a critical gap in our understanding of AMR dynamics in early life.

1.2.2 Functional roles in immunity and metabolism

The infant gut microbiome plays a foundational role in shaping host physiology during early life, influencing immune maturation (Belkaid & Hand, 2014; Gensollen et al., 2016), nutrient metabolism (Magnúsdóttir et al., 2015), and protection against pathogens (Magnúsdóttir et al., 2015). These functions are especially critical in infancy, when host systems are still developing and depend heavily on microbial signals for appropriate maturation (Amenyogbe et al., 2017; Arrieta et al., 2014).

Immune system maturation: Early microbial colonisation provides cues for immune education, enabling discrimination between pathogens and commensals. Beneficial taxa such as *Bifidobacterium* and *Lactobacillus* promote regulatory T-cell differentiation, modulate cytokine production, and enhance gut barrier integrity (Belkaid & Hand, 2014; Gensollen et al., 2016). Microbial metabolites, particularly short-chain fatty acids (SCFAs),

support the development of gut-associated lymphoid tissue and mucosal tolerance. Disruption to these processes, for example through CS delivery or antibiotic exposure, has been associated with increased risk of allergic and autoimmune conditions later in life (Arrieta et al., 2015; Cukrowska et al., 2020; Ryan et al., 2025). Much of this evidence come from high-income cohorts, and its generalisability to LMIC settings remains uncertain.

Metabolism contributions: The infant microbiota also contributes to nutrient metabolism, fermenting complex substrates into SCFAs that provide energy to epithelial cells and regulate systemic metabolism (Marcobal et al., 2010). In addition, microbial synthesis of vitamins such as K2 and B-group vitamins supports coagulation, energy production, and neurological development (Magnúsdóttir et al., 2015; Milani et al., 2017). These intrinsic functions highlight the microbiome's role as a metabolic organ during infancy. Disruption of these pathways may reduce the abundance of protective taxa and create niches for facultative anaerobes, many of which are frequent carriers of ARGs (Hidalgo-Cantabrana et al., 2017; Laursen et al., 2021). The modulatory role of diet and feeding practices on these processes is discussed further in Section 1.3.

Colonisation resistance: A key protective function of the early microbiome is colonisation resistance, whereby commensals such as *Bifidobacterium infantis* restrict pathogen expansion through nutrient competition, SCFA production, pH reduction, and bacteriocin synthesis (Henrick et al., 2019; Milani et al., 2017). Clinical studies suggest that *B. infantis* in preterm infants can reduce both pathogen abundance and gut ARGs loads (Henrick et al., 2019; O'Brien et al., 2022). Conversely, delayed colonisation of protective taxa or reduced diversity has been associated with increased susceptibility to infections, including *Clostridium difficile*, and higher ARG carriage (Gasparrini et al., 2019).

Synthesis: Immune, metabolic, and protective functions underscore the microbiome's central role in infant health. This thesis focuses specifically on colonisation and its relationship with AMR, recognising that immune and metabolic pathways, while important, fall outside its scope.

1.2.3 Why early life is a critical window

The neonatal and infant periods are widely recognised as a sensitive stage in which host microbiota interactions shape developmental processes. During this time, the gut microbiome undergoes rapid changes in response to environmental and host-derived cues. This plasticity, while essential for adaptation, also makes the infant gut particularly susceptible to disruption (Amenyogbe et al., 2017; Tamburini et al., 2016). Perturbations such as early antibiotic exposure, caesarean delivery, or formula feeding can delay microbial maturation, suppress colonisation by beneficial taxa, and promote early acquisition of ARGs (Amenyogbe et al., 2017; Gasparrini et al., 2019). Early-life as a critical window is graphically captured in Figure 1.2.

As discussed in Section 1.2.2, the infant immune system is still maturing and relies heavily on microbial signals for calibration. Early colonisers support the differentiation of regulatory T-cells, mucosal barrier development, and cytokine balance (Belkaid & Hand, 2014; Gensollen et al., 2016). When colonisation is disrupted, immune education may be compromised, increasing susceptibility not only to inflammatory disorders but also to infections. These vulnerabilities heighten the risk of AMR acquisition, as antibiotics administered during the first weeks of life have been shown to enrich ARGs in the gut microbiome, sometimes persisting for months or years (Saturio et al., 2023).

The ecological instability of the early microbiome further facilitates HGT between commensals and opportunistic pathogens, amplifying AR within the community. In LMICs, where infants face high infectious disease burdens and frequent empiric broad-spectrum antibiotic use, these risks are magnified (Russell et al., 2023; Sands, Carvalho, et al., 2022; Trosvik et al., 2024). The convergence of antibiotic pressure, immature immunity, and fragile microbial ecology creates conditions in which early-life AMR acquisition is not only likely but potentially enduring.

Increasing attention has been given to the concept of “microbial programming,” which suggests that exposures during this critical period set long-term trajectories for microbiota composition and function (Trosvik et al., 2024). Just as immune, metabolic, and neurological systems are shaped by early stimuli, so too is the resistome. Perturbations can leave lasting signatures that influence responses to future infections and antimicrobial therapies (K. Pärnänen et al., 2018). Despite these insights, most evidence derives from high-income settings. How early-life perturbations shape the persistence and dissemination of the infant resistome in LMICs remains insufficiently understood.

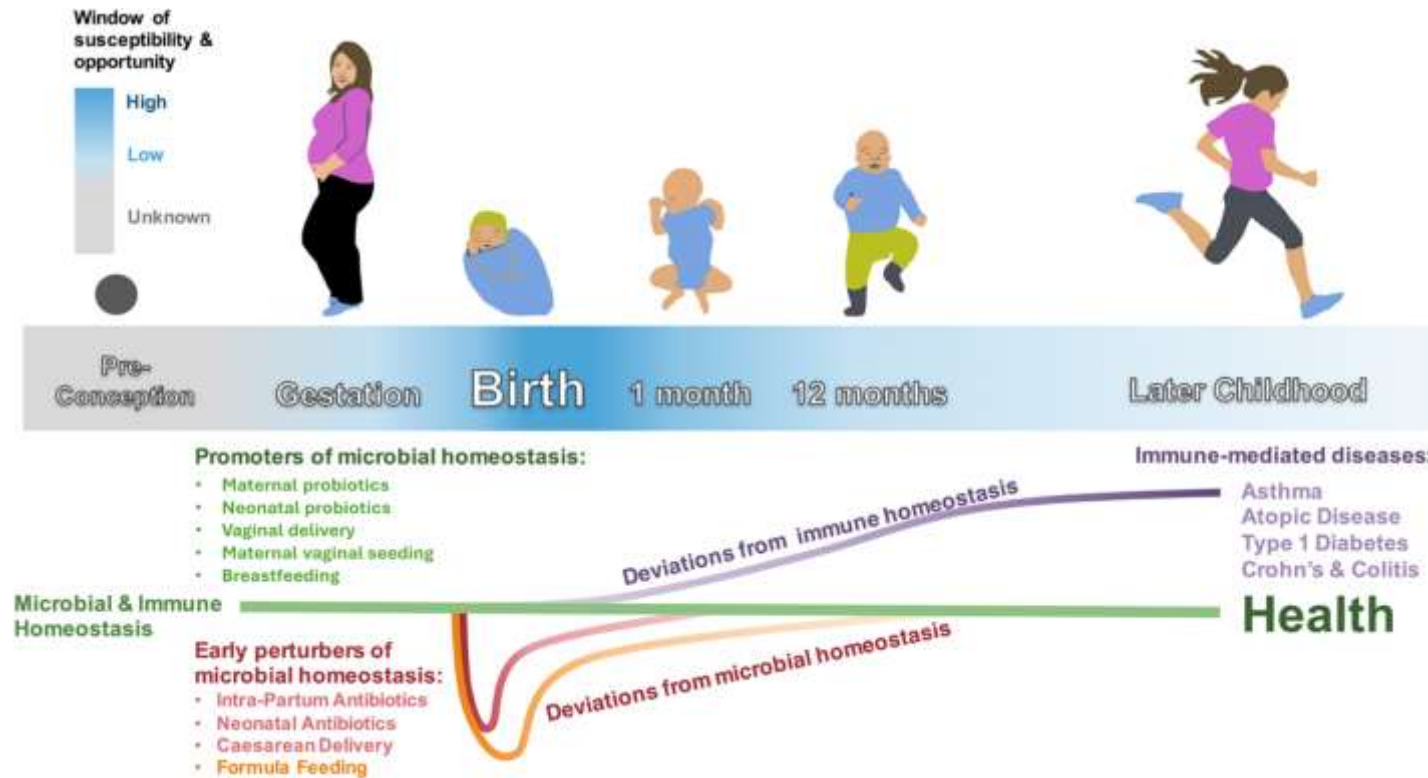


Figure 1.2. Conceptual model of early-life “window of susceptibility” linking perinatal microbial perturbations to later immune-mediated disease risk.

The schematic depicts key developmental periods from pre-conception through later childhood and illustrates how early-life exposures can shape trajectories of microbial and immune homeostasis. Promoters of microbial homeostasis (e.g., vaginal delivery, breastfeeding) are contrasted with early modulators (e.g., intrapartum and neonatal antibiotics, caesarean delivery, formula feeding). The coloured gradients indicate relative susceptibility/opportunity across time, while the diverging curves represent potential deviations from microbial and immune homeostasis that may persist and contribute to immune-mediated outcomes (e.g., asthma, atopic disease, type 1 diabetes, inflammatory bowel disease). *Adapted and modified from* (Amenyogbe et al., 2017).

1.3. Modulators of the microbiome and resistome in early life

The composition of the neonatal gut microbiome and resistome is shaped by external exposures from birth onward. Factors such as mode of delivery, infant feeding practices, antibiotic use, and environmental contact, act as modulators that influence which microorganisms colonise the infant gut, how microbial communities develop, and how ARGs are acquired or sustained (Bargheet et al., 2025; K. Pärnänen et al., 2018; Reyman et al., 2022; Thänert et al., 2022).

These effects vary with timing, intensity, and setting, and are particularly pronounced in LMICs, where high pathogen burden, frequent empiric antibiotic use, and limited infection control amplify microbiome disruption and resistome expansion (Russell et al., 2023; Sands, Carvalho, et al., 2022). An overview of these modulators, their primary mechanisms, and reported effects on microbiome and resistome development is summarised in Table 1.1.

Table 1.1 Modulators influencing early-life gut microbiome and resistome.

Modulator	Primary Mechanism	Microbiome outcomes	Resistome (ARGs) outcomes	Context (LMIC vs HIC)	Reference
Diet	HMOs, immune factors, nutrient profile	Breastfeeding: ↑ <i>Bifidobacterium</i> , <i>Bacteroides</i>	Breastfeeding linked to ↓ ARGs & MGEs	In LMICs, unsafe complementary foods amplify ARG acquisition	(Laursen et al., 2021; Lyons et al., 2020; Theophilus & Taft, 2023)
		Formula: ↑ <i>Enterobacteriaceae</i> , ↓protective taxa; Complementary feeding: ↑diversity	Formula feeding linked to ↑ ARG repertoire, even without antibiotics		(Baumann-Dudenhoeffer et al., 2018; Ma et al., 2020; K. M. Pärnänen et al., 2022)
Mode of delivery	Delivery route, maternal antibiotics, hospital environment	Vaginal: early colonisation by maternal gut/vaginal taxa	CS linked to ↑ ARG diversity & abundance	LMIC neonatal units: high ARG acquisition rates with CS + poor IPC	(Beaumont et al., 2025; Dominguez-Bello et al., 2010; Shao et al., 2019)
		CS: delayed anaerobes, ↑ skin/hospital taxa	NICU exposure compounds ARG acquisition		
Antibiotic exposure	Colonisation disruption, selective pressure,	↓Obligate anaerobes (<i>Bifidobacterium</i> , <i>Bacteroides</i>); ↑ facultative anaerobes (<i>Enterococcus</i> , <i>C. difficile</i>); ↓ diversity	↑ ARG abundance (β-lactams, macrolides, tetracyclines) ARG persistence via MGEs ↑ ESBL & carbapenemases in LMIC hospitals	Effects more severe with prolonged empiric therapy and weak IPC	(Dierikx et al., 2020; Luchen et al., 2023; Patangia, Grimaud, O'Shea, et al., 2024a; Saeed & Hall, 2023)

Modulator	Primary Mechanism	Microbiome outcomes	Resistome (ARGs) outcomes	Context (LMIC vs HIC)	Reference
Environment & household	Household and environmental transmission	Household density, sibling contact: ↑ diversity but also ↑ <i>E. coli</i>	ARG sharing within families (twins, mother–infant pairs)	LMICs: unsafe WASH major driver of ARG exposure	(Amato et al., 2023; Jokela et al., 2024; Lane et al., 2019; Laursen et al., 2015; Moore et al., 2015)
			crowding linked to ↑ ESBL carriage		
Geography & socioeconomic context	Sanitation, healthcare, regulation	LMICs: ↑ colonisation by resistant organisms at birth	LMICs: high background prevalence of ESBL, colistin-resistant strains	Within-country disparities strongly shape outcomes	(Anderson et al., 2024; Collignon et al., 2018; El-Maradny et al., 2025)
		HICs: risks from CS, formula, antibiotic overuse	HICs: ARG expansion via prescribing practices		

ARG (antibiotic resistance genes), LMIC (low- middle-income country), HIC (high-income country), HMO (Human Milk Oligosaccharides)

1.3.1 Diet

Diet is one of the strongest modulators of the infant gut microbiome. Breastfeeding promotes enrichment of protective taxa such as *Bifidobacterium* and *Bacteroides* (Laursen et al., 2021; Lokossou et al., 2022; Lyons et al., 2020), which metabolise Human Milk Oligosaccharides (HMOs) to produce SCFAs that strengthen gut barrier integrity and reduce expansion of facultative anaerobes. These taxa are associated with reduced ARG carriage, including lower prevalence of *Enterobacteriaceae* harbouring ESBL genes (Ma et al., 2020). Formula feeding, by contrast, is linked to increased microbial diversity but reduced *Bifidobacterium*, creating ecological niches that favour facultative anaerobes such as *Clostridium* and *Klebsiella* (Ma et al., 2020). Complementary feeding on the other hand diversifies the microbiome (Laursen et al., 2016; Palmer et al., 2007), but in LMICs may overlap with unsafe water and food exposures, amplifying ARG acquisition. Thus, diet can either reinforce microbiome resilience or facilitate resistome expansion depending on context.

1.3.2 Mode of delivery

Mode of delivery shapes the infant's earliest microbial exposures. Vaginal birth introduces maternal vaginal and gut microbiota, seeding colonisation with taxa such as *Bacteroides*, *Lactobacillus*, and *Bifidobacterium* that contribute to immune maturation and colonisation resistance (Dominguez-Bello et al., 2010; Shao et al., 2019). CS, in contrast, is associated with delayed acquisition of obligate anaerobes and increased colonisation by skin- and hospital-associated organisms (Dominguez-Bello et al., 2010; Shao et al., 2019).

Evidence suggests that infants born by CS may carry a higher abundance and diversity of ARGs during early life, potentially due to reduced colonisation resistance, intrapartum antibiotic prophylaxis, or hospital exposures (Shao et al., 2019). Longitudinal studies from

LMIC neonatal units highlight compounded risks, documenting high ARG acquisition where operative delivery, invasive interventions, and limited infection control coincide (Beaumont et al., 2025; Carvalho et al., 2022). These findings are particularly relevant given the global rise in caesarean deliveries, including in LMICs, where both medical and non-medical drivers are fuelling increasing rates (Betran et al., 2021).

1.3.3 Antibiotic Exposure

Antibiotics are among the major modulators of the early-life gut microbiome and driver of resistome expansion. Perinatal prophylaxis and empirical neonatal therapy can alter vertical microbial transmission, reduce taxonomic diversity, and favour the expansion of facultative anaerobes (Patangia, Grimaud, O'Shea, et al., 2024a; Saeed & Hall, 2023). These disturbances have been linked to altered immune maturation and metabolic programming and may persist long after treatment (Korpela et al., 2017; Leong et al., 2020).

From a resistome perspective, even short antibiotic courses enrich transferable ARGs, including those conferring resistance to β -lactams, macrolides, and tetracyclines (Karkman et al., 2019; K. Pärnänen et al., 2018). Many of these genes are located on MGEs, enabling HGT within the gut community and persistence for months to years (Gibson et al., 2016).

In LMIC neonatal units, where empiric therapy is widespread and AROs are prevalent, acquisition of ESBL- and carbapenemase-producing Enterobacteriaceae can occur within days of admission (Gibson et al., 2016; Guiton et al., 2022). Antibiotic exposure therefore accelerates both microbial disruption and resistome expansion in early life.

1.3.4 Environmental and Contextual Modulators

Beyond clinical and dietary factors, the infant gut microbiome and resistome can also be shaped by household, and environmental exposures. Horizontal transmission from mothers, caregivers, siblings, and close contacts introduces new taxa and AR determinants, with household density and daycare attendance associated with increased ARG carriage (Dominguez-Bello et al., 2010; Jokela et al., 2024). Environmental reservoirs, including contaminated soil, water, and surfaces, act as ongoing sources of resistant organisms, especially where water, sanitation, and hygiene (WASH) infrastructure is limited (Fuhrmeister et al., 2023). Early weaning in such contexts can further amplify risk, as infants are exposed to resistant bacteria through complementary foods prepared with untreated water.

These influences are cumulative: antibiotic-induced depletion of protective anaerobes may lower colonisation resistance, magnifying the impact of environmental exposures. In LMICs, where high background prevalence of resistant organisms intersects with weak IPC, environmental modulators significantly shape early resistome trajectories.

1.3.5 Geographic and Socioeconomic Context

Geographic and socioeconomic conditions frame the impact of all other modulators. Infants in LMICs are frequently exposed to multidrug-resistant organisms from birth, with studies reporting neonatal colonisation by ESBL- (Carvalho et al., 2022) and colistin-resistant bacteria in settings such as Nigeria (Portal et al., 2024). In contrast, HICs present lower baseline AMR pressures but face different risks, including higher rates of elective caesarean section, formula feeding, and broad-spectrum antibiotic prescribing (Collignon et al., 2018; Kuehl et al., 2020).

Within-country disparities further shape microbiome and resistome development: maternal education, household income, and healthcare access influence feeding practices, hygiene behaviours, and antibiotic use (El-Maradny et al., 2025). In resource-limited households, restricted access to clean water and sanitation can increase exposure to resistant enteric pathogens, while financial or geographic barriers to healthcare may encourage inappropriate antibiotic consumption (Anderson et al., 2024). These macro-level factors rarely act in isolation, instead compounding the pressures exerted by diet, delivery, antibiotics, and environment. By situating modulators within broader social and geographic contexts, it becomes clear that microbiome and resistome trajectories reflect not only biological processes but also structural determinants of health.

1.4. The early life resistome

1.4.1 Definition and significance

As previously mentioned in Section 1.2.1, the resistome comprises the full repertoire of ARGs within a microbial community, including both expressed and latent genes (D’Costa et al., 2006; Wright, 2007). These genes may be intrinsic or mobile, with the latter enabling the spread of resistance across diverse taxa (Davies, 2007; Martinez & Baquero, 2000). Importantly, ARG are not merely a modern byproduct of antibiotic use; they have ancient evolutionary origins and reflect a long history of microbial competition (Perry et al., 2016). Consequently, the infant resistome can reflect background microbial inheritance and environmental pressures as much as direct therapeutic selection.

In early life, the resistome develops in parallel with the gut microbiome during a period of rapid microbial assembly. The infant gut provides a permissive environment for ARG maintenance and exchange, characterised by high microbial density, close taxonomic

proximity, and repeated influence from maternal, environmental, dietary, and healthcare-associated sources. Notably, ARGs are detectable in the infant gut even without direct antibiotic exposure (Carvalho et al., 2022), indicating that resistome acquisition reflects both natural microbial inheritance and selective pressures from human practices. Once established, particularly when carried on MGEs, ARGs can persist for extended periods, shaping future responses to antibiotic treatment. Understanding the early-life resistome is therefore crucial for elucidating how resistance emerges before clinical infection, for anticipating pathways of dissemination, and for identifying opportunities to interrupt transmission in high-burden settings.

1.4.2 Detection and carriage of ARGs in infants

ARGs are detectable in the infant gut within days of birth, as evidenced by metagenomic signatures in meconium and early stool samples (Gasparrini et al., 2019; Moore et al., 2015). The presence of these ARGs in the earliest life stages suggests that the initial resistome is established through maternal and environmental transmission prior to or alongside clinical intervention. Once acquired, these ARGs, and the MGEs that facilitate their transfer, can persist in the infant gut for years (Ravi et al., 2015), establishing a long-term commensal reservoir that may shape the individual's resistome throughout early childhood.

Comparative cohort studies demonstrate variability across geographic settings. In HICs, the Finnish HELMi study found that ARG abundance was inversely associated with *Bifidobacterium* dominance in infancy, while the DIABIMMUNE cohort in Finland, Russia, and Estonia showed low baseline ARG diversity in antibiotic-naïve infants but notable expansion after antibiotic treatment (Jokela et al., 2024; Lebeaux et al., 2022). In contrast, studies in India and Nigeria report early colonisation with ARG-harbouring

Enterobacteriaceae even among community-born infants (Kothari et al., 2013; Portal et al., 2024), consistent with higher background exposure to resistant organisms in these settings. These contrasts highlight that resistome trajectories are context-dependent, shaped by local antibiotic practices, environmental exposures, and underlying microbiome composition.

1.4.3 Transmission routes and reservoirs

The early-life resistome is shaped by both vertical and horizontal transmission pathways, drawing from maternal, environmental, and healthcare-associated reservoirs. The importance of each transmission route varies across settings, shaped by maternal carriage, healthcare practices, and environmental exposures, with distinct risks observed in resource-limited contexts.

1.4.3.1 Vertical transmission

Maternal transmission is the earliest route of ARG acquisition. ARGs have been detected in meconium, the first-pass stool, suggesting possible in utero exposure or immediate perinatal transfer (Gosalbes et al., 2016; Shao et al., 2019). During vaginal delivery, neonates encounter maternal gut, vaginal, and perineal microbiota that can harbour genes such as *bla*_{TEM}, *ermB*, and *tetW*, many carried on MGEs capable of horizontal transfer (Bargheet et al., 2025; Gosalbes et al., 2016). The safety of this transmission route is linked to maternal health and the surrounding WASH infrastructure. In settings with poor WASH standards, mothers are more likely to be colonised by resistant 'ESKAPE' pathogens, a group of high-priority resistant bacteria (Rice, 2008), which are then passed to the neonate during birth (Carvalho et al., 2022). Caesarean delivery alters this dynamic, reducing direct maternal microbial transfer and increasing exposure to hospital-associated organisms such as ESBL-producing *K. pneumoniae* and *Enterobacter cloacae* (Nogacka et al., 2018).

Breast milk is another potential route of vertical transfer, containing commensal taxa and low-abundance ARGs, though selective enrichment of *Bifidobacterium* by HMOs generally favours colonisation resistance. Together, these vertical routes establish the initial ARG pool, but its long-term significance depends on the microbial balance within the infant gut.

1.4.3.2 Horizontal transmission

Horizontal transmission introduces ARGs postnatally from diverse reservoirs. Healthcare environments, especially NICUs, are high-density microbial hubs where infants may encounter AROs carrying genes such as *aac* (aminoglycoside resistance), *bla*_{KPC} (carbapenem resistance), and *bla*_{NDM} (metallo- β -lactamase activity) (Gibson et al., 2016; Xiao & Zhao, 2023). Colonisation with ESBL-producing *Enterobacteriaceae* or carbapenemase producers has been documented within days of NICU admission, particularly in LMICs where IPC practices may be limited (Sands, Spiller, et al., 2022).

Beyond healthcare, close contact with caregivers, siblings, and shared household surfaces facilitates ARG exchange, as shown by family studies documenting identical ARG in mother-infant dyads, twins, and siblings (Jokela et al., 2024; Moore et al., 2015). Environmental exposures, including untreated water, wastewater, animals, and contaminated surfaces, add further routes, with resistant *Enterobacteriaceae* commonly isolated from these reservoirs (Navab-Daneshmand et al., 2018; Pickering et al., 2019). Within the gut itself, commensal bacteria act as reservoirs and amplifiers of AR through HGT mechanisms such as conjugation (plasmid exchange), transduction (bacteriophage-mediated transfer), and transformation (uptake of free DNA), processes facilitated by high microbial density and turnover in early life (von Wintersdorff et al., 2016).

1.4.4 Clinical and public health implications

The early acquisition of ARGs in infancy has direct clinical consequences. Colonisation with ARG-harbouring taxa such as *K. pneumoniae*, *E. coli*, and *Enterococcus faecium* can precede invasive infections, particularly in vulnerable neonates in NICUs (Folgori et al., 2017; Zaidi et al., 2005). Although many ARGs remain phenotypically silent, selective pressures, especially antibiotic exposure, can rapidly expand resistant populations, complicate treatment and increasing mortality risk. Multi-country LMIC data from the BARNARDS study highlight how common ESBLs and carbapenemases are in neonatal sepsis isolates, with standard empiric regimens showing poor predicted coverage (Sands et al., 2021; Thomson et al., 2021). Findings from NeoOBS study similarly demonstrate high empiric use, limited diagnostic capacity, and poor treatment outcomes in LMIC neonatal units, underscoring the clinical relevance of early resistome dynamics (Russell et al., 2023). These realities align with global stewardship guidance such as the WHO AWaRe framework (WHO, 2021).

At the household and community level, infants function as reservoirs and vectors of ARGs. Close contact with caregivers, siblings, and shared environments facilitates bacterial exchange, with family studies documenting potential transmission of identical ARG across mother-infant dyads, siblings, and household members (Jokela et al., 2024; Moore et al., 2015). Broader environmental exposures amplify these risks: inadequate WASH infrastructure, animal contact, and contaminated food sources all contribute to ARG acquisition, particularly in densely populated or resource-limited settings. Research from Maputo, Mozambique illustrate how unsafe water and household wastewater increase gut ARG burdens in children (Berendes et al., 2019), while risk-factor analyses link ESBL-*E.*

coli carriage to crowding and local livestock production (Amato et al., 2023; Kurowski et al., 2021).

From a public health perspective, the persistence of ARGs poses a major challenge for antimicrobial stewardship. Longitudinal studies show that resistome shifts acquired in early life can persist for months or years (Gibson et al., 2016), reinforcing the need for interventions that begin earlier than is typically assumed. Global frameworks such as the WHO AWaRe antibiotic classification system (WHO, 2021) and the WHO/UNICEF Joint Monitoring Programme for WASH (World Health Organization (WHO) & the United Nations Children's Fund (UNICEF), 2021) emphasise that stewardship must be integrated with structural improvements in sanitation, hygiene, and healthcare delivery. In parallel, microbiome-supportive practices such as breastfeeding promotion and rational antibiotic use offer potential strategies to reduce resistome expansion in infancy.

1.5 Detecting ARGs in the microbiome

Detecting ARGs within the human microbiome presents unique methodological challenges. Unlike clinical diagnostics, which primarily focus on AR in specific pathogens, resistome analysis requires approaches that capture both culturable and unculturable organisms, as well as latent ARGs embedded in commensal taxa. Because the gut is a dense and diverse ecosystem, effective tools must identify not only resistant bacteria but also the genetic basis of resistance, its genomic context, and its potential for mobility. This section reviews the limitations of conventional culture- and PCR-based techniques, introduces metagenomic strategies for resistome profiling, and evaluates their relevance to early-life studies.

1.5.1 Limitations of traditional detection methods

Historically, AMR surveillance has relied on culture-based susceptibility testing and targeted molecular assays. While indispensable for clinical decision-making, these approaches offer a narrow and sometimes distorted view of the resistome, particularly in complex environment such as the infant gut.

Culture-based methods, including antimicrobial susceptibility testing (AST), infer AR phenotypes from bacterial growth in the presence of antibiotics. They are central to guiding treatment in infected patients but biased toward organisms that can be readily cultured under laboratory conditions. This is a major limitation for gut microbiome studies, as a large proportion of commensals, particularly obligate anaerobes and other fastidious taxa, cannot be recovered using conventional culture techniques (Browne et al., 2016). Consequently, many ARG in unculturable or low-abundance organisms remain undetected.

Targeted molecular methods such as PCR and qPCR partially overcome this limitation by detecting specific resistance genes (e.g., *bla*_{CTX-M}, *mecA*). These assays provide sensitivity and specificity for known and targeted ARGs but lack breadth, as they may miss novel or divergent variants depending on primer design and rarely resolve whether a gene is chromosomal or carried on MGE (Olsen & Riber, 2025; Saak et al., 2020). Moreover, because PCR amplifies its target, the apparent burden of an ARG in a complex community can be exaggerated, for example, whether *bla*_{NDM} is harboured by a dominant taxon or a transient, low-abundance species cannot be determined (Yamin et al., 2023).

1.5.2 Metagenomic approaches to ARG detection

Metagenomics has transformed AMR research by enabling comprehensive profiling of ARGs in complex microbial communities without the need for culture or prior assumptions about gene identity (X. Li et al., 2023; Olsen & Riber, 2025). In the infant gut, where many taxa are unculturable and resistance determinants may exist at low abundance or within commensals, this approach provides unique insights into early resistome assembly (Gasparrini et al., 2019; X. Li et al., 2024a).

The most widely applied strategy is shotgun metagenomic sequencing, in which total DNA from a sample is sequenced in an untargeted manner and aligned to curated resistance gene databases such as Comprehensive Antibiotic Resistance Database (CARD), ResFinder, ARG-ANNOT and AMRFinderPlus (Boolchandani et al., 2019; Nassereddine et al., 2024). This allows the detection of both abundant and rare ARGs, estimation of their relative abundance, and characterisation of their diversity. Two complementary analytical strategies are commonly employed:

- Read-based methods, which are computationally efficient and sensitive for detecting low-abundance ARGs (Y. Li et al., 2021).
- Assembly-based approaches, which facilitate genomic contextualisation, including co-location of ARGs with MGEs such as plasmids, integrons, and transposons (Abramova et al., 2024).

Genomic context is particularly important for evaluating the potential for HGT. Dedicated pipelines now attempt to predict mobility by annotating genes associated with conjugation/mobilization, but resolution remains imperfect, especially in metagenomes with limited sequencing depth or from under-sampled environments (Fan et al., 2021; Peng et al., 2021). Applied to early-life cohorts, metagenomics has shown that even antibiotic-naïve

infants carry diverse ARG repertoires within days of birth, including resistance determinants against β -lactams, macrolides, tetracyclines, and aminoglycosides (Gasparrini et al., 2019; K. Pärnänen et al., 2018).

It is important to note, however, that DNA-based metagenomics identifies the genetic blueprint of resistance rather than its functional activity (Quince et al., 2017). Whether detected ARGs are expressed cannot be determined without RNA-based meta-transcriptomics, which profiles gene expression, or functional assays (Sedova et al., 2025). Although such approaches are beyond the scope of this thesis, they represent the next step in linking resistome composition to actual resistance phenotypes in early life.

1.5.3 Strengths, limitations, and relevance to this study

Shotgun metagenomic sequencing offers a powerful, culture-independent approach to profiling the infant gut resistome. It enables detection of ARGs in both culturable and unculturable taxa, captures abundant and rare genes, and provides a broader and less biased view of resistance than traditional culture- or PCR-based methods. For early-life studies, this is particularly important given the dominance of obligate anaerobes and the presence of AR even in antibiotic-naïve infants.

Assembly-based strategies extend this by situating ARGs within their genomic context, including whether they are located on MGEs such as plasmids, key indicator of transfer potential. Emerging approaches such as chromosomal conformation capture (e.g., Hi-C) go further, attempting to link plasmids and resistance genes to specific host chromosomes. While promising, these techniques remain technically demanding, require stringent sample handling, and are not yet feasible for large-scale fieldwork in LMIC settings.

Nevertheless, both read- and assembly-based approaches are constrained by reliance on reference databases, which are heavily biased toward pathogens from HICs and may underrepresent ARGs circulating in LMICs. Assembly methods also demand greater sequencing depth and computational resources and may incorrectly collapse genomic contexts for genes present in multiple flanking environments. Meanwhile, functional metagenomics, considered the gold standard for demonstrating phenotypic activity, remains low-throughput and impractical for large cohorts.

In this study, I applied a largely assembly-based metagenomic approach, supplemented by read-based quantification using curated databases to balance feasibility and resolution. While functional assays and full plasmid-chromosome linkage are beyond scope, results are interpreted with attention to microbial and clinical context. This strategy enables robust characterisation of ARG diversity and mobility in Nigerian infants, while acknowledging the methodological limitations that define current knowledge gaps.

1.5.4 Summary of knowledge gaps

Despite substantial advances in understanding AMR and the human gut microbiome, key gaps remain in our knowledge of the early-life resistome, particularly in LMIC settings where the AMR burden is greatest. First, few longitudinal studies have characterised the resistome in neonates and infants from birth, leaving uncertainties about its natural developmental trajectory and the timing of peak acquisition of ARGs. Second, the interplay between microbiome assembly, environmental exposures, and clinical factors such as antibiotic use, feeding mode, and delivery mode remains poorly resolved in early life, especially in resource-limited contexts with distinct ecological and epidemiological pressures.

Third, while the potential routes of ARG transmission, from maternal sources, household contacts, and the environment are recognised, their relative contributions in LMIC settings have not been quantified with high-resolution, culture-independent methods. Fourth, most existing studies rely on targeted or culture-based approaches that may underestimate resistome diversity and miss novel or low-abundance ARGs; large-scale metagenomic analyses in early-life cohorts remain rare. Finally, few studies explicitly integrate resistome data with detailed environmental, demographic, and clinical metadata to identify modifiable drivers of AMR carriage in infancy.

Addressing these gaps requires well-designed, longitudinal cohort studies in LMIC contexts that combine robust microbiome and resistome profiling with detailed exposure assessment. This thesis aims to meet that need by applying metagenomic sequencing to infant stool samples collected over the first three months of life, integrating these data with rich contextual information to characterise the determinants, dynamics, and potential intervention points for AMR in early infancy.

1.6 Study rationale and objectives

1.6.1 Rationale for the study

AMR is a major global health challenge, yet its earliest origins in infancy remain poorly understood, particularly LMICs. The infant gut microbiome, still immature and immunologically naive, is especially vulnerable to colonisation by resistant organisms. Early colonisation events can establish long-term reservoirs of resistance genes, shaping both individual infection risk and population-level transmission dynamics.

Most existing resistome research has been conducted in HIC cohorts, where antibiotic stewardship, sanitation, and healthcare infrastructure differ from LMICs. In some sub-Saharan African countries, and Nigeria in particular, infants are frequently exposed to empirically prescribed broad-spectrum antibiotics, given unreliable diagnostics. These clinical pressures are compounded by environmental conditions that promote ARG acquisition, such as poor water quality, informal healthcare access, and unregulated antibiotic sales. Yet systematic, longitudinal investigations of how these exposures shape the developing resistome are virtually absent.

This study addresses that gap by applying longitudinal metagenomic profiling to a Nigerian infant cohort, integrating resistome data with clinical, demographic, and environmental exposures. As the first study of its kind in Nigeria, it will provide foundational evidence to inform local stewardship policies, strengthen surveillance capacity, and contribute to global understanding of AMR acquisition in early life.

1.6.2 Study objectives

This study investigates how early-life exposures influence the development of the gut microbiome and resistome in Nigerian infants recruited at a tertiary healthcare facility, followed over the first three months of life.

Primary objective

- To characterise the composition, diversity, and temporal dynamics of the gut microbiome and resistome in Nigerian infants over the first three months of life.

Secondary objectives

1. To compare DNA extraction protocols for optimising metagenomic recovery of ARGs from low-biomass infant stool samples in resource-limited settings.

2. To assess the influence of key clinical (delivery mode, antibiotic exposure) and environmental (household composition, WASH conditions) factors on microbiome and resistome development.
3. To identify exposures most strongly associated with ARG acquisition, persistence, or diversity, highlighting potential intervention points.
4. To explore associations between resistome profiles, microbiome composition, and MGE carriage over time.

1.6.3 Thesis structure and chapter overview

This thesis progresses from global and clinical context to methodological approach, empirical findings, and implications.

Chapter 1 has outlined the scope of the problem, from the worldwide burden of AMR to its particular impact on neonates and infants in LMICs, the biological and environmental drivers influencing early gut microbiome colonisation, and the significance of the infant resistome as a reservoir of ARGs. The limitations of existing detection methods and the advantages of metagenomic approaches were highlighted, alongside critical knowledge gaps in the epidemiology, transmission dynamics, and ecological determinants of the infant resistome.

Chapter 2 describes the study setting, cohort recruitment procedures, sample collection protocols, and laboratory workflows, including DNA extraction, sequencing pipelines, and the metagenomic analysis strategies used for both taxonomic and AMR profiling.

Chapter 3 reports a pilot study comparing DNA extraction protocols and evaluating their performance for downstream metagenomic analyses in infant stool samples. This chapter

also analyses the quality control measures incorporated during sample processing, providing an assessment of their effectiveness in handling low-biomass specimens under the constraints of a resource-limited setting.

Chapter 4 presents a detailed description of the study cohort, summarising demographic and clinical characteristics.

Chapter 5 examines the temporal dynamics of the infant gut microbiome and resistome, tracking longitudinal changes from birth through the first three months of life.

Chapter 6 explores associations between key exposures, (e.g. birth mode, antibiotic use, and feeding type) and differential microbiome and resistome trajectories.

Chapter 7 integrates findings, situates them within the global AMR context, discusses methodological limitations, and outlines implications for surveillance, neonatal health, and policy in LMICs.

Through this structure, the thesis moves from establishing the rationale and background to generating new empirical evidence on the determinants and dynamics of the early-life resistome in Nigeria.

Chapter 2: Materials and methods

This chapter describes the design, setting, participants recruitment and follow-up, sample collection, laboratory workflows, bioinformatics pipelines, and statistical analyses used in this study. Core laboratory, sequencing and computational procedures were standardised to support reproducibility, and analytical approaches were selected to account for repeated measures and to transparently report denominators where missing data occurred. All study procedures were conducted under institutional and national ethical approvals, with written informed consent obtained from participating mothers.

2.1. Study design and setting

This study was a prospective longitudinal infant cohort conducted in Kano, Northern Nigeria. Recruitment and baseline procedures were undertaken at Murtala Muhammed Specialist Hospital (MMSH), Kano, a high-volume tertiary hospital serving urban and peri-urban populations. The hospital-based design enabled systematic enrolment around delivery, timely capture of perinatal clinical metadata from routine records, and early postnatal sampling within the first days of life.

For the analyses presented in this thesis, the dataset includes dyads enrolled between February and September 2024, with infant follow-up defined through 90 days of age (D90). Although the parent cohort is designed for longer follow-up, analyses beyond D90 (including later feeding transitions such as weaning and broader environmental exposures) are outside the scope of this thesis and will be reported separately.

Infants were scheduled for stool sample collection at eight postnatal timepoints: D0, D1, D3, D5, D7, D14, D28, and D90. Under complete attendance, this schedule yields a theoretical

maximum of 720 samples for 90 enrolled infants. In practice, the analysed microbiome/resistome dataset comprises 87 infants who provided at least one stool sample; three enrolled infants withdrew before providing any stool samples, although their baseline details were retained in the study database. The number of samples contributing to each analysis therefore varies by timepoint and by outcome after quality control and decontamination; analysed denominators are reported alongside each result, with additional cohort and sample-flow details provided in Chapter 4.

2.2 Ethical approvals

2.2.1 Institutional review and approvals

This study received ethical approval from the Oxford Tropical Research Ethics Committee (OxTREC) at the University of Oxford (Reference: 36-22; approved January 2023) and from the appropriate ethics/governance authorities in Kano State, Nigeria (approved October 2022). National oversight requirements were met through the National Health Research Ethics Committee of Nigeria (NHREC) (Approval Number: NHREC/17/03/2018). The study was conducted in accordance with the principles of the Declaration of Helsinki and Nigeria's national ethical guidelines for research involving human participants.

2.2.2 Informed consent procedures

Written informed consent was obtained from each infant's primary caregiver (predominantly mothers) prior to enrolment. Consent materials were provided in English and explained verbally in Hausa by trained research assistants to support comprehension. Mothers were approached at a suitable time when clinically stable, and participation was voluntary. For

caregivers with limited literacy, consent was documented by thumbprint and countersigned in the presence of an impartial witness (an attending hospital nurse not involved in the research team). Participants were informed of the study aims, study procedures (including biological sample collection and follow-up visits), and their right to withdraw at any time without any effect on access to routine healthcare services.

2.2.3 Confidentiality and participant protection

All samples and study records were labelled using pseudonymised study identifiers. Personal identifiers were stored separately from research data in locked cabinets and in password-protected electronic databases with access restricted to authorised study personnel. Data handling complied with applicable institutional data protection policies. Sample collection and follow-up visits were conducted in a manner that respected privacy, dignity, and local cultural norms. Given the involvement of infants, procedures were designed to minimise risk, and mothers provided proxy consent on behalf of their infants.

2.2.4 Operational permissions

Operational permissions and logistical support were obtained from the Kano State Ministry of Health and the management of MMSH, which facilitated integration of the study with routine maternal and infant healthcare services.

2.3. Study population

2.3.1 Inclusion and Exclusion Criteria

The parent cohort enrolled 120 term neonates (gestational age 37-42 weeks) delivered at MMSH, Kano. For the purposes of this thesis, analyses are restricted to the first 90 enrolled

infants recruited between February and September 2024; of these, 87 infants provided at least one stool sample and contribute to microbiome/resistome analyses.

Infants were eligible if they were: (i) born at MMSH by vaginal or caesarean delivery, (ii) enrolled within 24 hours of birth, (iii) had written informed consent provided by the mother/caregiver, and (iv) resided within metropolitan Kano to enable follow-up visits. Gestational age was ascertained from antenatal records using last menstrual period dating and, where available, ultrasound reports.

Exclusion criteria were:

- Preterm delivery (<37 weeks gestation).
- Major congenital anomalies or gastrointestinal malformations.
- Birth complications requiring prolonged inpatient neonatal care (e.g., neonatal unit admission beyond the immediate postnatal period).
- Lack of feasibility for follow-up (e.g., residence outside the study catchment area).
- Refusal of follow-up visits.

Maternal dietary exclusions (e.g., vegan/vegetarian diets) were prespecified in the protocol to minimise confounding from atypical maternal dietary patterns; however, no eligible participants met these criteria in this setting, and no exclusions were made on this basis. These criteria were established through a comprehensive review of known confounders in neonatal outcomes, such as gestational age and pre-existing clinical complications (Grech et al., 2021), and were refined through extensive consultation with the BARNARDS II clinical team to ensure both scientific rigour and practical feasibility in the study setting.

2.3.2 Recruitment Process

Eligible mothers delivering at MMSH were identified within 24 hours postpartum by trained research assistants using a standardised eligibility checklist. Written informed consent was obtained as described in Section 2.2.2. Recruitment was conducted consecutively among eligible deliveries during the recruitment period until the target sample size for the parent cohort (n=120) was reached.

2.3.3 Participant Identification and Data Collection

At enrolment (Day 0), each infant was assigned a unique pseudonymised study identifier in the format MBK-B-XXXX, where XXXX was a sequential four-digit serial number (0001–0120). This identifier was used to label biological samples, questionnaires, and metadata across follow-up. Baseline demographic and clinical data were obtained from hospital records and caregiver interview. Structured questionnaires at follow-up visits captured infant feeding practices, medication exposure (including antibiotics), household/environmental exposures, and interim health events through Day 90.

2.3.4 Retention Strategies

To minimise loss to follow-up, the study team maintained regular contact with caregivers and used community-based tracking approaches when required, including for families without reliable mobile phone access. Modest, ethics-approved reimbursements and/or small tokens of appreciation were provided to compensate for time and transport, and were designed to be culturally appropriate, non-coercive, and not conditional on continued participation or sample provision.

2.4. Sample size calculation

In consultation with the University's statistical advisory service, a conventional single outcome power calculation was not undertaken because the primary outcomes include high-dimensional microbiome and resistome profiles with multiple correlated features, and because effect sizes for comparable early-life exposure contrasts in this setting were not well established at the time of study design. The cohort size was therefore determined pragmatically, balancing recruitment feasibility, sequencing capacity, and the need to support adjusted analyses across common exposure strata.

An initial target of 60 mother-infant dyads was inflated by 50% to allow for attrition and incomplete sampling, yielding an initial recruitment goal of 90. The parent cohort target was subsequently increased to 120 to support planned future analyses requiring additional stratification and to improve the precision of estimates across exposure groups. For this thesis, analyses are restricted to the first 90 enrolled infants followed through 90 days of life (D90); 87 of these infants provided at least one stool sample and contribute to microbiome/resistome analyses, with denominators varying by timepoint and outcome after quality control and decontamination (reported with each analysis).

This cohort size supports descriptive characterisation of early microbiome and resistome trajectories and provides sufficient sample counts in the most prevalent exposure categories (e.g., delivery mode and common feeding patterns) to fit prespecified adjusted models at key scheduled timepoints, while recognising that some stratified and interaction analyses are limited by sparse sample sizes. Samples from the additional enrolled infants (up to $n=120$) are being processed and will contribute to analyses beyond the scope of this thesis.

2.5 Sample and metadata collection

To support the longitudinal characterisation of gut microbiome development and ARG acquisition, the study collected both biological samples and structured exposure metadata over the first 90 days of life.

2.5.1 Sample collection and storage

Stool samples were collected at eight predefined time points during the first 90 days of life (D0, D1, D3, D5, D7, D14, D28, and D90). Day 0 and Day 1 samples were obtained in hospital shortly after birth, with subsequent collections performed during scheduled home visits (Figure 2.1). The full longitudinal sampling schedule and ancillary specimen collection are summarised in Figure 2.2.

Mothers were provided with sterile, pre-labelled polypropylene vials and instructed in hygienic collection techniques; where needed, trained research assistants assisted with collection. Samples were transported in insulated carriers with ice packs and delivered to the local laboratory within 4-6 hours of collection. On receipt, samples were logged and stored at -20°C. Microbial DNA extraction was performed within 24 hours of collection, with samples held at -20°C until extraction. Extracted DNA was stored at -20°C prior to batch shipment to the University of Oxford, where aliquots were transferred to -80°C for long-term storage. This workflow was informed by a validation experiment (Chapter 3) demonstrating that extraction within 24 hours yields DNA of sufficient quality for shotgun metagenomic sequencing.

Conditional intensified sampling around antibiotic exposure: For infants who received systemic antibiotics during follow-up, an intensified stool sampling protocol was

implemented to capture short-term perturbations. Additional samples were scheduled at 1, 3, 5, and 7 days after antibiotic initiation, where “post abx-Day 1” refers to sampling after the first antibiotic dose. These post-initiation samples were intended to fall within the 0–7-day window used for antibiotic window analyses in Chapter 6 and typically occurred while infants were still receiving treatment.

2.5.2 Exposure metadata collection

Structured exposure metadata were collected alongside biological samples to characterise determinants of infant gut microbiome and resistome development during the first 90 days of life. Metadata collection was synchronised with the scheduled sampling visits to support longitudinal analyses.

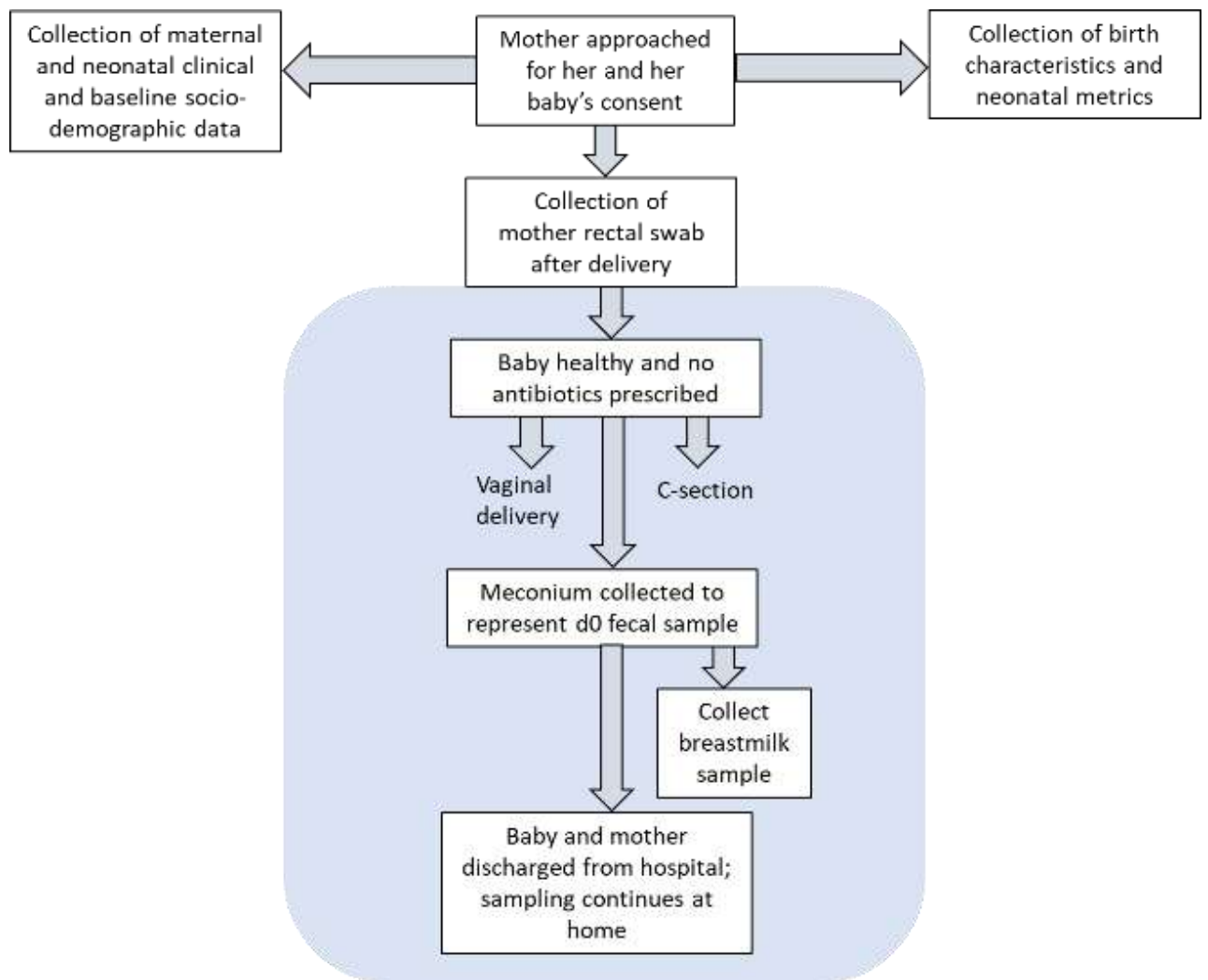


Figure 2.1 Baseline sampling workflow and transition to home follow-up.

Schematic of enrolment at birth for infants delivered vaginally or by C-section, restricted to clinically well neonates with no antibiotics prescribed at baseline. Meconium was collected to represent the Day 0 stool sample, and a breast milk sample was collected from breastfeeding mothers where applicable. Following hospital discharge, longitudinal sample collection continued during scheduled home visits.

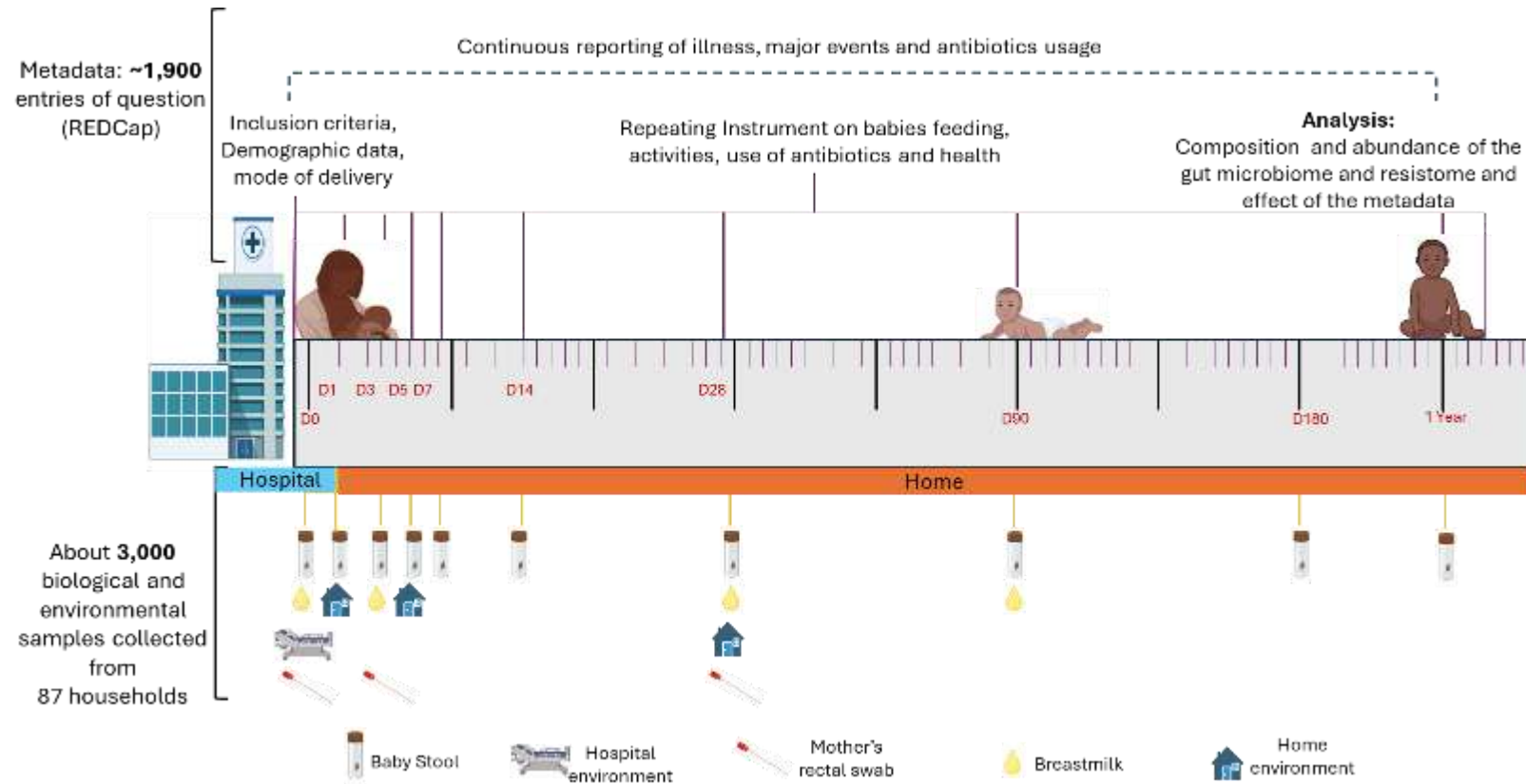


Figure 2.2 Longitudinal sampling schedule and specimen types across infant follow-up

Timeline of scheduled study visits from birth (D0) through Day 180 and 1 year in the parent cohort, showing the transition from in-hospital recruitment and early postnatal sampling to home-based follow-up. Infant stool samples were collected at prespecified ages (D0, D1, D3, D5, D7, D14, D28, and D90 for the thesis analyses), with additional ancillary samples collected at selected visits, including maternal rectal swabs, breast milk, and household/environmental specimens (icons).

At enrolment (within 24 hours of delivery), trained research assistants completed a standardised electronic case report form (CRF) in REDCap (v12) using encrypted tablets. Baseline variables included maternal demographics (e.g., age, education, parity), delivery details (including mode of delivery), household socioeconomic indicators, and infant birth characteristics (including birthweight and Apgar scores), abstracted from hospital records and caregiver interview where required.

At follow-up visits, repeat CRFs captured time-varying exposures and outcomes, including:

- Infant exposures and health: feeding practices (categorised using WHO-consistent definitions; operationalised in analyses as breastfeeding intensity groups (Aarts et al., 2000; Burgin, 1996; Geneva: World Health Organization, 2018)), illness episodes, and medication exposure. For antibiotics, the drug name, indication/source, start date, duration, and (where possible) dose were recorded, with caregivers encouraged to retain and present medication packaging to improve accuracy.
- Maternal factors: intrapartum antibiotic administration, postpartum antibiotic use.
- Household/environmental factors: family size, sanitation and water access, household composition, domestic animal exposures, and other contextual factors relevant to early-life microbial acquisition.

To support recall between visits, caregivers were provided with a simple household event grid (Appendix A.1: Supplementary Figure 2.1) to log key events and exposures using predefined codes. The grid was reviewed at each follow-up visit and cross-checked against CRF responses. Mothers were encouraged to retain and present medication packaging at follow-up visits to improve the accuracy of antibiotic exposure documentation. Research

staff also maintained brief field notes to document contextual information not captured by structured items (e.g., use of traditional remedies or water supplementation practices), where relevant to exposure interpretation.

All metadata were linked to samples using pseudonymised infant study identifiers and managed in REDCap (v12) with built-in range and consistency checks. Periodic data quality audits were undertaken prior to export for statistical analysis.

Structure of the questionnaire: Data were collected using a combination of electronic and paper-based tools to capture both structured exposures and contextual events across follow-up.

- Enrolment questionnaire (REDCap): Administered within 24 hours of delivery to confirm eligibility, obtain informed consent, assign the infant study identifier, and capture baseline maternal, household, and infant information. Variables included maternal demographics (e.g., age, education), obstetric history (e.g., parity), pregnancy and delivery factors (including delivery mode), and infant birth characteristics (including birthweight and early clinical observations).
- Household event grid (caregiver-held diary; Appendix A.1: Supplementary Figure 2.1): A simple calendar-style tool provided at enrolment to support recall between scheduled visits, particularly during longer intervals. Caregivers recorded key events using predefined codes (e.g., illness episodes, medication exposure, feeding changes, and notable milestones). The study team reviewed the grid at each visit, and caregivers received reminder contact to support completion where feasible.
- Repeat follow-up instruments (REDCap): Structured questionnaires administered at scheduled follow-up visits to capture time-varying exposures and outcomes. These included infant health events (start/end dates of illness), infant and maternal

medication exposure (with emphasis on antibiotic name, start date, and duration), feeding practices, gastrointestinal symptoms (e.g., stool frequency/consistency), and relevant household/environmental exposures. Caregivers were asked to retain and present medication packaging to improve accuracy of antibiotic documentation.

- Field notebook (research team): Used to record additional contextual information not captured in the structured questionnaires that could influence exposure interpretation, such as culturally mediated health practices (e.g., traditional remedies, water supplementation practices, and procedures such as uvulectomy) or household disruptions (e.g., relocation).

All instruments were linked using the infant's pseudonymised study identifier, with REDCap configured to include range and consistency checks to support data quality.

2.6 Laboratory methods

Laboratory procedures were conducted in two stages across two sites: DNA extraction was performed at MMSH, Kano, while library preparation and shotgun metagenomic sequencing were conducted at the University of Oxford. Protocols were informed by the pilot validation study conducted in MMSH (Chapter 3) and were implemented with contamination control and sample traceability measures throughout.

2.6.1 DNA extraction and technical quality control

Total genomic DNA was extracted from infant stool using the QIAGEN PowerSoil Pro DNA Isolation Kit (Qiagen, Luxembourg). Procedures were standardised across all timepoints and incorporated minor optimisations informed by the pilot study (Chapter 3), including a 10-

minute incubation at 65 °C to enhance lysis and extended bead beating (10 minutes; Vortex Genie® 2 with 24-ring adapter) to maximise DNA yield from infant stool.

DNA concentration was quantified using the Qubit 4.0 fluorometer (dsDNA HS assay).

Extracted DNA was stored at -20 °C for short-term preservation and shipment on dry ice to the University of Oxford, where aliquots were stored at -80 °C until library preparation.

Each extraction batch included negative controls (blank kit bead-tubes and elution-buffer-only controls) and a positive control (ZymoBIOMICS HMW DNA Standard; Zymo Research) processed alongside study samples. Additional procedures for contamination assessment and negatives-informed background control are described in Section 2.7 and implemented in Section 2.8.4.

2.6.2 Library preparation and shotgun sequencing

Shotgun metagenomic sequencing was selected over 16S rRNA amplicon sequencing to enable species- and strain-level taxonomic resolution and direct detection of ARGs.

Library preparation was performed at the University of Oxford using the Oxford Nanopore Technologies (ONT) Rapid Barcoding Kit 96 (SQK-RBK110.96) with R10.4.1 flow cell chemistry, following manufacturer guidance with minor optimisations for high-throughput processing. For each biological sample, three sequencing aliquots were prepared and sequenced independently; the resulting FASTQ files were subsequently concatenated to generate a single merged FASTQ per sample for downstream bioinformatic processing.

Briefly, DNA aliquots were arrayed in 96-well plates, barcoded, pooled, and bead-cleaned using AMPure XP magnetic beads. Libraries were quantified using the Qubit dsDNA BR assay prior to loading. Sequencing was performed initially on a GridION with progression to a PromethION P2 Solo. Low-biomass samples (primarily Day 0-1) were sequenced at

available DNA concentrations without amplification, in order to preserve representation of early colonisation signals.

2.6.3 Sequencing Output and Data Handling

Sequencing yields varied across timepoints, reflecting differences in biomass and extracted DNA concentration. Across the thesis dataset, the median yield per sample was approximately 316 Mb (IQR: 59-956 Mb) (Table 2.1), with a cumulative yield of approximately 494 Gb FASTQ. Rather than applying a fixed read-count threshold, samples were retained for downstream analyses if they generated sufficient cleaned sequence to complete the assembly- and mapping-based steps of the in-house bioinformatic pipeline (Section 2.8). A small number of very low-yield libraries, predominantly early timepoint samples, did not generate assembled contigs and therefore could not be processed through downstream steps; analysed denominators are reported per timepoint and analysis.

Raw sequencing outputs were securely transferred to University of Oxford servers and stored using pseudonymised study identifiers, with the re-identification key retained separately under restricted access (Section 2.2).

Table 2.1: Summary of sequencing yield and processing feasibility across analysed infant stool metagenomes

Metric			Value	Notes / interpretation
IQR	reads	per sample	189,000 – 870,000	Reflects variability in yield across timepoints, with lower read counts more common in early postnatal samples
Median	total bases	per sample	316 Mb	Overall sequencing depth across the thesis dataset; yields varied substantially by sampling day.
Operational processing criterion (in-house pipeline)		Assembly-dependent processing feasibility		Samples were retained for primary analyses if cleaned non-host reads supported per-sample assembly (metaMDBG) and completion of downstream steps (taxonomic classification, ORF calling, ARG annotation, read mapping, and reporting). A small number of very low-yield early samples did not generate contigs and therefore could not be processed through the full pipeline; denominators are reported per timepoint/analysis.

A small number of very low-yield early postnatal samples did not generate assembled contigs with metaMDBG and therefore could not be processed through downstream assembly-dependent steps (taxonomic classification, ORF calling, ARG annotation, read mapping, and report generation). Such samples were excluded from assembly-based analyses; analysed denominators are reported by timepoint and analysis.

2.7 Quality control procedures

Robust quality control (QC) procedures were implemented across sample handling, sequencing, and bioinformatic processing to minimise background contamination, manage host DNA, and ensure traceability of outputs. QC procedures were applied consistently across the dataset. Summary QC outcomes and the derivation of analysis-ready denominators are reported at the start of Chapter 5, with full per-sample QC metrics and diagnostics provided in Appendix A.3: Supplementary data 1.

2.7.1. Contamination controls: extraction and sequencing controls

Each extraction batch included negative controls processed alongside study samples, an extraction control (blank, no sample), and elution-buffer-only controls to monitor reagent-derived background signal. The sequencing negative control was the library kit elution buffer. A positive control (ZymoBIOMICS HMW DNA Standard, Zymo Research) was included in each sequencing run to monitor run-to-run consistency and barcode performance.

Several sequencing runs did not include a negative control, particularly the initial sequencing runs; therefore, contamination assessment and downstream background control did not rely exclusively on run-specific blanks. Instead, all available extraction and library blanks across runs were pooled to define a global background profile. This approach avoided applying subtraction only to the subset of runs with negatives (which would risk introducing an artificial batch effect) and enabled a consistent background model to be applied across the full dataset (implementation in Section 2.8.4).

2.7.2. Host DNA Removal

Given the high proportion of host-derived DNA expected in neonatal stool particularly at Day 0–1 due to low microbial biomass and meconium, reads were screened against the human reference genome (GRCh38) and host reads removed prior to taxonomic and resistome profiling. Host decontamination was performed using Hostile v1.1.0, following read-level cleaning (fastp; Section 2.8.1). Host-derived reads were excluded from all downstream analyses.

To support auditability and sensitivity analyses, host removal was quantified per library by recording the number of reads entering Hostile (“reads_in”) and the number retained after host removal (“reads_out”); the fraction removed was taken as an estimate of host burden. Two host-burden flags were prespecified for robustness checks: high-host libraries ($\geq 20\%$ reads removed as host) and extreme-host libraries ($\geq 70\%$ reads removed as host). These flags were used for interpretive caution and sensitivity analyses (Section 2.9.5), with analysed denominators reported per timepoint and analysis.

2.7.3. Sequencing depth normalisation

Sequencing yield varied across timepoints and participants, with early postnatal samples generally producing lower microbial sequence due to biological low biomass. Downstream analyses therefore used normalised abundance measures and compositional transformations appropriate to each outcome:

- Microbiome summaries: bacterial community profiles were analysed using bacteria-restricted relative abundance tables (normalised to sum to one after restricting to bacterial taxa).

- Resistome composition: where appropriate, resistome feature tables were analysed using centred log-ratio (CLR) transformations and Aitchison distances.
- Resistome burden outcomes: burden-style summaries were expressed using depth-normalised measures derived from read-to-feature counts (e.g., reads per million (RPM)) and/or log10-transformed totals with a pseudocount (see Section 2.9).

Robustness checks based on yield and host-burden flags are described in Section 2.9.5, as the detection of low-abundance genera/ARGs can remain sensitive to sequencing depth even after conversion to proportions.

2.7.4 Data Traceability and Replicates

All samples and outputs were tracked using pseudonymised infant study identifiers recorded in REDCap and propagated through laboratory and bioinformatic workflows. For each biological sample, three sequencing aliquots were prepared and sequenced independently (Section 2.6.2); replicate FASTQ outputs were reviewed for gross inconsistencies (e.g., unexpectedly low yield or barcode anomalies) and concatenated to generate a single merged FASTQ per sample prior to downstream processing (Section 2.8.1).

2.8 Bioinformatics and computational analysis

Bioinformatic processing of metagenomic data was performed using an in-house long-read pipeline that integrates host decontamination, per-sample assembly, taxonomic classification, and ARG profiling. The workflow was implemented in Snakemake and executed within Conda-managed environments. Pipeline code and documentation are

provided in the project repository (GitHub link provided in Appendix A.1, Supplementary data 1). An overview of the pipeline is shown in Figure 2.3.

A small subset of very low-yield early samples (primarily Day 0-1) did not complete assembly-dependent steps of the in-house workflow. Where needed for descriptive context, such samples were processed using BugSeq, a cloud-based bioinformatics platform designed for rapid, automated microbiology sequencing analysis (Fan et al., 2021), for limited taxonomic screening only. BugSeq-derived outputs were not used for analyses requiring contigs, ORF prediction, read-to-contig mapping, contig-linked ARG quantification, or host-assignment summaries; analysed denominators are reported per timepoint and analysis.

2.8.1. Preprocessing and quality control

Raw sequencing data underwent real-time basecalling and adapter trimming in MinKNOW using Guppy v6.4.2, 6.4.6, 7.3.9, 7.3.11, 7.4.12, 7.4.13 with the super-accuracy model for R10.4.1/E8.2 pores (dna_r10.4.1_e8.2_400bps_sup.cfg). Read-level cleaning was performed using fastp v0.23.4, trimming low-quality ends and removing reads failing prespecified criteria (mean Phred < 7; post-trim length < 40 bp; >50 ambiguous bases; and low sequence complexity). Summary metrics (read count, total bases, median read length, N50, and mean Phred quality) were recorded before and after cleaning to support QC auditing (QC outcomes reported in Chapter 5; full metrics in the Appendix A.3: Supplementary data 1). Where required, auxiliary utilities (e.g., seqkit v2.8.2) were used for FASTQ handling and summary.

Following read cleaning, host read removal was performed using Hostile v1.1.0 as described in Section 2.7.2, generating per-sample host-depleted long-read FASTQ files. Host-depleted reads were then assembled per sample using metaMDBG v1.1 to generate contigs, enabling

contig-linked taxonomic and ARG annotation. Samples that did not produce assembled contigs could not proceed through subsequent assembly-dependent steps; analysed denominators are therefore reported per timepoint and analysis.

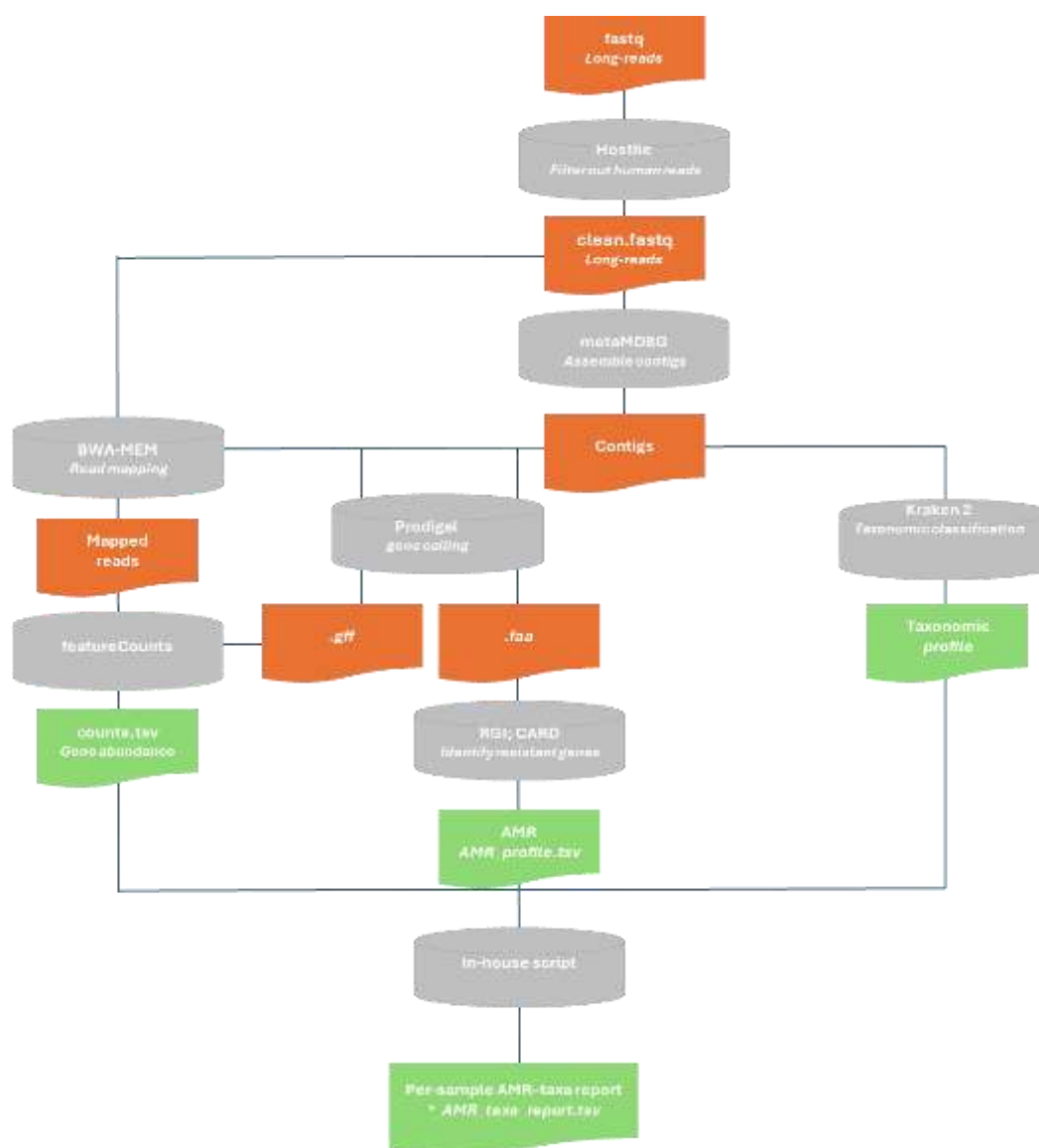


Figure 2.3: Overview of the long-read bioinformatics workflow for taxonomic and antimicrobial resistance profiling.

Raw Oxford Nanopore FASTQ reads were first decontaminated by removing human reads with Hostile to generate cleaned long reads. Cleaned reads were assembled per sample using metaMDBG to produce contigs, which were taxonomically classified with Kraken2 to generate a contig-based taxonomic profile. In parallel, open reading frames were predicted from contigs using Prodigal (GFF/FAA outputs) and antimicrobial resistance genes were identified using RGI against the CARD database to generate an AMR feature table. To estimate gene-level abundance, cleaned reads were mapped back to assembled contigs using BWA-MEM and summarised with featureCounts to produce a per-sample gene count table. An in-house integration script combined the taxonomic profile, AMR calls, and abundance outputs to generate the final per-sample AMR–taxa report used in downstream analyses.

2.8.2. Taxonomic profiling

Taxonomic classification was performed on assembled contigs using Kraken2 v2.1.4 with a RefSeq-derived database including bacterial genomes and plasmid sequences (database snapshot downloaded September 2024). Contig-level taxonomic assignments were summarised to generate per-sample taxonomic profiles for downstream ecological analyses.

Diversity metrics and community-composition analyses were computed in R (v4.3.1) using established packages including phyloseq and vegan, operating on bacteria-restricted relative-abundance tables as described in the chapter-specific methods (Chapters 5-6).

2.8.3. ARG and MGE detection

Genes were predicted from assembled contigs using Prodigal v2.6.3, producing GFF and FAA outputs. ARGs were identified using the Resistance Gene Identifier (RGI v6.0.4) against CARD, generating per-sample AMR feature tables. Unless otherwise stated in chapter-specific analyses, downstream resistome summaries were based on the CARD “Best Hit” ARO feature calls and curated mappings to drug class and mechanism (Chapter 5; Chapter 2 statistical methods).

To estimate gene-level abundance, cleaned reads were mapped back to assembled contigs using BWA-MEM, and alignments were processed with SAMtools v1.21. Gene-level counts were generated using featureCounts v2.0.8, assigning mapped reads to predicted ORFs based on Prodigal-derived annotations. Per-sample ARG abundance summaries were expressed as normalised measures derived from these counts (e.g., relative abundance and transformed totals with a pseudocount for burden-style outcomes), with the specific

transformations and modelling approaches described in Section 2.9 and in chapter-specific methods (Chapters 5-6).

2.8.4 Integration of taxonomic and resistome outputs

Contig-level taxonomic classifications were linked to ARG-bearing contigs to enable host assignment summaries for selected priority ARGs in chapter-specific analyses (e.g., *bla*_{CTX-M-15} host mapping). Final outputs from the pipeline included per-sample taxonomic profiles, AMR feature tables, gene count tables, and an integrated per-sample AMR–taxa summary report (`AMR_taxa_report.tsv`), which served as the primary input to downstream statistical analyses.

A “negatives-informed” background control was implemented at the feature-table level using pooled negative controls, as described in Section 2.7.1. Three complementary approaches were applied and compared: microDecon (proportional subtraction using the pooled blank composition under non-negativity constraints), decontam (prevalence mode) as a contaminant-flagging screen, and a rule-based prevalence/blank-dominance screen. To minimise over-correction, the microDecon-subtracted table served as the “correction method” for downstream analyses. Decontam and rule-based screening were used to generate contaminant flags for review and sensitivity checks.

To account for the fact that clinically plausible gut taxa may appear at low levels in negative controls (e.g., due to barcode bleed or low-level carryover), removal was restricted to features that were both blank-enriched and biologically unlikely to be neonatal gut organisms; plausibly gut-associated taxa were retained but could be flagged for interpretive caution in sensitivity analyses (Section 2.9.5).

2.9. Statistical analysis

Statistical analyses were designed to (i) describe maturation of the infant gut microbiome and resistome over the first 90 days of life, and (ii) evaluate associations with key early-life exposures, including mode of delivery, infant feeding practices, and antibiotic exposure. Analyses were conducted in R (v4.3.1). Chapter 2 provides the general analytic framework; chapter-specific modelling choices, outcome definitions, and clinically focused endpoints are detailed in Chapters 5 and 6.

2.9.1 Overview of data structure and general approach

The dataset comprises repeated stool metagenomes collected longitudinally from each infant at scheduled postnatal days. All analyses accounted for repeated measures where applicable. Unless otherwise stated, analyses were performed on bacteria-restricted relative abundance profiles for microbiome summaries and on ARG feature tables derived from CARD/ARO calls for resistome summaries (Sections 2.8 and Chapters 5-6). Analyses used normalised abundance measures and compositional methods appropriate to each outcome (Section 2.7.3).

2.9.2. Microbiome diversity and composition

Alpha diversity was summarised using Shannon diversity and observed richness, calculated from bacteria-only relative abundance tables (constructed after excluding host and non-bacterial reads, and normalising bacterial genus/species counts to sum to one). Observed richness was defined as the number of genera/species with non-zero relative abundance after bacterial restriction and preprocessing. Because detection of low-abundance taxa can remain

sensitive to sequencing yield, robustness checks based on host burden and yield restrictions were included where applicable (Section 2.9.5). Cross-sectional comparisons at scheduled timepoints used regression-based models where appropriate, and longitudinal trajectories were described using mixed-effects models with infant-level random intercepts to account for within-infant correlation.

Between-sample community composition was quantified using Bray-Curtis dissimilarity on bacteria-only relative abundance profiles and visualised using principal coordinates analysis (PCoA). Associations between composition and explanatory variables were tested using PERMANOVA (`vegan::adonis2`). Where repeated measures were included, permutations were constrained within infant to preserve the longitudinal structure. Differences in multivariate dispersion were assessed using `betadisper` as a diagnostic to support interpretation of PERMANOVA findings.

2.9.3. Resistome summaries and composition

ARG features were represented at CARD ARO level and, where required for interpretability, mapped to drug class and mechanism using curated annotations (Chapters 2 and 5). Resistome analyses were expressed in relative terms (no absolute quantification), consistent with the absence of spike-ins or qPCR calibration.

For interpretable burden-style summaries, total ARG abundance was analysed on a log₁₀ scale with a pseudocount (e.g., $\log_{10}(\text{total ARG} + 1)$), as implemented in Chapter 6. Feature-level and class-level summaries were described using normalised abundance measures derived from RPM read-to-feature counts.

Resistome compositional structure was analysed using log-ratio methods where appropriate (centred log-ratio transformation and Aitchison distance), with ordination by PCoA and hypothesis testing by PERMANOVA. Clinically focused endpoints (e.g., selected β -lactamase markers and *bla*_{CTX-M-15} carriage) were analysed using presence/absence definitions and logistic regression as described in Chapters 5-6.

2.9.4. Exposure-Outcome association modelling and covariate adjustment

Associations between early-life exposures and microbiome/resistome outcomes were evaluated using models appropriate to the outcome distribution:

- Continuous outcomes (e.g., Shannon diversity, log-transformed ARG burden): linear regression at scheduled timepoints and mixed-effects models for repeated measures where relevant.
- Binary outcomes (e.g., ARG carriage or taxon presence/absence): logistic regression at scheduled timepoints and mixed-effects logistic models where repeated measures were used.

Exposure definitions and modelling windows were aligned to the study questions. Chapter 6 prioritised clinically interpretable cross-sectional models at prespecified milestones (D5, D28, D90) and used additional longitudinal or exposure-anchored “acute window” analyses for antibiotic perturbation dynamics. Covariates were prespecified on biological grounds (infant sex, gestational age, birthweight), with additional variables included only where supported by data completeness and relevance to the specific chapter aims (Chapters 5–6). Missing data were handled using complete-case analysis within each model, with denominators reported alongside results.

Multiple testing was addressed using false discovery rate control (Benjamini–Hochberg) where families of related hypotheses were tested (e.g., multiple taxa or multiple ARG features), with emphasis placed on effect sizes and confidence intervals alongside adjusted p-values.

2.9.5. Sensitivity analyses

Sensitivity analyses were used to assess robustness to analytic choices and data limitations. These included: (i) alternative normalisation/transformation approaches for compositional data (e.g., CLR-based analyses for resistome composition), (ii) diagnostics for dispersion effects in PERMANOVA, and (iii) exclusion or separate handling of samples that could not complete assembly-based processing due to very low yield (Section 2.8.1). Very low-yield early samples processed outside the in-house assembly-dependent workflow (e.g., via BugSeq) were limited to descriptive taxonomic screening and were not used for assembly-linked ARG quantification or host-assignment analyses (Chapters 5–6).

To evaluate whether host-dominated libraries influenced key inferences, sensitivity checks also considered exclusion or stratified handling of libraries flagged as high-host ($\geq 20\%$ reads removed as host) or extreme-host ($\geq 70\%$ reads removed), as defined in Section 2.7.2. Where a “negatives-informed” background control was applied (Section 2.8.4), sensitivity checks considered whether conclusions were materially altered under contaminant-flagging criteria (decontam and rule-based screening) relative to the microDecon (correction method) table. Analysed denominators are reported by timepoint and model where applicable.

2.10 Data management and reproducibility

Data handling and analysis workflows were designed to support transparency, traceability, and reproducibility, consistent with FAIR (Findable, Accessible, Interoperable, Reusable) principles.

2.10.1 Data Management

Participants were assigned pseudonymised alphanumeric study identifiers at enrolment, which were used to label questionnaires, laboratory records, and all biological samples throughout follow-up (Section 2.3.3). Personally identifiable information (e.g., names and contact details) was stored separately from research datasets in REDCap on a secure institutional server with role-based access controls. The re-identification key linking identifiers to personal information was restricted to authorised study personnel.

Sequencing data generated at the University of Oxford were transferred using secure institutional data-transfer procedures and stored on access-controlled servers. Access to sequencing outputs and derived analysis files was limited to approved study investigators. Routine checks were undertaken to verify file integrity following transfer and during downstream processing.

2.10.2 Reproducibility

Bioinformatic processing was conducted using the in-house Snakemake workflow described in Section 2.8. Software environments were managed using Conda with versioned environment specifications to support reproducibility across computing systems. Pipeline

source code, configuration templates, and documentation are available in the project repository (GitHub link provided in Appendix A.1: Supplementary data 2).

Statistical analyses were conducted in R (v4.3.1, 4.5.0) using version-controlled scripts (RMarkdown), with session information captured to record package versions and computational settings. Intermediate outputs (e.g., taxonomic tables, ARG feature tables, distance matrices) were stored in structured formats (CSV/TSV) with accompanying metadata to support auditability and re-analysis.

Summary statistics and analysis code underpinning thesis figures and tables are intended to be deposited in an institutional open-access repository upon thesis submission. Raw sequencing reads are intended for deposition in the European Nucleotide Archive (ENA); accession details will be provided with relevant publications.

2.10.3 Quality Assurance

Consistency checks were performed to confirm alignment between REDCap metadata, laboratory logs, and sequencing outputs (e.g., sample identifiers, collection dates, and sequencing run information). Quality control checkpoints relating to contamination controls and host-read removal are described in Section 2.7 and were evaluated in the pilot validation study (Chapter 3). Because early-life samples vary substantially in microbial biomass and sequencing yield, downstream inclusion for assembly-dependent analyses was based on completion of the in-house processing workflow; analysis denominators are reported by timepoint and model where applicable.

Chapter 3: Validating DNA extraction for neonatal, low-biomass metagenomics in a resource-limited setting

A major part of this chapter was adapted and modified from the following first-author peer-reviewed publication:

Chinenye Akpulu*, Aditya Kumar Lankapalli, Rida Toufiq, Kate Cook, Edward AR Portal, Rashida Yakubu Khalid, Aisha Mukkadas, Safiya Gambo, Aliyu Aminu, Kenneth Iregbu, Kathryn Thomson, Timothy Walsh and Kirsty Sands (2025). Comparative evaluation of DNA extraction protocols for neonatal gut microbiota profiling in a resource-limited setting. *The Microbe*, 7: 100398. (co-authors contributions are listed in Appendix A. 2, Supplementary data 1).

This chapter reports a pilot comparison of DNA extraction protocols for infant stool performed in Kano, evaluating yield, read length/assembly potential, practical consideration, and downstream metagenomic performance. The kit of choice was selected based on the best combination of yield and read length at a sustainable cost and documented the operational implications (time-to-extraction) for the main study.

3.1 Introduction

The efficiency and representativeness of DNA extraction fundamentally constrain what can be recovered from a metagenomic dataset. Variation in extraction performance can bias total microbial DNA yield, distort observed taxonomic composition, and alter DNA integrity (fragment-size distribution), while co-extracted inhibitors can further reduce library conversion and sequencing performance (Costea et al., 2017; Lim et al., 2020; Quince et al., 2017). Together, these parameters influence downstream read-length distributions, assembly

contiguity, community composition estimates, and the recoverability of ARGs within their chromosomal or mobile genetic contexts (Costea et al., 2017; Gand et al., 2023; Salter et al., 2014).

This dependence is particularly acute for long-read sequencing. ONT workflows benefit from high-molecular-weight (HMW) DNA and stable fragment-size distributions, which support long reads, improved assembly, and ARG-context inference (Maghini et al., 2021; Moss et al., 2020). However, extraction imposes opposing requirements: stronger mechanical lysis improves recovery of resilient taxa (notably Gram-positive organisms) but increases DNA shearing; gentler conditions preserve longer fragments but risk under-recovering cell-wall-rich organisms, and such deficits cannot be fully corrected downstream (Gand et al., 2023; Moss et al., 2020). In neonatal cohorts, where microbial biomass and microbial DNA mass are often low and the proportion of host DNA can be high, DNA extraction is therefore a principal rate-limiting step for genome-resolved inference (Gallo et al., 2023; Kennedy et al., 2023; Turunen et al., 2024).

Early neonatal stool poses distinct laboratory challenges for metagenomic studies. Biomass is typically lowest for meconium, the first stool after birth, and rises rapidly over the first days to weeks of life (Dos Santos et al., 2021; Kennedy et al., 2023; Seki et al., 2021). Early life exposures, including intrapartum factors, antibiotic exposure, and mode of delivery, can alter early colonisation trajectories and delay establishment of typical anaerobic communities in some cohorts, potentially leaving some early samples borderline for metagenomic workflows (Cho et al., 2012; Eisenhofer et al., 2019; Jakobsson et al., 2014; Mitchell et al., 2020; Reyman et al., 2022; Shao et al., 2019). Human DNA frequently predominates in neonatal samples because epithelial cells and mucus constitute a larger fraction of early faeces; unless depleted or computationally removed, this reduces effective

microbial sequencing depth and can determine whether a sample remains analysable following host-read removal (Eisenhofer et al., 2019; Turunen et al., 2024). In addition, neonatal stool can contain inhibitors (e.g., bile salts and complex polysaccharides) that hinder lysis efficiency and disrupt end-repair and ligation during library preparation, thereby reducing library conversion and introducing bias (Maghini et al., 2021).

Community standardisation efforts for faecal microbiome methods, developed largely around adult stool inputs and predominantly short-read workflows, have demonstrated that DNA extraction choices can materially influence observed community structure (Costea et al., 2017; Wesolowska-Andersen et al., 2014). Protocols aligned with the International Human Microbiome Standards (IHMS), including Protocol Q, were developed primarily for adult stool and relatively large input masses in well-resourced laboratory settings. Neonatal low-biomass samples have fewer widely adopted, standardised extraction protocols tailored to long-read metagenomics, and performance is likely to be context-specific. Adapting adult standards to neonatal ONT work therefore introduces specific challenges. First, input assumptions: bead-beating durations and buffer volumes calibrated for adult faeces risk overdiluting small neonatal aliquots and may still under-recover DNA when only limited material is available (Stinson et al., 2018). Second, host fraction: adult protocols rarely report host DNA as a primary performance metric, yet in neonatal extracts the host fraction often dictates whether sufficient microbial reads remain after host filtering for meaningful downstream analysis (Bittinger et al., 2020). Third, implementation constraints: uninterrupted power, routine -80°C storage, and automated bead-beaters cannot be assumed in many clinical laboratories in LMIC settings. Under such conditions, workflow steps must match available equipment, staffing capacity, and cold-chain reliability, while still achieving lysis sufficient for resilient taxa without excessive DNA shearing.

This work was conducted under conditions typical of a resource-limited setting: a temporary laboratory environment designed from a shipping container, intermittent electricity, and constrained cold-chain capacity (Figure 4.9). Within these constraints, the pilot experiment was designed to address two linked questions: (i) which extraction approach can consistently recover sufficient, high-integrity microbial DNA from neonatal stool to support long-read metagenomics; and (ii) which operational parameters, particularly time-to-extraction and interim storage conditions, minimise avoidable loss of microbial DNA yield and integrity under field-to-laboratory timelines typical of MMSH. Evidence from stool preservation studies indicates that storage conditions and processing delays can alter measured community composition and reduce recoverable DNA (Lim et al., 2018; Stinson et al., 2018), reinforcing the need to define feasible operating procedures when immediate extraction is not guaranteed. Accordingly, the evaluation of extraction methods here was paired with the development of practical standard operating procedures for use in similar settings.

Comparative evaluations of DNA extraction methods often report yield, purity, and sometimes diversity metrics under controlled conditions, most commonly using adult stool. Operational reports, by contrast, focus on feasibility (cost, staffing, procurement, supply chains) but rarely connect those realities to long-read-relevant metrics, such as fragment length distributions and post-extraction host fraction, which strongly influence the value of ONT sequencing. For neonatal long-read work in LMIC hospitals, what remains relatively scarce is an integrated assessment that concurrently considers: (i) whether sufficient intact microbial DNA can be recovered to justify ONT library preparation; (ii) attainable read length and assembly-relevant proxies; (iii) realistic post-filter host DNA burden; and (iv) whether these outcomes are preserved under laboratory constraints on power, storage,

transport, and procurement. This chapter addresses that gap by evaluating extraction performance and feasibility together under MMSH operating conditions.

To address these gaps, three commercially available DNA extraction kits were compared: QIAamp Fast DNA Stool Mini (Fernández-Pato et al., 2024), DNeasy PowerSoil Pro (Costea et al., 2017; Maghini et al., 2021), and ZymoBIOMICS DNA Miniprep (Sudjai et al., 2022; Wongsurawat et al., 2019). These kits were selected because they are widely used in gut microbiome studies, are compatible with downstream long-read library preparation workflows, and are potentially deployable in LMIC laboratories, representing different design trade-offs between lysis strength, DNA integrity, hands-on time, and cost. Prior to selection, relevant protocols and kit comparisons were identified through a targeted PubMed search conducted in October 2022. Searches utilised various combinations of the following terms using Boolean operators (AND/OR): “stool DNA extraction kit”, “metagenomics”, “long-read microbiome sequencing”, “LMIC”, and “developing countries”, prioritising approaches that were frequently used, required minimal specialised equipment, and could accommodate low-biomass inputs.

Prior to selection, relevant protocols and kit comparisons were identified through a targeted PubMed search (queries included “stool DNA extraction kit metagenomics”, “long-read sequencing microbiome”, and “DNA extraction LMIC”), prioritising approaches that were frequently used, required minimal specialised equipment, and could accommodate low-biomass inputs.

In this study, ONT sequencing was used because it is comparatively portable and deployable, with lower infrastructure requirements than some short-read platforms and can enable genome-resolved analyses when extraction conditions support sufficient DNA integrity

(Chen et al., 2023; Elliott et al., 2020; Wasswa et al., 2022). Across multiple extraction methods and storage/processing conditions, ONT-derived yield and downstream taxonomic profiles were compared to identify a kit–condition combination that is analytically effective and logistically workable when immediate extraction cannot be guaranteed. Strengthening upstream sample preparation under these constraints was intended to enable reliable, high-quality long-read metagenomic data from neonatal cohorts in LMIC settings, reduce methodological barriers, and support more inclusive and scalable microbiome research.

Aim and objectives

The aim of the pilot experiment was to identify a neonatal stool DNA extraction approach that supports ONT long-read metagenomics and is deliverable at the study site MMSH, Kano, balancing analytical performance with operational feasibility.

Objectives.

1. Quantify DNA yield to assess suitability for ONT library preparation.
2. Describe read-length distributions and assembly-relevant indicators, recognising that observed differences may reflect combined effects of extraction and handling conditions.
3. Compare per-sample consumables, hands-on time, and failure rates, to identify a method that is robust in routine practice.
4. Examine how time-to-extraction and interim storage conditions are associated with yield and HMW preservation within ward-to-laboratory intervals typical of MMSH.

3.2 Methods

3.2.1. Ethics approval and study setting

Ethical approval was obtained from the National Health Research Ethics Committee of Nigeria (NHREC/2022/3448) and the Oxford Tropical Research Ethics Committee (OxTREC/REF:36-22). Parents or legal guardians provided written informed consent prior to enrolment.

The work was conducted at MMSH, Kano, Nigeria (intermittent power, routine -20°C cold chain, and constrained procurement, see section 4.2.1 for mor details). The methods below were selected to be feasible under these conditions.

3.2.2. Participants, sampling and design

Stool was collected from five neonates during the first postnatal day in the hospital maternity unit (Day 0-Day 1 of life). Fresh specimens were placed into sterile containers, labelled, maintained on ice, and transferred directly to the on-site laboratory.

To mimic realistic handling constraints, each specimen was homogenised and split across six storage/preservation conditions prior to extraction (Figure 3.1):

1. No additive; processed same day (fresh, no stabiliser).
2. DNA/RNA Shield (Zymo Research), 24 h at room temperature (RT).
3. DNA/RNA Shield, 24 h at 4 °C.
4. DNA/RNA Shield, 24 h at -80 °C.

5. DNA/RNA Shield, 48 h at RT.
6. Charcoal swab storage, 6 weeks at 4 °C.

This produced six stool aliquots per neonate. Each aliquot was extracted using each of three kits (Section 3.2.3), yielding 90 DNA extracts in total (5 neonates $\times$ 6 conditions $\times$ 3 kits). All extracts contributed to the extraction-yield and feasibility comparisons. A subset of extracts was selected for downstream sequencing as described in Section 3.2.6.

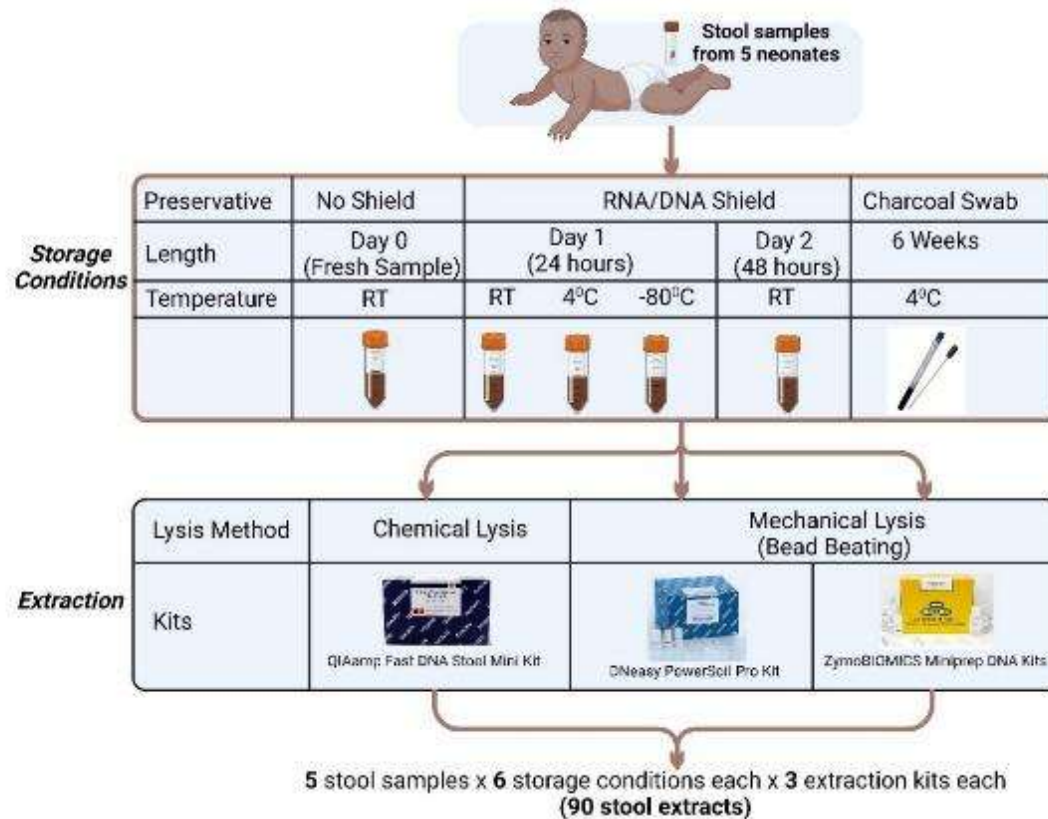


Figure 3.1 Study design for neonatal stool storage and extraction comparison.

Stool from five neonates (Day 0–Day 1) was allocated to six storage/preservation conditions: no additive (same-day processing); DNA/RNA Shield stored for 24 h at RT, 4°C, or –80°C; DNA/RNA Shield stored for 48 h at RT; and charcoal swab storage for 6 weeks at 4°C. Aliquots from each condition were extracted using three kits representing contrasting lysis strategies: QIAamp Fast DNA Stool Mini (SM), DNeasy PowerSoil Pro (PS), and ZymoBIOMICS DNA Miniprep (ZB), yielding 90 extracts in total ($5 \times 6 \times 3$). RT = room temperature. Reproduced from Akpulu et al., 2025.

3.2.3 Selection of DNA extraction kit

The three commercially available kits selected as described in 3.1 were:

- QIAamp Fast DNA Stool Mini (Qiagen); referred to as protocol SM.
- DNeasy PowerSoil Pro (Qiagen; formerly MoBio); referred to as protocol PS.
- ZymoBIOMICS DNA Miniprep (Zymo Research); referred to as protocol ZB.

All three kits use chemical lysis and silica spin-column purification (Table 3.1). PS and ZB incorporate bead-beating (mechanical lysis), whereas SM relies primarily on vortex-based disruption with a high-temperature denaturation step.

Table 3.2 Features and operational requirements of three commercial DNA extraction kits evaluated for neonatal stool.

DNA Extraction Methods	SM	PS	ZB
Mechanical lysis	Vortex	Bead-beating	Bead-beating
Chemical lysis	Yes	Yes	Yes
Thermal lysis	95°C for 10 mins	65°C for 10 mins (optional)	No
DNA purification method	Spin column-based	Spin column-based	Spin column-based
Cost of kit (50 Preparations)	~£290	~£374	~£417
Processing time per 12 samples	~1.5 - 2 hours	~1.5 - 2 hours	~2 – 2.5 hours
Equipment needed	Micro-centrifuge Heat block	Micro-centrifuge Heat block (optional) Vortexer	Micro-centrifuge Vortexer

SM = QIAamp Fast DNA Stool Mini (Qiagen); PS = DNeasy PowerSoil Pro (Qiagen); ZB = ZymoBIOMICS DNA Miniprep (Zymo Research). Costs and runtimes reflect container-laboratory conditions at MMSH. Differences map onto expected trade-offs between inhibitor tolerance, HMW DNA preservation, and feasibility under limited infrastructure.

3.2.4. DNA extraction procedures

For solid stool, approximately 220–250 mg was processed per extraction. For stool preserved in DNA/RNA Shield, 200 μ L of the suspension was used.

- Protocol SM (QIAamp Fast DNA Stool Mini). Stool was homogenised in InhibitEX buffer followed by proteinase K and chaotropic lysis. After incubation, ethanol was added and lysate loaded onto silica columns. Columns were washed as per kit instructions and DNA eluted in 200 μ L elution buffer.
- Protocol PS (DNeasy PowerSoil Pro). Material was transferred into bead tubes with lysis buffer, bead-beaten (Vortex-Genie 2; 10 min), and incubated at 65°C for 10 min before clarification. Inhibitor removal buffers were applied, DNA bound to spin columns, washed, and eluted in 100 μ L.
- Protocol ZB (ZymoBIOMICS DNA Miniprep). Material was combined with lysis buffer in bead tubes and bead-beaten. After clarification, lysate was passed through a Zymo-Spin III-F filter. DNA was bound to silica, washed, eluted in 100 μ L nuclease-free water, and further cleaned using the Zymo-Spin HRC column to reduce inhibitors.

3.2.5. DNA quantification

DNA concentration was measured using a Qubit 4 fluorometer (Life Technologies) with dsDNA HS or BR assays as appropriate. DNA yield (ng) was calculated as concentration multiplied by elution volume.

3.2.6. Library preparation and sequencing

As SM frequently produced insufficient yield across storage/preservation conditions, downstream sequencing focused on extracts generated using PS and ZB from conditions selected to represent: (i) same-day processing, and (ii) longer-term storage (week-6 charcoal swab condition). Prior to library preparation, extracts were concentrated using SPRI (solid-phase reversible immobilisation) bead cleanup (AMPure XP) and eluted in 30–70 μL , targeting approximately 15–20 ng/ μL .

Sequencing libraries were prepared using the ONT Rapid Barcoding Kit (SQK-RBK114.96) and run on GridION using R10.4.1 flow cells (Oxford Nanopore Technologies, UK), following the manufacturer's instructions. The Rapid PCR-free Barcoding workflow was selected to accommodate the lower DNA input concentration typical of neonatal low-biomass extracts, while reducing per-sample costs. This approach follows the ONT protocol for high-throughput, cost-effective library preparation from limited starting material.

3.2.7. Bioinformatics processing

Raw signals were basecalled, demultiplexed, and adapter-trimmed using Guppy v6.4.2. Initial QC summaries were generated with SeqKit v2.6.1. Low-quality or very short reads were removed using NanoQ v0.10.0, applying filters of $Q < 10$ and read length < 100 bp.

Filtered reads were classified using Kraken2 v2.1.3 against the PlusPFP database (release 2024-01-12). Abundances were re-estimated using Bracken and exported as BIOM format via kraken-biom v1.0.1 for downstream analysis.

3.2.8. Statistical analysis and diversity

Analyses were performed in R v4.4.1. BIOM tables were imported into phyloseq v1.42.0 and visualised using ggplot2 v3.5.1. To support like-for-like comparisons in this small pilot, samples were rarefied to equal depth (`phyloseq::rarefy_even_depth`), and alpha diversity metrics (Observed richness, Shannon, Simpson, Inverse Simpson) were calculated using microbiome v1.20.0.

Beta diversity was assessed using Bray–Curtis distances and visualised using PCoA ordination. Group differences by kit and storage condition were assessed using PERMANOVA (`adonis2`; `vegan` v2.6.6.1), with dispersion differences evaluated using `betadisper` followed by `permutest`.

For univariate outcomes, between-kit differences in DNA yield were analysed using Friedman tests across kits (SM, PS, ZB) while preserving within-infant pairing, followed by paired Wilcoxon signed-rank post-hoc comparisons with Holm adjustment for the three pairwise contrasts. Within-kit storage/timing effects (PS and ZB analysed separately) were assessed with Friedman tests across time points (same-day; Day 1 DNA/RNA Shield; Day 2 DNA/RNA Shield; week-6 charcoal swab), followed by paired Wilcoxon post-hoc tests with Holm adjustment. Effect sizes for paired contrasts are reported as Hodges–Lehmann median differences with 95% confidence intervals (Walsh averages). Paired Wilcoxon tests were also used for read N50 and read-quality comparisons between kits. Non-parametric procedures were chosen a priori given the small sample size and non-normal outcome distributions.

Exploratory differential abundance analyses accounting for repeated measures was conducted using MaAsLin3 with infant-level random intercepts, treating findings as hypothesis-generating given the pilot sample size.

3.2.8. Data availability

Raw reads are available under NCBI BioProject PRJNA1260441. Software versions are specified above. Analysis scripts and parameter files will be archived alongside the thesis and deposited in a public version-controlled repository (GitHub) following thesis submission.

3.3. Results

3.3.1. Protocol-specific DNA yield

Across all handling conditions (Figure 3.2), PowerSoil Pro (PS) and ZymoBIOMICS (ZB) produced the highest DNA yields under same-day processing with no added preservative (Day 0), while Stool Mini (SM) yielded DNA consistently at/near the Qubit limit of quantification (LoQ) and was effectively non-performant for neonatal stool under the conditions tested.

On Day 0, median (IQR) yields were PS 742.7 ng (561.3-2000.0), ZB 285.7 ng (285.0-2000.0), and SM 0.0 ng (0.0-0.0) ($n = 5$ per kit; constant input mass; Qubit HS assay). After 24-48 h storage in DNA/RNA Shield across temperature conditions, yields for PS and ZB fell into the tens to low hundreds of ng, with the lowest yields observed after 48 h at room temperature (Day 2 RT). Week-6 charcoal swab storage at 4 °C produced higher but more variable yields for PS and ZB (Figure 3.2).

Using Friedman's test with blocks defined by paired infant $\times$ condition (5 infants $\times$ 6 conditions = 30 blocks), there was strong evidence of overall yield differences between kits ($\chi^2 = 39.47$, $p = 2.69 \times 10^{-9}$; Kendall's $W = 0.66$). Paired Wilcoxon signed-rank contrasts with Holm adjustment supported PS > SM and ZB > SM (both adjusted $p < 0.001$), whereas PS

vs ZB was not statistically different (adjusted $p = 0.236$). Full per-condition contrasts (including Hodges–Lehmann median differences with 95% CIs) are provided in Appendix A. 2, Supplementary Table 3.1.

3.3.2. Effects of storage and time-to-extraction

Within-kit comparisons ($n = 5$ infants per time point) showed the highest yields with same-day extraction (Day 0), reduced yields after 24 h in DNA/RNA Shield (Day 1; pooled per infant as the median across $-80\text{ }^{\circ}\text{C}/4\text{ }^{\circ}\text{C}/\text{RT}$), the lowest yields after 48 h at room temperature (Day 2 RT), and higher but more dispersed yields at Week 6 (charcoal swab; $4\text{ }^{\circ}\text{C}$) (Figure 3.2).

Friedman tests run separately by kit detected a time/storage effect for both bead-beating protocols: PS $\chi^2 = 9.72$, $p = 0.021$ ($W = 0.648$) and ZB $\chi^2 = 12.60$, $p = 0.0056$ ($W = 0.840$). SM showed no evidence of a time effect, consistent with persistently negligible yields across all conditions ($\chi^2 = 3.00$, $p = 0.392$; $W = 0.200$). On this basis, protocol SM was considered unsuitable for neonatal low-biomass stool in this setting and was excluded from subsequent downstream sequencing and community-level analyses, which therefore focus on the bead-beating protocols PS and ZB.

Pairwise Wilcoxon signed-rank tests (Holm-adjusted) did not identify statistically significant contrasts within PS or ZB across time points, although the largest paired shifts were observed between Day 0 and Day 1/Day 2. Per-time medians (IQRs) by kit are provided in Appendix A. 2, Supplementary Table 3.2.

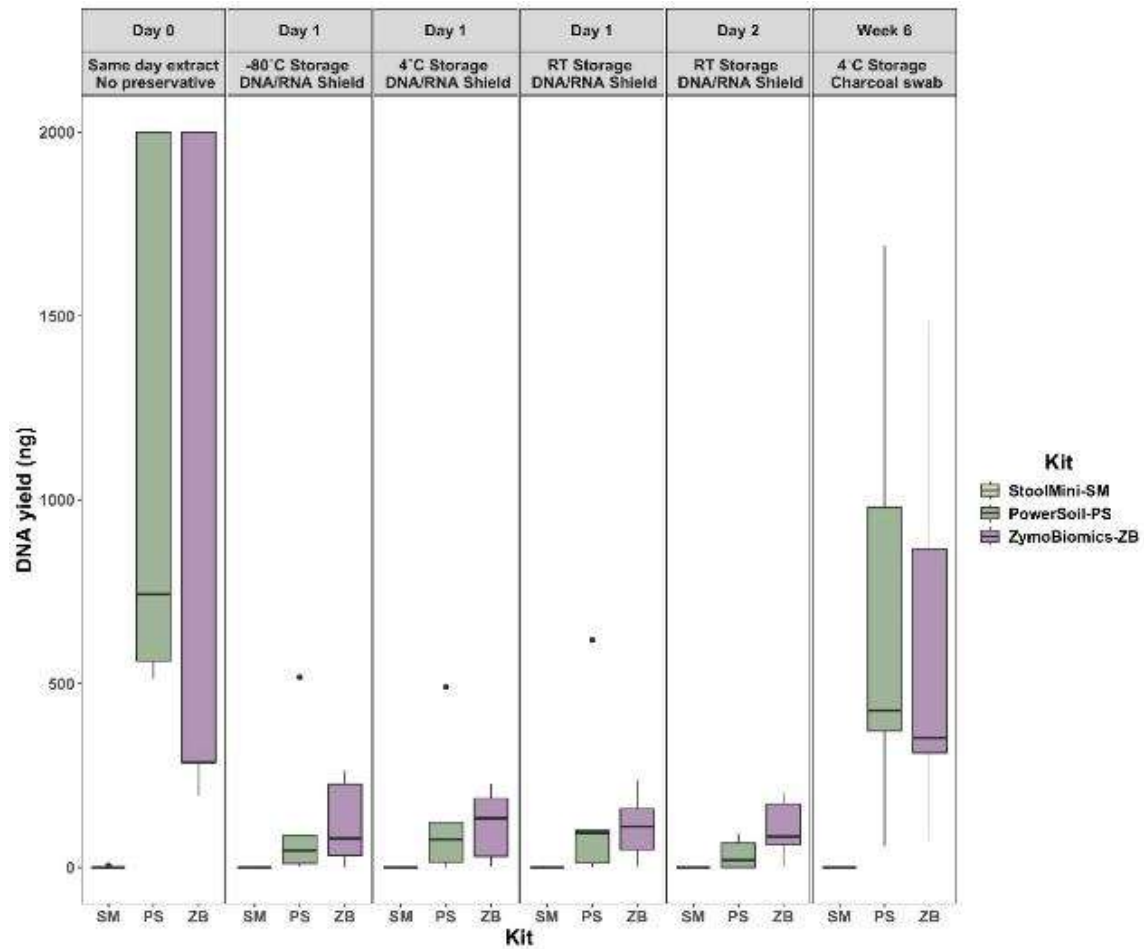


Figure 3.2 DNA yield by extraction kit across handling conditions.

Boxplots show total DNA yield (ng; Qubit HS assay) from five neonatal stool specimens extracted using three protocols: SM (QIAamp Fast DNA Stool Mini), PS (DNeasy PowerSoil Pro), and ZB (ZymoBIOMICS DNA Miniprep). Aliquots were processed under six handling conditions: Day 0 (same-day extraction; no preservative added); Day 1 (24 h in DNA/RNA Shield stored at -80°C , 4°C , or room temperature); Day 2 (48 h in DNA/RNA Shield at room temperature); and Week 6 (charcoal swab stored at 4°C). PS and ZB yielded highest DNA under same-day processing with reduced yields after 24–48 h in DNA/RNA shield; SM yielded DNA at/near the limit of quantification across all conditions.

3.3.3. ONT read-length distributions and assembly-relevant metrics

Sequencing was performed for PS and ZB extracts at Day 0 and Week 6 only; SM and the Day 1/Day 2 Shield conditions were not sequenced due to insufficient practical input mass (Section 3.3.1). Read-level summaries are shown in Figure 3.3 (N50), Appendix A.2

Supplementary Figure 3.1 (read counts and total bases), and Appendix A.2 Supplementary Figure 3.2 (mean read Q score).

Across both time points, PS tended to yield longer reads (higher N50) than ZB (Day 0 $\approx$ 5.6 kb vs 4.1 kb; Week 6 $\approx$ 4.6 kb vs 3.8 kb), but paired Wilcoxon tests found no statistically significant PS–ZB differences (Day 0: $p = 0.31$; Week 6: $p = 0.63$). Within each kit, the Day 0–Week 6 change in N50 was also non-significant (PS: $p = 0.63$; ZB: $p = 0.19$).

Median read quality was broadly similar between kits (Day 0 $\sim$ 12.7–12.8). Week-6 ZB libraries showed slightly higher median Q ($\sim$ 13.1) than PS ($\sim$ 12.7), but the paired contrast was not significant ($p = 0.63$). Read counts and total sequenced bases were higher at Week 6 than Day 0, particularly for ZB; paired tests approached significance for ZB (reads $p = 0.0625$; bases $p = 0.0625$) but were not significant for PS. Overall, these data suggest a modest tendency toward longer fragments with PS, while Week-6 throughput tended to be higher with ZB, with no strong evidence of systematic differences at this sample size. Full paired contrasts and 95% CIs are provided in Table S3.3.

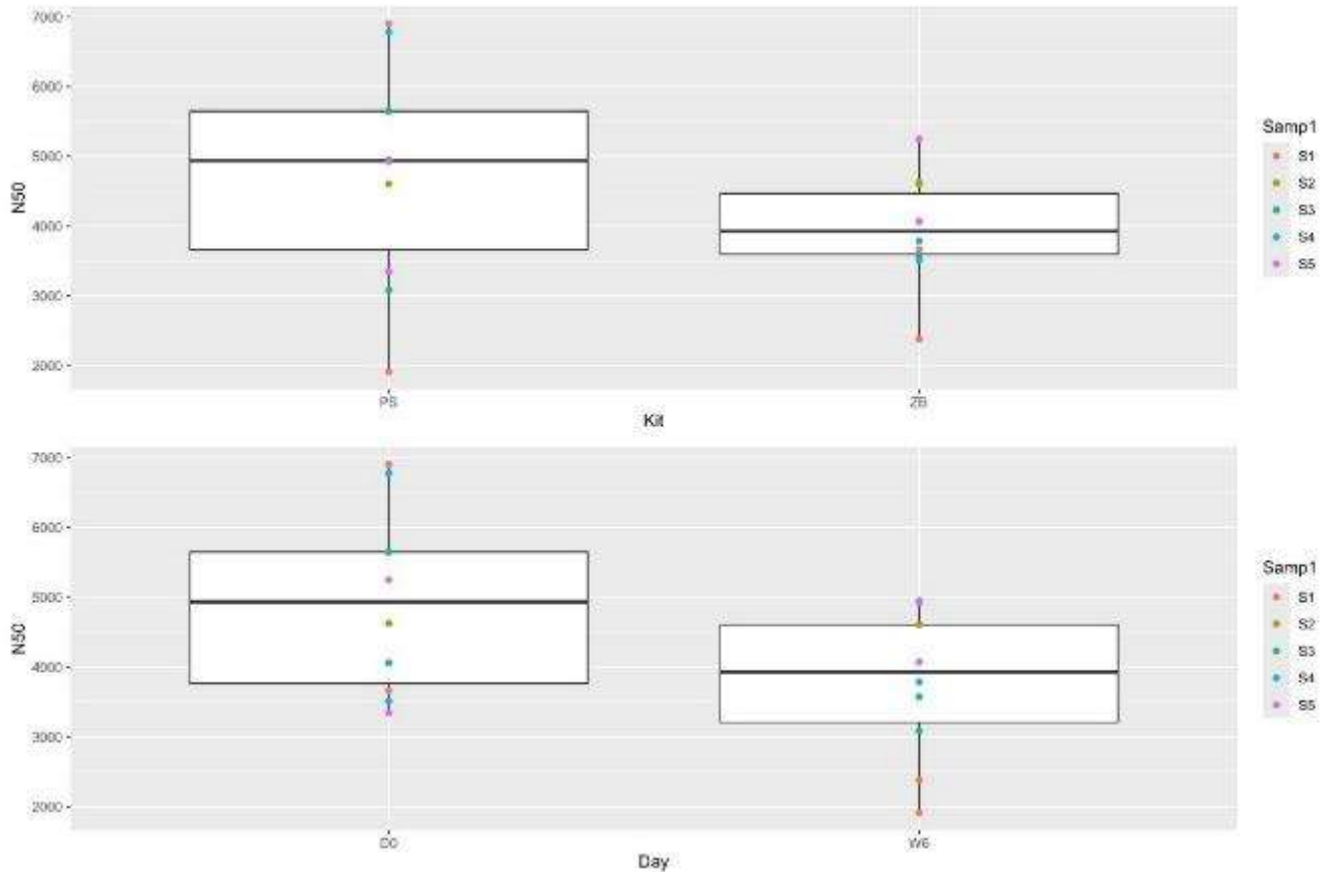


Figure 3.3 Comparison of read-level N50 across DNA extraction kits and storage times

Read N50 (bp) for ONT libraries prepared from PS and ZB extracts sequenced at Day 0 (same-day; no preservative) and Week 6 (charcoal swab stored at 4 °C). Points indicate matched infants (S1–S5). (A) N50 by extraction kit (PS vs ZB). (B) N50 by time point (Day 0 vs Week 6). Paired comparisons were assessed using Wilcoxon signed-rank tests.

3.3.4. Community profiles signal

Species-level profiles (Kraken2/Bracken; PlusPFP database) were dominated by facultative Enterobacteriaceae typical of early neonatal gut communities (Shao et al., 2019; Stewart et al., 2018), with *E. coli* frequently abundant and variable contributions from *Klebsiella*, *Enterococcus*, and *Staphylococcus*. In some infants, signal from early anaerobes (e.g., *Bacteroides/Phocaeicola*) and *Bifidobacterium* was detectable at low-to-moderate abundance. Across infants, between-infant variability was greater than any visually apparent

kit effect, and within-infant Day 0-Week-6 shifts were heterogeneous rather than uniform (Figure 3.4).

To formally evaluate kit and time associations, a MaAsLin3 model including Day (Week 6 vs Day 0) and Kit (ZB vs PS) (with repeated measures handled as specified in Methods) was fitted at the species level (Figure 3.5). Consistent with the visual summaries, few abundance associations survived FDR correction, while a small number of prevalence associations were detected. Overall, the pilot provides no evidence of a persistent kit-driven shift in community composition under the sequencing depth achieved, with modest day-linked prevalence changes detectable after multiple testing control.

Rarefied alpha diversity metrics (Observed richness, Shannon, Simpson, Inverse Simpson) did not show consistent differences between kits (Figure 3.6A). Bray-Curtis PCoA showed substantial overlap by kit (Figure 3.6B) and no clear separation by time point (Figure 3.6C), supporting the conclusion that inter-individual variability dominates in this small pilot dataset.

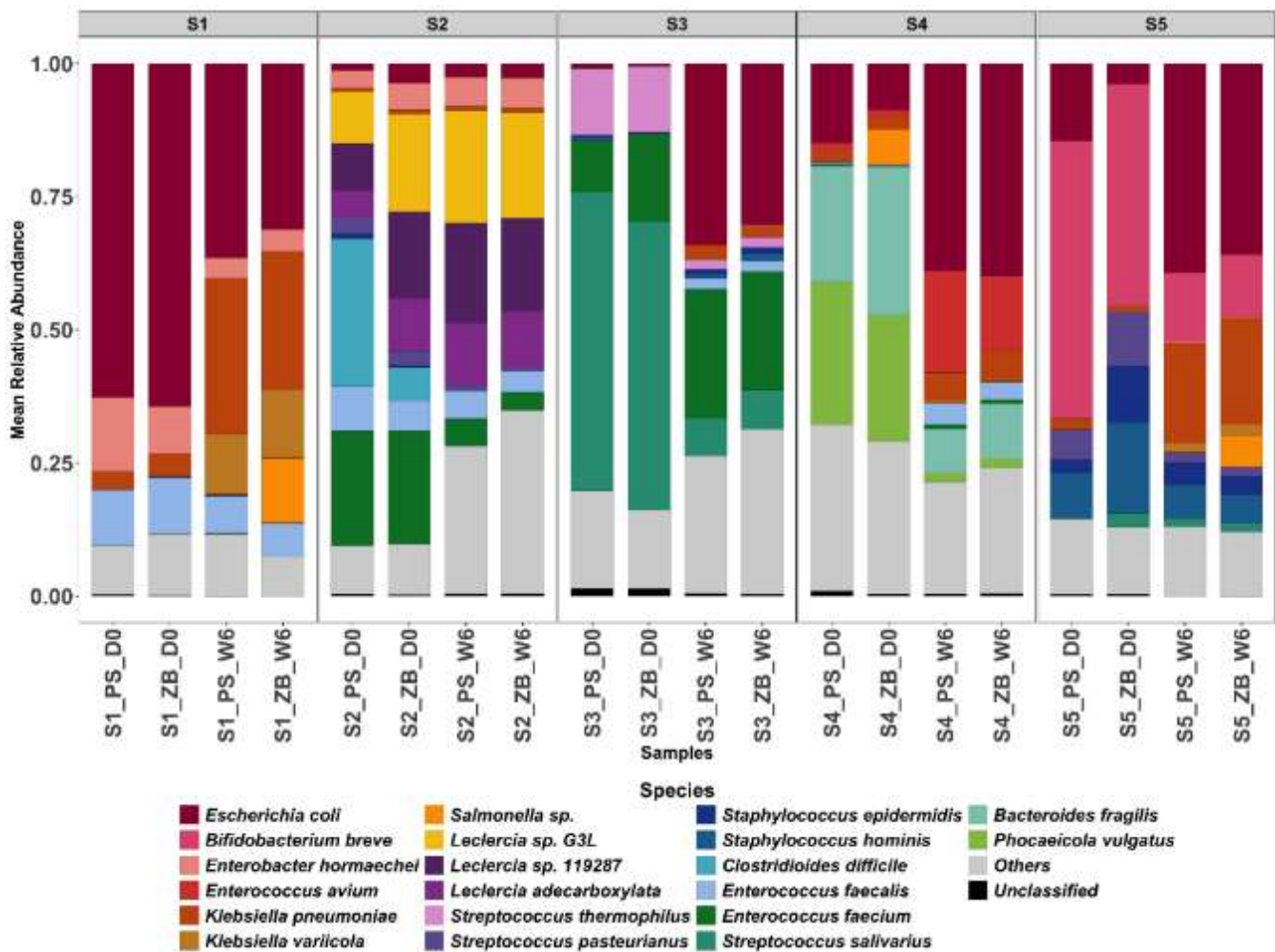


Figure 3.4 Species-level taxonomic profiles of neonatal stool libraries by kit and time point

Stacked bar plots show species-level relative abundance estimated from Kraken2/Bracken classifications (PlusPFP database) for PS and ZB libraries sequenced at Day 0 and Week 6 (five infants; 20 libraries). Bars are grouped by infant (S1–S5) and labelled by kit and time point (PS/ZB; Day 0/Week 6). “Others” aggregates lower abundance taxa; “Unclassified” indicates reads without species-level assignment.

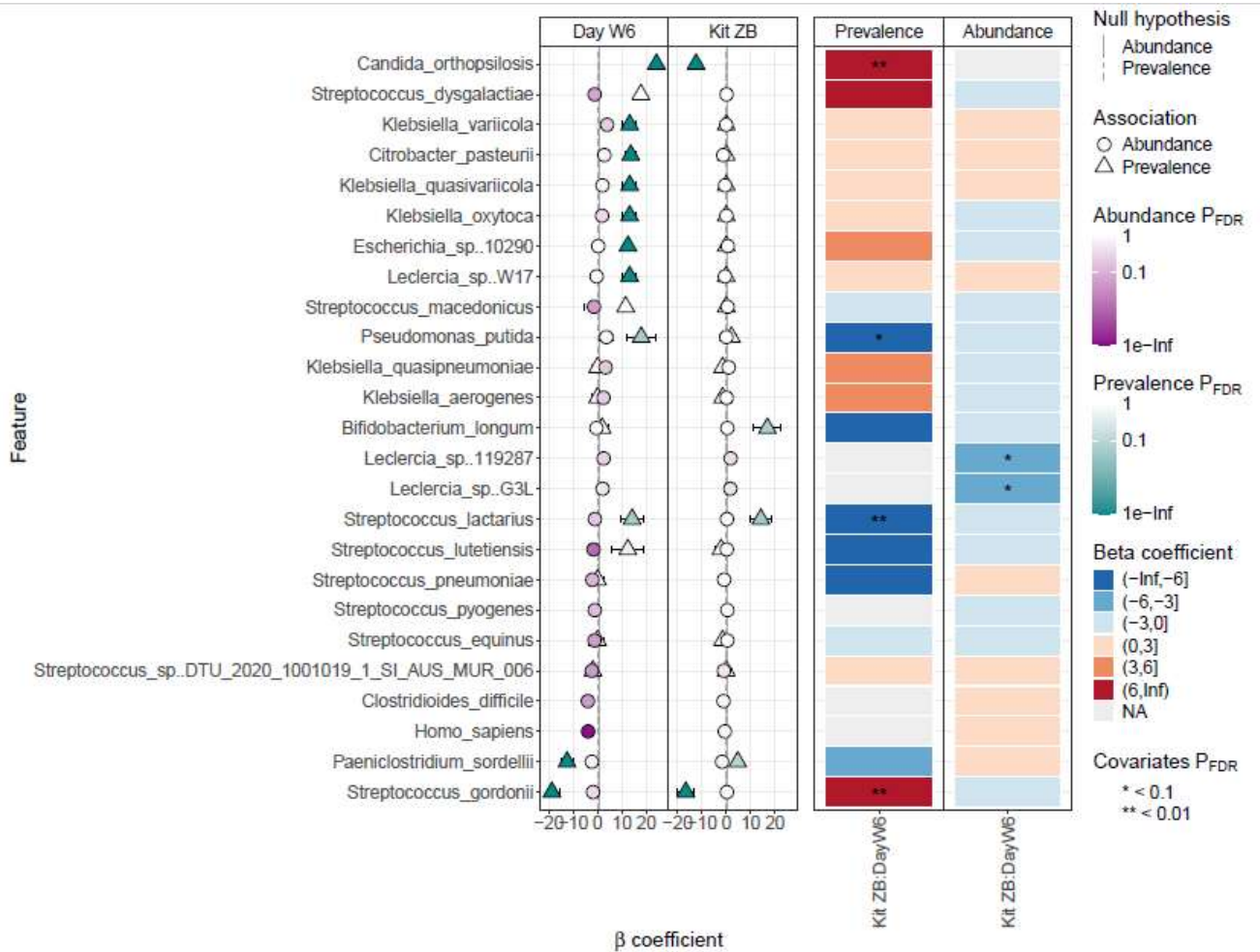


Figure 3.5 MaAsLin3 associations of species abundance and prevalence with kit and storage time

MaAsLin3 results evaluating associations with Day (Week 6 vs Day 0) and Kit (ZB vs PS) at the species level using PS/ZB libraries sequenced at Day 0 and Week 6 (five infants; repeated measures as specified in Methods). Points show estimated coefficients (β) for abundance (circles) and prevalence (triangles). Heat bars indicate FDR-adjusted significance for prevalence and abundance; starred covariates denote FDR thresholds (* < 0.1; ** < 0.01). *Figure plotted by Aditiya Lankapalli*

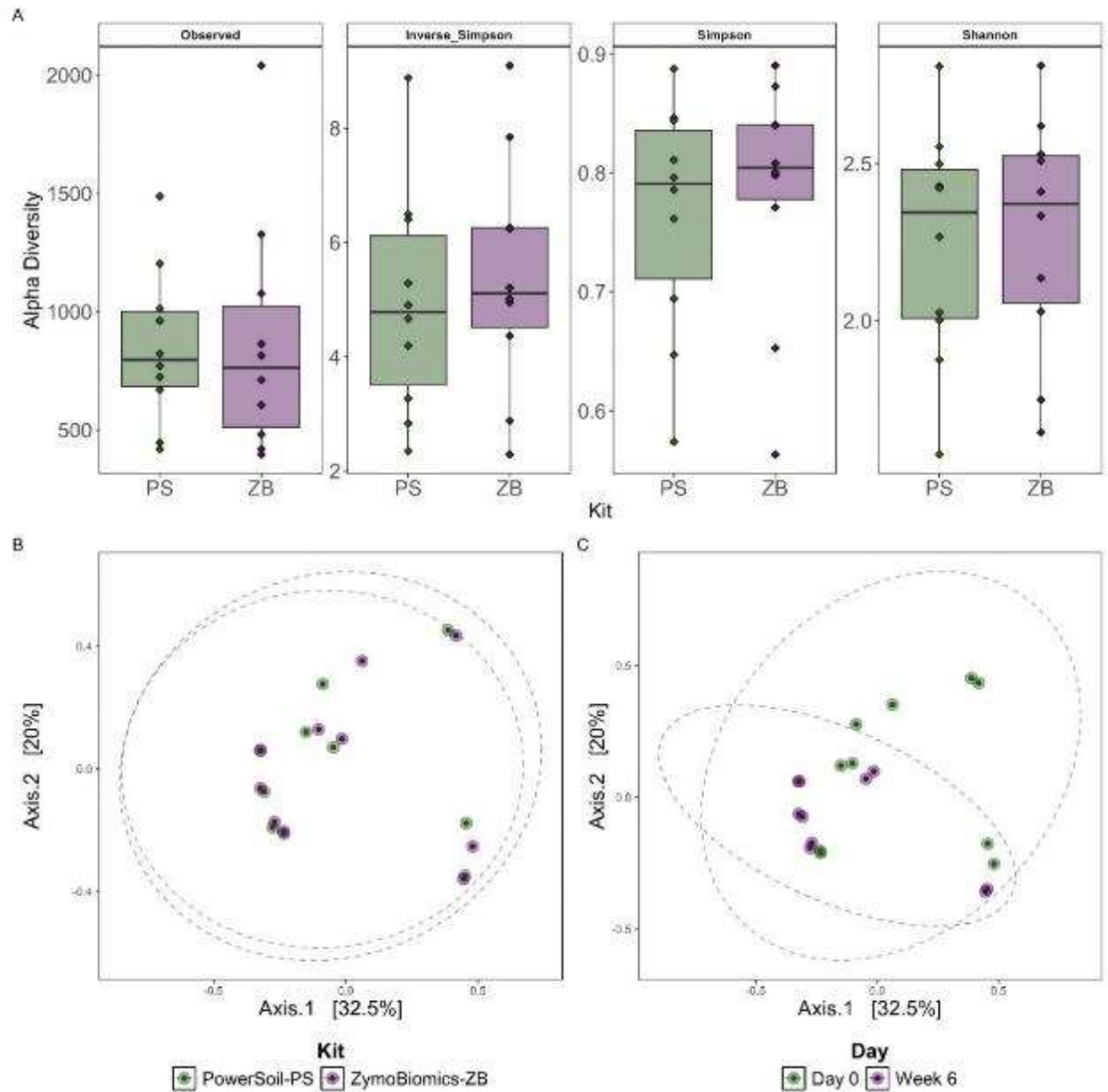


Figure 3.6 Alpha and beta diversity by extraction kit and time point.

(A) Alpha diversity (Observed richness, Shannon, Simpson, Inverse Simpson) for PS vs ZB libraries. (B) Bray–Curtis PCoA coloured by extraction kit (PS vs ZB) with 95% confidence ellipses. (C) Bray–Curtis PCoA coloured by time point (Day 0 vs Week 6) with 95% confidence ellipses. Each point represents one library; paired samples correspond to the same infant across kit and/or time point.

3.3.5. Operational feasibility and cost

This pilot was conducted in a hospital-adjacent laboratory designed from a shipping container with intermittent power and limited equipment. In that context, protocols were compared for execution reliability, production of library-eligible DNA, and per-sample cost.

In 12-sample batches, observed hands-on time aligned with kit workflows: PS and SM each required ~1.5-2.0 h, and ZB ~2.0-2.5 h. All three workflows used spin-column purification and required only a micro-centrifuge; SM required a 95 °C incubation, while PS used an optional 65 °C step and ZB required no heating step. Critically, SM did not generate usable yields for neonatal stool under any condition tested and therefore did not support sequencing. In contrast, PS and ZB routinely produced concentrates (after SPRI cleanup) sufficient for ONT Rapid Barcoding libraries for Day-0 and Week-6 conditions, whereas Day-1/Day-2 Shield extracts were generally below practical input requirements for long-read libraries.

Using list prices during the pilot, approximate extraction-only consumable costs were: SM ~£5.80, PS ~£7.48, and ZB ~£8.34 per prep. In practice, procurement constraints (lead times, customs clearance, and supply continuity) were operationally important considerations alongside the modest per-prep price differences, reinforcing the need to pre-position at least one bead-beating kit (PS or ZB) for sustained neonatal metagenomics at the study site.

3.4. Discussion

This pilot aimed to identify a neonatal stool DNA extraction approach that enables ONT long-read metagenomics and is feasible under field constraints at MMSH, balancing analytical performance with operational practicality. Using stool from five neonates processed across multiple storage and timing conditions, three commercial kits were

compared; only the bead-beating protocols, PowerSoil Pro (PS) and ZymoBIOMICS Miniprep (ZB), consistently produced DNA of sufficient to support library preparation, and these were therefore taken forward for ONT sequencing. Overall, the pilot indicates that mechanical lysis is required to obtain library-eligible DNA from neonatal low-biomass stool in this setting; delays between collection and extraction are associated with substantial loss of recoverable DNA; and, within this small dataset, biological differences between infants were larger than any consistent differences attributable to the bead-beating kit.

The poor performance of SM is biologically plausible for early neonatal stool as these samples are typically low biomass with a high host-DNA fraction and can contain organisms that are inefficiently lysed without mechanical disruption (Bajorek et al., 2019; Kennedy et al., 2023; Koenig et al., 2011). Under these conditions, bead-beating is needed to release sufficient microbial DNA for long-read library preparation, whereas chemistry-only extraction risks leaving microbial DNA below practical input thresholds. This interpretation is consistent with published extraction method evaluations and stool DNA standards that incorporate mechanical lysis to improve recovery of hard-to-lyse organisms and increase total yield. Salonen *et al.* showed up to 35-fold yield differences across faecal DNA methods and highlighted the impact of mechanical lysis on recovery profiles; subsequent reviews and experiments similarly conclude that bead-beating increases total DNA and improves representation of Gram-positive lineages (Costea et al., 2017; Hsieh et al., 2016; Salonen et al., 2010).

Storage and processing delay materially affected yield for both bead-beating protocols. Same-day extraction (Day 0) produced the highest yields, with marked declines after storage in DNA/RNA Shield at Day 1–2. Although pairwise post-hoc contrasts were not statistically significant after Holm adjustment, effect-size estimates for Day 0 versus Day 1-2 were large

and directionally consistent, indicating a practically important reduction in recoverable DNA with delay. The lack of pairwise significance most likely reflects limited power ($n = 5$ paired infants) and conservative multiplicity control rather than absence of a true effect. These findings are compatible with handling and storage literature showing that delays and suboptimal storage can affect DNA recovery and downstream metagenomic performance (Choo et al., 2015; Holzhausen et al., 2021; Nel Van Zyl et al., 2020).

Importantly, aliquots stored in DNA/RNA Shield did not preserve library-eligible input in this neonatal context. Although Shield is often reported to stabilise adult-stool profiles (Kim et al., 2023; Preywissh et al., 2025), in this pilot Shield-stored neonatal aliquots generally produced much lower recoverable DNA than same-day extraction and were rarely suitable for ONT library preparation. A plausible explanation is that, in very low-biomass samples, preservative-associated dilution and chemistry–workflow interactions can reduce effective microbial concentration and exacerbate stochastic effects, potentially increasing susceptibility to background signal (Hasrat et al., 2021; Rajar et al., 2022). Specifically, the addition of a liquid stabilisation buffer to a small neonatal stool aliquot creates a dilute slurry; if the DNA recovery from this volume is already near the kit's lower limit, the reduction in biomass-to-buffer ratio can lead to the observed precipitous drop in total yield.

By contrast, Week-6 charcoal swab storage sometimes yielded sufficient DNA, but with substantial variability across infants. The higher yields observed in some Week-6 charcoal swabs compared to the Day 1–2 Shield samples likely stem from the concentrated state of the sample during storage. Unlike the Shield-aliquots, which were stored as a dilute liquid slurry, the charcoal swabs maintained the stool matrix in a concentrated form on the swab tip for the duration of storage. While PBS was used to elute the sample prior to extraction, the immediate processing of this concentrated suspension likely facilitated better recovery

of the limited DNA available, compared to the prolonged storage in a larger volume of stabilization buffer which may have hindered efficient recovery. The inconsistency observed elsewhere in these results may reflect swab/charcoal matrix and handling effects, for example, DNA sticking to the material and being released unevenly during extraction, or gradual degradation during prolonged storage. Additionally, the variable initial composition of the samples may have contributed to differential microbial overgrowth rates, further driving the observed inconsistency. Because temperature logs were not available and storage was not replicated under tightly controlled conditions, the mechanism cannot be confirmed; however, the variability supports treating charcoal swab storage as a fallback option, rather than equivalent to same-day extraction.

At the sequencing step (Day 0 and charcoal-swab stored samples), read-length distributions tended to favour PS (higher N50), whereas ZB showed higher throughput (reads and total bases) in the Week-6 storage condition; mean read quality was broadly similar. These contrasts did not reach significance in paired exact tests ($n = 5$), but the pattern suggests a task-dependent trade-off: where downstream aims rely on longer reads (e.g., assembly contiguity or resolving ARG/MGE context), PS's read-length tendency may be advantageous; where throughput is the limiting factor for detection, ZB's higher read counts under the Week-6 storage condition may be helpful, although this could reflect combined extraction and library-prep variability rather than a pure kit effect.

Across PS and ZB libraries, communities were dominated by facultative Enterobacteriaceae typical of early life (Shao et al., 2019; Stewart et al., 2018), with substantial between-infant heterogeneity. In models accounting for repeated sampling, few FDR-significant kit or day effects were detected for abundance; prevalence signals were more common but sparse and feature-specific. Concordantly, α - and β -diversity showed no consistent separation by bead-

beating kit at the sequencing depth obtained, suggesting no systematic, cohort-wide kit driven shift in overall composition within this pilot. This is consistent with the expectation that rapid biological change and inter-individual variation in early life can outweigh moderate protocol effects (Stewart et al., 2018).

Given the absence of a consistent compositional shift between bead-beating kits in the sequenced subset, the practical decision point becomes operational feasibility at the study site. PS and ZB produced broadly comparable yields with imprecise estimates in this small pilot, but PS offered practical advantages for routine deployment: shorter hands-on time per batch, fewer steps (reducing handling error risk), and a tendency toward longer reads that support genome-resolved analyses. ZB also performed well but required more handling steps and longer processing time, which can create bottlenecks in constrained workflows and increase vulnerability to interruptions (e.g., power instability).

This pilot's principal strength is its site-realistic design, evaluating extraction performance under the same infrastructural constraints expected for the main cohort and also assessing outcomes directly relevant to long-read metagenomics (yield and long-read performance proxies). Limitations are primarily those of pilot scope: small sample size ($n=5$), sequencing restricted to selected conditions, and a single-site context. Finally, species-level ONT taxonomic assignment can be sensitive to database choice and the distribution of closely related taxa within Enterobacteriaceae in samples; genus-level summaries may therefore be more robust unless supported by assembly-based evidence (Portik et al., 2022; Van Uffelen et al., 2024).

In summary, under resource-constrained, on-site conditions, mechanical lysis was required to recover usable DNA from neonatal stool, and same-day extraction was the most reliable strategy for obtaining high-molecular-weight DNA eligible for Nanopore library

preparation. Between bead-beating kits, compositional differences were modest and task-dependent (PS tending toward longer reads; ZB occasionally higher throughput), and secondary to operational logistics and site-specific feasibility. On balance, PS was selected as the default extraction protocol for the main study, while ZB was retained as a validated alternative should procurement or scheduling constraints arise.

Chapter 4: Cohort profile of early-life gut microbiome and resistome longitudinal study: northern Nigerian birth cohort

4.1 Introduction

This chapter focuses on the profile of the Northern Nigerian Birth Cohort, describing maternal, infant, and household characteristics that contextualise microbiome and resistome findings. Subsequent chapters analyse taxonomic and resistome trajectories; their interpretation depends on who was enrolled, which exposures were common, and how follow-up and sampling performed in practice. This chapter provides that bridge from design (Chapter 2) to inference (Chapters 5-6), anchoring later models in the observed exposure distribution.

As outlined in Chapter 1, maternal health status (Milton, Modibbo, et al., 2022), delivery mode (Dominguez-Bello et al., 2010; Reyman et al., 2019), neonatal antibiotic exposure (Arboleya et al., 2015; Patangia, Grimaud, O'Shea, et al., 2024a; Saeed & Hall, 2023), infant feeding practices (Laursen et al., 2021; Lyons et al., 2020), and household living conditions (Laursen et al., 2015; Tun et al., 2017) are all established drivers of early-life microbial assembly. However, these determinants are not interchangeable across settings; the health-system structure, and WASH conditions in Northern Nigeria differ markedly from the high-income cohorts that dominate the global literature (Oyedemi et al., 2022). Even within Nigeria, significant regional heterogeneity exists. While urban cohorts in the Southwest have provided early insights into infant gut maturation (Oyedemi et al., 2022), comparative studies between urbanised Nigerians and traditional rural populations, such as the Bassa (F. A. Ayeni et al., 2018), suggest that high levels of environmental microbial exposure can alter successional trajectories. In such high-exposure settings, environmental acquisition may act

as a “leveller” that overrides individual clinical modulators. Consequently, a specific investigation of the Kano context is required to understand how these drivers operate within the unique ecological and AMR landscape of Northern Nigeria.

The cohort was recruited in Kano State, Northern Nigeria, where environmental constraints intersect with health-system and sociocultural factors relevant to early-life microbial exposures (Angell et al., 2022; Edefo, 2025; J Olumade et al., 2020). The Nigerian health system is organised across three tiers: primary health care (PHC) delivered mainly by Local Government/State PHC Boards; secondary care in state general hospitals; and tertiary care in federal teaching/specialist hospitals, following a hierarchical referral pathway from primary to specialised facilities (Christian et al., 2023). Alongside the public sector is a large mixed private market that includes hospitals, clinics, community pharmacies, and patent medicine vendors, important access points for antibiotics and first-contact care (Ajibola et al., 2018).

Public facilities charge user fees for many services (exemptions/subsidies exist but vary strikingly by state, programme, and stock), while health insurance coverage remains limited despite expansion under the National Health Insurance Authority (Christian et al., 2023). The Basic Health Care Provision Fund channels earmarked federal funds to PHC, but implementation is uneven and often constrained by supply chains and staffing. Out-of-pocket spending accounted for approximately 76% of current health expenditure in 2022. Many households rely on out-of-pocket payment, creating sensitivity to transport costs, availability of medicines/consumables, and informal charges (Christian et al., 2023).

WASH access is heterogeneous; exposure to informal refuse sites, open drains, or intermittent water supply can increase contact with environmental microbes. Nationally, access to WASH has improved in Nigeria, but gaps persist especially for children under the

age of 18 years (United Nations Children’s Fund (UNICEF) & World Health Organization (WHO), 2023; Victor et al., 2025). A recent government and WHO/UNICEF assessments indicate substantial shortfalls in basic and especially safely managed services (Federal Ministry of Water Resources (FMWR), 2022; United Nations Children’s Fund (UNICEF) & World Health Organization (WHO), 2023). At state level, Kano’s drive toward better sanitation, such as the campaign against open defecation and dump refuse has intensified yet sustaining hygiene behaviours and expanding reliable urban water and sanitation remain priorities given continued service gaps documented in WASHNORM and urban studies of Kano (Federal Ministry of Water Resources (FMWR), 2022; Lawal Bello et al., 2024).

Family structure often includes extended-family compounds and, in some households, polygyny arrangements, expanding caregiving networks and increasing person-to-person contact through shared sleeping spaces, and shared WASH facilities. Peripartum practices, including the use and timing of prophylactic antibiotics, and decisions around delivery facility, are shaped by cultural expectations (Sharma et al., 2019).

Longitudinal infant microbiome studies from Nigeria are rare. This cohort contributes dense early-life sampling (birth to 90 days), paired hospital and household environmental context, and community-validated retention in a public referral setting with predominantly out-of-pocket medicine purchase. These features allow us to examine exposures that are common in low-resource urban environments but largely absent from high-income cohorts.

In the sections that follow, I describe the study site and healthcare context, outline the implementation of the cohort, and present baseline maternal, infant, and household characteristics. I also highlight the completeness of sample collection and key exposures during the first 90 days of life. While detailed methods are provided in Chapter 2 (Materials and Methods), this chapter emphasises the cohort profile and the realities of study

implementation, including challenges and lessons learned during fieldwork. The objectives of this chapter are to: (i) characterises the study population; (ii) profile key exposures that could plausibly modulate the microbiome/resistome; and (iii) document sample collection including data missingness patterns.

4.1.1 Study aims

This section described the implementation of the Northern Nigerian Birth Cohort, including recruitment, retention, data and sample collection, and participant withdrawals. The focus was primarily on outcomes of field implementation, completeness of follow-up, and contextual realities of conducting longitudinal research in Kano.

Key objectives were:

- To describe baseline maternal, infant, and household characteristics.
- To monitor sample yields and retention across dense early-life timepoints.
- To document exposures pertinent to microbial colonisation and resistome development.

4.2 Study site context

4.2.1 Kano city

This study was conducted within Kano metropolis in north-western Nigeria (Figure 4.1). Kano is the second largest city, with a population of approximately 4.4 million residents living within administrative area of ~449 km² (National Population Commission & National Bureau of Statistics, 2023). “Kano metropolis” is used to denote the contiguous urban area spanning eight Local Government Areas (LGAs): Dala, Fagge, Gwale, Kano Municipal,

Kumbotso, Nassarawa, Tarauni, and Ungogo, each subdivided into numerous neighbourhoods.

The population is predominantly Hausa-Fulani and Muslim, and family and community norms influence healthcare trajectories in ways relevant to peripartum and infant care. For example, decisions about facility delivery or emergency obstetric procedures may involve husbands or senior male relatives, and many women prefer female providers for intimate care when available; patterns reported across northern Nigeria and pertinent to this referral hospital setting (Abubakar et al., 2020; Audu et al., 2014; Sharma et al., 2019).

Family structure also shapes exposure environments. In this context, polygynous households and extended-family compounds are common (National Population Commission (NPC) & ICF, 2019; Nigerian National Bureau of Statistics & World Bank Living Standards Measurement Study, 2024) and shape day-to-day microbial exposures. Larger, multi-nuclear households mean more close contacts (adult wives, co-wives' children, grandparents, cousins), shared sleeping spaces, and shared WASH facilities (National Population Commission (NPC) & ICF, 2019; Nigerian National Bureau of Statistics & World Bank Living Standards Measurement Study, 2024), which can increase opportunities for respiratory and enteric pathogen transmission while also broadening the infant's microbial contact network (Shannon et al., 2018). Caregiving is often distributed across multiple adults and older siblings, increasing hand-offs and feeding by several carers; this may diversify microbial sources (hands, utensils, complementary foods) and the chances of antibiotic sharing or informal advice-driven use (Lane et al., 2019; Wiley et al., 2024). Compound layouts may place food preparation, animal pens (poultry/small ruminants), water storage, and latrines in close proximity, with variable hand-hygiene infrastructure, influencing both environmental contamination and the resistome via contact with soil, surfaces, and faecal

matter (Mills et al., 2023; Pickering et al., 2019). At the same time, extended kin networks provide social support that can facilitate rapid care-seeking, postpartum help, and adherence to treatment, partially offsetting risk by improving supervision and access to services.

Economically, Kano is a long-standing commercial hub, known for trade in the livestock, groundnuts, textiles, and manufacturing areas. Household conditions are heterogeneous: surveys from low-income settlements in Lagos and Kano and recent demographic and health surveys for the North-West zone indicate high poverty prevalence alongside constrained access to improved water and sanitation (Chiripanhura & Niño-Zarazúa, 2016). In this environment, less than half of households in the North-West have adequate water and sanitation, sustaining elevated enteric risks (National Population Commission (NPC) & ICF, 2019). Kano's Open-Defecation-Free (ODF) campaign has expanded coverage in multiple LGAs, but safely managed services and reliable water remain uneven, particularly in dense urban wards (see Section 4.1 for national/state WASH context and references).

Infant feeding practices reflect these constraints. Extended breastfeeding is common, whereas exclusive breastfeeding (EBF) (breast milk only, no water or other liquids/foods (WHO & UNICEF, 2024)) under six months remains below ~30% in recent regional surveys (National Population Commission (NPC) & ICF, 2019; WHO & UNICEF, 2024). Complementary feeding often begins early, with reliance on locally prepared pap (fermented maize porridge). In households facing intermittent water or electricity, pap may be prepared with stored water, mixed by hand, and held at ambient temperature; limited refrigeration and sub-optimal storage containers can increase microbial loads. These pathways are directly relevant when interpreting early-life gut colonisation and resistome exposure.

4.2.2 Enrolment hub: Murtala Muhammed Specialist Hospital, Kano

Given these socioeconomic and cultural conditions, the enrolment facility is central to understanding maternal and infant exposures. Recruitment was conducted at MMSH, a high-volume state referral tertiary-level hospital in Kano that provides subsidised maternity services under Kano state policies, serving a broad catchment that includes many low-income families (Milton et al., 2021; Milton, Modibbo, et al., 2022).

MMSH is a government-owned tertiary healthcare facility managed by the Kano State Hospitals Management Board. It was selected as the enrolment hub due to its established collaboration with the BARNARDS I study (2016), a prior neonatal infection surveillance and microbiology capacity-strengthening programme conducted at MMSH, which provided continuity for research delivery and leveraged existing clinical and laboratory partnerships (Akpulu, Yakubu, et al., 2025). This study was implemented as a nested cohort within the second phase of the BARNARDS programme at MMSH, leveraging its established recruitment pathways, sample-handling workflows, and clinical metadata systems. The hospital maternity complex includes two main wards: the labour ward (for vaginal deliveries) and the postnatal ward (for mothers following caesarean section), each with a capacity of 24 beds (Figure 4.2). With a high occupancy rate managing about 650 deliveries per month, the facility operates under significant pressure where demand for neonatal care frequently exceeds capacity.

Although the special care baby unit (SCBU) has benefited from periodic equipment upgrades, including the 2016 incubator and resuscitator donations, recurrent challenges such as staffing constraints, infrastructure wear, and intermittent supply stock-outs persist. These systemic pressures directly influence clinical practice; for instance, empiric antibiotic regimens are widely used for suspected neonatal sepsis (Ogunkunle et al., 2024), with local

prescribing choices at MMSH shaped by both emerging susceptibility data and patient affordability (Thomson et al., 2021).

Within Kano's wider healthcare landscape, MMSH's role is distinct from that of the federally funded Aminu Kano Teaching Hospital (AKTH). While AKTH primarily manages complex tertiary referrals and medical training (Aminu Kano Teaching Hospital (corporate author), 2025), MMSH serves as a critical state-owned hub with a subsidised fee structure. This makes it a primary point of care for a broad, metropolitan, and largely low-income catchment. Consequently, the MMSH cohort represents a high-density clinical environment characterised by high patient turnover and frequent antimicrobial exposure, factors that are central to the early-life resistome dynamics investigated in this study.

(A) Map of Nigeria highlighting Kano State (red). (B) Map of Kano State showing the location of Kano City within the state (boxed area; inset). (C) Kano metropolitan area showing the study catchment across selected local government areas; numbered markers indicate study recruitment/collection locations. The blue location marker is the enrolment site, MMSH. Participants were recruited from across Kano metropolitan LGAs and peri-urban areas; for example, Gezawa (the farthest LGA) lies approximately 31 km (straight-line) from MMSH.



Figure 4.2: Clinical environment at Murtala Muhammed Specialist Hospital (MMSH), Kano.

(A) Postnatal ward showing shared inpatient bays and bed capacity. (B) Maternity/postnatal cubicles illustrating the open, high-throughput ward layout used for routine care during recruitment. (C) Example bedside/cubicle space typical of the inpatient setting in which sampling and clinical data collection were performed.

4.3. Study implementation

This section describes the implementation of the cohort, including recruitment, retention, data and sample collection, and participant withdrawals. While methodological procedures are detailed in Chapter 2 (Materials and Methods), the focus here is on operational outcomes, completeness of follow-up, and practical considerations from the longitudinal fieldwork in Kano.

4.3.1 Operational objectives

I pre-specified the following objectives for the first 90 days post-delivery:

- Enrolment: consecutively enrol eligible term mother-infant dyads at MMSH within 24h of delivery from February to September 2024.
- Retention: achieve $\geq 85\%$ completion of the Day 90 visit among enrolled dyads, with reasons for attrition recorded.
- Sample yield: collect ≥ 2 infant stool samples within 90 days at prespecified timepoints (Chapter 2); record laboratory receipt and library generation.
- Exposure capture: measure delivery mode, maternal/infant antibiotic exposure, feeding pattern, and household/WASH proxies with high item completeness ($\geq 90\%$ baseline; $\geq 80\%$ Day 90).

4.3.2 Recruitment

Recruitment took place at MMSH maternity wards between 1 February and 30 September 2024. Eligible full-term mother-infant dyads were identified within 24 hours of delivery.

Inclusion and exclusion criteria, as well as consent procedures, are described in Chapter 2 (Sections 2.3.1 and 2.3.2).

Mothers were approached at their bedside, provided with information sheets in non-technical language, and given time to decide on participation. To enhance awareness, Ethics Committee-approved recruitment posters were displayed in the maternity wards (Figure 4.3). The poster summarised eligibility, study commitments, and the benefits of participation in simple, non-technical language to support informed decision-making.

During the enrolment window there were 3,287 deliveries at MMSH. We enrolled 90 mother-infant dyads (2.7% of deliveries). Of these, 87/90 (96.7%) provided at least one infant stool sample at baseline (Figure 4.5).

linacres Oxford Institute for Antimicrobial Research (IOI)
Sir William Dunn School of Pathology,
University of Oxford
Oxford, OX1 3RE

OxTREC approval reference: 36-22
Expiry date: 31/12/2025



The INBUGS-NG Study

The Infant Gut Bacterial Study in Nigeria



Gut bacteria live in our tummy and provide some important health benefits. However, some gut bacteria are harmful, and resistant to antibiotics.

This study aims to understand the development of your baby's gut bacteria by looking at some factors such as your baby's diet, use of antibiotics, and more.

Who can participate in the Study?

- Pregnant women ready to deliver in MMSH hospital, and their newborn baby(s)
- You are not vegetarian or vegan.
- You and your family live in Kano City.

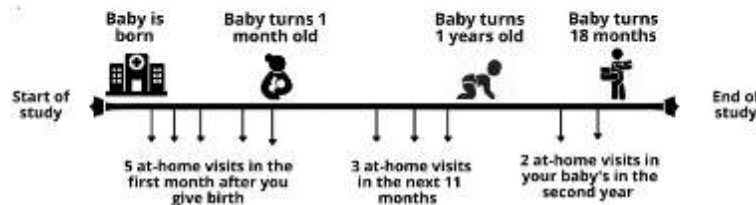
What information will we collect?

- Demographic data
- Information of your baby's birth
- Samples collected after your baby is born

What samples will we collect?



How long will you be in the study?



What will the study provide you?

- We will give you a **baby hamper** at the point of enrolment and at every at-home visit until your baby is two years of age.
- We will do a **free blood culture** when required if your baby is admitted to the hospital.
- We will give you a **phone credit** to communicate with the research team between follow-up visits.

Please email Chinenye Akpulu (chinenye.akpulu@linacre.ox.ac.uk) if you have any questions about the study.

Figure 4.3 Participant recruitment poster used in MMSH maternity wards for the Study.

The leaflet summarises the study aim (infant gut microbiome and antimicrobial resistance), eligibility criteria, the types of data and biological/environmental samples collected (infant stool; maternal breast milk and rectal swab; environmental samples), the follow-up schedule (home visits from birth to 18 months), and participant support provided (baby hamper, phone credit, and blood culture when clinically indicated), with IOI/OxTREC ethics reference details and study contact information shown (we have signed consent for use of photograph for research purposes).

4.3.3 Data collection

Data collection used a combination of structured electronic questionnaires and complementary non-electronic tools to strengthen completeness and contextual detail. Among mothers who provided at least one biological sample, questionnaire completion was 97% at baseline and 88% at Day 90: patterns of missing data broadly mirrored stool-sample availability (Table 4.1). In parallel, the team maintained notebooks to document contextual observations (e.g., uvulectomy, female genital cutting, traditional remedies). These notes were recorded opportunistically rather than systematically and are used to aid interpretation only; they were not analysed inferentially.

4.3.4 Biological sample types

4.3.4.1 Infant stool samples

Stool was the primary specimen, targeted at eight scheduled timepoints (D0, D1, D3, D5, D7, D14, D28, D90). Of the 90 dyads, 87 contributed at least one stool sample, with retention at Day 90 of 79 dyads (88%). Completeness across timepoints ranged from 32% at Day 0 to 81% at Day 14 (Table 4.1). The lower completeness at Day 0 reflects the practical and biological constraints of immediate post-delivery sampling, particularly that meconium was not always passed within the Day-0 collection. For the 26 infants who received antibiotics, additional stool samples were collected in predefined post-exposure windows to characterise short-term dynamics (Figure 4.4).

Table 4.1 Stool sample collection across scheduled timepoints (N = 90)

Timepoint	Expected Samples (N)	Collected Samples (N)	Percent Completion
			(%)
Day 0	90	29	32
Day 1	90	51	57
Day 3	90	58	64
Day 5	90	67	74
Day 7	90	72	80
Day 14	90	73	81
Day 28	90	67	74
Day 90	90	63	70

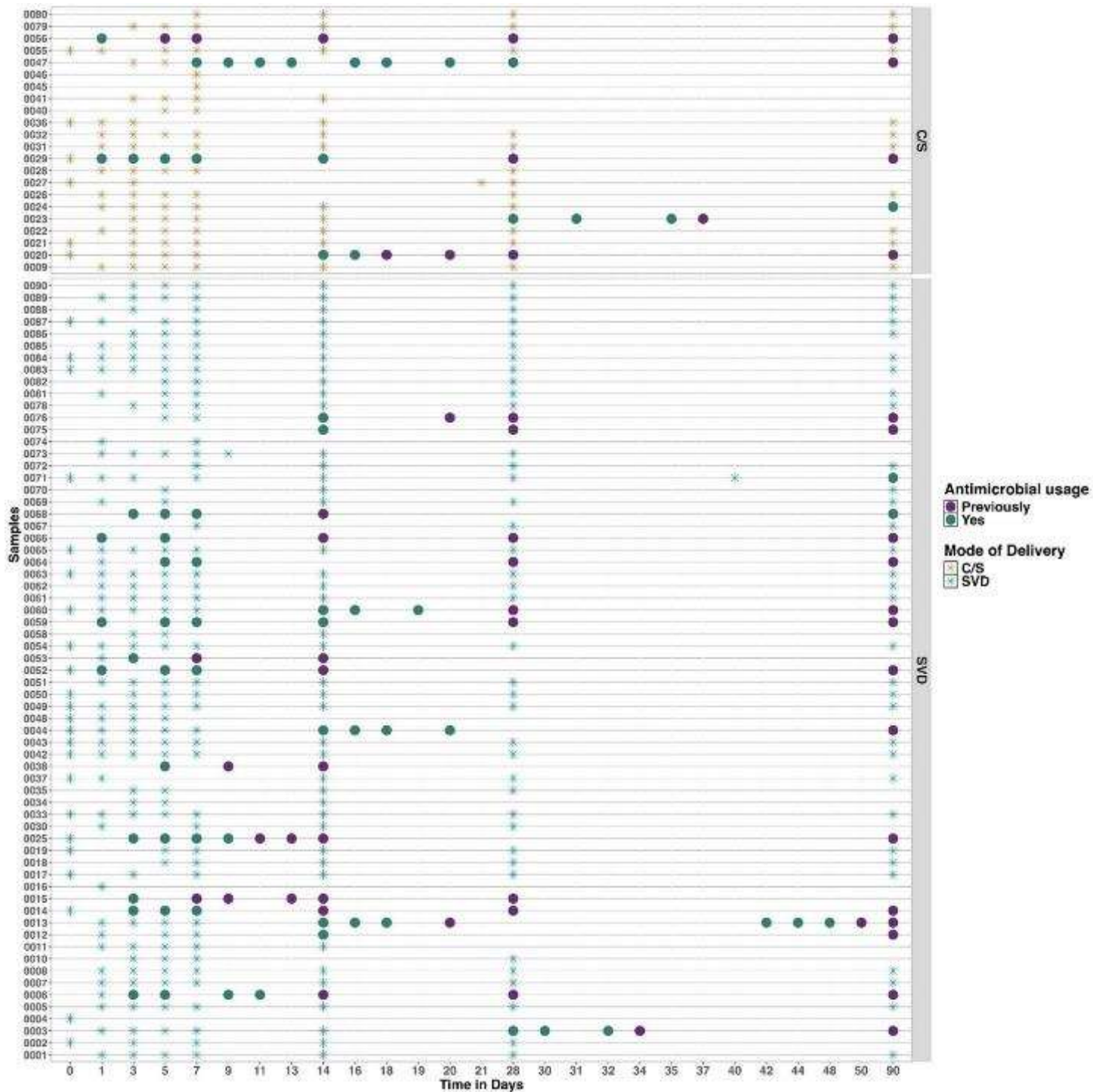


Figure 4.4 Infant stool sampling timeline by delivery mode and antibiotic exposure status

Each row represents one infant, and points are plotted at the day of life when a stool sample was collected. Panels stratify infants by mode of delivery (C/S vs SVD). Asterisks denote samples collected at the scheduled study timepoints (D0, D1, D3, D5, D7, D14, D28, D90). Filled circles indicate samples collected in relation to infant systemic antibiotic exposure, coloured to distinguish samples collected during antibiotic use (“Yes”) versus after prior exposure (“Previously”). This visualises both routine longitudinal sampling and additional post-exposure sampling windows.

4.3.4.2 Maternal rectal swabs

Rectal swabs were collected within 24 hours of delivery, 86/90 (95.6%) completion at Day 0; 71/90 (78.9%) at D3; and 75/90 (83.3%) by D28 (Table 4.2).

4.3.4.3 Breastmilk samples

Breast milk was collected at D0, D3, D28, and D90 from breastfeeding mothers. Completeness was 49/90 (54.4%) at D0, 66/90 (73.3%) at D3, 74/90 (82.2%) at D28, and 65/90 (72.2%) at D90 (Table 4.2). Occasional refusals or insufficient volume reduced yields at later timepoints.

Table 4.2 Breast milk and maternal rectal swab sample yields across scheduled timepoints (N = 90)

Time point	Maternal rectal				
	Breast Milk			swab	
	Expected	collected	% Completion	collected	% Completion
Day 0	90	49	54	86	96
Day 3	90	66	73	71	79
Day 28	90	74	82	75	83
Day 90	90	65	72	NA	NA

4.3.4.5 Environmental samples

Environmental exposure sampling included targeted swabs and water samples in hospital (≤ 48 h) and household settings ($\geq D7$). Table 4.3 reports count only; denominators vary by context and consent. Highest counts were obtained for caregiver phone swabs, household drinking water, and doorknobs.

Overall, sample completeness was high, with stool, breast milk, and maternal rectal swabs all achieving $\geq 70\%$ collection by Day 28, and environmental samples providing rich contextual data from both hospital and household settings.

Table 4.3 Environmental sample yields across scheduled timepoints

Sample type	Day 0	Day 1	Day 3	Day 5	Day 7	Day 14	Day 28
Baby bed rails sample	1	11	1	1	22	1	25
Caregiver mobile phone sample	26	45	4	1	25	2	63
Delivery bed sample	46	2	0	0	0	0	0
Drinking water source sample	1	35	5	1	35	5	74
Hospital bed sample	28	13	2	1	0	0	0
Around sink and tap used for baby wash sample	0	8	4	1	9	1	14
Doorknobs of room home sample	0	33	6	1	35	4	74
Runoff from dumpsite sample	0	28	5	3	26	5	73

Note: Frequencies varied depending on household context, consent, and availability. Day 0 and Day 1 samples were predominantly hospital-based, while later collections occurred during household visits.

4.4 Cohort profile and implementation outcomes

4.4.1 Attrition and cohort retention

Of the 90 enrolled dyads, 3/90 (3.3%) withdrew before providing any samples, leaving 87/90 (96.7%) with baseline data (Figure 4.5). Subsequent losses were one before Day 3, one before Day 7, and five before Day 14; none occurred thereafter. Retention to Day 90 was 79/90 (87.8%). Withdrawals were primarily logistical (permanent relocation, declining further participation, scheduling difficulties). Samples/data collected before withdrawal were retained unless removal was requested (none). Relocations within metropolitan Kano did not end participation, though interim relocations caused missing timepoints. The spatial distribution of enrolled households across Kano metropolis is shown in Figure 4.1.

4.4.2 Household characteristics

Median monthly income was NGN 60,000 (IQR NGN 40,000-100,000). 47.8% of households reported $\leq$ NGN 50,000, and 83.4% $\leq$ NGN 100,000 (Table 4.6). Household size ranged 3-29 persons (mean 6, SD 4) many with multiple children <14 years (mean 4), consistent with extended family structures and dense contact networks.

Environmental conditions were frequently adverse: 79/87 (90%) households were within $\leq$ 20m of a dumpsite or persistent stagnant water, and most of the households discharged greywater openly. Figure 4.4 illustrates typical scenes (open drains, stagnant wastewater, refuse). Electricity was unreliable (78.4% occasional supply; 19.8% generator backup). Cooking fuels included three-stone/open fires (36.2%), charcoal (18.1%), and coal (7.8%).

Table 4.4: Distribution of Household Monthly Income (N = 90)

Income Category (NGN)	Approx. USD/month*	N (%)
< 20,000	<14	9 (10%)
20,001-50,000	14-26	34 (38%)
50,001-100,000	26-34	32 (36%)
100,001–200,000	34-138	12 (13%)
> 200,000	>138	3 (3%)

**Approximate conversions shown for interpretability, calculated using exchange rates on 1 July 2023 (USD 1 = NGN 1,451.09).*

4.4.3. Maternal characteristics

Baseline characteristics are summarised in Table 4.5 (baseline N=87 unless noted). Maternal age ranged 17-42 years (mean 28.1, SD 5.9). Parity distribution is shown in Table 4.6; 47% reported 1-2 prior pregnancies, and 7% reported ≥ 9 . Parity increased with age: mothers reporting 5-8 pregnancies had a mean age of 31.7 years, while those with ≥ 9 pregnancies (7%) had a mean age of 37.2 years.

Formal education was reported by 70% (higher 20%, secondary 41%, primary 10%); 13% reported no formal education. Health insurance coverage was low (no insurance 79/87, 91%), consistent with out-of-pocket financing (Section 4.1). Female genital mutilation (FGM) responses were yes 16%, no 15%, and missing 70%. Given the emotive and sensitive

nature of the topic and the high non-response, these data are subject to self-report and social-desirability bias and do not support prevalence estimation in this cohort. The figures are therefore not treated as prevalence estimates but reported only descriptively with no inferential analyses.

Table 4.5: Baseline Characteristics of Mothers and Infants (N = 87)

Variable	Summary(N=87)
Maternal Age(years)	28.13[6.05]
Previous Pregnancies	3.24[2.55]
Monthly Household Income*	60000[40000-100000]
Gestation Age(weeks)	38.48[2]
Birth Weight(grams)	2998.85[352.23]
APGAR Score***	8.8[1.17]
Sex: Female**	43(49.4%)
Sex: Male**	44(50.6%)
Delivery: Caesarean Section**	22(25.3%)
Delivery: Vaginal**	65(74.7%)
Insurance: No Insurance**	79(90.8%)
Insurance: Government Insurance**	6(6.9%)
Insurance: Private Insurance**	2(2.3%)

Summary statistics are expressed as mean [SD], median (IQR)*, and number [%] **. *** Three babies with missing APGAR score

Table 4.6 Maternal parity distribution and age characteristics in the study cohort (n = 90)

Previous Parity^a group	N	Mean age	SD age	Min age	Max age
0	5	20.2	3	18	24
1-2	42	24.9	4.2	17	35
3-4	16	30.3	5.1	23	37
5-8	21	31.7	4.6	20	37
9+	6	37.2	3.2	32	40

^a Parity refers to the number of previous births reaching viable gestational age, excluding the current neonate

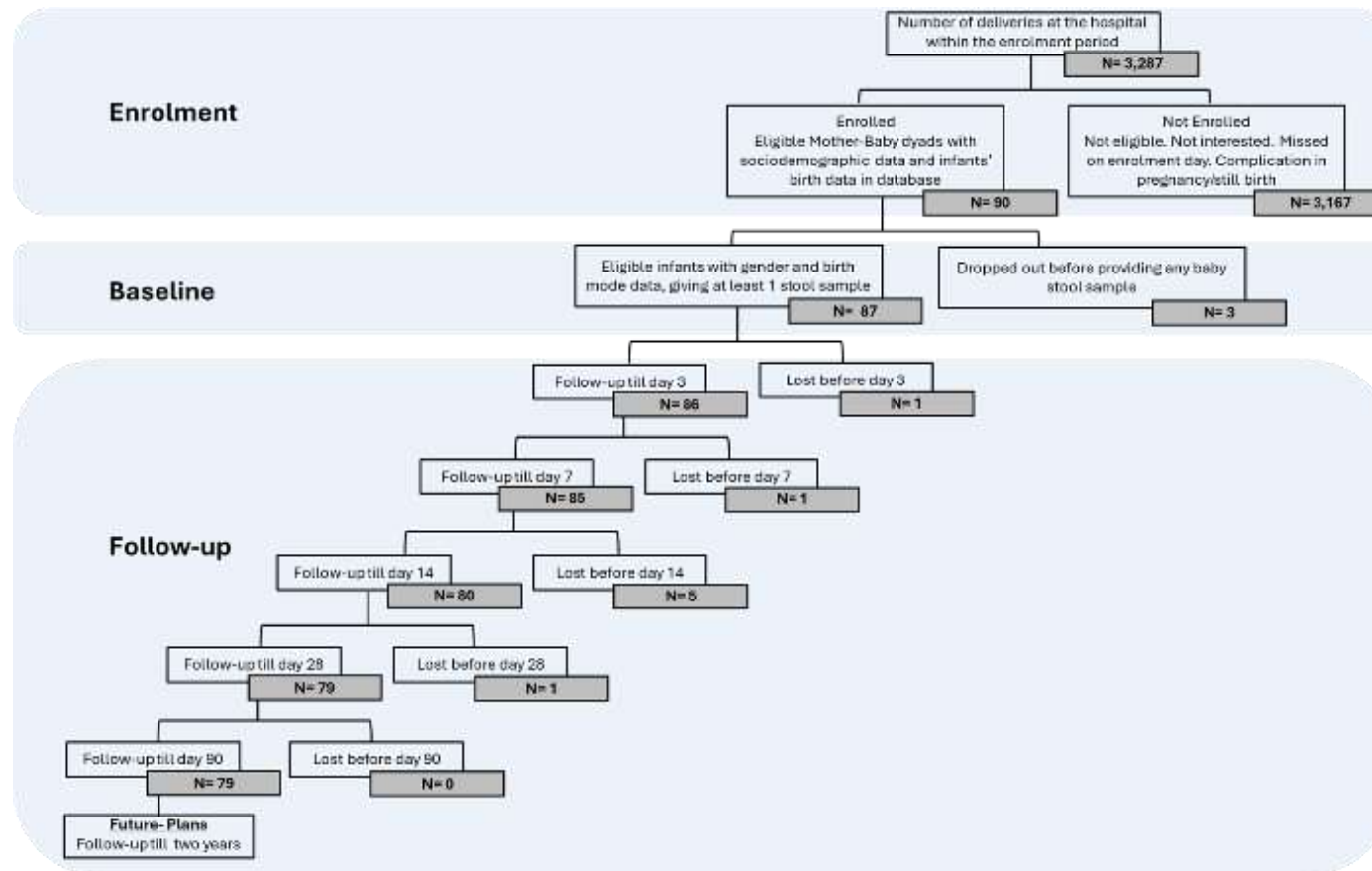


Figure 4.5 Participant enrolment, attrition and follow-up flow diagram for the study cohort

Flow chart summarising the number of deliveries at MMSH during the enrolment period (N = 3,287), numbers enrolled (N = 90) versus not enrolled (N = 3,167; not eligible, not interested, missed on non-enrolment days, or complicated pregnancy/stillbirth), and retention across infant follow-up. Of the enrolled dyads, 87 infants contributed baseline metadata and at least one stool sample, while 3 dyads withdrew before providing any infant stool specimen. Follow-up retention was 86 infants at Day 3 (1 lost), 85 at Day 7 (1 additional lost), and 80 at Day 14 (5 additional lost), with 79 retained through Day 28 and Day 90.



Figure 4.6 Typical environment and neighbourhood sanitation conditions in the study community

Representative photographs from participant communities illustrating environmental exposures relevant to infant enteric microbial (A) household frontage with visible solid waste accumulation and stagnant wastewater near the entrance; (B) nearby open waste dumping/accumulated refuse within the residential environment; (C) open drainage channel with stagnant/flowing wastewater adjacent to households. Images are included to contextualise the hygiene and sanitation setting in which enrolled infants lived and were followed up.

4.4.4 Birth characteristics

The average gestational age at birth was 38.48 weeks (SD 2.0), with a mean birth weight of 2,998.85 g (SD 352.23). As preterm births were excluded by design, all deliveries occurred at or beyond 37 weeks.

The APGAR score (Appearance, Pulse, Grimace, Activity, Respiration) assesses newborn condition at 1 and 5 minutes on a 0-10 scale; scores ≥ 7 at 5 minutes are generally reassuring, 4-6 indicate moderate abnormal, and 0-3 as low or non-specific sign of illness (Committee Opinion No. 644, 2015). The majority of neonates ($n = 63$, 73%) had APGAR scores between 9 and 10 at five minutes. A smaller subset ($n = 4$, 19%) scored between 7 and 8, while one neonate had a score of 4, following a 7-day interval between Premature Rupture of Membranes (PROM) and delivery. Among the 21 PROM cases, the time from membrane rupture to delivery ranged from 5 minutes to 7 days. Most delivered within 24 hours ($n = 17$, 81%), while 4 cases (19%) were classified as prolonged PROM (>24 hours). All PROM-related births were vaginal.

Newborn sex distribution was approximately equal, with 43/87 (49%) female and male 44/87 (51%) male. Vaginal delivery was the most common mode, accounting for 65/87 (75%) of births. Of these, 63 were spontaneous, one was induced, and one was assisted with forceps. C-section was performed in 22/87 (25%) cases, of which 15/22 (68%) were emergency procedures and 7/22 (32%) were planned operations.

4.4.5 Antibiotic use

4.4.5.1 Maternal antibiotic use

Antibiotic use among mothers was documented both during labour and postpartum, with postpartum prescribing being more widespread. Intrapartum antibiotics were administered to 7/90 mothers (8%), all via intravenous injection. The most common antibiotics were metronidazole only, or in combination with ceftriaxone and/or gentamicin. In seven of these cases, delivery occurred via caesarean section. Indications were not routinely documented, but the pattern and timing of administration suggest standard perioperative prophylaxis.

Postpartum antibiotics were prescribed to 83/90 mothers (92%) at hospital discharge. Among these, 79 received dual antibiotic therapy, and five were given three or more antibiotics concurrently (Figure 4.7A). Of the dual therapies, 74 mothers (94%) were prescribed a combination of amoxicillin-clavulanate and metronidazole, reflecting a commonly used empirical regimen for postpartum care. A smaller number received combinations such as metronidazole with amoxicillin (due to cost) or metronidazole with ceftriaxone (inpatients). Most of the antibiotics were prescribed for out-of-pocket purchase. 6/87 mothers (7%) who did not take postpartum antibiotics reported they were unable to afford the medications.

Traditional medicine use was also common within the cohort. A total of 37/87 mothers (43%) reported occasional use of *Agbo*, a local multi-herbal preparation traditionally consumed to support postpartum recovery, among other health purposes. Additionally, 7/87 mothers (8%) reported consuming bitter kola, another traditional plant-based remedy. These traditional treatments were typically taken alongside conventional antibiotic regimens.

4.4.5.2 *Infant antibiotic use*

By three months, 26/87 (30%) infants received at least one antibiotic course during follow-up. Most exposed infants (20/26; 77%) received a single course. Amoxicillin-based formulations (including amoxicillin, amoxicillin trihydrate, and amoxicillin–clavulanate) were the most frequently reported agents. Dual therapy was reported for 7/26 (27%), typically involving amoxicillin combined with flucloxacillin or other β -lactams (e.g., ampicillin or cloxacillin). Triple therapy occurred in 2/26 (8%), including combinations such as amikacin + amoxicillin + cefuroxime. Less commonly reported agents included ceftazidime, cefixime, ofloxacin, and metronidazole (Figure 4.7B).

Antibiotic exposure, the indication for use, and source of access (prescribed vs over-the-counter) were determined by caregiver report supported by inspection of available antibiotic packets. Overall, most antibiotic use was reported as prescribed (21/26; 81%), while a minority was obtained over the counter without a prescription (5/26; 19%). Caregiver-reported reasons for antibiotic use were most commonly fever-related illness (13/26; 50%), followed by neonatal complications recorded as birth asphyxia $\pm$ hypocalcaemia (3/26; 12%) and rash/skin symptoms (3/26; 12%), with smaller numbers reported for respiratory symptoms (“cough/catarrh”), gastrointestinal symptoms, and suspected infection (including “EOS” as recorded). Notably, the primary reason for the non-prescribed cases was not different from the prescribed. These reasons reflect caregiver descriptions and available documentation and may not correspond to clinically confirmed diagnoses.

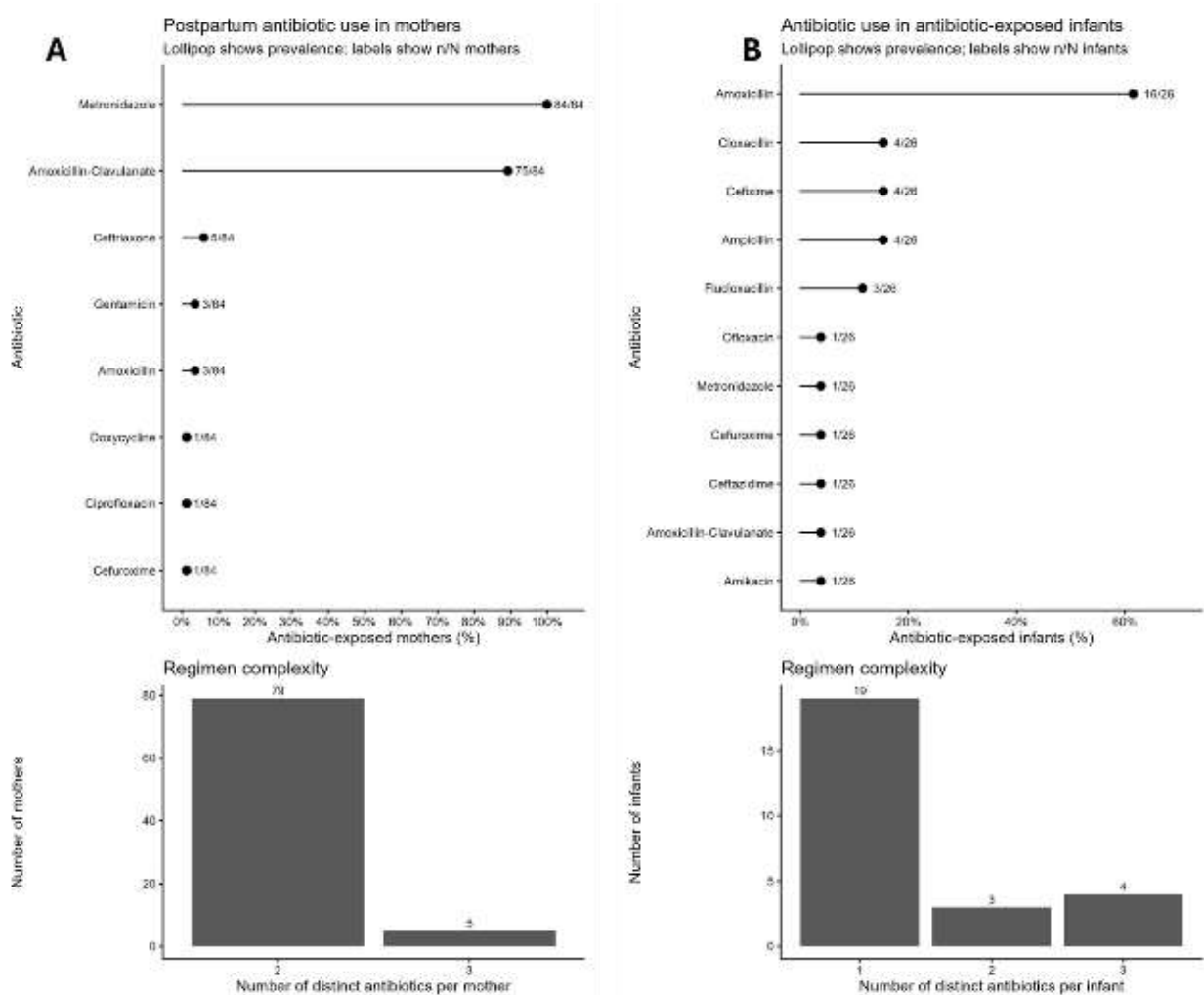


Figure 4.7. Antibiotic use profiles and regimen complexity in mothers and infants.

(A) Postpartum antibiotic use among mothers ($n = 84$). Lollipop plots show the proportion of antibiotic-exposed mothers receiving each antibiotic; point labels indicate the number of mothers exposed (n/N). The lower panel summarises regimen complexity as the distribution of the number of distinct antibiotics recorded per mother. (B) Antibiotic use among antibiotic-exposed infants ($n = 26$; exposure defined as any antibiotic recorded up to day 90). Lollipop plots show the proportion of antibiotic-exposed infants receiving each antibiotic (labels show n/N), and the lower panel summarises the distribution of the number of distinct antibiotics recorded per infant.

4.4.6 Infant Feeding

Infant feeding practices were dominated throughout follow-up by high breastfeeding with water supplementation (High_BF; breastmilk as the primary feed plus water) (Figure 4.8).

Among infants with feeding data at Day 90 ($n = 63$), High_BF remained the predominant category, with exclusive breastfeeding (EBF) consistently forming a smaller but stable minority. At the earlier scheduled visits, EBF accounted for a modest proportion of infants, while formula-only and mixed breastmilk–formula feeding (BF_F) were uncommon and largely confined to the early postnatal period. By Day 28 and Day 90, solid-food exposure (BF_SF) and water-only feeding were rare. Overall, Figure 4.8 indicates that the cohort’s “breastfeeding” experience is best characterised as breastfeeding with early and persistent water supplementation.

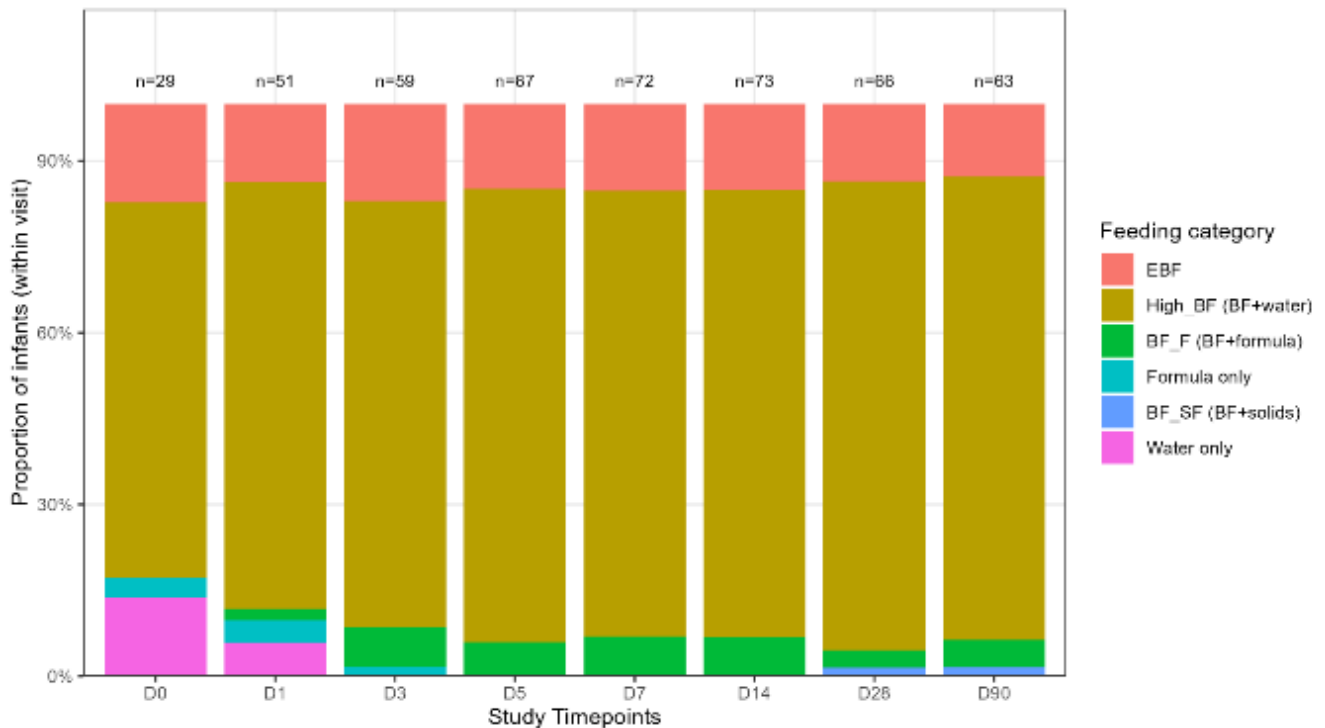


Figure 4.8: Infant feeding categories across scheduled study timepoints (D0-D90)

Stacked bars show the proportion of infants in each feeding category at each visit: exclusive breastfeeding (EBF), breastfeeding with water supplementation (High_BF; BF+water), breastfeeding plus formula (BF_F), breastfeeding plus solids (BF_SF), formula only, and water only. Numbers above bars indicate the number of infants with recorded feeding data at each timepoint.

4.4.7 Traditional and cultural practices

Several traditional and cultural health-related practices were reported among infants in the cohort. Uvulectomy, the removal of the uvula, was observed in 26% ($n = 23$) of infants, with a higher prevalence among males (15/47, 32%) compared to females (8/43, 19%). FGM was reported in 33% ($n = 14$) of female infants, representing 16% of the total cohort. The use of Qur'anic water was reported for 38% ($n = 34$) of infants, with a similar prevalence in males (19/47, 40%) and females (15/43, 35%). These practices may influence infant health and early microbial exposures through oral or mucosal routes, and their potential impact on microbiome composition warrants consideration in subsequent analyses.

4.4.8 Biological samples

Details of biological sample collection, including yields for infant stool, maternal rectal swabs, breast milk, and environmental samples, are presented in Section 4.3.4 (Tables 4.1-4.3). By Day 28, $\geq 70\%$ completeness was achieved for stool, breast milk, and maternal rectal swabs, with longitudinal stool sampling capturing both scheduled and antibiotic-associated timepoints. These environmental samples provided valuable context for hospital and household exposures, with caregiver phone swabs, drinking water, and doorknobs yielding the highest returns. All stool/breastmilk samples were pre-processed in the MMSH container laboratory; DNA extractions were stored at $-20\text{ }^{\circ}\text{C}$ (median ~ 3 weeks) and shipped to Oxford for shotgun metagenomics (Chapter 2). Figure 4.9 illustrates the laboratory setup and workflow.



Figure 4.9: On-site container laboratory infrastructure at MMSH, Kano.

Photographs showing the field laboratory used for sample processing and interim storage during the study: (A) exterior view of the container-based laboratory located within the hospital; (B) interior workspace illustrating bench layout and core equipment used for routine molecular workflows. This facility enabled near-real-time processing under site constraints (space, power stability, and cold-chain capacity).

4.5 Discussion

Evidence from clinical studies and animal models indicates that early-life disruptions of the gut microbiota are associated with later-life health outcomes (Cox et al., 2014; Olin et al., 2018; Tanaka & Nakayama, 2017). This chapter profiles the Northern Nigerian mother-infant cohort of this study, with follow-up to 90 days postpartum. The data describe maternal and infant demographics, household and environmental exposures, perinatal events, antibiotic use, feeding practices, traditional health-related practice and completeness of biological sample collection. The longitudinal design enables investigation of temporal changes and early-life exposures that are critical in shaping microbial colonisation and resistance patterns. The decision for a longitudinal study design was necessary to capture the dynamics of the developing gut microbiome over time as it evolves in early life, and to allow sample collection post antibiotic exposure. While the design offered the strength of temporal investigation of the study cohort, it also introduced study attrition, need for consistent follow-up, and the influence of confounders that may be missed over the course of the study. The rigorous data collection protocol including repeated home visits and biological sample collections was implemented to ensure high-quality longitudinal data. The multiple study follow-ups, and frequent sample and data collection did on occasion however raise concerns among community members (see section 4.60 above).

To our knowledge, this is the first longitudinal study in Nigeria to systematically investigate factors influencing the composition and temporal dynamics of the infant gut microbiota and resistome over the first 90 days of life. A distinctive feature of this cohort environment is the high potential for environmental exposure: approximately 90% of households resided within 20 m of a refuse dump or stagnant water. Such proximity plausibly increases contact with faecal and environmental microbes through contaminated water and soil and through

mechanical vectors (e.g., flies), creating multiple potential pathways relevant to AMR persistence and ARG carriage.

A recent multi-city hospital study in Nigeria (including Kano) collected 1,396 flies and found diverse bacteria with clinically relevant AMR genes on flies across all sites, suggesting the vector role of flies in healthcare environments (Cook et al., 2025). Parallel wastewater and pharmaceutical-effluent studies both locally and globally also report ARGs and resistant bacteria in receiving waters, showing that community ARG signatures track sanitation and socio-environmental context (Hendriksen et al., 2019; Munk et al., 2022; Obayiuwana & Ibekwe, 2020). Additional, metagenomic studies show that human and environmental resistomes are interconnected, with exchange across habitats (Pehrsson et al., 2016). Consistent with these pathways, several studies report higher ESBL-Enterobacteriaceae carriage in children living in communities with animal-waste or wastewater exposure (Amato et al., 2023).

Households in this cohort were large (range 3-29 members) and often multigenerational, with a high proportion of members under 14 years. Such structures increase close contact microbial sharing and have been associated with shifts in infant gut microbiota (Christensen et al., 2022; Laursen et al., 2015). Christensen *et al.* (2022) showed siblings to be major early-life determinants of airway and gut microbiota, with pronounced diversity shifts in the first year (Christensen et al., 2022). They found a significant shifts in alpha (within-sample diversity) and beta diversity (between-sample compositional differences), with the largest differences in the first year of life. In crowded homes with limited sanitation and constrained healthcare access, infectious illnesses can spread rapidly, potentially increasing, and sometimes inappropriately directing, antibiotic use, including non-prescription access documented in many LMIC settings (Auta et al., 2019). Reliance on solid fuels for cooking further contributes to indoor air pollution and respiratory morbidity, which may indirectly

influence antibiotic-use patterns(Mukesh Dherani et al., 2008). Unlike many HIC cohorts with engineered waste management and piped water, this cohort reflects common LMIC realities in which infants face higher baseline environmental microbial exposures from birth.

Beyond infrastructural exposures, several culturally embedded practices reported in this cohort may also be microbiologically relevant, notably uvulectomy and ingestion of “Qur’anic water”. Uvulectomy in this setting is typically performed by traditional practitioners rather than medical personnel, often outside aseptic conditions and sometimes with shared or non-sterile instruments(A. Adamu et al., 2021; Adebola et al., 2016; Musa, 2023). Aside from mucosal injury, these circumstances increase the risk of introducing exogenous oral microbiota and of post-procedure antibiotics, each of which could transiently alter oral-gut seeding and short-term selection of resistance determinants. The “Qur’anic water” practice usually involves rinsing Qur’anic verses from a wooden slate by Mallams(Brigaglia, 2017), locally known as *ruwan rubutu*, and fed to the baby over several days (Simane-Netshisaulu et al., 2022) as a source of blessing and protection. Nevertheless, water quality or source of water is usually uncertain, or containers not hygienically maintained, this adds additional contact points through which environmental microbes may enter infant feeds. I interpret these practices as contextual exposures that could shape early colonisation and resistome trajectories in this setting; I did not assay these liquids or procedures directly and therefore avoid causal claims.

As summarised in Table 4.5, mothers were relatively healthy at enrolment by design, with ages spanning the typical reproductive years and heterogeneous parity. Education levels were moderate overall. In many sub-Saharan African settings, higher maternal education is associated with more timely care-seeking for childhood illness and better engagement with antenatal/postnatal services, suggesting that maternal education may act as a proxy for health-seeking behaviour in these analyses (Tesema & Seifu, 2023). Similarly, health-

insurance coverage was low in this cohort. In contexts where health insurance is limited and out-of-pocket payments dominate healthcare financing, as the case in the Nigeria healthcare system, cost can delay or deter formal care, shaping where, when and which antibiotics are obtained (Awoyemi et al., 2023; World Bank, 2024). Non-prescription access to antibiotics is common across many LMICs and is repeatedly linked to inappropriate use, an important background consideration for resistome work (Auta et al., 2019).

All infants in the cohort were born at term with birthweights in the expected range and generally reassuring 5-minute Apgar scores. This term-only design limits confounding from prematurity and NICU-related care, exposures known to enrich the early-life gut resistome and shift community assembly patterns relative to healthy term infants (Gasparrini et al., 2019). Restricting to term births yields a more homogeneous baseline physiology (gut maturation, feeding readiness, growth rate) and largely avoids NICU-related exposures (ventilation, parenteral nutrition, prolonged empiric broad-spectrum antibiotics). This increases internal validity and focuses specifically on non-prematurity drivers of early microbial assembly, delivery mode (and intrapartum antibiotics), postpartum maternal and infant antibiotics (timing/class), feeding category (WHO exclusive vs predominant), and household/WASH context, without conflation by prematurity care pathways. It also enables comparison of standard early-term trajectories against deviations (e.g., post-antibiotic “shocks”), and improves interpretability of effect sizes (e.g., C-section signal) that would otherwise be reduced or amplified in a mixed term-preterm cohorts. Finally, a term-only frame supports clearer interaction analysis (e.g., antibiotics $\times$ feeding; antibiotics $\times$ WASH proxies) because baseline risk and exposure timing are less heterogeneous. Delivery mode remains a key contextual modulator (Bäckhed et al., 2015; Dominguez-Bello et al., 2010; Reyman et al., 2019; Shao et al., 2019); with 25 C-section births in this cohort, mode of

delivery is treated as a main effect and tested interactions with intrapartum/infant antibiotics and feeding to assess effect modification.

A notable minority of mothers experienced PROM, with a smaller subset meeting criteria for prolonged PROM. Prolonged ROM is a recognised risk factor for early-onset neonatal infection, and its presence frequently alters intrapartum management (e.g., empirical maternal antibiotics) and postnatal observation or treatment thresholds, pathways that can increase early antibiotic exposure for mothers and/or newborns (Al-lawama et al., 2019; National Institute for Health and Care Excellence (NICE), 2021). Because intrapartum antibiotic prophylaxis (IAP) and very early neonatal antibiotics have documented associations with reduced *Bacteroidetes/Actinobacteria*, increased *Proteobacteria*, and higher prevalence of selected β -lactamase genes in full-term infants (Arbolea et al., 2022; Nogacka et al., 2018), PROM-related care might be relevant to interpreting early microbiome and resistome trajectories in this cohort.

Documented intrapartum exposure was relatively uncommon in this cohort; postpartum prescribing was very common, typically dual therapy (combination of amoxicillin-clavulanate and metronidazole), purchased out-of-pocket. From a microbiome/ARG perspective, peripartum antibiotics can influence infant trajectories through several pathways: (i) intrapartum prophylaxis alters early colonisation (lower *Bacteroidetes/Actinobacteria* and expansion of *Proteobacteria* in term infants (Arbolea et al., 2022; Nogacka et al., 2018)); and (ii) maternal postpartum courses may modulate maternal gut and milk-mediated exposures (some antibiotics e.g., amoxicillin, metronidazole are detectable in human milk at low levels (Bethesda (MD): National Institute of Child Health and Human Development; 2006-. Metronidazole, 2024; Thomas et al., 2023)).

Antibiotic stewardship and access are two sides of the same coin. In a predominantly out-of-pocket system, easy OTC access can facilitate timely treatment but also appropriate indication, spectrum and duration, including occasional use of restricted antibiotics in infants (e.g., fluoroquinolones). In this cohort, antibiotic use sits at the intersection of access (timely, affordable treatment) and stewardship (appropriate indication, spectrum, and duration). With low insurance coverage and high out-of-pocket purchase, care pathways bifurcate: some courses were prescribed in the hospital; and others purchased OTC from pharmacies or patent-medicine vendors. This financing reality shapes what is bought (cheaper/substandard substitutes when amoxicillin-clavulanate is unaffordable or out of stock), how much (shortened courses), and how (oral self-start vs brief parenteral doses), and appropriate antibiotics for infants, all of which can bias selection pressures on the infant resistome. Beyond biological effects, the finding that some infant courses were obtained OTC aligns with multi-country estimates that a large share of paediatric antibiotic use in LMICs occurs without prescription, raising concerns about indication, dosing, and duration (Auta et al., 2019; Tigoi et al., 2025; Zewdie et al., 2024). Notably, fluoroquinolones are generally restricted in children and reserved for limited indications (Street & Ababa, 2021; WHO, 2013), therefore its appearance in this cohort and community use points at stewardship and access challenges.

Most infants were breastfed at three months, but only a small fraction met the WHO definition of EBF. Caregivers often give water in hot climates due to perceived infant thirst and long-standing cultural norms (Edney et al., 2022). However, breastmilk meets infants' energy and fluid needs even in hot environments (Geneva: World Health Organization, 2018); introducing water both displaces milk intake and adds preparation and water-quality steps that can seed environmental microbes (Lamberti et al., 2011). Large longitudinal cohorts show that breastfeeding strongly shapes early gut community assembly, whereas

shifts away from exclusive breast milk (or cessation of breastfeeding) alter trajectories toward more adult-like profiles (Bäckhed et al., 2015; Stewart et al., 2018). Beyond composition, diet relates to the resistome: in multinational analyses, formula/mixed feeding has been associated with a higher ARG load compared with predominant breastfeeding, and infant ARG profiles track maternal sources (K. Pärnänen et al., 2018; K. M. Pärnänen et al., 2022). From a clinical perspective, EBF is also protective against diarrhoeal illness relative to partial or no breastfeeding (Lamberti et al., 2011).

4.6. Fieldwork experience: challenges, lessons learnt, and things I wish I knew

Every DPhil journey is unique, and I have unconventionally chosen to include my fieldwork experiences as part of my thesis.

My fieldwork experiences were both humbling and educational. It was more than just recruiting participants and collecting data, it was a journey of discovering a part of Nigeria I thought I knew, and the privileges I had taken for granted. Additionally, there were lessons I wish I had known before my first trip but have realised that these challenges were part of the learning curve that shaped my DPhil journey. Reflecting on this experience, I would like to share insights that may be helpful to future students.

Contingency Planning

Planning is a key focus in fieldwork preparation, training, travel logistics, risk assessment, and management strategies. While these are essential, my fieldwork experience taught me that flexibility is just as important. No matter how detailed your plan is, the field presents real-life challenges. Being adaptable, open-minded, and willing to learn can help navigate these uncertainties and reduce stress along the way.

My initial plan was to live within the community where I was conducting my research to fully integrate into the environment. However, my first few weeks in a family house proved challenging due to unreliable electricity, household rodents, and an unclean water source. As a result, I had to arrange alternative accommodation during fieldwork. Having a contingency plan in place would have made the transition smoother and integration into the community easier.

When I arrived in Kano for fieldwork, healthcare workers at the enrolment site (MMSH) went on strike due to poor staff welfare, delaying the project for about a week. Shortly after the strike ended, a local government election led to a 24-hour curfew for several days. Elections in this region are often volatile, further stalling the project initiation for weeks. On those days, my team and I were unable to work. My key takeaway is to plan for unexpected disruptions and mentally prepare for situations beyond your control.

Community expectations

Despite my efforts to clearly communicate the primary academic focus of my research, it became evident that the community held expectations for direct or indirect tangible benefits. Many participants anticipated financial assistance or medical treatment, particularly in times of illness. Despite repeated clarification of the study's objectives, there remained a prevailing perception that research initiatives should provide immediate practical benefits.

Additionally, scepticism arose regarding the collection of biological samples and the frequency of home visits. Some individuals expressed concerns about the potential misuse of their personal data and the long-term implications of providing biological samples. Others were apprehensive about the researchers' intentions, fearing that their participation would not result in meaningful benefits for the community. These concerns contributed to some

individuals declining participation and, in some cases, enrolled participants withdrawing from the study.

These challenges underscore the importance of building trust, improving communication, and ensuring transparency throughout the research process to foster community engagement and participation.

Logistical challenges

Fieldwork presents several logistical challenges, particularly the need for reliable power and connectivity. With heavy reliance on electronic devices for data collection and communication, power banks are essential to keep equipment charged throughout the day. A stable network and Wi-Fi are also crucial for real-time communication, data access, and secure uploads.

Additionally, a steady power supply is vital for preserving biological samples at appropriate temperatures. One of the most distressing moments of my research was losing breastmilk samples from 30 participants due to a prolonged power outage. Investing in solar-powered appliances could prevent such losses, ensuring the integrity of collected samples and protecting months of hard work. Addressing these challenges will greatly enhance research efficiency and reliability in low-resource settings.

Power imbalances/exacting mindset

It is important to recognise the power dynamics between yourself as the researcher, the research assistants, and the community members you work with. Keeping this in mind fosters a more collaborative, rather than extractive, research process.

As you settle into the community, the lines between researcher and participant may blur. You may feel inclined to share your thoughts on various topics, but doing so indiscriminately can

introduce bias into your data. It is crucial to remain diplomatic, especially on sensitive issues.

For example, while breastfeeding is common in the community, practices do not always align with WHO's guidance on exclusive breastfeeding (initiation within the first hour of birth and no other foods or liquids, including water, for six months). In Kano, giving water in addition to breastmilk is commonly reported, often based on the belief that babies become thirsty in the hot weather, and is reinforced by prelacteal traditions (e.g., offering water or herbal infusions soon after birth). Some families practise "drinking the Qur'an" by feeding their babies with the "Qur'anic water", as described earlier, as a source of blessing and protection. Expressing strong personal opinions on religious or cultural practices such as this can lead to unnecessary conflict. If you must engage, do so diplomatically and base your discussion on evidence rather than personal viewpoints.

Psychological strain

Enrolling the babies felt like having 120 children who are my responsibility. Tragically, early in the project, we lost a mother-baby pair, losing the mother during childbirth and the baby a week later due to lack of care. Another mother passed away 11 months into the study. Despite the short time we knew them, these losses were profoundly felt and emotionally taxing, serving as a stark reminder of life's fragility and the unexpected sorrow that can come from even brief encounters. For longitudinal-based research on human participants, I believe that debriefing sessions and mental health support should be part of the contingency plan for researchers.

Chapter 5: Temporal dynamics of the infant gut microbiome and resistome during early life

5.1 Introduction

Early infancy represents a critical developmental window during which the gut microbiome assembles from a near-sterile state at birth to a complex, diverse microbial community by the end of the first year of life (Bäckhed *et al.*, 2015; Dominguez-Bello *et al.*, 2019). This period of pronounced microbial dynamism is characterised by rapid shifts in bacterial community composition and diversity, in contrast to the relatively stable adult gut microbiome (Stewart *et al.*, 2018; Yatsunenکو *et al.*, 2012). Alongside these bacterial compositional changes, ARGs are acquired, establishing the infant gut resistome that can persist into later life and serve as a reservoir of ARGs impacting infections in later life (Ferretti *et al.*, 2025; Patangia *et al.*, 2024). Understanding the temporal dynamics of both microbiome and resistome development during this formative period is essential for characterising early-life colonisation patterns and identifying potential windows for intervention.

Neonatal gut colonisation often precedes invasive infection, driven by barrier immaturity, immune system development, and high healthcare exposure (Beaumont *et al.*, 2024; Schwartz *et al.*, 2023). Infants who acquire, for instance, ESBL determinants, shortly after birth often become persistent reservoirs of multidrug-resistant bacteria. These reservoirs can subsequently seed bloodstream or urinary tract infections. Conversely, transient detections or lack of colonisation are associated with lower infection risk. Predicting these outcomes requires longitudinal profiling that tracks not only the presence of ARGs but also their acquisition timing, persistence, and bacterial hosts.

Longitudinal studies in HICs have described broad patterns of early gut microbiome development and reported ARG carriage in infants (Enav et al., 2022; Ferretti et al., 2025; Laforest-Lapointe & Arrieta, 2017; Patangia, Grimaud, Wang, et al., 2024). However, fewer studies provide densely longitudinally sampled metagenomes, reflecting the dynamism of the early-life microbiome (Ferretti et al., 2025; Patangia, Grimaud, Wang, et al., 2024). This gap is most acute in sub-Saharan Africa, where available data still rely on culture-based surveillance or cross-sectional sampling (Bulabula et al., 2017; Edwards et al., 2023; Neemann et al., 2020). While Maghini et al. (2025) recently provided a landmark metagenomic profile of African adults, revealing region-specific signatures, comparable high-resolution data from African infants remain critically lacking.

This gap is particularly consequential for West African settings such as Northern Nigeria, where contextual factors differ markedly from those shaping early-life microbiome development in HICs. At MMSH, Kano, postpartum antibiotic prescribing is common, neonates with suspected sepsis are frequently treated empirically without confirmed diagnosis, and ward and household environments may serve as ARG reservoirs due to poor WASH infrastructure and high-density living (Edwards et al., 2023; Nieto-Rosado et al., 2024; Sands, Carvalho, et al., 2022; Sands et al., 2021; Thomson et al., 2021). Culture-based studies from similar hospitals have documented high rates of colonisation with ESBL-producing Enterobacteriaceae, acquired AmpC producers, and, in some settings, carbapenemase-producing organisms among mothers and infants (Carvalho et al., 2022; Edwards et al., 2023; Neemann et al., 2020; Sands, Carvalho, et al., 2022; Sands et al., 2021). Understanding how the gut microbiome and resistome evolve from birth through the first three months in this context is therefore directly relevant to characterising neonatal colonisation risk and informing AMR control strategies.

This chapter addresses these gaps by characterising the joint temporal development of the gut microbiome and resistome in a Nigerian birth cohort, using longitudinal long-read metagenomes collected from birth through Day 90. The microbiome and resistome are treated as linked trajectories: age-related shifts in bacterial community structure provide the taxonomic context in which ARG burden, diversity, and class composition evolve. The study design provides dense coverage of the first postnatal weeks, when colonisation is potentially most dynamic. Repeated within-infant sampling, enables cohort-level descriptions of age-associated change alongside individual trajectories of ARG acquisition, persistence, and loss. Four temporal dimensions are prioritised: (i) the bacterial composition of this cohort over the first three months of life, (ii) first detection of clinically relevant ARGs during the first postnatal week, (iii) persistence of ARGs versus loss across successive sampling points after initial detection, and (iv) the proportion of infants who remain never-colonised within this period. These dimensions shape interpretation of early-life resistance burden and inform plausible windows for preventive or decolonisation interventions (Bargheet et al., 2023; Lebeaux et al., 2021).

By deriving taxonomic profiles, ARG features, and contig-linked host assignments from the same metagenomes, this study relates shifts in microbiome composition to changes in resistome load and ARG class balance (Magiorakos et al., 2012). These descriptive analyses establish baseline trajectories against which Chapter 6 evaluates how modulators, such as delivery mode, antibiotic exposure, and feeding practices, drive variation in microbiome/resistome co-development. The outcomes defined here, acquisition timing, persistence, multi-class carriage, and dominant hosts, serve as the dependent variables for subsequent analyses, identifying the drivers of heterogeneity in this high-exposure setting.

5.2 Aims and Objectives

The overarching aim of this chapter is to describe how the gut microbiome and gut resistome develop from birth to Day 90 in this Nigerian birth cohort, and to define temporal and compositional outcomes that support subsequent exposure-outcome analyses in Chapter 6. Specifically, the chapter addresses the following objectives:

- 1) Microbiome trajectory: To characterise the assembly of the gut microbiome from birth through Day 90, quantifying age-associated shifts in alpha-diversity alongside community structure and identifying the "early core" genera that colonise infants in this cohort.
- 2) Resistome trajectory: To describe the overall development of the gut resistome over the same period, including changes in total ARG burden, ARG richness and resistome diversity, alongside shifts in the distribution of major resistance classes and mechanisms, with emphasis on clinically interpretable categories.
- 3) Acquisition and persistence dynamics: To quantify the temporal dynamics of high-priority ARGs, estimating time to first detection (acquisition timing) and subsequent patterns of persistence versus clearance across successive sampling points during the first three months of life.
- 4) Multi-class carriage and bacterial hosts: To characterise the structure and genomic context of resistance by quantifying therapy-relevant multi-class ARG colonisation (e.g., concurrent β -lactam and aminoglycoside resistance) and identifying the dominant bacterial hosts of selected high-priority markers (e.g., *bla*_{CTX-M-15} in Enterobacteriaceae) using contig-level taxonomic assignments.

5.3 Chapter-specific methods

Laboratory procedures, sequencing, quality control and the core bioinformatic workflow for taxonomic and ARG profiling are described in Chapter 2. This chapter draws on longitudinal infant stool metagenomes collected between birth (Day 0) and Day 90; the analytic denominator varies by endpoint because a small number of low-yield libraries could not complete assembly- and contig-dependent steps required for some outputs, and denominators are therefore reported for each analysis. Microbiome summaries in this chapter are based on genus-level bacterial profiles and expressed as relative abundances (proportions of the total bacterial community). Resistome outcomes are reported on a relative scale (presence/absence, feature-count summaries, or normalised abundance measures (detailed below)). Key QC outcomes supporting data readiness are summarised at the start of the Results (Section 5.4.0), with full QC outputs provided in the Appendix A.3.

5.3.1 ARG feature representation

ARGs were annotated against CARD and represented at the level of the CARD "Best Hit" Antibiotic Resistance Ontology (ARO) features. For interpretability, ARO-level calls were mapped to drug class and resistance mechanism using the curated mapping described in Chapter 2, with refinements applied where clinically relevant in this chapter (e.g., separating ESBL- and AmpC-related β -lactamases from other β -lactam-associated mechanisms). Analyses in this chapter used three complementary representations derived from the ARO table:

- (i) Binary carriage (presence/absence) for clinically focused markers and time-to-event analyses;

(ii) Presence-weighted feature counts to characterise the breadth of ARG classes and mechanisms detected per sample. In this approach, each distinct ARO feature detected in a sample contributed a maximum of one count to its assigned drug class and mechanism, irrespective of read depth or gene copy number. This method emphasises repertoire diversity, the number of distinct resistance genes present, rather than the abundance or coverage of individual features, reducing sensitivity to sequencing depth variation and highlighting the range of resistance functions detected;

(iii) Normalised abundance summaries (used where specified in Results) derived from read-to-feature counts and expressed on a within-sample normalised scale (Chapter 2), reflecting the relative coverage of ARG features when quantitative comparisons are warranted.

5.3.2 Presence-weighted class and mechanism summaries

To describe resistome composition by age while emphasising repertoire breadth rather than coverage, presence-weighted class and mechanism stacks were generated. Within each sample, each distinct ARO contributed a maximum of one count to its assigned drug class and mechanism, irrespective of read depth. This approach reduces sensitivity to sequencing depth differences and highlights the diversity of resistance functions detected per sample.

For main class-level panels, drug classes were restricted to those contributing the majority of ARO detections across the dataset under the presence-weighted definition, with remaining classes grouped into an “Other” category. Mechanism-level summaries were constructed similarly, highlighting clinically interpretable categories (e.g., ESBLs and AmpC β -lactamases). The full all-classes resistome panel, without class restrictions, is provided in the supplementary material (Appendix A.3, figure S5.9). The main text instead features a therapy-relevant panel, generated by excluding multidrug efflux-dominated and non-

specific categories and re-normalising the remaining classes. Here, “therapy-relevant” refers to systemic antibiotic classes used in clinical care (e.g., β -lactams, aminoglycosides, fluoroquinolones, tetracyclines, glycopeptides, sulfonamide/trimethoprim, and macrolides), with biocide/antiseptic categories reported separately where relevant.

5.3.3 Clinically focused ARG definitions

Clinically focused measures were defined using curated ARO lists and regular-expression matching of ARO names where appropriate. Definitions were prespecified to support prevalence and trajectory summaries.

- ESBLs: CTX-M family AROs and ESBL-type SHV variants; narrow-spectrum TEM-1 and OXA-1 were excluded from ESBL definitions.
- Carbapenemases: NDM, KPC, VIM, IMP, and OXA-48-like enzymes; OXA-181 and OXA-244 were grouped with OXA-48-like enzymes.
- Aminoglycoside resistance determinants: subdivided into aminoglycoside-modifying enzymes (AMEs) and other aminoglycoside-associated mechanisms. A clinically focused AME subset was additionally defined for gentamicin-relevant AAC and APH families observed in the cohort (e.g., *aac(3)-II*, *aac(6')-Ib*, and selected APH families).

To avoid inflating rifamycin resistance, *rpoB*-only annotations without confirmed resistance variants were excluded from class summaries; acquired ARR enzymes were retained. Similarly, AmpC β -lactamases were categorised into acquired (CMY/DHA; FOX handled in sensitivity analyses) and chromosomal-like (ACT/MIR) families. While it is acknowledged that the distinction is not absolute across all genomic contexts, these groups

were reported separately and combined where appropriate to facilitate clinical interpretation.

5.3.4 Time-to-event and persistence

To describe the timing of first detection of key ARGs, per-infant time-to-event variables were constructed for first detection between birth and Day 90. For each ARG group, infants contributed time from their first sampled day until the first visit at which at least one relevant ARO was detected. Infants without detection during follow-up were censored at their last sampled day. These analyses interpret non-detection as absence, or being below the limit of detection at the available sequencing depth. Denominators and sampling frequency are reported alongside estimates.

Kaplan-Meier curves (Kaplan & Meier, 1958) were used to summarise time to first detection; medians with 95% confidence intervals were reported where the median was reached during follow-up. In addition, age-specific prevalence (n/N) with exact binomial confidence intervals was tabulated for selected ages (D7, D14, D28, D90), using the first sample at or beyond the target day as the assessment point.

Persistence was summarised descriptively using within-infant run lengths, defined as the number of consecutive sampled visits with detection for a given ARG group, and by the proportion of infants with single-visit versus multi-visit detections.

5.3.5 Co-presence and enrichment between resistance classes

To evaluate whether resistance classes co-occurred more often than expected under independence, class-by-class contingency tables were constructed within age strata. For a given pair of classes A and B, observed co-presence was compared with the expected co-

presence under independence, $\Pr(A)\Pr(B)$, summarised as log2 enrichment: $\log_2 \left[\frac{\Pr(A \cap B)}{\{\Pr(A)\Pr(B)\}} \right]$. Enrichment or depletion was assessed using Fisher's exact tests for each class pair, with p-values adjusted for multiple testing using Benjamini-Hochberg false discovery rate control (nominal FDR 0.05) (Benjamini & Hochberg, 1995). Heatmaps display observed co-presence and enrichment values and indicate class pairs meeting the FDR threshold.

5.3.6 Contig-linked taxonomic context for selected ARGs

To investigate likely bacterial context for selected priority ARGs, contig-level taxonomic assignments derived from the metagenomic assemblies were linked to ARG-bearing contigs using an ARO–contig linkage table (Chapter 2). For each scheduled day, hosts were summarised as the genera and species whose contigs carried at least one instance of a target ARO (e.g., *bla*_{CTX-M-15} and selected AMEs). Bubble plots summarised, by day of life, the distribution of contig-linked taxa among positive samples. These summaries are interpreted as contig-linked taxonomic context rather than definitive cellular host attribution.

5.4 Results

5.4.0 Data readiness and QC outcomes (summary)

Sequencing and QC outcomes confirmed that the infant stool metagenomes were of sufficient quality for longitudinal analysis. Initial cleaning with fastp removed only 1–2% of reads from infant samples, while negative controls lost substantially more (~12%), indicating the successful removal of short-fragment background noise. Host-DNA burden was generally low (median 2.5%), with only a small minority of libraries (~1%) being

heavily host-dominated, a pattern consistent with the low-biomass nature of infant sampling. Read-length distributions were consistent across samples (N50 ~5 kb; Appendix A.3, Supplementary Figure 5.2), supporting uniform downstream processing.

Despite host removal, substantial microbial depth was preserved; three-quarters of the samples retained >200,000 non-host reads, with a minimum of 5,508 reads in the lowest-yield library. To mitigate potential batch effects, background contamination was managed by pooling all available blanks into a composite profile. A dual-method approach using microDecon and decontam identified contaminants that were predominantly environmental taxa. This decontamination process created only minor shifts in Shannon diversity (~0.1–0.2 units) and did not alter fundamental sample relationships. These outcomes justify the use of the corrected dataset for the microbiome and resistome trajectories that follow (detailed metrics are provided in Appendix A.3, Supplementary data 1).

Of the 90 infants enrolled, 87 provided at least one stool sample for metagenomic analysis, and 85 had sufficient sequencing depth for resistome analysis. Sample sizes varied by timepoint (range: n=10 at Day 0 to n=69 at Day 14) due to missingness.

5.4.1 Early-life gut microbiome trajectory

5.4.1.1 *Diversity and community structure*

Genus-level α -diversity was significantly associated with age (Wald χ^2 $P = 0.033$) but characterised by high inter-infant heterogeneity (Figure 5.1). Diversity was lowest at birth (Day 0), frequently reflecting communities dominated by a single genus (Shannon index median 1.0; IQR 0.47–1.54). Following a peak in median diversity observed at Days 3 (1.46; IQR 0.98–1.90), median scores fluctuated between 1.23 and 1.48 through Day 90, showing no consistent monotonic trend.

Beta-diversity (Bray-Curtis PCoA) showed that age explained only a minor portion of the community variance ($R^2 = 0.122$, $p = 0.001$). Although a transition along the first principal axis was visible from birth to 3 months (Figure 5.2), the high degree of scatter at every timepoint confirms that infant gut communities remain highly individualised. Dispersion was slightly reduced by Days 28 and 90, yet the lack of clear clustering indicates that these communities have not converged toward a uniform state by the end of the study period (Figure S5.6).

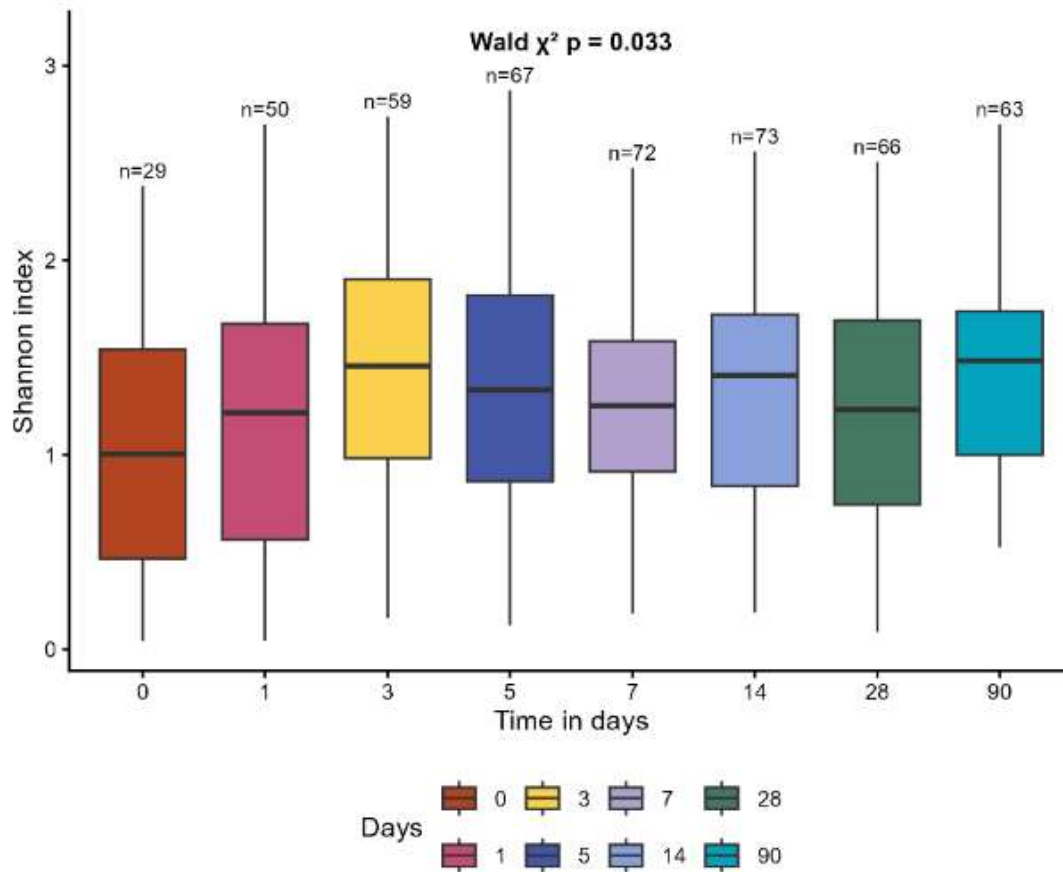


Figure 5.1: Gut microbiome α -diversity by day of life.

Shannon diversity index for genus-level bacterial communities in infant stool samples collected at scheduled postnatal days (D0, D1, D3, D5, D7, D14, D28, D90). Boxplots show the median (centre line) and interquartile range (box); whiskers extend to $1.5 \times \text{IQR}$.

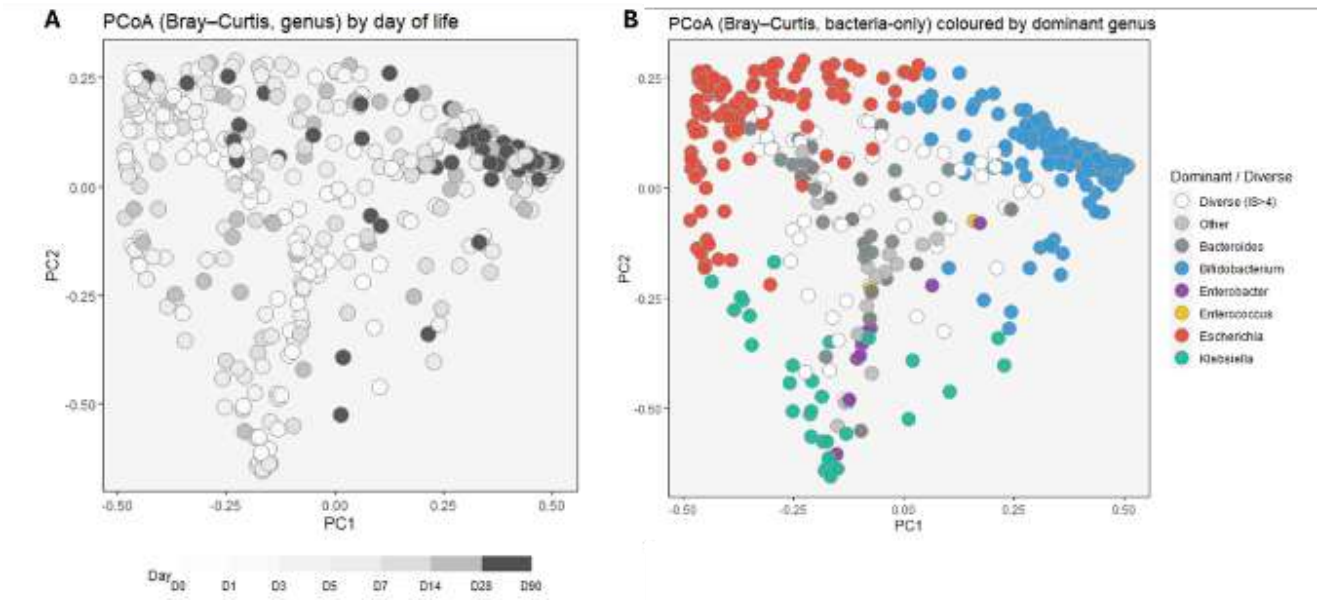


Figure 5.2: Age-associated structuring of infant gut microbiome composition.

(A) PCoA of Bray-Curtis dissimilarities calculated from genus-level relative-abundance profiles for all infant stool samples collected at scheduled days of life (D0-D90). Points represent individual samples and are shaded by sampling day. (B) The same ordination coloured by the dominant genus in each sample (genus with the highest relative abundance). Samples without a single dominant genus are labelled “Diverse (IS > 4)”, where IS denotes the inverse Simpson index ($1/D$) exceeding 4.

5.4.1.2 Composition and dominant taxa

Genus-level composition plots highlighted the specific bacteria driving these patterns (Figure 5.3). During the first week of life, communities were commonly dominated by genera within the Enterobacteriaceae family, with *Escherichia* and *Klebsiella* frequently accounting for the majority of reads; other genera, including *Enterobacter*, *Citrobacter*, and

Salmonella, appeared intermittently. Gram-positive genera such as *Staphylococcus*, *Streptococcus*, and *Enterococcus* were also common early colonisers but typically contributed a smaller share of the overall community. By Day 14 and especially by Days 28 and 90, composition diversified: a larger fraction of samples showed substantial contributions from genera more typical of later infancy, including *Bifidobacterium* and *Bacteroides* taxa (Chapter 1, Section 1.2), and fewer samples remained overwhelmingly dominated by a single Enterobacteriaceae-associated genus.

PCoA analyses confirmed an enrichment of *Escherichia*- or *Klebsiella*-dominated communities among earlier samples (Figure 5.2B). In contrast, *Bifidobacterium*-dominated and more compositionally diverse communities were more common at later timepoints. Crucially, pronounced inter-infant variability persisted at all ages; some infants retained Enterobacteriaceae-rich profiles at Day 90, while others transitioned earlier to *Bifidobacterium* or mixed communities. Illustrative within-infant trajectories for infants with complete sampling across visits are shown in (Figure S5.7).



Figure 5.3: Genus-level gut microbiome composition by day of life.

Stacked barplots showing the relative abundance of the 20 most abundant bacterial genera in each infant stool sample, grouped by scheduled day of life (D0-D90). During the first week (D0-7), samples are frequently dominated by one or two Enterobacteriaceae genera, particularly *Escherichia* and *Klebsiella*, with smaller contributions from genera such as *Staphylococcus*, *Streptococcus* and *Enterococcus*. By later timepoints (D28-90), more samples show substantial contributions from the genus *Bifidobacterium*, although at all ages there is marked between-infant variability in which genus dominates. *Figure plotted by Aditiya Lankapalli*

5.4.2 Infant gut resistome trajectory: burden, diversity and antibiotic-class composition

5.4.2.1 *Total ARG burden, richness, and diversity*

ARG burden remained largely stable across the first 90 days, characterized more by substantial inter-infant heterogeneity than by temporal shifts (Figure 5.4A). While a mixed-effects model (Section 2.9.2) identified a statistically significant decline in burden after the first postnatal week ($\sim 0.2 \log_{10}$ units; $p = 0.0029$), temporal patterns were more pronounced in resistome diversity. ARG richness peaked during the first week before declining modestly at later timepoints ($p = 0.01$; Figure 5.4B). Shannon diversity exhibited the strongest temporal structure ($p = 10^{-7}$), increasing initially and then decreasing by D90, a trend indicating consolidation toward fewer dominant ARG features as the resistome matures (Figure 5.4C).

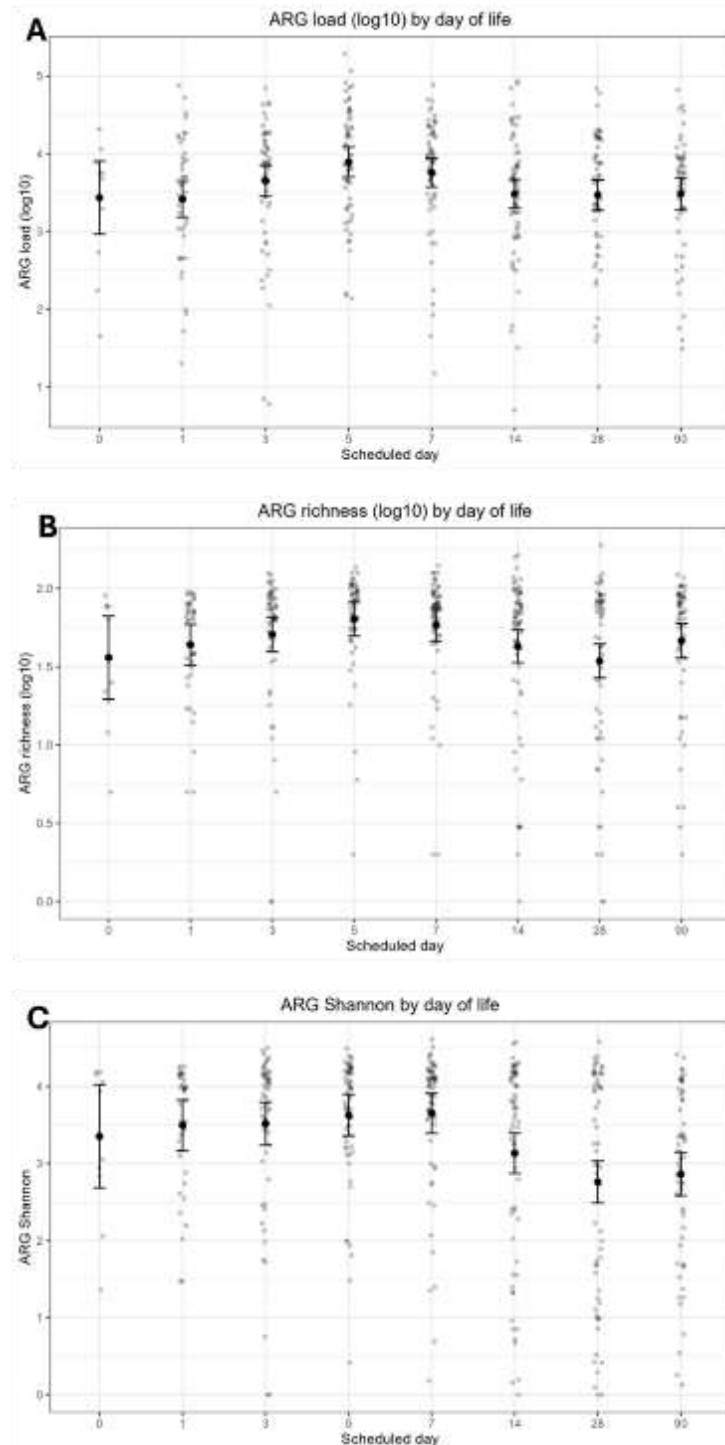


Figure 5.4 Infant gut resistome burden and diversity across scheduled days of life (D0–D90).

(A) ARG load (log10), expressed as the log10-transformed abundance of ARG-assigned signal per sample. (B) ARG richness (log10), defined as the log10 number of distinct ARG features (CARD ARO “Best Hit” calls) detected per sample. (C) ARG Shannon diversity, reflecting the evenness and complexity of the ARG profile within each sample. Grey points represent individual infant samples at each scheduled timepoint. Black points with error bars show model-estimated means and 95% confidence intervals from linear mixed-effects models with day of life modelled as a categorical predictor and an infant-level random intercept to account for repeated measures.

5.4.2.2 *Resistome compositional structure*

Ordination of resistome profiles showed substantial overlap between timepoints, with only a weak age-associated signal (Figure 5.5). In a PCoA of Aitchison distances, samples from different ages intermingled throughout ordination space rather than forming discrete day-specific clusters. PERMANOVA indicated that day of life explained a small but statistically detectable proportion of compositional variance ($R^2 = 0.047$; $p = 0.001$). Dispersion diagnostics suggested modest differences in spread across days (Figure S5.8), indicating that the age signal reflects small shifts in centroid location alongside minor changes in within-day variability.

5.4.2.3 *Therapy-relevant drug-class composition*

When ARGs were summarised at the drug-class level, β -lactam resistance classes were consistently dominant from birth through Day 90 (Figure 5.6). In the presence-weighted class summary restricted to therapy-relevant systemic antibiotic classes (with multidrug efflux and non-specific/unspecified categories excluded and the remainder renormalised), β -lactam ARGs contributed the largest share at every timepoint. Fluoroquinolone and aminoglycoside ARGs represented the other primary components in the first week of life. Over time, the relative contribution of tetracycline and glycopeptide-associated ARGs increased, with tetracycline-associated ARGs becoming a major component by Day 28. Biocide/antiseptic-associated ARGs contributed smaller, relatively stable proportions throughout the study period.

A companion plot including all classes (including multidrug efflux and unspecified categories) is provided in the Supplement (Figure S5.9) to show the full resistome background.

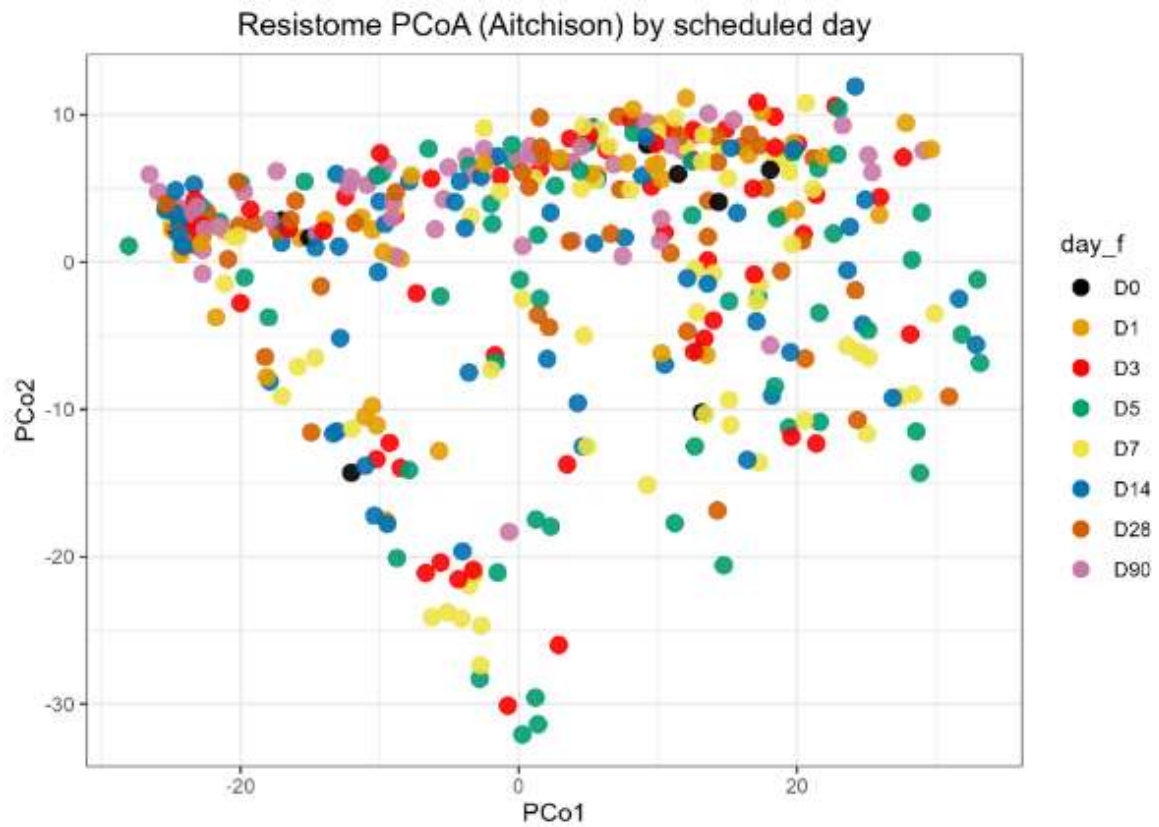


Figure 5.5 Principal coordinates analysis of gut resistome composition by day of life.

PCoA of Aitchison distances between infant stool resistome profiles, with each point representing one sample and coloured by scheduled day of life (D0, 1, 3, 5, 7, 14, 28, 90). Samples from different ages occupy overlapping but partially separated regions of ordination space, indicating modest age-related structuring of resistome composition. In PERMANOVA, day of life explained a small but statistically significant proportion of variance ($R^2 \sim 0.05$, $p = 0.001$), with some evidence of differences in dispersion between days ($p = 0.003$).

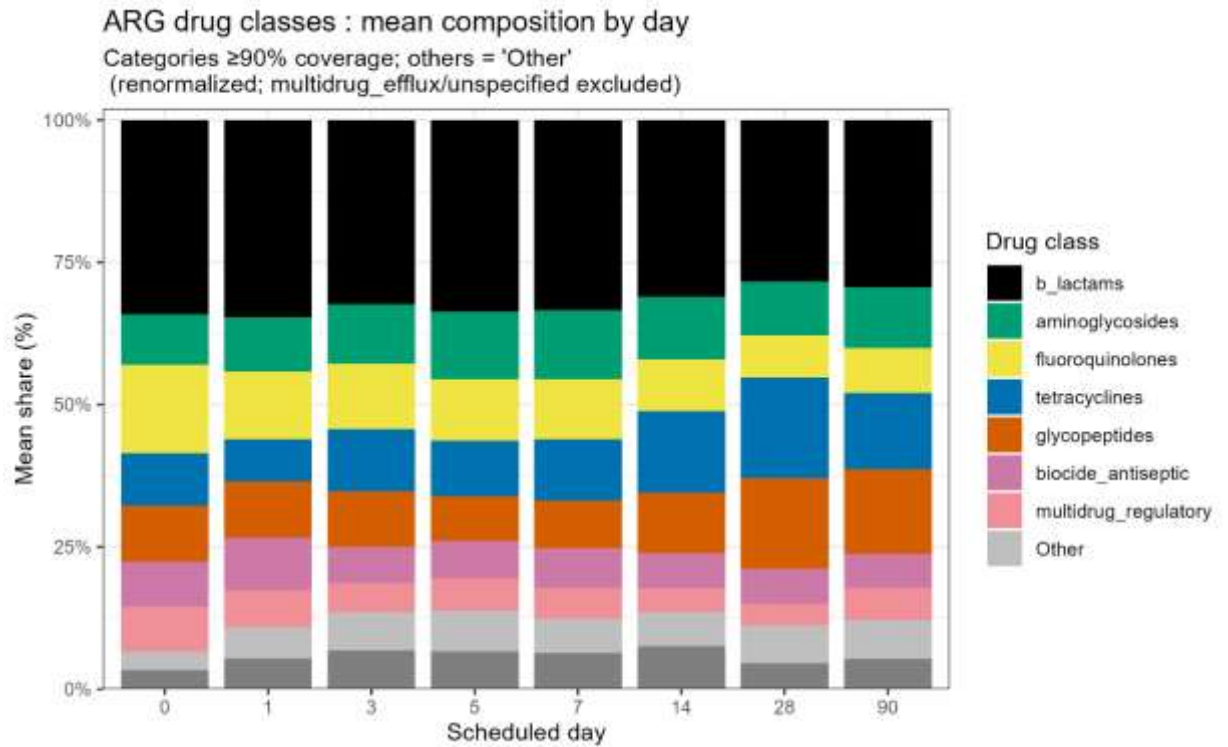


Figure 5.6 Antibiotic-resistance drug-class composition by day of life.

Stacked barplots showing the mean proportional share of major ARG drug classes at each scheduled day. Bars are presence-weighted (each detected ARG feature contributes once per sample before aggregation). Multidrug efflux and non-specific/unspecified categories are excluded for interpretability, and remaining classes are re-normalised to sum to 100%.

5.4.3 Early acquisition of key ARGs and the “never-colonised” group

β -lactam- and aminoglycoside-associated ARGs were detected in the majority of infants from the earliest sampling points and remained common throughout follow-up (Figure 5.7). β -lactam ARG prevalence was already high at Day 0 (90%; 9/10) and increased further across the first postnatal week, remaining consistently high at Days 14, 28 and 90 (85-90%). Aminoglycoside ARGs followed a similar pattern, with lower prevalence at Day 0 but a rapid rise during Days 1-7, reaching 94% prevalence by Day 7. Prevalence estimates at later

visits fluctuated modestly, largely reflecting changing denominators (i.e., fewer available samples at some visits), rather than a sustained decline in carriage.

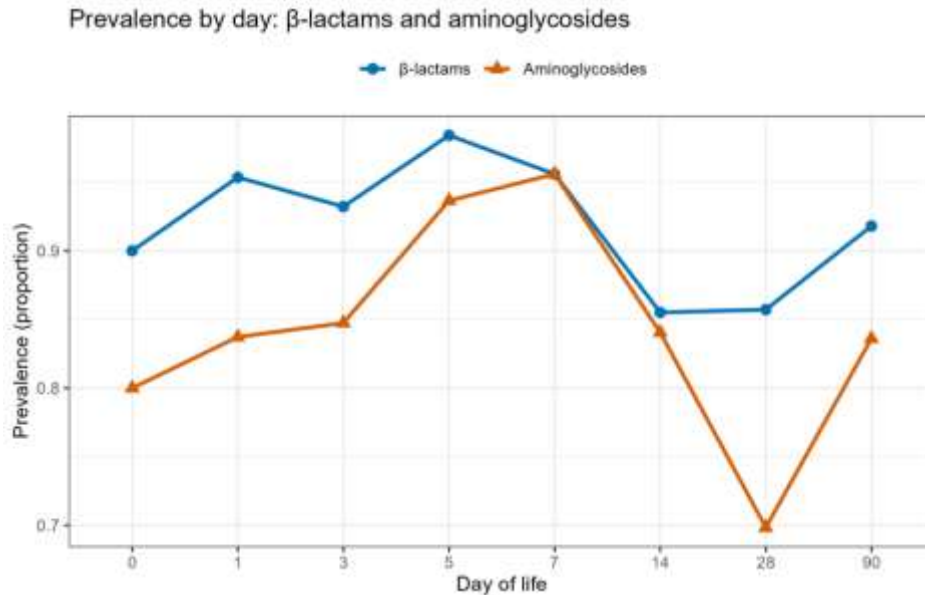


Figure 5.7 Prevalence of β -lactam and aminoglycoside resistance genes acquisition.

Point prevalence of any β -lactam ARGs (blue) and any aminoglycoside ARGs (orange) in infant stool samples at birth (D0) and on Days 1, 3, 5, 7, 14, 28 and 90. Points show n/N positive samples at each timepoint (β -lactam: D0 n/N=9/10, D1(41/43), D3(55/59), D5(62/63), D7(65/68), D14(59/69), D28(54/63), D90(56/61) ; aminoglycoside: D0 n/N=8/10, D1(36/43), D3(50/59), D5(59/63), D7(65/68), D14(58/69), D28(44/63), D90(51/61)).

Time-to-first-detection curves showed that acquisition occurred early for both major classes (Figure 5.8A-C). For β -lactam-associated ARGs, cumulative detection rose steeply during the first postnatal week and then approached a plateau by approximately Days 7-14, indicating that most acquisitions occurred very early. Aminoglycoside-associated ARGs exhibited a similarly rapid early rise, with most first detections occurring during the first week. Co-detection of both classes within the same sample accumulated quickly as well, reaching very high levels by Day 7 and nearing saturation by Day 14. These trajectories suggest that exposure to β -lactam- and aminoglycoside-resistant organisms is common in

the immediate perinatal and early neonatal period, rather than being a feature that emerges only later in hospitalisation.

The “never-colonised” summary emphasised how universal these two classes were by the first quarter of life. None of the infants in the cohort were free of β -lactam and aminoglycoside ARGs at all sampled timepoints. However, many infants never acquired higher-risk ARGs over the study period: around one-third had no ESBL genes (30/85; 35%), a similar number remained *bla*_{CTX-M-15}-negative (31/85; 36%), approximately 60% (51/85) never carried acquired AmpC genes, and almost all infants (79/85; 93%) remained negative for carbapenemase genes throughout follow-up. These proportions are summarised in Table 5.1. Thus, while β -lactam and aminoglycoside ARGs were effectively ubiquitous from early life onwards, clinically important extended-spectrum and carbapenemase genes showed a much more restricted, infant-specific pattern of acquisition.

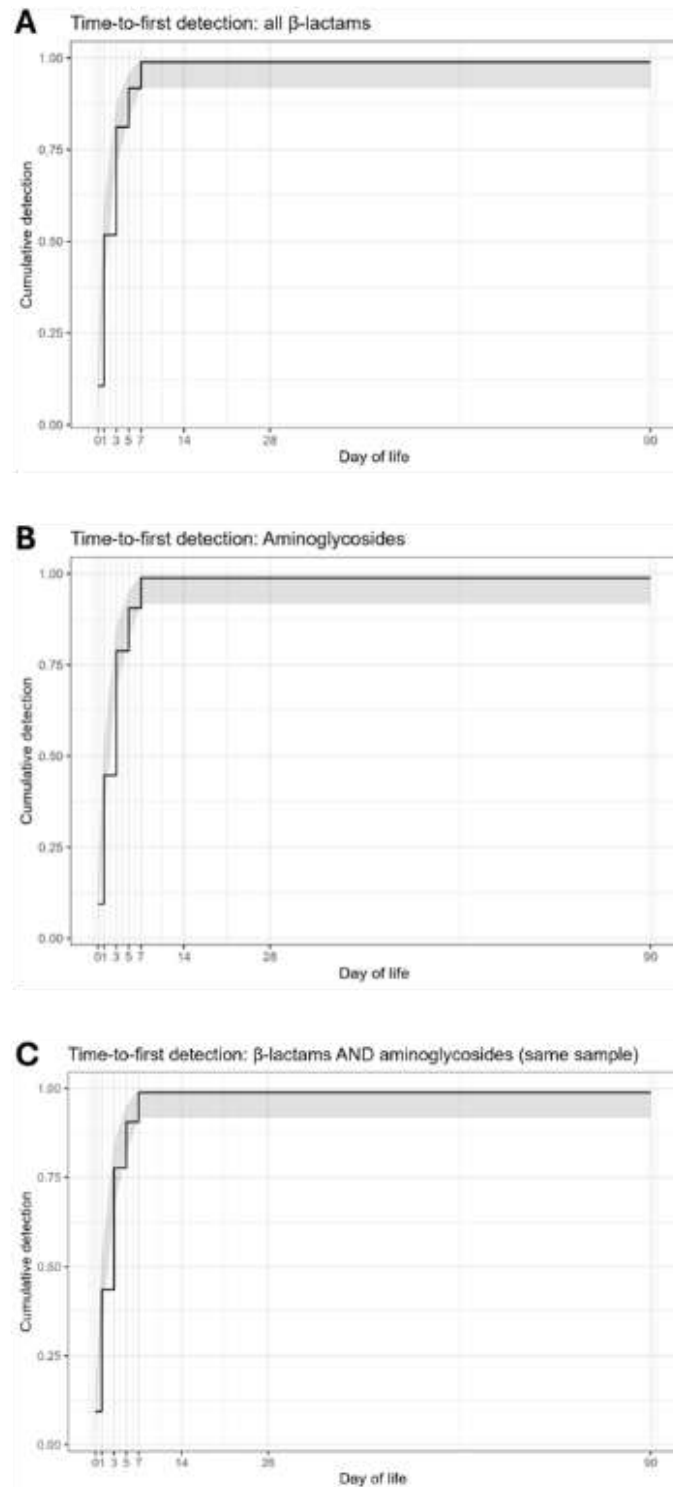


Figure 5.8 Time to first detection of β -lactam and aminoglycoside resistance genes.

(A) Kaplan-Meier (KM) estimate of time to first detection of any β -lactam ARG in infant stool, treating infants without a positive sample as censored at their last follow-up. (B) KM estimate of time to first detection of any aminoglycoside ARG on the same follow-up visits. (C) KM estimate of time to first co-detection of β -lactam and aminoglycoside ARGs in the same sample. In all panels, shaded bands indicate 95% confidence intervals.

Table 5.1 Number of infants negative for selected major ARGs at any sampled timepoint

ARG groups	Number never positive
Aminoglycoside	0
β -lactamase	0
Acquired AmpC	51
Chromosomal AmpC	24
Carbapenemase	79
ESBL	30
<i>bla</i> _{CTX-M-15}	31

Note: Category counts are not mutually exclusive; umbrella categories (e.g., ESBL) include the constituent sub-types (e.g., bla_{CTX-M-15})

5.4.4 Dual β -lactam/aminoglycoside and co-carriage

To translate class-level carriage into clinically interpretable combinations, β -lactam and aminoglycoside ARGs were considered jointly. ARG from these two classes confer resistance to commonly used empirical regimens for suspected neonatal infection (e.g., a β -lactam plus an aminoglycoside), so their co-detection provides a pragmatic summary of how often infants carried resistance potential spanning both components.

Across scheduled timepoints, the dominant state was co-carriage of both classes in the same sample (Figure 5.9). Co-carriage was already common at birth and increased further across the first postnatal week. By Days 5-7, the “both classes present” state accounted for most samples (85-88%), while samples carrying only one class were uncommon and the “neither class detected” rare. This pattern is consistent with the time-to-first detection analysis, which showed that most infants reached first co-detection during the first week of life and that new co-detections after Day 14 were infrequent.

More clinically specific β -lactamase combinations were then evaluated by pairing aminoglycoside ARG carriage with AmpC and ESBL markers (Table 5.2). Co-carriage of AmpC and aminoglycoside determinants was already frequent by Day 7 (64%) and remained common through Day 90 (77%). Co-carriage of ESBL and aminoglycoside determinants was also common, affecting a substantial fraction of infants from the first postnatal week onward (50-64%). These summaries indicate that dual-class resistance potential (β -lactam plus aminoglycoside) is established very early in life in this cohort, and that combinations involving broader-spectrum β -lactamase markers (AmpC and ESBL) are already present in many infants by clinically relevant early-life milestones.

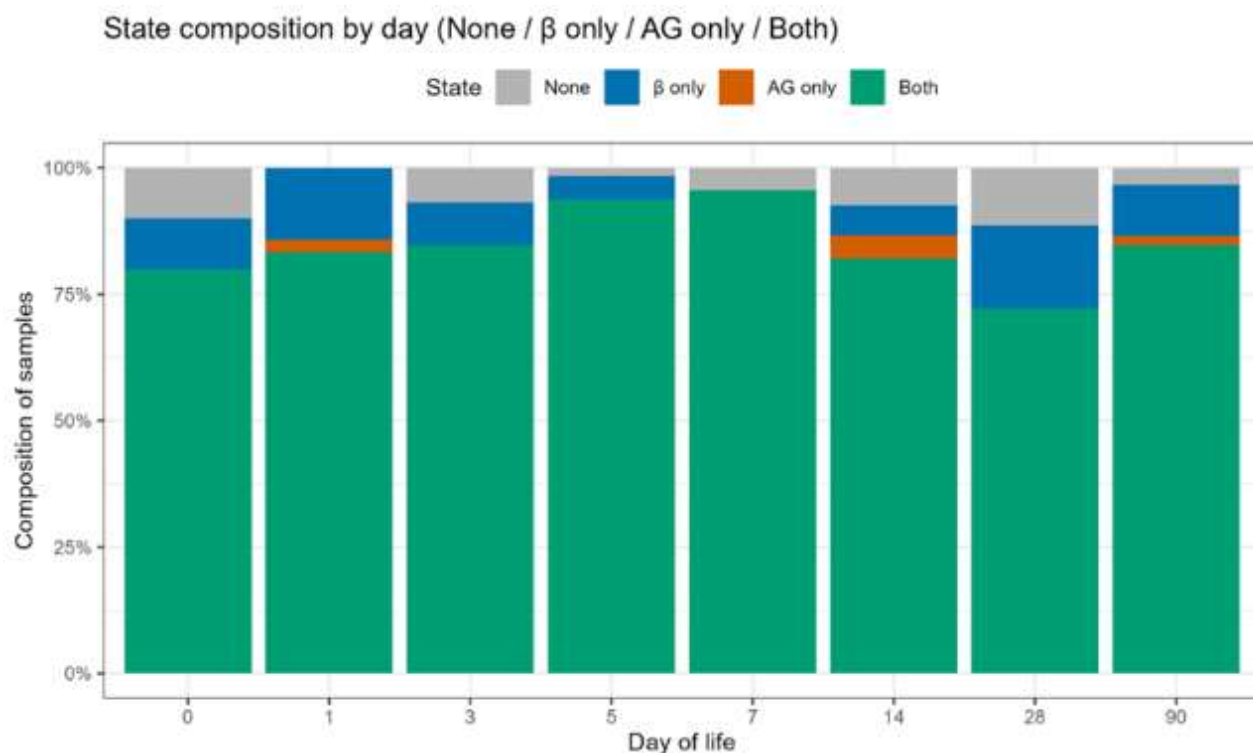


Figure 5.9 Co-carriage states for β -lactam and aminoglycoside resistance classes over time.

Stacked barplots show, for each scheduled day of life, the proportion of infant stool samples falling into four states based on the presence of β -lactam and aminoglycoside resistance genes: neither class detected (“None”), β -lactam only, aminoglycoside only, or both classes present in the same sample (“Both”). The denominator at each timepoint is the number of infants with a sequenced stool sample at that visit (D0=10, D1=42, D3=59, D5=63, D7=68, D14=69, D28=63, D90=61)

Table 5.2 Prevalence of combined β -lactam and aminoglycoside resistance

ARG						
combinations	Day of life	N	Positive	Prevalence	Lower	Upper
AmpC and AG	7	84	54	0.64	0.53	0.74
AmpC and AG	14	80	59	0.74	0.63	0.83
AmpC and AG	28	73	54	0.74	0.62	0.84
AmpC and AG	90	61	47	0.77	0.65	0.87
ESBL and AG	7	84	42	0.50	0.39	0.61
ESBL and AG	14	80	45	0.56	0.45	0.67
ESBL and AG	28	73	41	0.56	0.44	0.68
ESBL and AG	90	61	39	0.64	0.51	0.76
Beta-lactam and AG	7	84	83	0.99	0.94	0.99
Beta-lactam and AG	14	80	79	0.99	0.93	0.99
Beta-lactam and AG	28	73	73	1	0.95	1
Beta-lactam and AG	90	61	61	1	0.94	1

For each ARG combination and scheduled day of life, N is the number of infants with a stool sample, “Positive” is the number with co-detection of the specified β -lactam resistance group (narrow-spectrum β -lactamase, AmpC or ESBL, as indicated) and any aminoglycoside resistance gene in the same sample, “Prevalence” is the corresponding proportion, and “Lower” / “Upper” give the bounds of the exact 95% binomial confidence interval.

5.4.5 Enrichment of resistance class co-presence

The previous section focused on β -lactam–aminoglycoside co-carriage because it aligns with commonly used empirical regimens and provides an interpretable “two-class” clinical summary. This section extends the analysis to the broader set of common, therapy-relevant resistance classes to determine whether same-sample co-presence reflects non-independence (i.e., certain classes occurring together more often than expected given their marginal prevalences).

5.4.5.1 *Observed co-presence across major classes*

For each scheduled day of life, pairwise co-presence was summarised for the major therapy-relevant classes (β -lactams, aminoglycosides, fluoroquinolones, tetracyclines, and trimethoprim–sulfamethoxazole). The observed co-presence heatmap (Figure 5.10) shows that multi-class carriage was common from early life. Co-presence was already moderate at Day 0 for several pairs and increased across the first postnatal week. Across timepoints, pairs involving β -lactams and/or aminoglycosides consistently showed the highest co-presence. For example, by Day 7, β -lactam + aminoglycoside co-presence reached 85-90%, β -lactam + fluoroquinolone reached ~80%, and aminoglycoside + fluoroquinolone reached ~75%. In contrast, pairs involving trimethoprim-sulfamethoxazole (SXT) showed lower co-presence (typically 40-60%), particularly at the earliest timepoints. Of the 10 pairwise combinations examined, those involving β -lactams or aminoglycosides (7 of 10 pairs) consistently showed higher co-presence than those involving only tetracyclines, fluoroquinolones, and SXT.

5.4.5.2 *Enrichment beyond independence*

High co-presence prevalence does not, however, on its own demonstrate non-independence between classes. To test whether drug classes were found together more often than expected by chance, a contingency-table approach was used (Section 5.3.5). For each pair of classes and day of life, the observed proportion of samples positive for both classes was compared with the expected proportion under independence, calculated from the marginal prevalences of each class. The ratio of observed to expected co-presence was summarised as enrichment ($\log_2[\text{obs}/\text{exp}]$), and p-values from Fisher's exact tests were adjusted for multiple comparisons across pairs and days using Benjamini-Hochberg FDR control.

The resulting enrichment heatmap (Figure 5.11) shows that many combinations were significantly enriched beyond chance, particularly from the second week onwards, although effect sizes were generally modest ($\log_2$ enrichment typically 0.1-0.5, corresponding to fold-changes of 1.1-1.4). Enrichment was most consistent for combinations linking β -lactams or aminoglycosides to other classes, including β -lactam + aminoglycoside, aminoglycoside + fluoroquinolone, aminoglycoside + SXT, fluoroquinolone + SXT, and tetracycline + SXT, with multiple timepoints showing positive $\log_2$ enrichment and adjusted p-values below 0.05.

At early timepoints (D0-D3), enrichment signals were weaker and less stable, reflecting both lower marginal prevalences for some classes and wider confidence intervals; nevertheless, several pairs already showed a tendency towards co-clustering rather than independent occurrence. Full numerical results, including observed and expected co-presence, enrichment, p-values and adjusted p-values for each drug-class pair by day of life, are provided in Table S5.1.

These indicate that the infant gut resistome is not simply a random pairing of independently acquired resistance classes. Resistance to different antibiotic groups tends to appear in patterns, with certain combinations, particularly those linking β -lactam and aminoglycoside resistance to other key classes, occurring together more often than would be predicted from their individual prevalences.

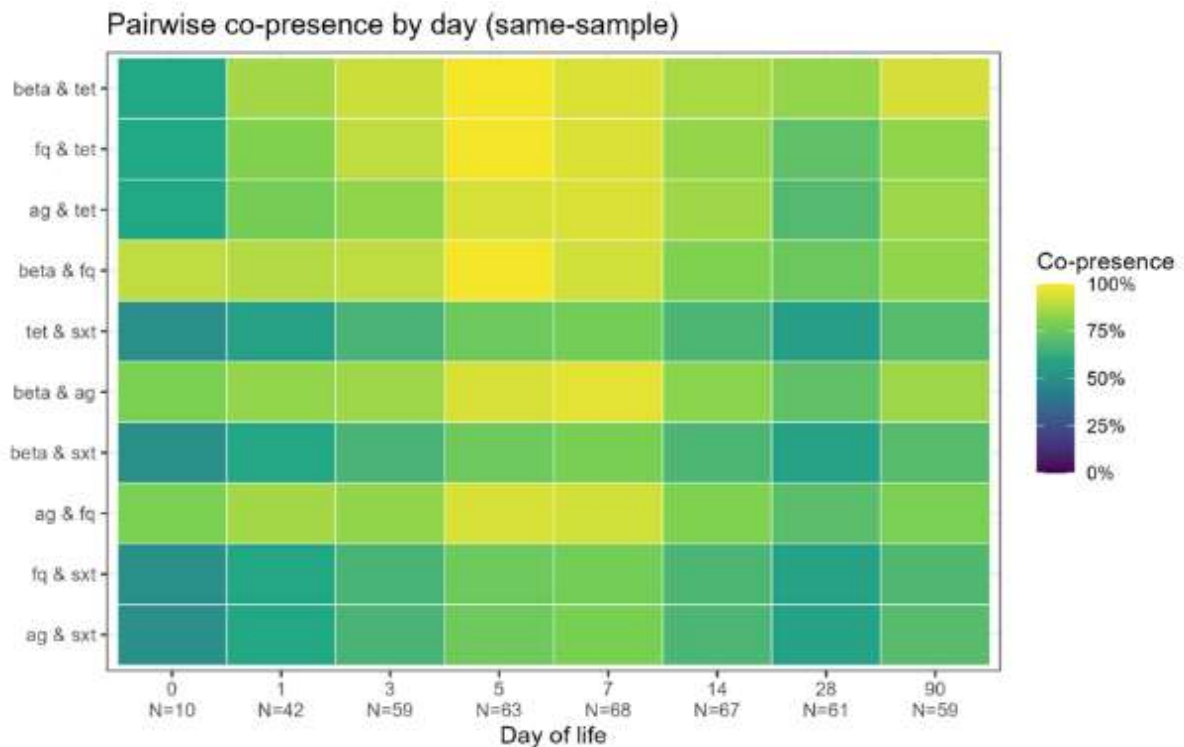


Figure 5.10: Pairwise co-presence of major resistance classes by day of life.

Heatmap showing, for each scheduled day of life, the proportion of infant stool samples in which combination of a given resistance-class pair were detected in the same sample. Rows correspond to all pairwise combinations of β -lactam (beta), aminoglycoside (ag), fluoroquinolone (fq), tetracycline (tet) and trimethoprim-sulfamethoxazole (sxt) resistance classes; columns correspond to days of life. Colour intensity represents the observed co-presence prevalence (0-100%). Sample sizes (number of infants with a sequenced stool at each visit) are shown beneath the x-axis.

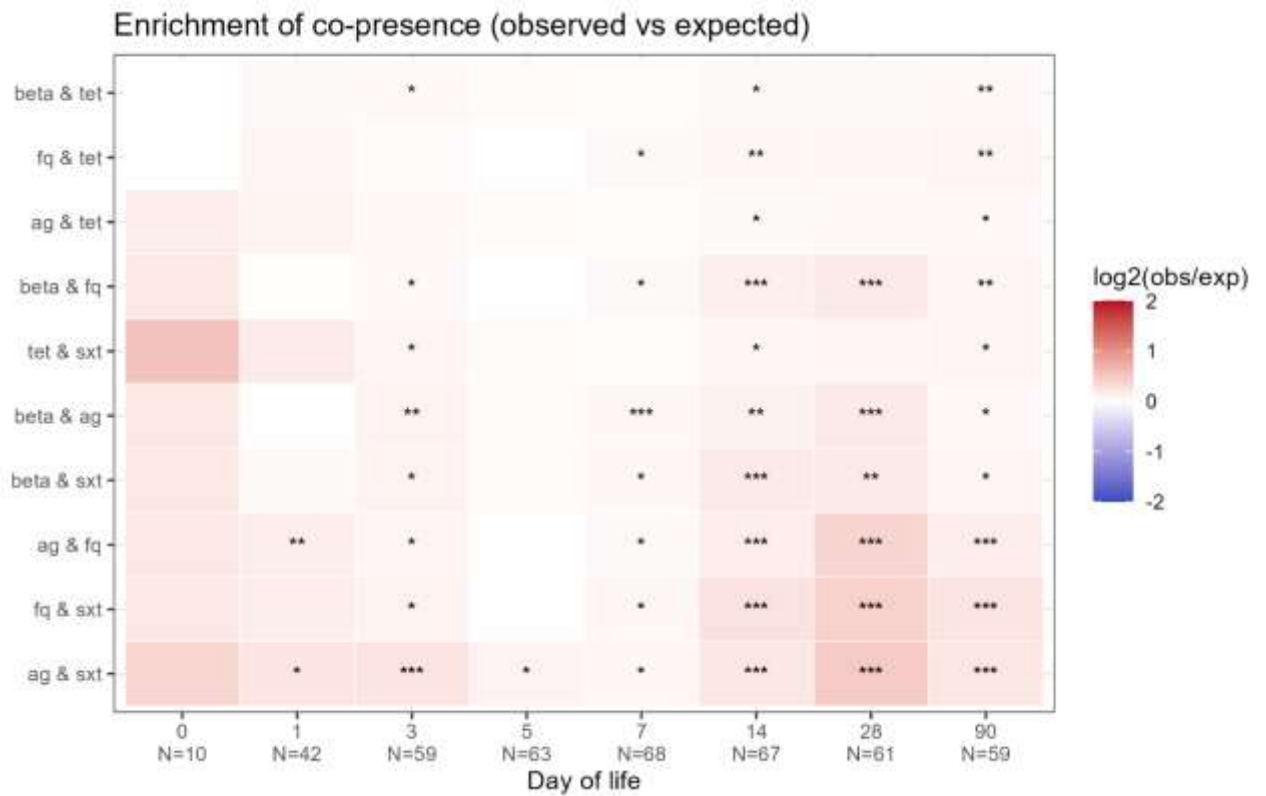


Figure 5.11: Enrichment of resistance-class co-presence beyond independence

Heatmap summarising, for the same resistance-class pairs and days of life as in Figure 5.10, the enrichment of same-sample co-presence relative to an independence model. Each cell shows the $\log_2$ ratio of observed to expected co-presence, where expected prevalence is calculated as the product of the marginal prevalences of the two classes on that day. Positive values indicate that the pair co-occurs more often than expected under independence, values near zero indicate approximate independence, and negative values indicate less frequent co-occurrence than expected. Symbols mark pairs with Benjamini-Hochberg-adjusted p-values below the pre-specified significance threshold for the independence test.

5.4.6 Host context and dynamics of *bla*_{CTX-M-15} and aminoglycoside resistance

This section examines two clinically salient resistance signals in this setting: (i) ESBL carriage, focusing on *bla*_{CTX-M-15}, and (ii) aminoglycoside resistance determinants, focusing on AMEs and the subset flagged as “gentamicin-resistance-associated”. Because long-read assemblies link ARG calls to contigs with taxonomic labels, host assignment can be

summarised to identify the genera and species most frequently associated with these determinants over time.

5.4.6.1 Host context of ESBL and *bla*_{CTX-M-15}

Across both early (D0-D7) and later (D14-D90) windows, contig-linked taxonomic assignment showed that ESBL ARGs were predominantly carried by *Escherichia* and *Klebsiella* (Figure 5.12A). Among ESBL-positive samples with an assigned host, *Escherichia* accounted for approximately 70-75% of detections in both periods, while *Klebsiella* contributed approximately 38% of early and 27% of late samples. *Citrobacter*, undifferentiated Enterobacteriaceae, *Pectobacterium* and other genera each contributed at most about 5-10% of ESBL-positive samples. There was no strong shift in the balance between *Escherichia* and *Klebsiella* between the early and late windows, although *Klebsiella* appeared slightly more prominent in early samples.

At species-level, *bla*_{CTX-M-15} host assignments indicated a multi-host pattern rather than restriction to a single lineage (Figure 5.13). The majority of assigned signal mapped to *E. coli* across timepoints, but with detections distributed across multiple reference backgrounds, and with additional contributions from *Klebsiella* (e.g., *K. pneumoniae* and closely related species) at several visits. Overall, *bla*_{CTX-M-15} appears to circulate across a mixed Enterobacteriaceae host within the cohort, consistent with dissemination via mobile genetic elements and/or repeated acquisition from shared reservoirs.

5.4.6.2 Host context of gentamicin-resistance-associated AMEs

Gentamicin-resistance-associated AMEs showed a broader host range than ESBLs, spanning both Gram-negative and Gram-positive genera (Figure 5.12B). In the early window, AME-positive samples commonly mapped to *Escherichia* and *Klebsiella*, with additional contributions from *Enterococcus* and *Staphylococcus*, consistent with the known distribution of aminoglycoside resistance determinants across multiple gut-associated taxa. In later samples (Day 14-90), *Escherichia* remained the most frequent host among endpoint-positive samples, while contributions from other genera varied over time. This pattern indicates that gentamicin resistance potential in this cohort is not confined to a single “pathogen” reservoir but is distributed across common early colonisers.

Host mapping by AME family further supported this mixed ecology (Figure 5.14). AAC-type enzymes were predominantly assigned to Enterobacteriaceae (especially *E. coli* and *Klebsiella*), whereas APH-type enzymes showed more frequent assignments to Gram-positive genera (notably enterococci and staphylococci) in addition to Enterobacteriaceae. Both AAC and APH signals were detectable from the earliest sampled days and persisted across multiple genera through Day 90.

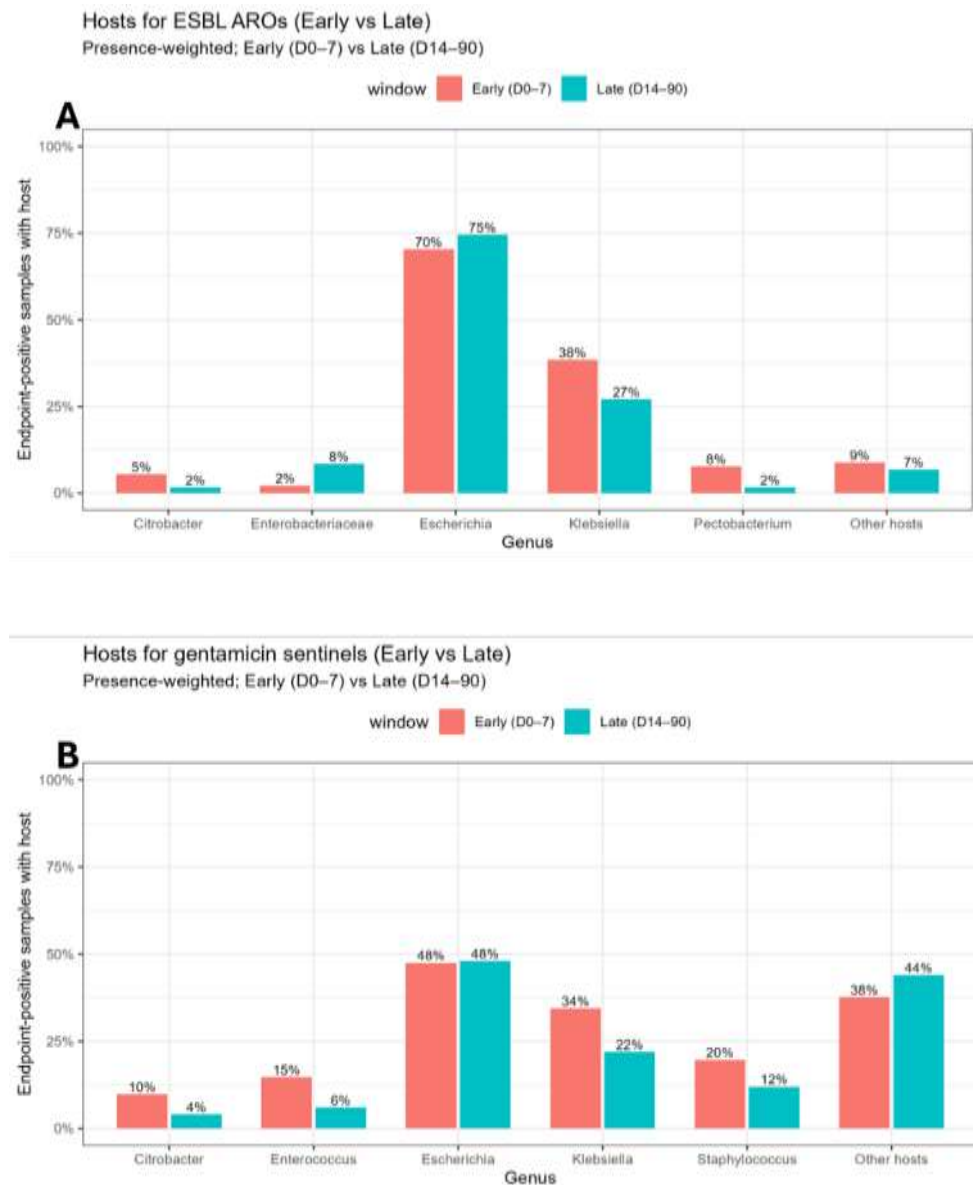


Figure 5.12: Host context of ESBL ARGs and gentamicin-resistance-associated AMEs in early and later samples.

(A) Among samples positive for any ESBL ARG (*bla*_{CTX-M} or ESBL-*bla*_{TEM}/*bla*_{SHV}), the proportion of endpoint-positive samples in which at least one ESBL-carrying read was assigned to each bacterial genus, comparing the first postnatal week (Early, D0–D7) with later samples (Late, D14–D90).

(B) Equivalent analysis for samples positive for any gentamicin-resistance-associated AME sentinel. Bars show presence-weighted percentages within each time window (a sample can contribute to multiple genera if ARG-carrying reads map to more than one host); the category “Enterobacteriaceae” refers specifically to reads that were resolved to the family level but remained unclassified at the genus level, while “Other hosts” combines all remaining genera occurring at low frequency.

Note: Percentages across genera within a time window do not sum to 100% because individual ARG-positive samples often had ARG-carrying reads assigned to multiple genera.

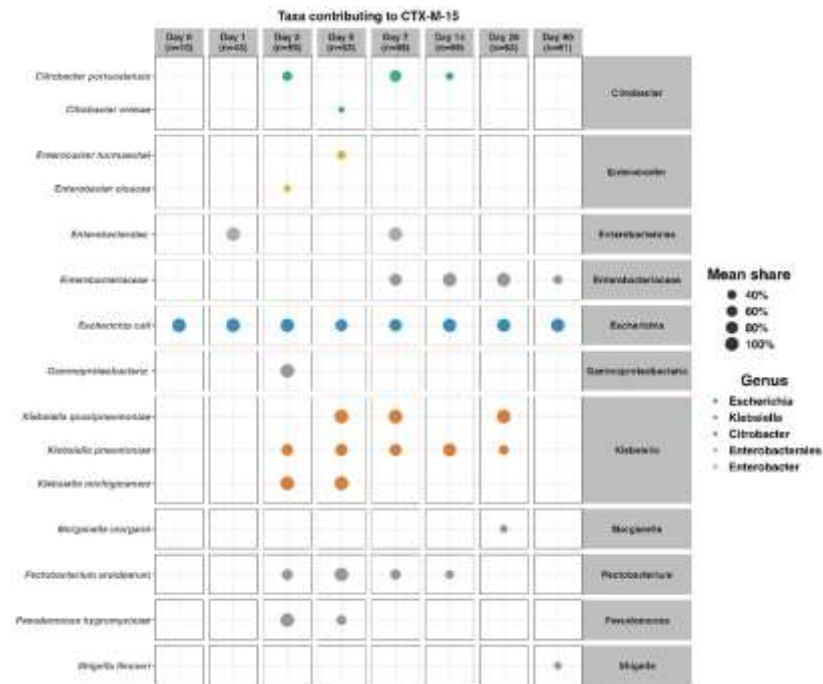


Figure 5.13: Taxa contributing to *bla*_{CTX-M-15} over time

Presence-weighted summary of putative bacterial hosts carrying *bla*_{CTX-M-15} at each scheduled day of life. Within each day, rows list taxonomic assignment –at the species, genus, or family level depending on the resolution –to which *bla*_{CTX-M-15} reads were assigned by the long-read host-resolution pipeline. Point size indicates the mean proportion of *bla*_{CTX-M-15} reads attributed to that taxon among *bla*_{CTX-M-15}-positive samples; point colour indicates genus (*Escherichia*, *Klebsiella*, *Citrobacter* or Enterobacteriales/Enterobacteriaceae). Taxa contributing $\geq 5\%$ on at least one day are shown.

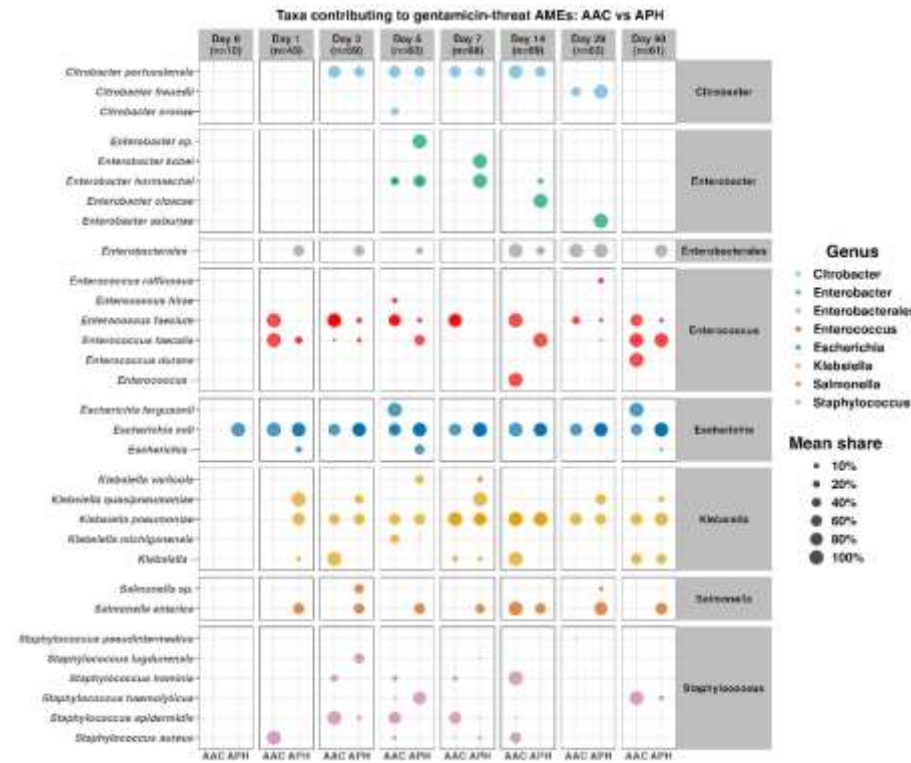


Figure 5.14: Taxa contributing to gentamicin-resistance-associated aminoglycoside-modifying enzymes (AAC and APH) over time.

Presence-weighted summary of recognised bacterial hosts carrying aminoglycoside-modifying enzymes predicted to compromise gentamicin activity (AAC and APH families) at each scheduled day of life. Within each day, rows list taxonomic assignments—at the species, genus, or family level depending on the resolution achieved by the long-read host-resolution pipeline—to which the AME reads were assigned. Point size shows the mean proportion of AME reads attributed to that taxon among positive samples; point colour indicates genus. Taxa contributing $\geq 5\%$ on at least one day are shown.

5.4.6.3. Persistence of β -lactam and aminoglycoside resistance once acquired

To understand how long colonisation with β -lactam and aminoglycoside ARGs persisted once acquired, time-to-loss analyses were performed for infants who acquired resistance to each class at least once. For β -lactam ARGs, the median run length of the longest positive occurrence was 5 consecutive visits (IQR width 3 visits), corresponding to a median of 85 days with β -lactam ARGs detectable (Table 5.4). Aminoglycoside ARGs showed less persistence over time, with a median of 4 visits (IQR width 3) and median duration of 27 days. When β -lactam and aminoglycoside ARGs are both present in a sample, the longest co-colonisation occurrence also spanned a median of 4 visits and 25 days. These persistence estimates assume continuous carriage between positive samples; however, our sampling intervals (ranging from 1-2 days early to 28-90 days later) mean we cannot distinguish between true persistence and repeated re-acquisition of the same ARG class. Intermittent loss and re-acquisition from environmental or maternal sources would appear as persistence in this analysis. This limitation is discussed further in Section 5.5.

Retention estimates after acquisition were consistent with these patterns. By 28 days after first detection, an estimated 79% (95% CI roughly 70-88%) of infants remained colonised with bacteria carrying β -lactam ARGs, compared with about 62% (52-74%) for aminoglycoside ARGs and 61% (51-73%) for concurrent β -lactam and aminoglycoside ARGs combination (Table 5.4). By 90 days after first detection, retention falls only to ~57% (36-91%) and ~41% (29-57%) for β -lactam- and aminoglycoside-colonised infants respectively, with similar values for combined colonisation. The survival curves in Figure 5.15 illustrate these dynamics, with β -lactam resistance showing a shallower cumulative loss trajectory than aminoglycoside resistance.

Transition analyses between visits revealed that antibiotic ARGs remained stable over short intervals but were gradually lost across longer timepoints. During the first postnatal month, ARGs showed high persistence between adjacent visits: $\geq 85\text{-}95\%$ of infants remained positive for β -lactam, while $\geq 75\text{-}90\%$ maintained aminoglycoside resistance or concurrent β -lactam + aminoglycoside carriage. In contrast, retention rates declined extended intervals, such as D14-D28 and D28-D90 (Table S5.2). This decline was particularly pronounced for aminoglycosides, suggesting a gradual clearance or replacement of resistant strains as the gut microbiome matures.

Table 5.4: Run length and retention of β -lactam and aminoglycoside resistance.

(i)	Analysis	ARG Class	Median		Median	IQR
			visits	IQR visits	days	days
	Run length Summary	Aminoglycosides	4	3	27	76
	Run length Summary	Beta-lactams	5	3	85	76
	Run length Summary	Both	4	3	25	76
(ii)	Analysis	ARG Class	day	% still pos.	lower	upper
	Retention Rate	Aminoglycosides	28	61.88	51.74	74.01
	Retention Rate	Aminoglycosides	90	40.84	29.14	57.24
	Retention Rate	Beta-lactams	28	79.11	70.46	88.82
	Retention Rate	Beta-lactams	90	57.34	35.93	91.49
	Retention Rate	Both	28	60.92	50.81	73.03
	Retention Rate	Both	90	39.38	27.68	56.03

For infants ever colonised with β -lactam, aminoglycoside, or both classes of ARGs, the table summarises (i) median number of consecutive positive visits and median duration (in days) of the longest positive occurrence, with IQR widths, and (ii) the percentage of infants still positive for each class at D28 and D90 after first detection, with 95% confidence intervals. A supplementary

transition table (Table S5.2) provides, for each interval, the proportion of infants remaining positive vs loss between visits.

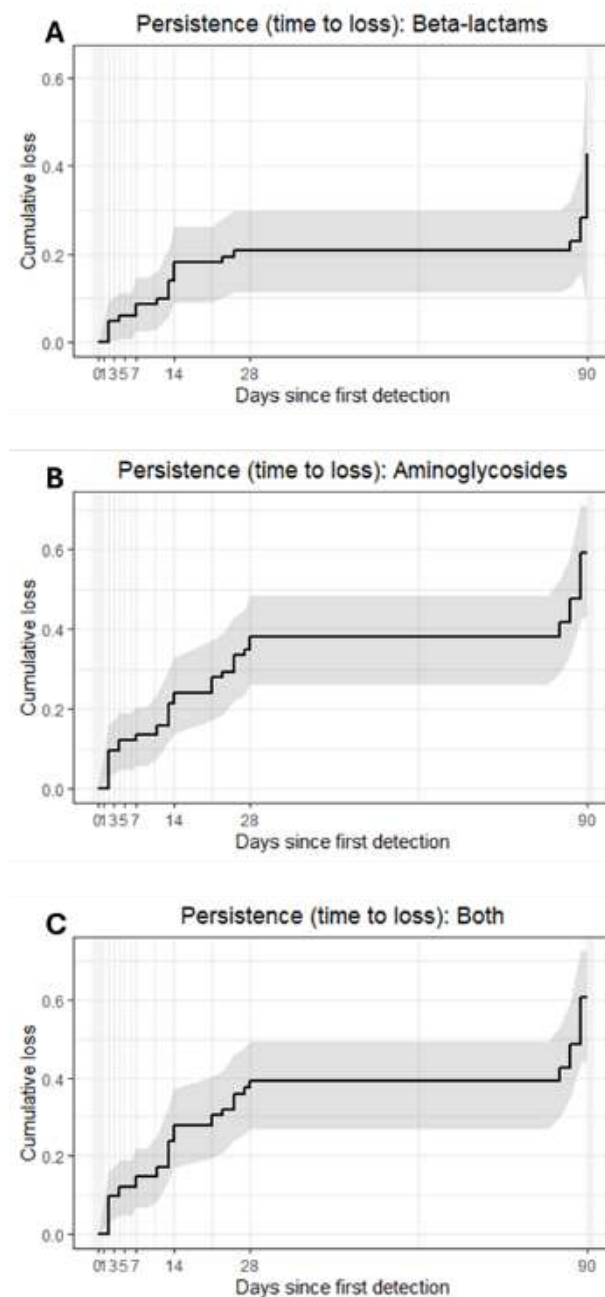


Figure 5.15: Persistence of β -lactam and aminoglycoside resistance once acquired.

Cumulative incidence curves for loss of colonisation with (A) β -lactam ARGs, (B) aminoglycoside ARGs, and (C) concurrent β -lactam plus aminoglycoside ARGs (“Both”) among infants who ever acquired ARGs to the relevant class. Time zero is the first positive sample for that class. The y-axis shows the cumulative proportion of colonised infants who had lost the class by each time point; shaded areas indicate 95% confidence intervals from non-parametric survival estimates.

5.4.6.4 *Aminoglycoside-modifying enzyme family profiles over time*

To characterise whether aminoglycoside resistance composition changed over time, aminoglycoside-modifying enzymes were grouped into four families, AAC, APH, ANT and AAD, and their prevalence and acquisition timing were compared across days of life (Figure 5.16A-B).

Across the cohort, APH genes dominated the AME class at all ages. Around half of infants were APH-positive at birth, rising to roughly 70-75% by D3-D7 and plateaued at $\geq 60\%$ thereafter, with a slight increase again by D90. In contrast, AAC and ANT families were less common but still widespread, with prevalence trajectories that increased over the first week and then plateaued. AAC genes were detected in about one-third of infants by D5-D7 and in just under 30% by D90, while ANT genes reached prevalences of $\sim 15\text{-}20\%$ at their peak. AAD genes were the rarest group: prevalence remained below 10% at all timepoints, with only a small increase by D90. Thus, by three months of age a large majority of infants acquired at least one APH ARG, whereas AAC, ANT and particularly AAD ARGs were present in smaller subsets of the cohort.

Time-to-first-detection curves mirrored these patterns: APH genes were acquired earliest and by the largest fraction of infants, while AAC and ANT accumulated more gradually and remained confined to subsets of the cohort; AAD detections were infrequent throughout follow-up. Collectively, these results indicate that early-life aminoglycoside resistance in this cohort is dominated by phosphotransferase mechanisms, with additional layers contributed by acetyl- and nucleotidyl/adenylyltransferases in smaller subsets of infants.

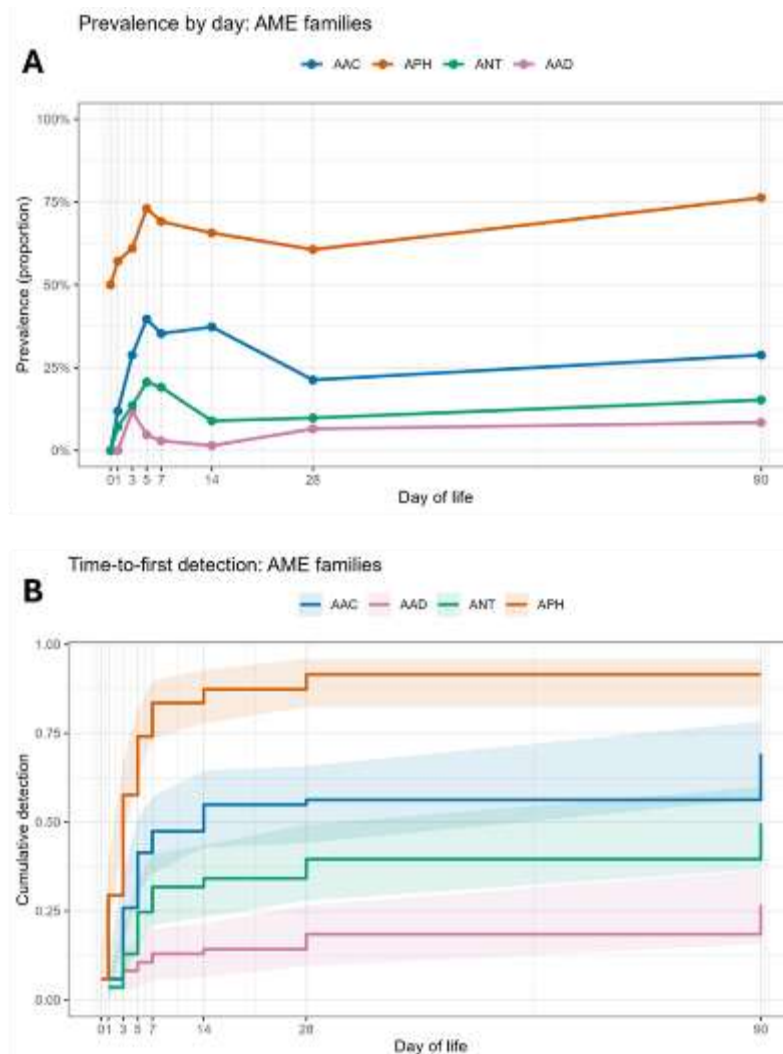


Figure 5.16: Aminoglycoside-modifying enzyme family dynamics in the first 90 days of life.

(A) Prevalence of each AME family by sampling day, expressed as the proportion of infants carrying at least one gene from that family at each visit (AAC, APH, ANT, AAD). (B) Kaplan-Meier curves for time to first detection of each AME family, with shaded bands indicating 95% confidence intervals.

5.4.7 Rare but high-stakes ARG sentinels

To assess carriage of particularly concerning resistance mechanisms, a panel of “high-risk” ARG sentinels was defined, comprising broad and narrow vancomycin clusters (*van* lenient vs stringent), methicillin resistance gene (*mecA*), macrolide/lincosamide/streptogramin B resistance gene (*ermT*), putative plasmid-mediated colistin resistance (*mcr*), 16S rRNA

methylase *rmtB* and *cfr*-family genes associated with linezolid and phenicol resistance. The lenient *van* definition flags any *van*-associated marker as a broad screen, whereas the stringent definition requires co-detection of a ligase/cluster marker plus at least one accessory component (*vanH* and/or *vanX*) within the same sample.

Across all samples, the lenient *van* group were more frequent (~380 positive samples) compared to the stringent *van* group (~80 positives). The other sentinels were rare: *ermT* and *mecA* together accounted for about 50 positive samples, while *mcr*, *rmtB* and *cfr*-family genes were each detected only in a few samples.

Prevalence by day of life highlighted a striking contrast between the broad *van* signal and the other sentinels (Figure 5.17). Even at birth, around one quarter of infants had at least one read containing a *van* gene in the lenient cluster, and by D1-D7 prevalence rose to 70–90% of the samples, remaining high thereafter through D90. In contrast, the stringent *van* group remained uncommon throughout: prevalence increased from ~3% at D0 to around 10–20% during the first month and reached only about one third of samples by D90. This contrast suggests that while *van* gene fragments are common in the infant gut microbiome, complete functional operons capable of conferring glycopeptide resistance are present in only a minority of infants during early life.

Macrolide-associated *ermT* and *mecA* followed similar low-frequency patterns. Both were virtually absent in D0 samples, rose gradually to single-digit or 10-14% prevalence by D7-D14, and then fluctuated around ~5–10% at later visits. Signals for *mcr*, *rmtB* and *cfr*-family genes were extremely sparse, with at most one or two positive samples at any given day of life.

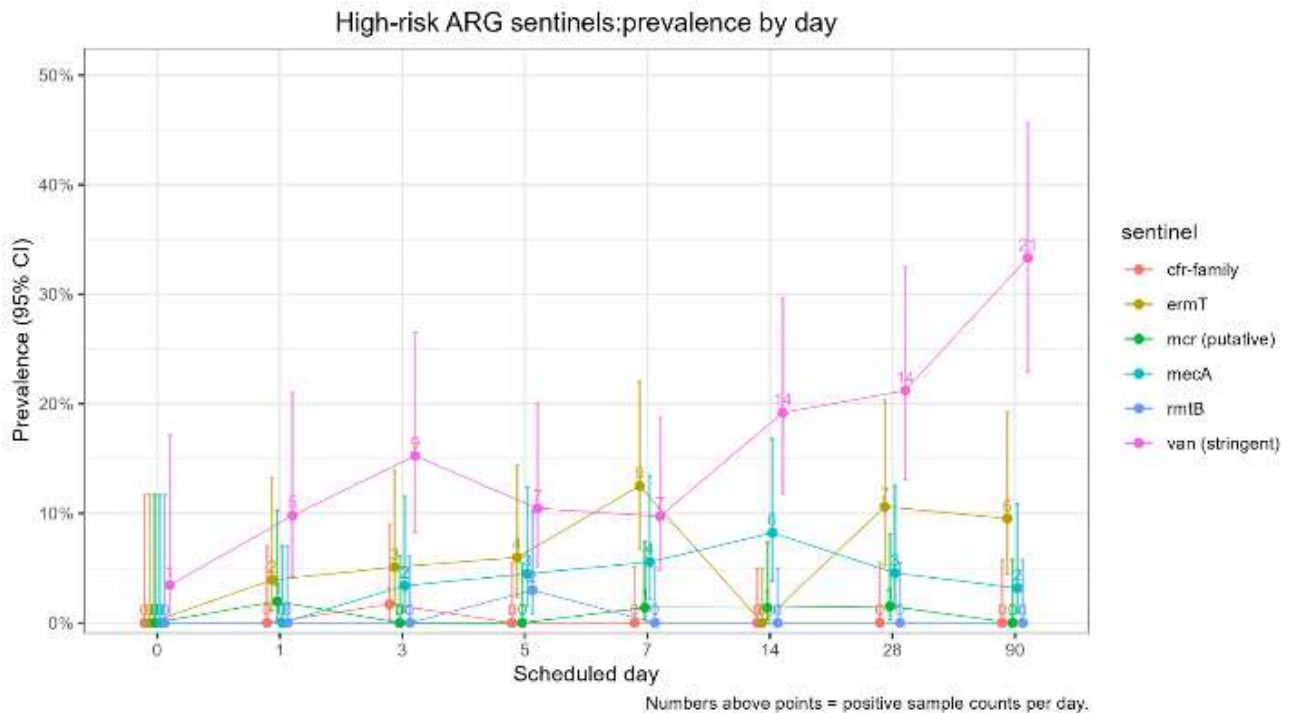


Figure 5.17: Prevalence of high-risk ARG “sentinels” across early life.

Line plots show day-of-life prevalence with 95% binomial CIs; numbers above points are positive sample counts (denominator = samples available at that visit). Sentinels: van (stringent, operon-based), *rmtB*, *mecA*, *mcr* (putative), *ermT*, and *cfr*-family.

Time-to-first-detection analyses were consistent with these prevalence patterns (Figure S5.10). Few infants showed the presence of the “Stringent” van by D28 with approximately half of the infants still negative by D90. For *mecA* and *ermT* genes, only a subset of infants ever acquired the genes during the follow-up and their retention reduced over the study period. Finally, given the very small numbers of positives, formal survival curves were not plotted for *mcr*, *rmtB* or *cfr* genes, but inspection of the raw data confirmed that detections were isolated events in very few infants (see number of positives in Figure S5.10).

5.5 Discussion

This chapter characterised the temporal co-development of the infant gut microbiome and resistome over the first 90 days of life (Section 5.2). The results reveal an immediate and near-universal establishment of therapy-relevant ARGs by the end of the first week. While taxonomic succession followed a canonical progression from facultative to obligate anaerobes, this was marked by significant between-infant heterogeneity and the prolonged persistence of potential pathogen reservoirs. These findings highlight a critical mismatch between current empiric treatment guidelines and the observed colonisation patterns in this setting. This mismatch implies a high potential for first-line treatment failure and suboptimal clinical outcomes, a discrepancy partly driven by the frequent detection of resistance markers for multiple drug classes.

5.5.1 Early microbiome assembly and resistome establishment

The temporal trajectories in this chapter reveal distinct patterns of microbiome maturation and resistome establishment. Taxonomic composition followed an expected global trajectory: initial dominance by facultative anaerobes (*Escherichia*, *Klebsiella*) during the first week, followed by gradual diversification into obligate anaerobes such as *Bifidobacterium* by Days 28–90 (Section 5.4.1). While marked between-infant variability persisted, this succession aligns with diverse geographic settings where facultative anaerobes create the anaerobic niche necessary for subsequent colonisers (Saturio et al., 2021).

Although α -diversity increased from birth through the first week (Shannon index peak Days 3–5), it plateaued thereafter rather than increasing linearly through Day 90 (Section 5.4.1.1). Consistent with other LMIC settings, this indicates that environmental or nutritional

constraints may limit standard maturation trajectories (Amato et al., 2023; Jokela et al., 2024). The striking contrast with HICs is in the timing and extent of resistome establishment. While microbiome diversity evolved gradually, the resistome was established immediately: therapy-relevant β -lactam and aminoglycoside ARGs were detected from birth, with dual-class carriage approaching 100% by Day 7 (Sections 5.4.3, 5.4.4). This differs fundamentally from the Melbourne Infant cohort, where ARG acquisition is more gradual and dominated by tetracycline and macrolide ARGs (Leo et al., 2023). In this Nigerian cohort, the early-life resistome is dominated immediately by β -lactam and aminoglycoside ARG markers underpinning WHO-recommended first-line empiric therapy (ampicillin and gentamicin) for the treatment of neonatal sepsis (Fuchs et al., 2018; World Health Organization, 2013). This suggests an 'initial inoculum' that occurs during delivery and immediate postnatal care. In this setting, the inoculum is shaped by intense colonisation pressure from resistant maternal and environmental reservoirs, a factor likely exacerbated by high maternal peripartum antibiotic use.

Aminoglycoside ARGs on the other hand, showed significantly shorter persistence than β -lactam ARGs (median 27 vs. 85 days; Section 5.4.6.3), likely reflecting their distribution across early colonisers (*Staphylococcus*, *Enterococcus*) that decline as obligate anaerobes establish. However, as our sampling intervals cannot distinguish true persistence from repeated re-acquisition, this distinction remains a critical consideration for the intervention strategies, such as maternal decolonisation versus environmental remediation, explored further in Chapter 7.

5.5.2 Multi-class resistance clustering and host-context dynamics

AR classes emerged in structured clusters, with several pairs co-occurring significantly more often than predicted by their individual prevalences (Section 5.4.5). This pattern implicates MGEs as shared "vehicles" capable of trafficking multiple ARGs across diverse bacterial hosts. Three lines of evidence support MGE-mediated co-selection. First, co-presence heatmaps reveal that class pairs involving β -lactams and aminoglycosides were both pervasive (85-90% by Day 7) and statistically enriched beyond independence ($\log_2$ enrichment typically 0.1-0.5; Section 5.4.5.2). While effect sizes were modest, consistency across timepoints indicates structured linkage likely maintained by physical co-location on MGEs or by shared selective pressures (X. Li et al., 2024a). Additionally, host-context analysis confirms that these ARGs are not confined to single lineages (Section 5.4.6). While *bla*_{CTX-M-15} was primarily associated with *Escherichia* and *Klebsiella* (70–75% and 27–38% of ESBL-positive samples, respectively), it was distributed across multiple *E. coli* strains and *Klebsiella* species.

Simultaneously, gentamicin-associated AMEs mapped across both Enterobacteriaceae and Gram-positive genera (*Enterococcus*, *Staphylococcus*). This distribution is driven by the dominance of phosphotransferase (APH) genes, which reached 70% prevalence by the first week (Section 5.4.6.4). This APH-heavy profile contrasts with a global meta-analysis of infant resistomes where Patangia et al. (2024) observed that geographic and socioeconomic factors often correlate with lower ARG abundance in resource-limited settings, whereas this cohort exhibits rapid, near-universal acquisition. This multi-host distribution is a hallmark "field signature" of mobile resistance, where ARGs appear across taxonomically distant taxa because the underlying platform is highly transferable (Baquero, 2004; Partridge et al., 2018).

Furthermore, the early detection of these clusters (Day 0-1) suggests that infants are being colonised by pre-stabilised MDR modules. The fact that these multi-class clusters appear across taxonomically diverse 'founder' species within hours of birth suggests that horizontal transfer, likely occurring within maternal or environmental reservoirs, has already linked these ARGs onto mobile platforms before they reach the infant. This immediate presence of multi-host resistance is more consistent with the acquisition of integrated MGE cassettes than with the independent acquisition of multiple, unlinked resistant strains (Partridge et al., 2018).

This co-selection mechanism has critical implications for antibiotic stewardship. When ARGs are physically linked, selection pressure from a single antibiotic can inadvertently maintain resistance to multiple unrelated classes (collateral damage) (Lopatkin et al., 2017). In this cohort, the enriched co-presence of β -lactam and aminoglycoside resistance suggests that empiric gentamicin use could maintain ESBL carriage even without β -lactam exposure, and vice versa.

The modest shifts in resistome class composition over time (increasing tetracycline and glycopeptide contributions at later visits; Section 5.4.2.3) likely reflect underlying bacterial community changes, as these increases coincide with *Bifidobacterium* emergence (Section 5.4.1). However, the persistent dominance of β -lactam and aminoglycoside classes despite microbiome maturation suggests that core resistance potential is maintained across community transitions, likely through MGE-mediated horizontal transfer. While complete plasmid reconstruction was not performed, the combination of statistical enrichment, multi-host distribution, and early detection is consistent with MGE-mediated resistance clustering. This interpretation is further supported by Chapter 6, which shows that antibiotic exposure amplifies pre-existing resistance patterns rather than initiating them *de novo*, consistent with selection acting on MGE-linked modules already present.

5.5.3 Rare but high-stakes resistance markers

The sentinel analysis distinguished between common but potentially non-functional gene fragments and clinically relevant operons (Section 5.4.7). While the lenient *van* cluster was detected in 70–90% of infants by Day 7, the stringent cluster (functional operons) reached only 10–20% by the first month. This suggests that while *van* gene fragments are widespread, likely reflecting commensal organisms carrying incomplete operons, functional glycopeptide resistance remains a minority trait. This aligns with findings from high-income cohorts where *van* genes are often detected in the infant gut without corresponding phenotypic resistance (K. Pärnänen et al., 2018).

Other high-stakes markers, including *mecA*, *mcr*, and *rmtB* (16S rRNA methylase), were rare or absent throughout the 90-day period. This indicates that while first-line resistance (β -lactam, aminoglycosides) is ubiquitous, last-resort resistance remains uncommon in this community-based cohort. The near-absence of *mcr* in this cohort is consistent with findings in Kolkata, where *mcr-1.1* was detected in pregnant mothers but remained absent in their neonates (Naha et al., 2023). The isolated detection of *mcr-10* in this study aligns with broader genomic surveillance in Nigeria (Portal et al., 2024), suggesting that while plasmid-mediated reservoirs exist, they have yet to achieve high neonatal penetration in this setting.

Similarly, the rarity of *mecA* (5–10% prevalence) is consistent with the lower burden of community-acquired MRSA in LMICs compared to the high prevalence seen in intensive care environments (Beaumont et al., 2024). However, the early detection of stringent *van* clusters in a subset of infants warrants surveillance, as vancomycin-resistant *Enterococcus* (VRE) is an emerging threat in African healthcare facilities (Ssekatawa et al., 2021; Iweriebor et al., 2015). These findings suggest that while the immediate clinical risk from

last-resort pathogens is low, the infant gut already serves as a sentinel reservoir for emerging resistance determinants.

5.5.4 Clinical significance: from gut colonisation to infection risk

The resistome trajectories identified in this chapter have direct clinical significance because they define the resistance landscape that exists in the infant gut before symptomatic infection occurs. The link between gut colonisation and subsequent infection is well-established: Schwartz et al. (2023) demonstrated that gut pathogens frequently precede bloodstream infections in the NICU, with antibiotic pressure often selecting for these resistant strains. In LMIC settings, this burden is substantial; a recent meta-analysis Beaumont et al. (2024) found that nearly one-third of infants are colonised with resistant Enterobacteriaceae early in life, highlighting the gut as a critical reservoir for the antibiotic-resistant sepsis seen in these cohorts.

Standard WHO-recommended empiric regimens for neonatal sepsis rely on ampicillin combined with gentamicin (World Health Organization, 2013). However, by the end of the first week, the majority of infants in this cohort already carried β -lactam and aminoglycoside resistance determinants (dual-class carriage approaching 100% by Day 7), often in the same bacterial species (*E. coli*, *Klebsiella*) that cause neonatal sepsis. Co-carriage of AmpC β -lactamases plus aminoglycoside resistance was present in 64% of infants by Day 7, rising to 77% by Day 90; ESBL plus aminoglycoside resistance was present in 50% by Day 7, rising to 64% by Day 90 (Table 5.2). When these resistance determinants are present in *E. coli* and *Klebsiella*, which together account for 40-60% of culture-confirmed neonatal sepsis cases in sub-Saharan Africa (Okomo et al., 2019), the potential for empiric treatment failure is substantial.

The timing is as critical as the prevalence. Detection at Day 0-1 suggests that the earliest colonising exposures, spanning delivery, immediate postnatal care, and early ward contact, are sufficient to establish therapy-relevant resistance. This compressed timeline creates a fundamental mismatch: guidelines assume that early-onset infections are reliably covered by narrow-spectrum first-line regimens, but in this setting, infants are colonised with resistant organisms from birth. This mismatch has implications for both clinical management and public health strategy, as discussed in detail in Chapter 7.

5.6 Conclusion

This chapter has characterised the rapid assembly of the neonatal microbiome and the parallel, "compressed" establishment of a multi-drug resistant gut reservoir. The data demonstrate that while taxonomic succession follows expected global patterns, the acquisition of therapy-relevant β -lactam and aminoglycoside resistance is near-universal within the first week of life. This resistome is shaped by high colonisation pressure and likely maintained by MGE-mediated co-selection, creating a fundamental mismatch with current international treatment guidelines. These findings establish the biological baseline for Chapter 6, which examines how these established resistance patterns are further amplified by direct antibiotic exposure.

Chapter 6: Early-life modulators of the infant gut microbiome and resistome: an association study

6.1 Introduction

As detailed in Chapter 1, the early-life gut microbiome is a critical determinant of both immediate infant health and long-term outcomes, influencing susceptibility to infection, immune maturation, metabolic programming, and growth (Arrieta et al., 2014; Tamburini et al., 2016). As seen in Chapter 5, the first weeks and months of life represent a period of rapid microbial community assembly, during which the gut is highly dynamic and vulnerable to perturbation (Bäckhed et al., 2015; Stewart et al., 2018). Alongside shifts in taxonomic composition, it is now well established that infants' guts harbour a substantial burden of ARGs, even in the absence of direct antibiotic exposure (Carvalho et al., 2022).

A set of recurrent modulators shape microbiome and resistome trajectories (Section 1.3, Table 1.1). Diet, particularly breastfeeding intensity, has been associated with enrichment of protective taxa such as *Bifidobacterium* and *Bacteroides* and suppression of facultative anaerobes like Enterobacteriaceae, which frequently serve as reservoirs for ARGs (Lyons et al., 2020; Theophilus & Taft, 2023). Exclusive breastfeeding has been shown to reduce both total ARG abundance and the prevalence of clinically relevant resistance markers compared to formula feeding or mixed feeding, likely through combined effects of prebiotic human milk oligosaccharides, antimicrobial factors, and reduced exposure to environmental microbes via feeding equipment (Nadimpalli et al., 2024; K. Pärnänen et al., 2018).

Mode of delivery influences the initial inoculum: while vaginal birth facilitates the vertical transmission of maternal gastrointestinal and vaginal flora, including commensal

Lactobacillus, *Prevotella*, and *Bacteroides* species, caesarean delivery is associated with disrupted seeding, delayed colonisation by obligate anaerobes, and increased exposure to skin- and hospital-associated organisms (Dominguez-Bello et al., 2010; Reyman et al., 2019; Shao et al., 2019). These delivery-mode signatures can persist for months to years and have been linked to increased carriage of ARGs, particularly in healthcare settings with high background AMR prevalence (Gasparrini et al., 2019).

Furthermore, systemic antibiotic exposure can profoundly disrupt this successional process, reducing diversity, depleting beneficial obligate anaerobes, and selecting for transferable ARGs with effects that may persist well beyond the treatment period (Luchen et al., 2023; Patangia, Grimaud, O'Shea, et al., 2024a). Neonatal antibiotic exposure has been associated with increased abundance of Enterobacteriaceae, enrichment of β -lactam and aminoglycoside resistance genes, and altered metabolic profiles that may influence long-term health outcomes including obesity, asthma, and inflammatory bowel disease (Dierikx et al., 2020; Stewart et al., 2018).

These clinical and dietary influences operate within broader environmental and structural contexts, including the high background AMR pressure and constraints in WASH infrastructure described for this cohort in Chapters 1 and 4. While these environmental factors are not modelled as direct covariates, they are essential for the ecological interpretation of the findings presented herein.

This chapter builds upon that foundational work described in the study design (Chapter 2), the baseline characteristics (Chapter 4), and the descriptive temporal characterisation of microbiome and resistome dynamics (Chapter 5). Here, the independent and potentially interactive effects of mode of delivery, systemic antibiotic exposure, and breastfeeding intensity are examined in relation to variability in gut profiles, while accounting for perinatal

confounders (gestational age, birthweight, and sex). Outcomes include bacterial α -diversity, community composition, and colonisation by selected commensal and opportunistic species. Resistome outcomes include total ARG burden, resistome composition, and carriage of clinically prioritised β -lactamase markers, specifically CTX-M-15, the most prevalent ESBL genotype globally and a primary driver of multidrug-resistant infections in sub-Saharan Africa (Bevan et al., 2017; Musicha et al., 2017).

Despite extensive study of these modulators in high-income settings, several gaps remain that this chapter is designed to address. Most existing work has focused on taxonomic composition, often using 16S rRNA sequencing, with relatively little attention to the resistome. Where resistance has been studied, analyses are frequently limited to targeted genes in cross-sectional designs and rarely include African populations, where exposure patterns, healthcare practices and background AMR burden differ from high-income settings (Musicha et al., 2017; Okomo et al., 2019). The landmark study by Maghini et al. (2025) provided the first comprehensive characterisation of the African gut microbiome, demonstrating distinct taxonomic profiles compared to Western populations, with higher *Prevotella* and lower *Bacteroides* abundance reflecting differences in diet, environment, and host genetics. However, that study focused on adults using single timepoint sampling, and significant gaps remain in understanding early-life gut colonisation across diverse African settings.

The present chapter uses the rich, longitudinal dataset to address these gaps by asking three related questions:

1. How are mode of delivery, neonatal antibiotics and breastfeeding pattern associated with the colonisation of key commensal and opportunistic taxa in this high AMR Nigerian setting?

2. How do these same modulators relate to total ARG burden, resistome composition and carriage of clinically important β -lactam resistance determinants over the first 90 days of life?
3. Do these early-life "hits" combine to concentrate the carriage of clinically important β -lactamase determinants, such as *bla*_{CTX-M-15}, during the neonatal period?

6.2 Aims

This chapter quantifies the associations between key early-life modulators, mode of delivery, neonatal antibiotic exposure, and infant feeding patterns, and the developmental trajectories of the gut microbiome and resistome over the first 90 days of life in this Nigerian birth cohort. The work therein considers:

- 1) Modulators of microbiome and resistome development (D0-D90): To determine how mode of delivery, infant antibiotic exposure, and feeding pattern are associated with:
 - (a) microbiome diversity and resistome burden across the first 90 days of life, with formal inference prioritised at clinically interpretable timepoints (D5, D28, D90);
 - (b) overall microbiome and resistome composition over time; and
 - (c) colonisation by pre-specified commensal and opportunistic taxa during early life.
- 2) Infant antibiotics: scheduled-timepoint differences and acute perturbation: To assess associations between infant antibiotic exposure and microbiome/resistome outcomes, including clinically important β -lactam resistance markers (focusing on *bla*_{CTX-M-15}), using:
 - (a) cross-sectional models at D5, D28 and D90 based on prior antibiotic exposure at each sampling timepoint; and

(b) acute-window analyses among exposed infants indexing samples to first antibiotic initiation (pre-exposure, 0-7 days, 8-30 days).

3) Combined modulators and high-risk resistance colonisation (cumulative exposure and effect modification). To examine whether combined early-life exposures are associated with high-risk resistance colonisation, using:

(a) a cumulative exposure score reflecting caesarean delivery, infant antibiotic exposure, and low/no breastfeeding in relation to *bla*_{CTX-M-15} detection and (as a sensitivity outcome) a composite ESBL marker outcome at days 28 and 90; and

(b) exploratory effect modification analyses assessing whether antibiotic associations differ by mode of delivery or feeding pattern, using a small number of pre-specified interaction questions and/or stratified summaries where cell sizes permit.

Hypotheses: It was hypothesised that C-section and infant antibiotic exposure would be associated with differences in early microbiome maturation and higher carriage of clinically relevant β -lactam resistance markers at scheduled timepoints, and that antibiotic exposure would be associated with measurable short-term changes in microbiome diversity and resistome burden in the period after the first antibiotic course.

6.3 Chapter-specific methods

6.3.1 Early-life modulators

Mode of delivery was analysed as a binary exposure (C-section vs vaginal delivery). Infant feeding was classified at each sampling timepoint using maternal report for the interval preceding sample collection and collapsed into a three-level breastfeeding intensity variable

consistent with WHO definition of breastfeeding categories and prior epidemiologic practice in infant feeding research (Aarts et al., 2000; Burgin, 1996; Dharel et al., 2020): exclusive breastfeeding (EBF), predominantly breastfed with water and/or supplements (High_BF), and low/no breastmilk (Low_no_BF). This grouping was used to reflect breastfeeding intensity in the first 90 days while retaining adequate sample sizes within categories. For selected analyses requiring a simpler, clinically interpretable contrast (and to avoid sparse Low_no_BF sample size), feeding was additionally represented as a binary EBF vs non-EBF variable (non-EBF: combining High_BF and Low_no_BF).

Infant antibiotic exposure was defined at each sampling timepoint as any prior systemic antibiotic administration from birth up to sample collection timepoint (Yes/No). For acute-window analyses among antibiotic-exposed infants, the day of life of first antibiotic initiation (abx_start_day) was identified. To define specific windows of effect, each sample was timed relative to the day of initiation (days_since_abx_start), categorising them into pre-exposure, 0-7 days (inclusive of the initiation day), and 8–30 days post-initiation (sub-acute window). Antibiotic regimen intensity at initiation (start day) was defined as narrow β -lactam only (specifically penicillins: amoxicillin, ampicillin, cloxacillin, or flucloxacillin, as monotherapy or in combination) versus broader/intensive (receipt of any broader-spectrum agent and/or combination therapy). The windowing approach was prespecified to capture an acute post-initiation period (first week) and a sub-acute period (up to one month), adapting from infant/NICU studies that index samples to antibiotic timing (Benitez et al., 2025; Coenen et al., 2020; Kleine Bardenhorst et al., 2021; Reyman et al., 2022).

Prespecified covariates were sex, gestational age at birth (weeks), and birthweight (grams), selected a priori as markers of neonatal maturity and early physiological vulnerability that may confound associations between clinical modulators and microbiome/resistome outcomes.

6.3.2 Outcomes

Microbiome outcomes were: (i) Shannon diversity at species level; calculated from relative abundance profiles; (ii) overall community composition using Bray-Curtis dissimilarities on relative abundance, visualised by ordination and tested using PERMANOVA; and (iii) carriage of a pre-specified panel of commensal (*Bifidobacterium spp.*, *Bacteroides fragilis*, *Phocaeicola vulgatus*) and opportunistic pathogen taxa (*E. coli*, *K. pneumoniae*, *Enterobacter hormaechei*, *Citrobacter portucalensis*, *Enterococcus faecium*, *Staphylococcus aureus*), classified as presence/absence based on non-zero relative abundance at each timepoint.

Resistome outcomes were: (i) total ARG burden, defined as the sum of normalised ARG counts per sample and analysed as $\log_{10}(\text{total ARG} + 1)$ to reduce right-skewness and allow interpretation on an approximate fold-change scale; and (ii) resistome composition analysed using centred log-ratio (CLR) transformation and Aitchison distances for ordination and PERMANOVA (implemented as described in Chapter 2). Given the clinical focus on β -lactam resistance in this cohort, β -lactam-specific outcomes included carriage of *bla*_{CTX-M-15} (binary) and, where reported as a sensitivity analysis, a composite ESBL marker outcome using the predefined marker panel described in Chapter 5.

6.3.3 Statistical analysis

6.3.3.1 Scheduled timepoint analyses (D5, D28, D90)

Cross-sectional analyses were performed at D5, D28 and D90 (one observation per infant per timepoint) to provide clinically interpretable adjusted comparisons at key milestones (early neonatal period, end of neonatal period, and early infancy). Continuous outcomes (Shannon diversity; $\log_{10}$ total ARG burden) were analysed using linear regression, and

binary outcomes (taxon carriage; CTX-M-15 carriage; ESBL marker carriage where applicable) were analysed using logistic regression. Models adjusted for sex, gestational age and birthweight, and included the other modulators measured at that timepoint (delivery mode, prior antibiotic exposure, and feeding category). These therefore provide mutually adjusted estimates for the primary modulators at each timepoint.

In addition to these scheduled-timepoint models, outcomes were summarised descriptively across all scheduled sampling days (D0-D90) to characterise trajectories and to contextualise the timepoint-specific estimates.

6.3.3.2 Acute-window antibiotic analyses

Acute changes around antibiotic exposure was assessed among antibiotic-exposed infants only using an exposure-anchored window approach that indexed each stool sample to the infant's first day of antibiotic initiation (Coenen et al., 2020; Kleine Bardenhorst et al., 2021). Samples were classified as pre-exposure, 0-7 days after initiation (inclusive of the start day), or 8-30 days after initiation. To account for repeated sampling within infants, linear mixed-effects models with an infant-specific random intercept were fitted, with antibiotic window as the primary predictor and adjustment for day of life, mode of delivery, feeding category, and the previously prespecified perinatal covariates (sex, gestational age, and birthweight). In secondary analyses, a window $\times$ regimen intensity term was included to evaluate whether window-associated changes differed between infants treated with narrow β -lactam monotherapy versus broader/intensive regimens. Model results are presented as adjusted mean outcomes by window with 95% confidence intervals, and as pairwise contrasts between windows derived from the fitted model.

6.3.3.3 Combined exposure analyses

Combined-exposure analyses used a cumulative “multi-hit” (Bogaert et al., 2023) score defined at the timepoint of interest as the count of: C-section delivery, prior antibiotic exposure at sampling, and Low_no_BF feeding status. The score was categorised as 0, 1, or ≥ 2 hits to preserve stability given sparse 2- and 3-hit strata. Logistic regression models were fitted at D28 and D90 with CTX-M-15 carriage (and the composite high-risk definition where applicable in Results) as the outcome, adjusting for sex, gestational age and birthweight.

6.3.3.4 Missing data and checks

Prevalence estimates are reported as n/N , where N denotes samples with non-missing outcome. Regression analyses were complete-case with respect to outcome and model covariates at the relevant timepoint; “Model N” denotes observations included in the fitted model. Collinearity diagnostics were checked for primary models and did not indicate material multicollinearity. Where sample size are sparse (e.g., very small strata or outcome), results were presented descriptively or interpreted with emphasis on imprecision.

6.4 Results

6.4.0 Results overview: distribution of key modulators and rationale for antibiotic windowing

Mode of delivery, infant antibiotic exposure, and feeding pattern varied across infants and scheduled visits and are summarised here before modulator-specific analyses. Among the 87 infants contributing to the analysed dataset, 22 were delivered by caesarean section and 65 by vaginal delivery (25.3% vs 74.7%). Twenty-six infants received systemic antibiotics

at least once during follow-up and 61 did not (29.9% vs 70.1%) (Table 6.1, Panel A). At scheduled sampling timepoints, prior antibiotic exposure before sample collection increased over follow-up (D5: 11/67 [16.4%]; D28: 15/66 [22.7%]; D90: 21/63 [33.3%]) (Table 6.1, Panel B).

Feeding was classified at each sampling timepoint into EBF, High_BF, or Low_no_BF (Section 6.3). At key scheduled visits, feeding was predominantly High_BF (D5-D90: 79.1-81.8%), with smaller proportions EBF (12.7-14.9%) and Low_no_BF (4.5-6.3%) (Table 6.1, Panel C).

Table 6.1: Distribution of key early-life modulators in the analysed cohort and at scheduled sampling visits.

Panel A: Infant-level (N = 87):

Distribution of key early-life modulators in the analysed cohort and at scheduled sampling visits.

Mode of delivery	Vaginal	65/87 (74.7%)
	C/S	22/87 (25.3%)
	Systemic antibiotics ever	
	Yes	26/87 (29.9%)
	No	61/87 (70.1%)

Panel B: (sample-level; scheduled visits): Prevalence of prior systemic antibiotic exposure by the sampling timepoint (Yes/No) at D5, D28 and D90 among samples assessed at each visit.

Day	N assessed	Prior antibiotics Yes (n/N, %)	Prior antibiotics No (n/N, %)
D5	67	11/67 (16.4%)	56/67 (83.6%)
D28	66	15/66 (22.7%)	51/66 (77.3%)
D90	63	21/63 (33.3%)	42/63 (66.7%)

Panel C: (sample-level; scheduled visits): Feeding status at each sampling timepoint classified into three breastfeeding intensity categories: exclusive breastfeeding (EBF), predominantly breastfed with water and/or supplements (High_BF), and low/no breastmilk intake (Low_no_BF).

Day	N assessed	EBF (n/N, %)	High_BF (n/N, %)	Low_no_BF (n/N, %)
D5	67	10/67 (14.9%)	53/67 (79.1%)	4/67 (6.0%)
D28	66	9/66 (13.6%)	54/66 (81.8%)	3/66 (4.5%)
D90	63	8/63 (12.7%)	51/63 (81.0%)	4/63 (6.3%)

Values are n/N (%), where N denotes samples with non-missing feeding classification at the specified timepoint.

Figure 6.1 summarises individual-level antibiotic exposure histories for the 26 infants who received systemic antibiotics during follow-up (up to D97). Initiation was heterogeneous in timing: while many courses began in the first postnatal week, start days were distributed across the early postnatal period and were not aligned to a single scheduled sampling day. Regimen structure was also heterogeneous, including overlapping combination therapy and sequential courses of varying duration. Given this timing and regimen heterogeneity, antibiotic effects are presented both as scheduled-timepoint “prior exposure” contrasts and using the acute-window approach (Section 6.4.2), where samples are indexed to each infant’s first antibiotic initiation and grouped into prespecified windows.

To provide a single adjusted snapshot across modulators at each key milestone, overview regression models were fitted at D5, D28, and D90 including mode of delivery, prior antibiotic exposure, and feeding category (with adjustment for sex, gestational age, and birthweight). In these overview models, caesarean delivery was associated with lower Shannon diversity at D5, with no clear evidence of differences at D28 or D90 (Table 6.2). Caesarean delivery was also associated with higher total ARG burden at D5 and D28, with no clear evidence of differences by D90 (Table 6.2). Estimates for prior antibiotic exposure and feeding category were generally imprecise in these overview models and are therefore interpreted primarily in the modulator-specific sections that follow.

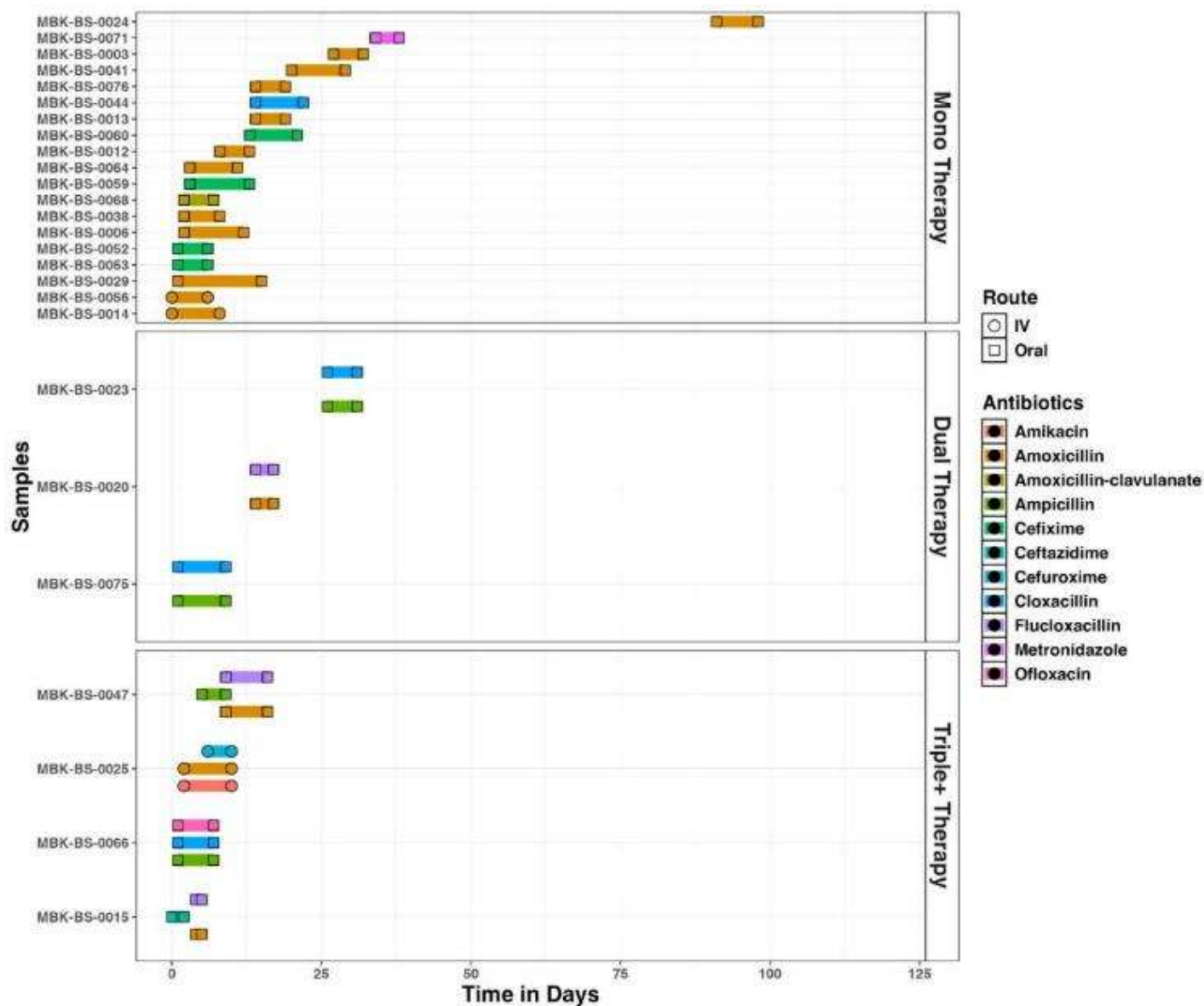


Figure 6.1: Individual-level systemic antibiotic exposure histories during follow-up, demonstrating heterogeneous timing and regimen complexity.

Each row represents one antibiotic-exposed infant, plotted over day of life. Coloured horizontal segments indicate start–stop periods for specific antibiotic course; multiple segments within an infant indicate sequential courses and/or overlapping co-administration. Point symbols indicate route of administration (IV vs oral). Infants are grouped into panels according to regimen structure (mono-, dual-, or triple- therapy). The irregularity of initiation times across the postnatal period, use of multi-therapy and or sequential regimens motivate the exposure-anchored acute-window approach, which indexes stool samples to the first day of antibiotic initiation rather than relying solely on scheduled sampling timepoints.

Table 6.2: Mutually adjusted associations of early-life modulators with microbiome diversity and total resistome burden at scheduled sampling milestones (D5, D28, D90).

Outcome	Model						
	Day	N	C/S vs Vaginal	Prior antibiotics (Yes vs No)	High_BF vs EBF	Low_no_BF vs EBF	
Shannon diversity	D5	67	-0.665 (-1.028 to -0.302); p=0.001	-0.338 (-0.726 to 0.050); p=0.087	-0.659 (-1.080 to -0.239); p=0.003	-0.198 (-0.911 to 0.514); p=0.579	
Shannon diversity	D28	66	0.013 (-0.373 to 0.399); p=0.946	0.010 (-0.372 to 0.392); p=0.958	-0.101 (-0.581 to 0.379); p=0.675	0.185 (-0.653 to 1.023); p=0.660	
Shannon diversity	D90	63	-0.036 (-0.350 to 0.278); p=0.818	-0.005 (-0.269 to 0.259); p=0.969	-0.264 (-0.659 to 0.131); p=0.186	0.152 (-0.437 to 0.742); p=0.606	
log10(total ARG + 1)	D5	63	0.092 (0.044 to 0.139); p=<0.001	0.037 (-0.013 to 0.087); p=0.146	-0.009 (-0.066 to 0.048); p=0.759	-0.045 (-0.150 to 0.061); p=0.400	
log10(total ARG + 1)	D28	61	0.070 (0.028 to 0.112); p=0.002	0.039 (-0.004 to 0.082); p=0.075	-0.009 (-0.073 to 0.055); p=0.780	-0.027 (-0.124 to 0.069); p=0.573	
log10(total ARG + 1)	D90	59	0.001 (-0.028 to 0.030); p=0.951	0.023 (-0.001 to 0.047); p=0.063	-0.029 (-0.064 to 0.005); p=0.095	0.002 (-0.050 to 0.053); p=0.948	

Linear regression models were fitted separately at D5, D28, and D90 to estimate adjusted differences in Shannon diversity and total ARG burden (log10 [total ARG + 1]). Estimates are presented as adjusted mean differences with 95% confidence intervals and p-values for: C-section (C/S) versus vaginal delivery; systemic antibiotics (Yes vs No, by the sampling timepoint); and feeding category contrasts with exclusive breastfeeding (EBF) as the reference (High_BF and Low_no_BF). All models adjust for sex, gestational age, and birthweight, and include the concurrent modulators measured at the same timepoint (delivery mode, prior antibiotic exposure, and feeding category). “Model N” denotes complete cases included in each fitted model.

6.4.1 Mode of delivery: early microbiome and resistome difference

6.4.1.1 Is mode of delivery associated with altered microbiome diversity and resistome burden?

To assess whether mode of delivery was associated with early microbiome maturation and resistome burden, Shannon diversity and total ARG burden ($\log_{10}[\text{total ARG} + 1]$) were compared between C-section and vaginally delivered infants across all scheduled sampling days (D0, 1, 3, 5, 7, 14, 28 and 90). For formal inference at clinically interpretable milestones (D5, D28 and D90; Section 6.3.3.1), models were mutually adjusted as prespecified (Section 6.3.3.1). Descriptive trajectories for both alpha diversity and resistome burden were visualised using boxplots stratified by delivery mode (Figure 6.2A-B).

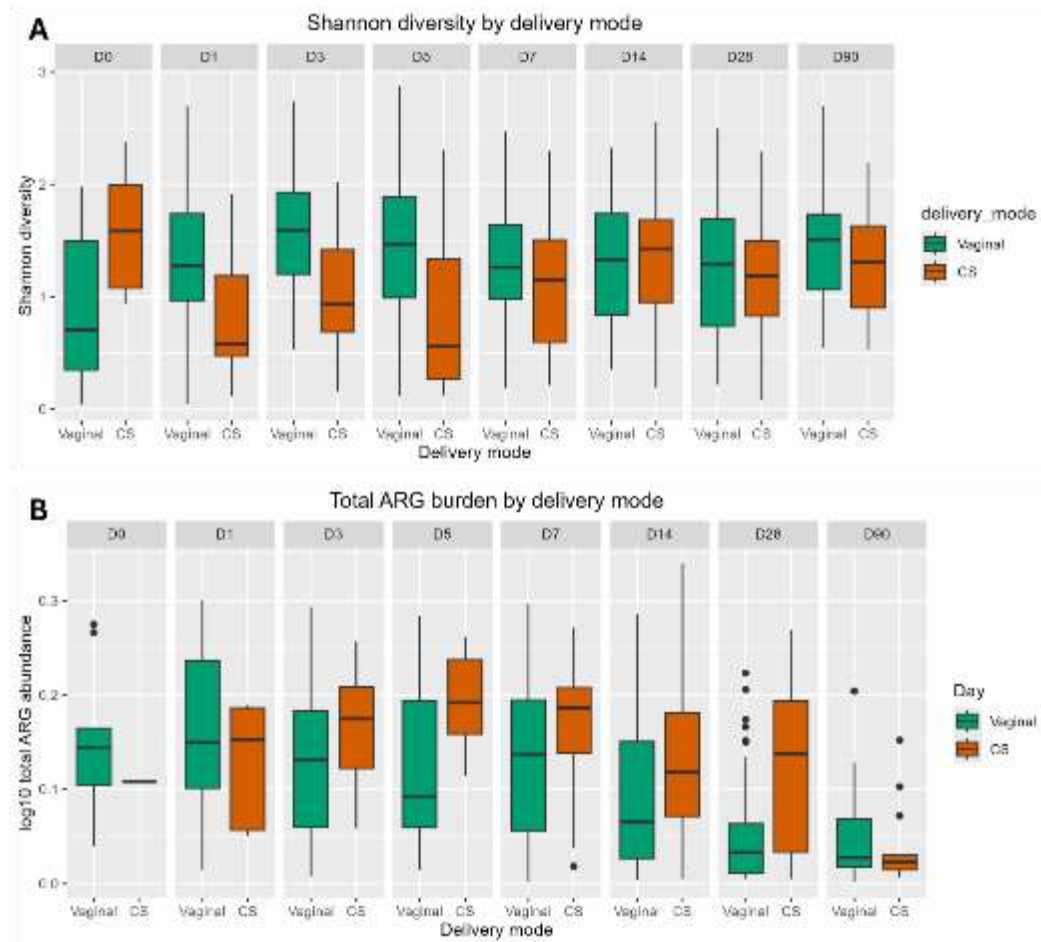


Figure 6.2: Longitudinal distributions of microbiome diversity and total resistome burden by mode of delivery across sampling days (D0-D90)

A) Shannon diversity (species level) stratified by delivery mode (vaginal vs C-section) at each timepoint. (B) Total ARG burden expressed as $\log_{10}(\text{total ARG} + 1)$ at the same visits, stratified by delivery mode. Boxplots summarise the median and interquartile range; whiskers indicate $1.5 \times \text{IQR}$ and points denote outliers. These plots are descriptive (unadjusted) and include all available samples at each timepoint.

Shannon diversity was lower in caesarean-delivered infants at several early neonatal timepoints, with convergence by later infancy. In adjusted mixed-effects models allowing delivery-mode effects to vary by sampling day, the caesarean–vaginal difference was negative at D3 and D5 (e.g., D3: $\Delta = -0.55$, 95% CI -0.88 to -0.21 ; D5: $\Delta = -0.65$, 95% CI -0.98 to -0.33), with little evidence of a difference by D28 ($\Delta = -0.01$, 95% CI -0.33 to

0.31) or D90 ($\Delta = -0.13$, 95% CI -0.47 to 0.20). Visual inspection of the full 0–90-day trajectories is consistent with these model-based contrasts, with early separation narrowing after the first postnatal week and substantial overlap by days 28 and 90.

In contrast, total ARG burden ($\log_{10}[\text{total ARG} + 1]$) was higher in C-section infants during early life, with attenuation by D90 (Figure 6.2B). In adjusted mixed-effects models, delivery-mode differences were most evident from D3 through D28 (e.g., D5: $\Delta = +0.085$, 95% CI $+0.042$ to $+0.129$; D28: $\Delta = +0.070$, 95% CI $+0.029$ to $+0.112$), while no evidence of a difference was observed by D90 ($\Delta = +0.002$, 95% CI -0.043 to $+0.046$). These analyses indicate that caesarean delivery is associated with reduced microbiome diversity at select early neonatal timepoints and an elevated resistome burden through the first month, with convergence by three months.

6.4.1.2 Does mode of delivery change overall community composition?

To evaluate whether mode of delivery was associated with overall gut community structure, taxonomic microbiome and resistome composition were examined. For the microbiome analysis, genus-level relative abundance profiles were visualised in PCoA space using Bray–Curtis dissimilarities, with delivery mode overlaid to identify clustering patterns (Figure 6.3A). Resistome analysis employed centred log-ratio transformation of sample-wise ARG abundances, followed by PCoA using Aitchison distance (Figure 6.3B). Differences in community composition and resistome by delivery mode were tested at key milestone timepoints (D5, D28, D90) using adjusted PERMANOVA models.

Results indicated that delivery mode was associated with a significant shift in taxonomic composition during the first month of life. At D5, delivery mode explained ~6.2% of between-infant Bray-Curtis variation ($R^2 = 0.062$, $P = 0.001$), and a similar magnitude

persisted at D28 ($R^2 = 0.056$, $P = 0.002$). By D90, there was no evidence of a delivery-mode effect on microbiome composition ($R^2 = 0.016$, $P = 0.363$), consistent with the diversity results indicating convergence by three months.

A parallel pattern was observed for the resistome, with delivery mode explaining a measurable fraction of resistome compositional variation early in life and at D28. In adjusted PERMANOVA models, delivery mode was associated with resistome composition at D1 ($R^2 = 0.078$, $P = 0.001$), D3 ($R^2 = 0.048$, $P = 0.010$), D5 ($R^2 = 0.059$, $P = 0.001$), D7 ($R^2 = 0.050$, $P = 0.003$), and D28 ($R^2 = 0.099$, $P = 0.002$), but not at D14 ($R^2 = 0.020$, $P = 0.18$) or D90 ($R^2 = 0.017$, $P = 0.288$). By three months, resistome profiles overlapped substantially across delivery modes. These analyses indicate that mode of delivery exerts a significant effect on both microbiome and resistome composition in early life, with attenuation by D90.

6.4.1.3 Effect of mode of delivery colonisation by key taxa

Moving from global composition patterns to clinically relevant organisms, a pre-specified panel of commensal taxa and common neonatal pathogens/hospital-associated organisms (6.3.2) was examined. For each taxon, delivery-mode effects on $\log_{10}$ relative abundance were estimated at D5, D28 and D90 (Figure 6.4). In parallel, colonisation prevalence (presence/absence) was evaluated at D5 and D28 using adjusted logistic regression (Table 6.3).

At D5, caesarean delivery was associated with depletion of multiple obligate anaerobes. Relative abundances of *B. longum*, *P. vulgatus* and *B. fragilis* were lower in C-section infants (e.g., $\Delta\log_{10}$ -2.56 for *B. longum*; $FDR < 0.01$). Consistent with this, the odds of colonisation with *B. longum* and *P. vulgatus* were reduced in the C-section group (aOR 0.11 and 0.07, respectively; $FDR < 0.05$), with *B. fragilis* showing a similar but borderline pattern. In

contrast, some hospital-associated organisms showed higher relative abundance in C-section infants at D5 (e.g., *E. faecium* and *C. portucalensis*; $\text{FDR} \leq 0.02$), although corresponding prevalence differences were imprecise.

By D28, differences in commensal taxa had reduced but remained evident for selected species (e.g., *B. bifidum* and *B. fragilis*). Over the same period, several Enterobacteriaceae and *E. faecium* showed higher relative abundance in C-section infants, with wide confidence intervals for prevalence contrasts. By D90, most delivery-related differences in potential pathogens were no longer evident, whereas reduced abundance of certain anaerobic commensals (notably *B. fragilis* and *P. vulgatus*) persisted, suggesting delayed colonisation of these taxa despite convergence in global community structure by three months.

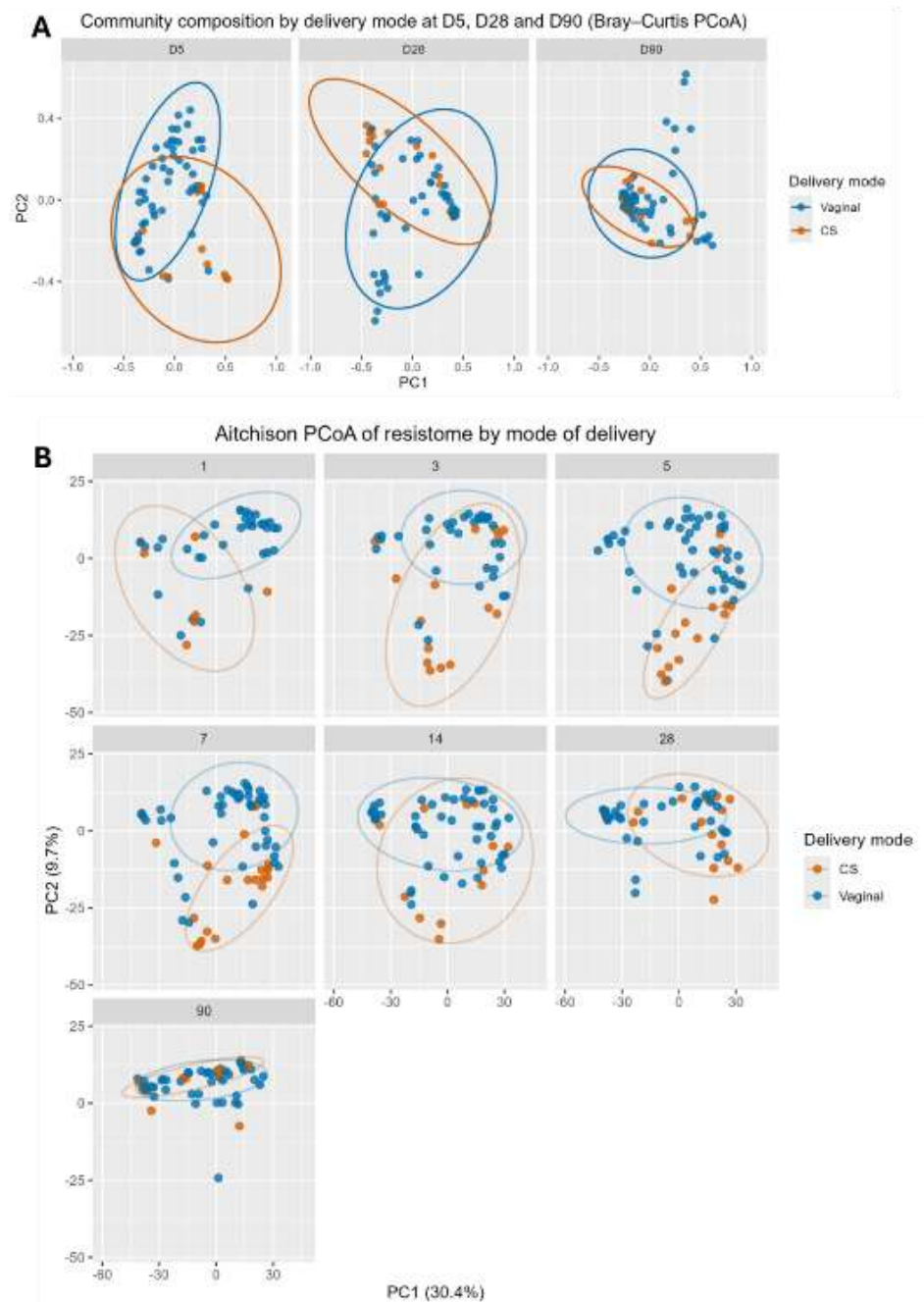


Figure 6.3: Delivery-mode differences in microbiome and resistome community composition across infancy.

(A) Bray-Curtis PCoA of gut microbiome community composition at days 5, 28 and 90 of life, coloured by mode of delivery. Each point represents one infant sample; ellipses indicate the 95% confidence region for each delivery group.

(B) Aitchison PCoA of the gut resistome at days 1, 3, 5, 7, 14, 28 and 90, again stratified by mode of delivery with 95% confidence ellipses. Axes show the first two principal coordinates (PC1 and PC2) with the percentage of variance explained.

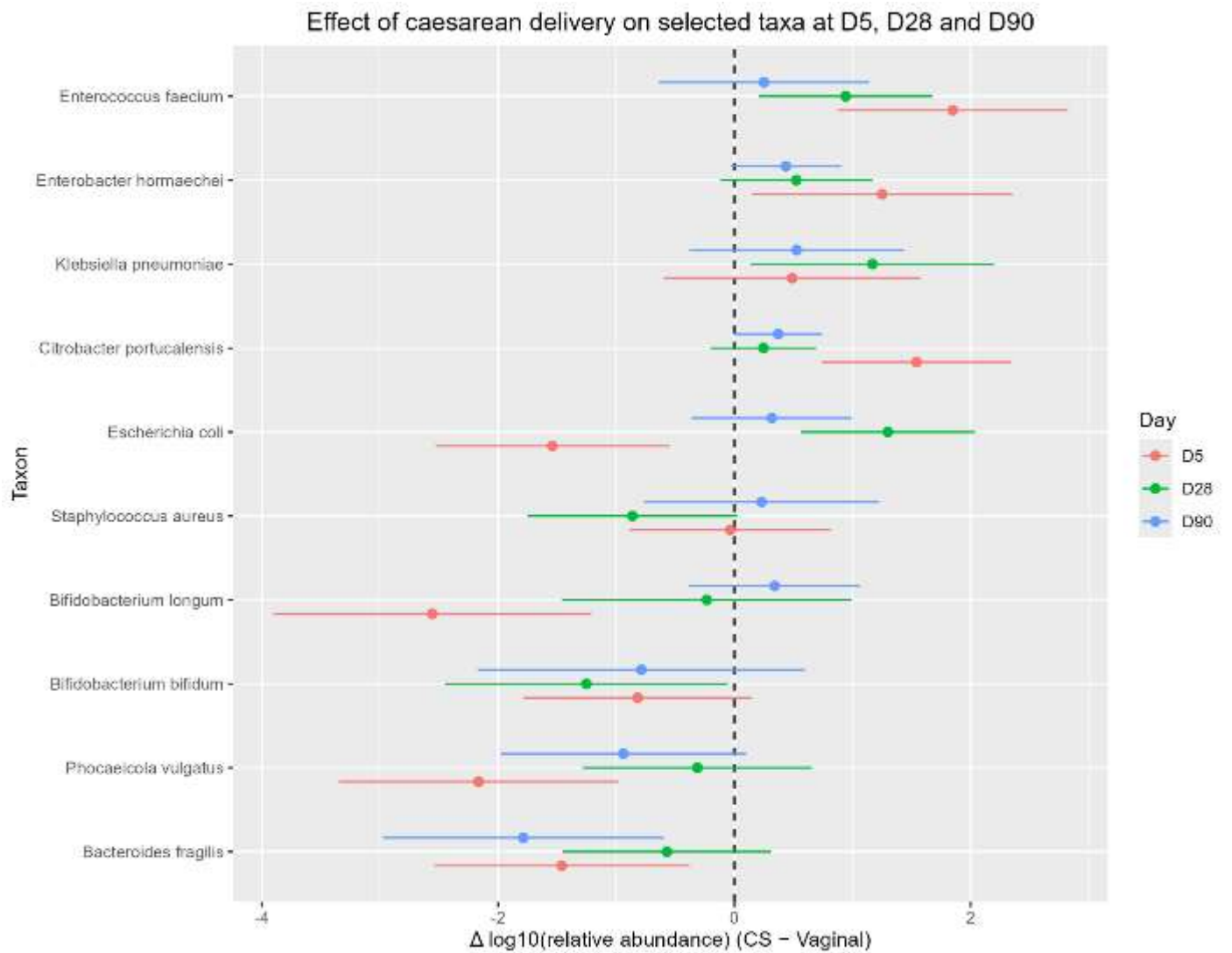


Figure 6.4: Mode of delivery associations with the relative abundance of selected gut taxa.

Each point represents the estimated difference in $\log_{10}(\text{relative abundance})$ between CS-delivered and vaginally delivered infants (CS – vaginal) for the indicated taxon at day 5 (red), day 28 (green) and day 90 (blue); horizontal lines show 95% confidence intervals. Values to the right of the vertical zero line indicate higher abundance in CS infants, and values to the left indicate higher abundance in vaginally delivered infants.

Table 6.3 Adjusted odds ratios for colonisation by selected taxa in caesarean-delivered versus vaginally delivered infants.

Taxon carriage was defined as presence/absence based on non-zero relative abundance at the relevant sampling timepoint. For each taxon and timepoint, logistic regression models estimated the odds of carriage in caesarean-delivered infants relative to vaginally delivered infants (odds ratio, OR), with 95% confidence intervals and p-values.

Taxa	n	Day of Life	OR	OR_lower	OR_upper	p_value	p_FDR
	(Samples)						
Commensals:							
<i>Bacteroides fragilis</i>	67	D5	0.09	0.01	0.45	0.008	0.052
<i>Bacteroides fragilis</i>	66	D28	1.31	0.39	4.56	0.661	0.734
<i>Bifidobacterium bifidum</i>	67	D5	0.31	0.06	1.20	0.112	0.250
<i>Bifidobacterium bifidum</i>	66	D28	0.72	0.21	2.48	0.591	0.734
<i>Bifidobacterium longum</i>	67	D5	0.11	0.02	0.45	0.004	0.038
<i>Bifidobacterium longum</i>	66	D28	4.47	0.57	113.47	0.230	0.328
<i>Phocaeicola vulgatus</i>	67	D5	0.07	0.01	0.33	0.003	0.038
<i>Phocaeicola vulgatus</i>	66	D28	0.53	0.13	1.94	0.351	0.468
Opportunistic/pathogens:							
<i>Citrobacter portucalensis</i>	67	D5	2.32	0.67	8.64	0.190	0.317
<i>Citrobacter portucalensis</i>	66	D28	2.90	0.79	11.40	0.111	0.250
<i>Enterobacter hormaechei</i>	67	D5	1.39	0.38	5.86	0.632	0.734
<i>Enterobacter hormaechei</i>	66	D28	2.90	0.83	12.10	0.112	0.250
<i>Enterococcus faecium</i>	67	D5	3.70	1.07	14.36	0.045	0.149
<i>Enterococcus faecium</i>	66	D28	2.16	0.63	8.03	0.229	0.328
<i>Escherichia coli</i>	67	D5	0.10	0.01	0.86	0.043	0.149
<i>Escherichia coli</i>	66	D28	-	-	-	-	-
<i>Klebsiella pneumoniae</i>	67	D5	0.25	0.04	1.56	0.127	0.254
<i>Klebsiella pneumoniae</i>	66	D28	4.90	0.74	99.08	0.163	0.296
<i>Staphylococcus aureus</i>	67	D5	0.80	0.20	3.53	0.757	0.797
<i>Staphylococcus aureus</i>	66	D28	0.22	0.05	0.89	0.037	0.149

Odds ratio not estimable at *Escherichia coli* D28 due to (quasi-)complete separation; colonisation was almost ubiquitous across groups.

6.4.2 Infant antibiotics and acute-window changes in microbiome/resistome endpoints

6.4.2.1 *Cross-sectional differences in diversity and resistome burden*

To assess whether prior systemic antibiotic exposure was associated with microbiome diversity and resistome burden at scheduled follow-up visits, infants were classified at each sampling timepoint according to whether they had received any systemic antibiotics before or on the day of sampling (“antibiotics prior” vs “no antibiotics prior”). This indicator captures heterogeneous exposure histories (including infants who may still have been receiving therapy at the time of sampling), as described in the antibiotic exposure timelines (Figure 6.1). Analyses were restricted to samples collected at days 5, 28 and 90, representing early (first postnatal week), intermediate (end of the neonatal period) and later (approximately three-month) infant timepoints (Section 6.3.3.1), and aligning with the delivery-mode analyses in Section 6.4.1.

At each timepoint, Shannon diversity and total ARG burden ($\log_{10}[\text{total ARG} + 1]$) were modelled separately using adjusted linear regression (Section 6.3.3.1). These models complement descriptive boxplots of diversity and ARG burden stratified by antibiotic exposure at each scheduled visit (Figure 6.5).

At day 5, infants with prior antibiotic exposure tended to have lower Shannon diversity than those without prior exposure (adjusted coefficient -0.33 ; $P = 0.095$). By day 28, there was no clear association between prior antibiotic exposure and Shannon diversity (coefficient -0.04 ; $P = 0.81$). By day 90, the antibiotics coefficient was essentially zero (-0.01 ; $P = 0.95$), consistent with convergence of overall diversity by three months of age.

Patterns for total ARG burden differed. At all three timepoints, the adjusted coefficients for antibiotic exposure were positive, indicating higher ARG abundance among infants who had received antibiotics, although estimates were small and imprecise. At day 5, antibiotic-

exposed infants had on average 0.03 higher $\log_{10}$ total ARG burden than unexposed infants ($P = 0.17$), equivalent to approximately a 1.07 $\times$ difference. At day 28, the difference was slightly larger (0.04; $P = 0.083$), and at day 90 a similar estimate was observed (0.02; $P = 0.066$; $\sim 1.05\times$). Across these milestone models, other covariates (including feeding pattern, sex, gestational age and birthweight) contributed less consistently to variation in total ARG burden at the available sample sizes.

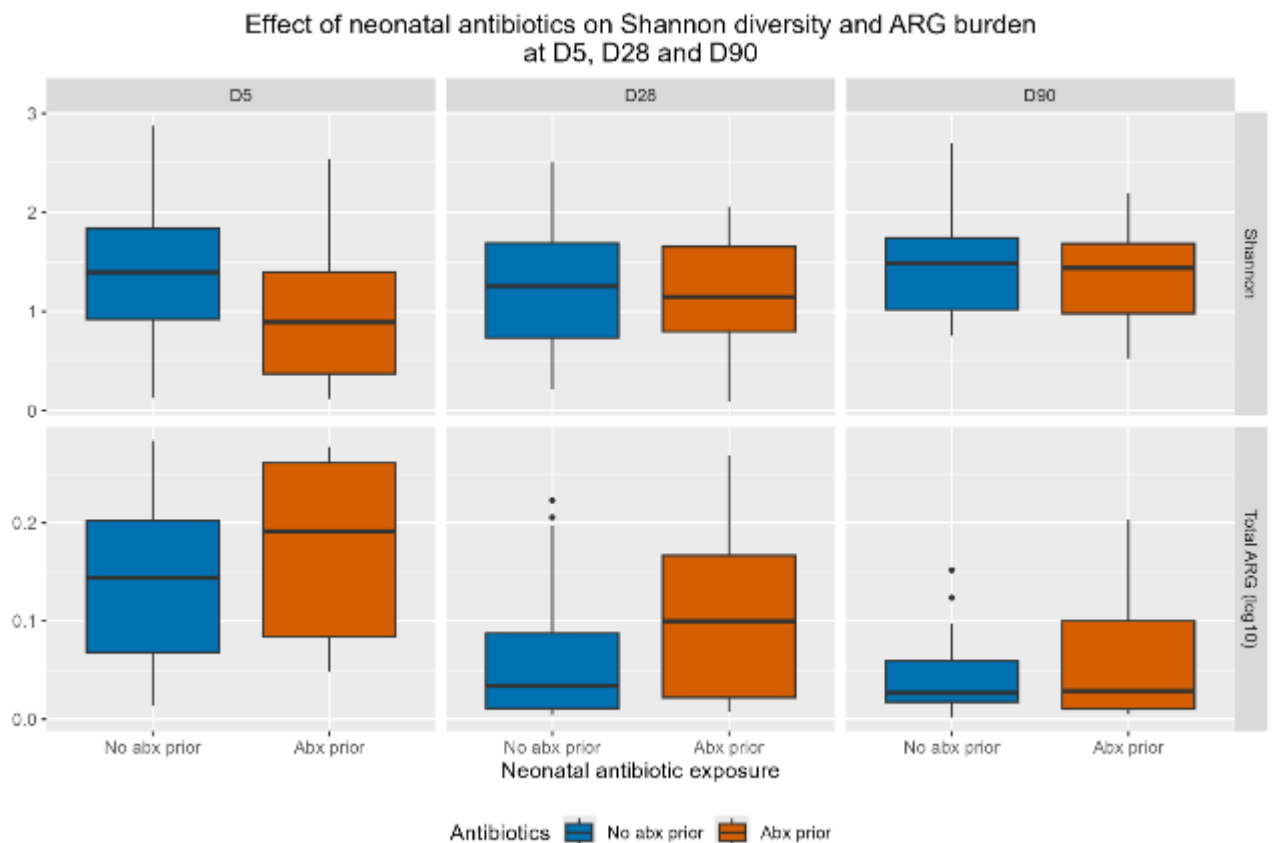


Figure 6.5: Descriptive distributions of microbiome diversity and total resistome burden by antibiotic exposure.

Boxplots compare Shannon diversity (top row) and total ARG burden between infants with prior antibiotics exposure by the sampling timepoint (“Abx prior”) versus those with no prior exposure (“No abx prior”) at D5, D28 and D90. Boxes show the interquartile range with the median; whiskers indicate 1.5 $\times$ IQR and points denote outliers. These panels provide an unadjusted visual summary.

6.4.2.2 Acute changes around antibiotic exposure: windows (Pre / 0-7 / 8-30 days), regimen intensity, taxa, and β -lactam ARG carriage

To examine short-term changes around neonatal antibiotic exposure, samples were classified into three exposure-anchored windows defined by days since first dose: Pre, 0-7 days (inclusive of the initiation day), and 8-30 days post-initiation. Analyses were restricted to infants who received systemic antibiotics, with samples indexed to each infant's first antibiotic start day. For Shannon diversity and $\log_{10}$ total ARG burden, mixed-effects models were fitted with window and regimen intensity (narrow β -lactam only vs broader/intensive) as predictors, including a random intercept for infant and adjustment for day of life, delivery mode, feeding category, and other prespecified covariates.

Visual inspection of within-infant trajectories showed heterogeneity in both Shannon diversity and total ARG burden across the three windows, with no consistent, uniform shift shared across most infants (Figure 6.6A-B). In adjusted interaction models, there was no clear evidence of an overall window effect for Shannon diversity (Type III test for window: $p = 0.537$) and no strong evidence that the window pattern differed by regimen group (window $\times$ regimen: $p = 0.110$). Regimen-specific contrasts suggested limited separation between Pre and 0-7 days in either regimen group, while the narrow-only group showed higher Shannon diversity in the 8-30-day window relative to 0-7 days (0-7 vs 8-30 estimate -0.386, Tukey-adjusted $p = 0.025$), compatible with partial recovery by several weeks after first exposure.

For total ARG burden, the adjusted model provided borderline evidence that values differed across windows (Type III test for window: $p = 0.0685$) and weak evidence for a window $\times$ regimen interaction ($p = 0.0720$) (Figure 6.6B). Regimen-stratified contrasts indicated no clear separation across windows among narrow-only infants. Among broader/intensive infants, estimated Pre values were higher than those in the post-exposure windows (Pre vs

0–7 estimate 0.106, Tukey-adjusted $p = 0.045$; Pre vs 8–30 estimate 0.108, $p = 0.052$). This contrast is sensitive to small and uneven denominators contributing to the broader/intensive Pre window (shown in Figure 6.6B) and should not be interpreted as a treatment-driven reduction in ARG burden; rather, it is consistent with clinical heterogeneity and differences in which infants contribute samples to each window. Overall, after covariate adjustment and accounting for repeated sampling, acute shifts in Shannon diversity and total ARG burden around first antibiotic exposure were not strong or uniform at the cohort level, and regimen-level separation in these global outcomes was limited at the available sample sizes (Figure 6.6A-B). Mixed-model window contrasts are provided in Table S6.1.

Because Shannon diversity and total ARG burden can mask clinically relevant compositional change, selected commensals and opportunistic taxa were examined across the same acute windows, stratified by infant-level regimen group (Figure 6.7A-B). Across commensal taxa (including *Bifidobacterium* species and *Bacteroides/Phocaeicola*) and opportunistic taxa (including Enterobacteriaceae, enterococci and staphylococci), windowed profiles showed overlapping confidence intervals across many strata, with patterns varying by taxon and by infant group. Several taxa were detectable before antibiotics and remained detectable in the post-exposure windows, indicating that first antibiotic exposure commonly occurs against a background of established early colonisation. Where differences were apparent, they were not consistently aligned in the same direction across taxa or regimens, and limited denominators in some strata, particularly within the broader/intensive group, constrain strong inference. These panels are therefore primarily descriptive, highlighting that any acute antibiotic-associated change is more likely to present as shifts in specific taxa than as a large, consistent change in Shannon diversity alone.

Finally, given the clinical focus on β -lactam resistance, carriage of two sample-level endpoints was summarised, any β -lactam ARG and *bla*_{CTX-M-15}, collapsed to infant $\times$ window

presence/absence (any detection within the window), with denominators reported for each stratum. Carriage of any β -lactam ARG was high across windows, offering limited discrimination across regimen strata. In contrast, *bla*_{CTX-M-15} provided more clinically informative resolution: this marker was already present in a substantial proportion of infants in the Pre window and remained detectable across the 0–7 and 8–30 day windows, with wide uncertainty in smaller strata (Figure 6.8). This pattern indicates that clinically relevant resistance markers frequently pre-date the first antibiotic course and persist across the subsequent month, and should be interpreted in light of clinical heterogeneity, small denominators in some strata, and the observational nature of regimen assignment (Figures 6.7–6.8).

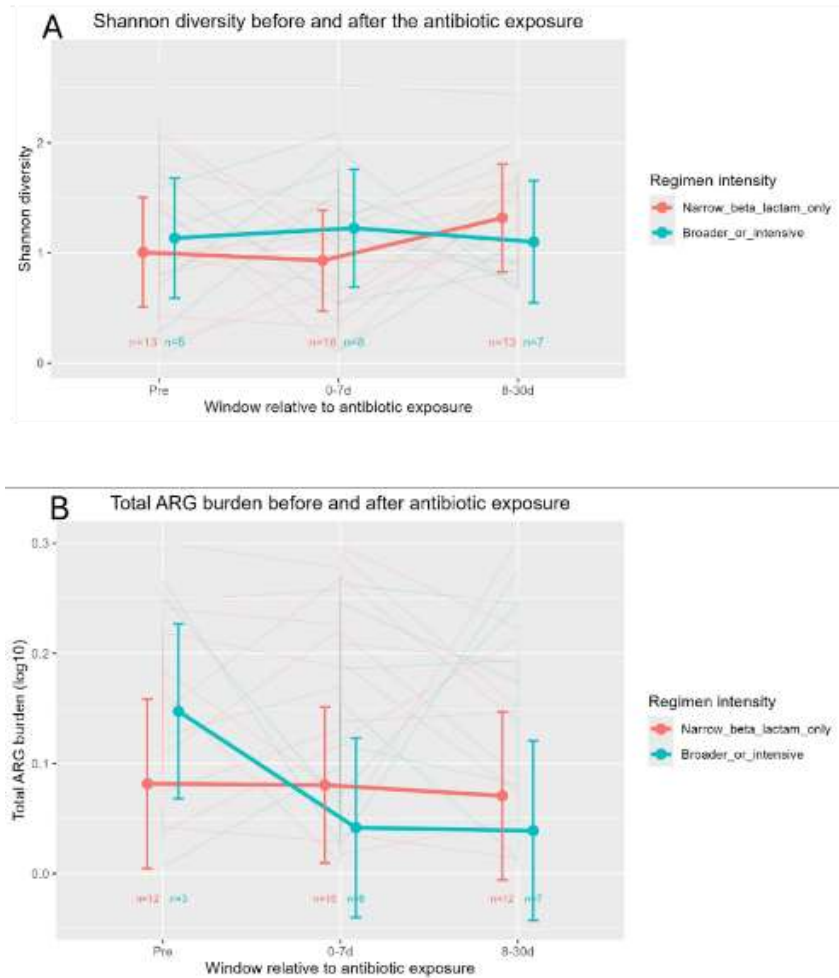


Figure 6.6: Acute-window mixed-model estimates of microbiome diversity and total resistome burden around first antibiotic initiation, stratified by regimen intensity.

Among antibiotic-exposed infants, stool samples were indexed to the first day of systemic antibiotic initiation and classified into three windows: Pre (before exposure), 0-7 days (inclusive of the start day), and 8-30 days post-initiation. Panels show adjusted mean outcomes by window and regimen intensity (infant-level classification: narrow β -lactam monotherapy vs broader/intensive regimens). (A) Shannon diversity. (B) Total ARG burden. Points indicate adjusted means and vertical bars indicate 95% confidence intervals from linear mixed-effects models including an infant-specific random intercept and adjustment for day of life, delivery mode, feeding category, sex, gestational age, and birthweight. Faint lines show within-infant trajectories across windows. Denominators (n) shown beneath each window indicate the number of samples contributing to each regimen $\times$ window stratum.

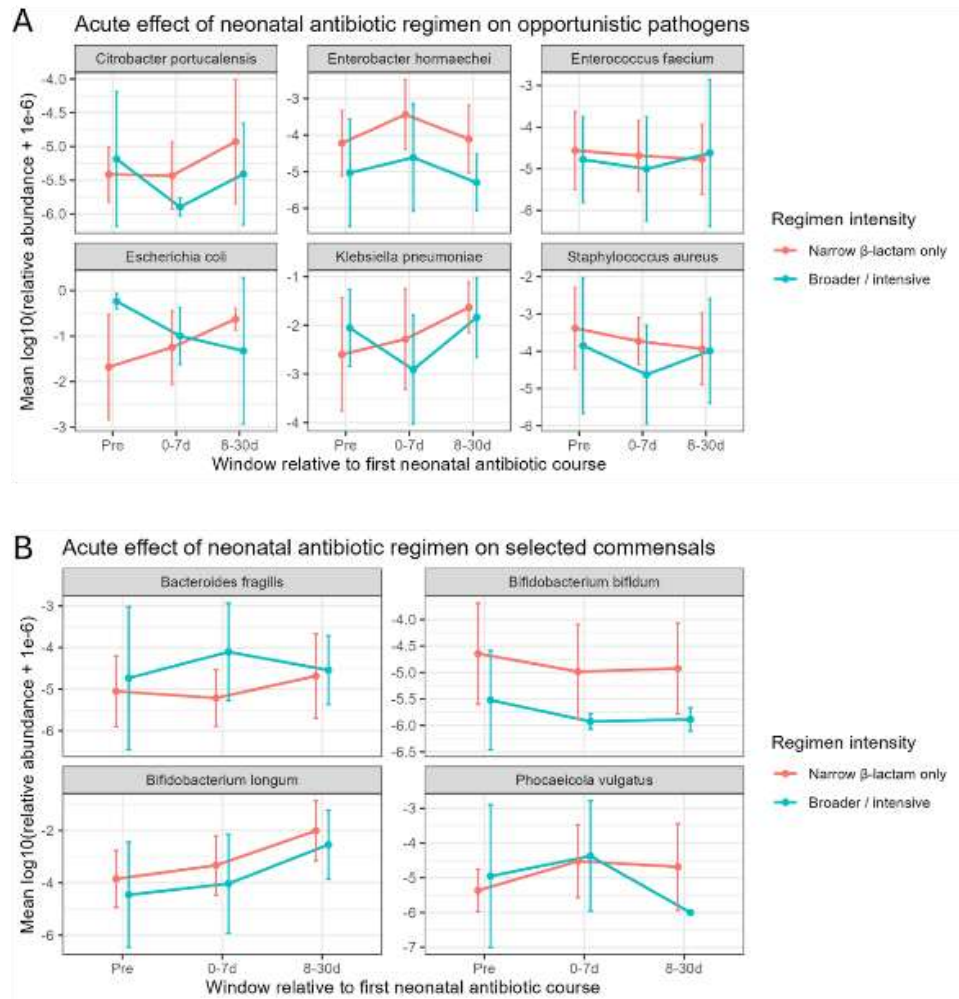


Figure 6.7: Taxon-specific patterns around first antibiotic exposure by regimen intensity: opportunistic/pathogen taxa and selected commensals.

Among antibiotic-exposed infants, samples were indexed to the first day of antibiotic initiation and grouped into Pre, 0-7 days (inclusive), and 8-30 days post-initiation windows. Panels summarise mean taxon abundance within each window, stratified by infant-level regimen intensity (narrow β -lactam vs broader/intensive regimens).

(A) Opportunistic/pathogen taxa. (B) Selected commensals. Points show window-specific means of $\log_{10}(\text{relative abundance} + 1 \times 10^{-6})$ and vertical bars indicate 95% confidence intervals. These panels are intended as descriptive summaries of taxon-level shifts around antibiotic initiation. Denominators vary across windows and regimen groups.

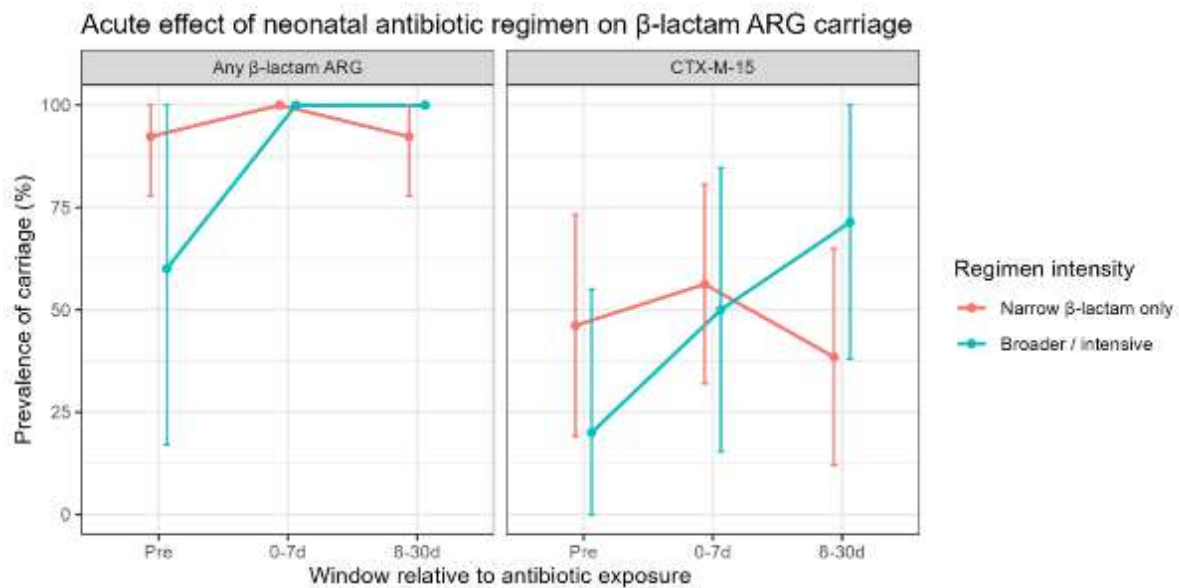


Figure 6.8: Windowed prevalence of β -lactam resistance markers around first antibiotic exposure, stratified by regimen intensity.

Among antibiotic-exposed infants, samples were indexed to the first day of antibiotic exposure and classified into Pre, 0-7 days, and 8-30 days post-exposure windows. Prevalence is shown for any β -lactam ARG (left) and CTX-M-15 carriage (right), stratified by infant-level regimen intensity (narrow β -lactam vs broader/intensive regimens). Carriage was collapsed to infant $\times$ window presence/absence, such that an infant was classified as positive for a given window if the ARG marker was detected in any sample contributing to that window.

6.4.3 Feeding pattern and clinically prioritised microbiome/resistome markers

6.4.3.1 Association of breastfeeding with α -diversity and total ARG burden at D5, D28 and D90

Feeding was predominantly breastfeeding-based throughout the first three months, but WHO-defined EBF was uncommon, with most infants receiving breastmilk as the main feed while also being given water and/or other supplements (Chapter 4; Figure 4.9; Table 6.1C for denominators by timepoint). To assess whether EBF was associated with early gut community diversity and resistome burden, EBF and non-EBF infants (non-EBF combining High_BF and Low_no_BF) were compared at the scheduled sampling timepoints (D5, D28

and D90) using covariate-adjusted linear regression models (Figure 6.9A–B; Table 6.4), as prespecified in Section 6.3.3.1. Models adjusted for delivery mode, prior antibiotic exposure at sampling, sex, gestational age and birthweight.

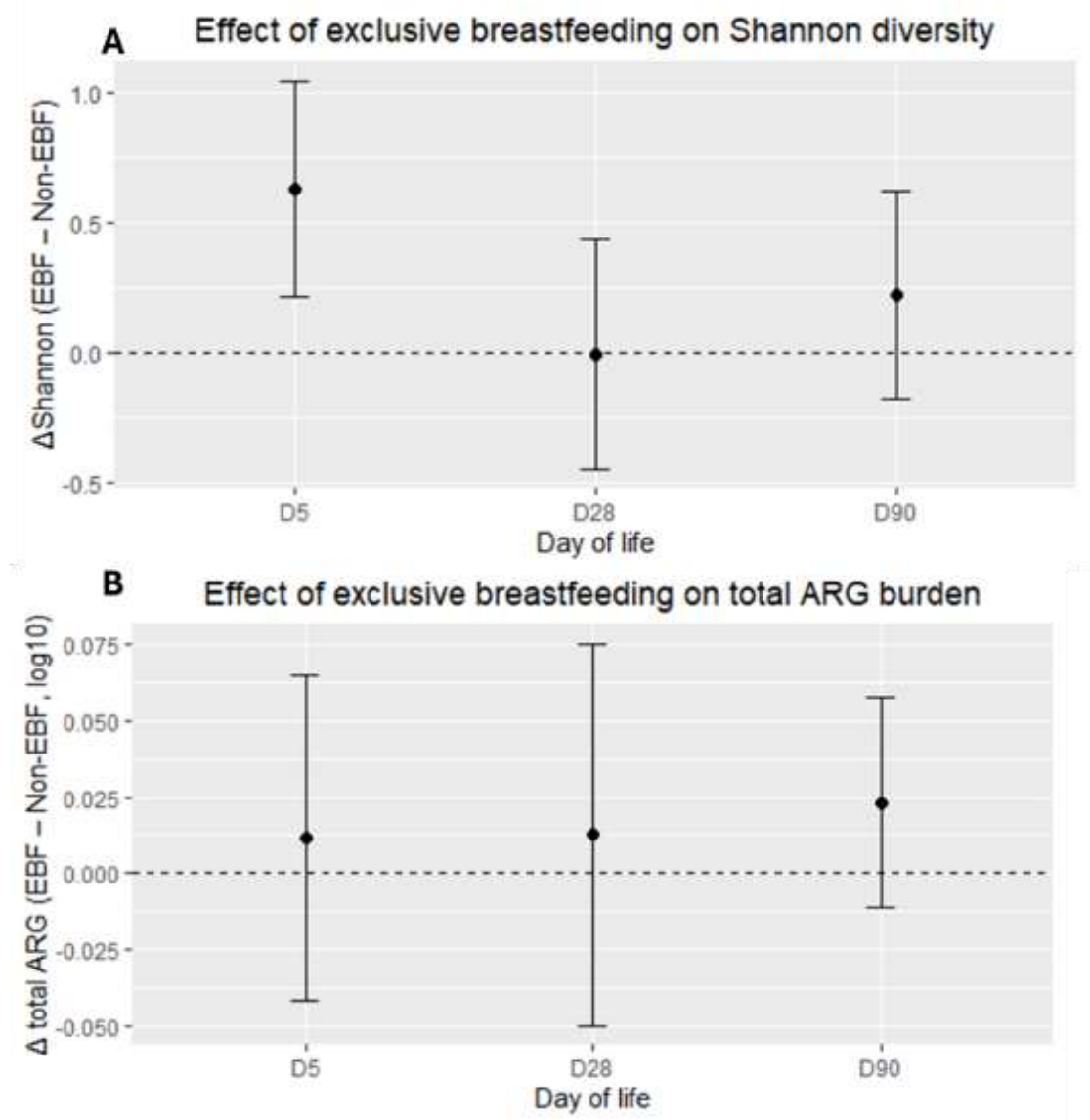


Figure 6.9: Adjusted associations of exclusive breastfeeding with microbiome diversity and total resistome burden.

Points show the adjusted difference between exclusive breastfeeding (EBF) and non-EBF infants (Δ = EBF - non-EBF) at D5, D28 and D90 for Shannon diversity (top) and total ARG burden (bottom). Vertical bars indicate 95% confidence intervals from timepoint-specific linear regression models fitted separately at each day. Positive values indicate higher outcomes in EBF infants relative to non-EBF infants; the dashed horizontal line denotes no difference.

At D5, EBF infants had higher Shannon diversity than non-EBF infants (adjusted difference +0.63, 95% CI +0.22 to +1.05; $p = 0.003$). At D28 and D90, there was no clear evidence of a difference in Shannon diversity between feeding groups (Table 6.4). Across all timepoints, total ARG burden ($\log_{10}[\text{total ARG} + 1]$) did not differ between EBF and non-EBF infants (Figure 6.9B; Table 6.4), and the early diversity difference observed at D5 was not apparent at later timepoints.

Table 6.4: Adjusted associations of exclusive breastfeeding with microbiome diversity and total resistome burden

Outcome	Day	Contrast	Adjusted estimate	95% CI	p-value	n
Shannon	D5	EBF – non-EBF	0.633	0.218 to 1.049	0.0033	67
Shannon	D28	EBF – non-EBF	-0.003	-0.447 to 0.440	0.988	66
Shannon	D90	EBF – non-EBF	0.226	-0.174 to 0.627	0.263	63
Total (log10)	ARG D5	EBF – non-EBF	0.012	-0.042 to 0.065	0.664	67
Total (log10)	ARG D28	EBF – non-EBF	0.013	-0.050 to 0.075	0.686	66
Total (log10)	ARG D90	EBF – non-EBF	0.023	-0.011 to 0.058	0.182	63

Timepoint-specific linear regression models were fitted separately at timepoints to compare EBF) versus non-EBF for Shannon diversity and total ARG burden. Values are adjusted mean differences (EBF - non-EBF) with 95% confidence intervals, p-values, and the number of samples included in each model. Positive estimates indicate higher outcomes in EBF infants relative to non-EBF infants.

6.4.3.2 Breastfeeding intensity and early microbiome signatures

Because WHO-defined EBF was uncommon and most infants were predominantly breastfed while also receiving water, breastfeeding intensity was additionally evaluated using the three-category variable (EBF, High_BF, Low_no_BF; defined in Section 6.3.1). Feeding group differences were assessed at D5, D28 and D90 using covariate-adjusted linear regression models with EBF as the reference category and pairwise contrasts across the three feeding categories (Figure 6.10A–B; Table S6.2), adjusted as prespecified for delivery mode, prior antibiotic exposure at sampling, sex, gestational age and birthweight. Due to the small size in the Low_no_BF group (Table 6.1C), these contrasts were underpowered and are interpreted as exploratory, with results subject to wide confidence intervals.

At D5, infants in the High_BF group had lower Shannon diversity than EBF infants (adjusted difference High_BF vs EBF = -0.66 , 95% CI -1.07 to -0.24 ; $p = 0.0026$). In contrast, Low_no_BF infants did not show clear evidence of a difference from EBF at D5 (-0.31 , 95% CI -1.03 to $+0.41$; $p = 0.396$), and the High_BF and Low_no_BF groups were not clearly separated (High_BF vs Low_no_BF = -0.35 , 95% CI -0.98 to $+0.29$; $p = 0.275$). By D28 and D90, there was no clear evidence of differences in Shannon diversity across feeding groups, and point estimates did not support a consistent gradient pattern (all $p > 0.1$; Table S6.2). Overall, detectable feeding-group separation in α -diversity was confined to the first postnatal week and did not persist into later infancy.

Across all three timepoints, total ARG burden ($\log_{10}[\text{total ARG} + 1]$) was similar across feeding categories, with no statistically supported pairwise differences at D5, D28 or D90 (all $p > 0.1$; Table S6.2). Point estimates were close to zero and confidence intervals spanned modest increases and decreases, indicating that, within this cohort, variation in breastmilk exposure category did not translate into measurable differences in overall resistome burden at the scheduled sampling visits.

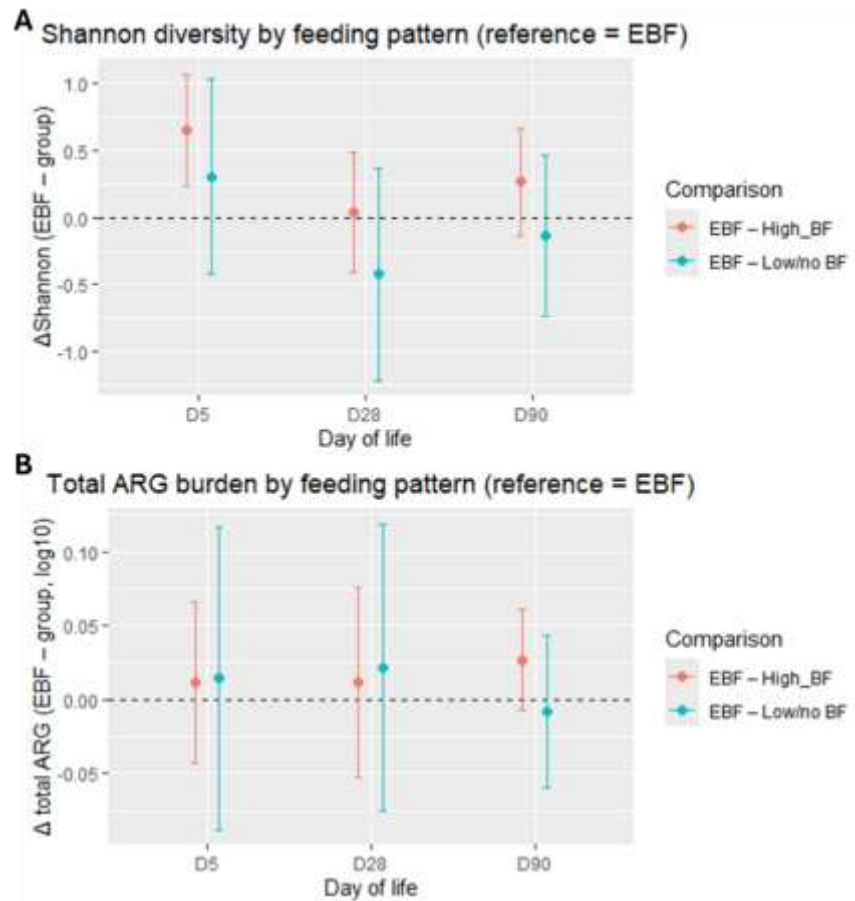


Figure 6.10: Adjusted associations of breastfeeding intensity with microbiome diversity and total resistome burden at scheduled sampling milestones (reference = EBF).

(A) Shannon diversity (species level). (B) Total ARG burden. The dashed horizontal line denotes no difference from EBF; positive values indicate higher outcomes in EBF relative to the comparison feeding group (EBF group).

Because Shannon diversity and total ARG burden may be insensitive to breastfeeding-related taxon shifts, marker genera were additionally examined to assess whether canonical breastfeeding-associated taxa (particularly *Bifidobacterium*) and common early-life pathobiont genera showed evidence of patterning across EBF, High_BF and Low_no_BF (Figure 6.11). *Bifidobacterium* increased across infancy in all feeding groups and was prominent by D90. Early trajectories differed, with an earlier rise apparent in EBF at D3

relative to High_BF. The Low_no_BF group showed a comparatively high median abundance at D14; however, this estimate was based on a small denominator ($n = 5$) and is interpreted descriptively. In contrast, *Escherichia* showed high relative abundance at birth across feeding strata, with lower and more similar medians by D28. *Klebsiella* and *Veillonella* displayed episodic elevations that were most evident within the Low_no_BF group, including higher *Veillonella* medians at later timepoints (D28 and D90). Finally, *Bacteroides/Phocaeicola* remained low during the first 90 days across feeding groups, with sporadic peaks consistent with substantial inter-individual variability.

6.4.3.3 Feeding pattern and resistome endpoints

To complement analyses of overall ARG burden, the descriptive prevalence of clinically relevant ARG markers across infancy was summarised by breastfeeding intensity (EBF, High_BF, Low_no_BF) and timepoint (D0–D90). Markers were considered detected when relative abundance was non-zero, using the study pipeline’s normalised outputs (Chapter 2). Prevalence estimates (with denominators) are provided in the accompanying table (Table S6.3), and temporal patterns are shown in Figure 6.12.

Across all feeding strata, aminoglycoside resistance markers were frequently detected from the earliest timepoints. ANT-family markers showed consistently high prevalence across infancy, with only modest fluctuation between feeding groups. APH-family markers were also common throughout follow-up; while prevalence varied by day within each feeding category, there was no consistent ordering that separated EBF from the two non-EBF groups across time. AAC-family markers showed greater temporal fluctuation, with lower prevalence at birth and a more prominent rise in early infancy; however, this did not translate into a consistent gradient across breastfeeding intensity categories.

For β -lactamase markers, TEM and SHV markers were widely detected across infancy in all feeding groups. Some timepoint-specific differences were apparent (most notably at the earliest timepoints, where Low_no_BF occasionally showed higher prevalence than predominantly breastfed groups), but these contrasts were not sustained across subsequent visits. CTX-M-15 was detected in a substantial minority of samples across early life, again without consistent separation of EBF from High_BF and Low_no_BF across timepoints. Overall, Figure 6.12 primarily indicates temporal patterning across infancy rather than persistent between-group separation by breastfeeding intensity for these markers.

Stratified descriptive summaries restricted to caesarean-delivered infants were considered; however, feeding categories were highly imbalanced within the caesarean subgroup (with EBF uncommon and Low_no_BF sparse at most timepoints; Table 6.1C), limiting precision and interpretability.

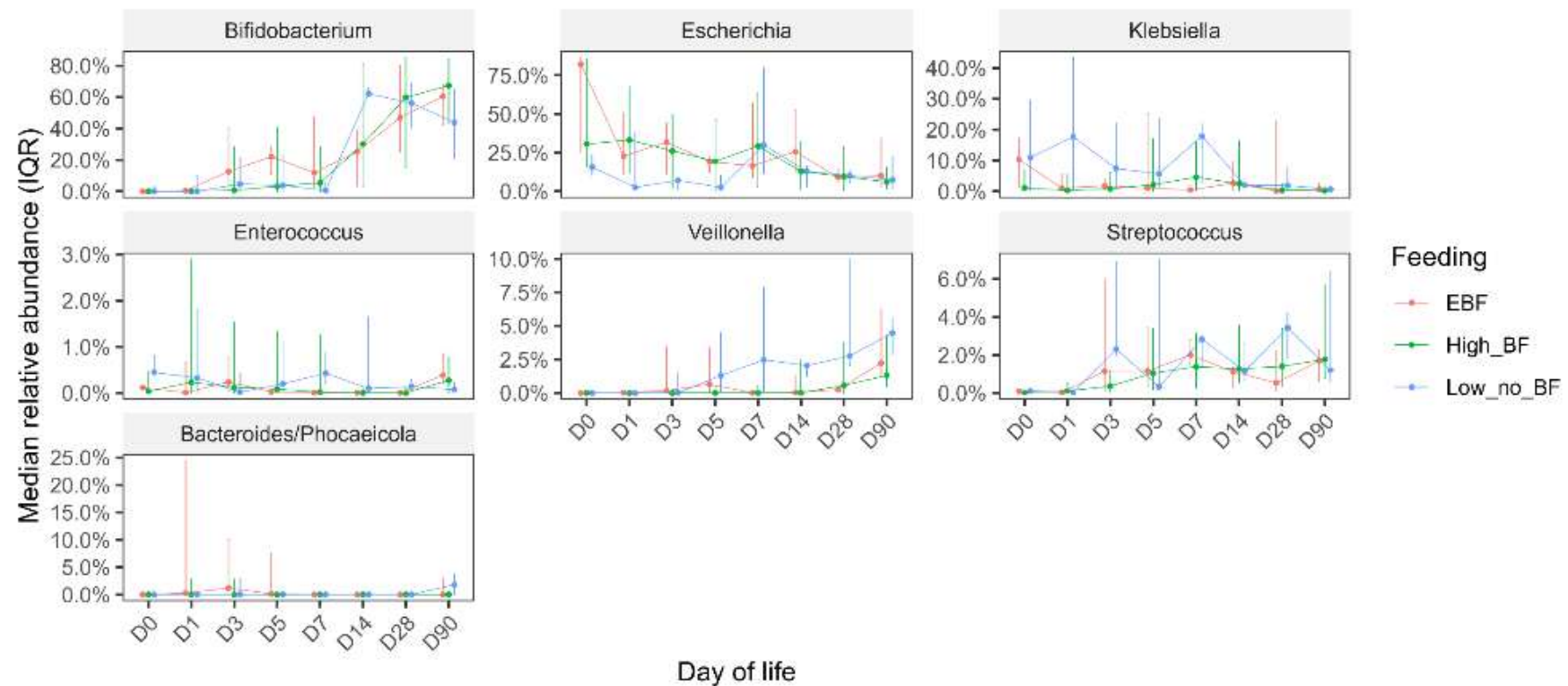


Figure 6.11: Longitudinal trajectories of marker genera by breastfeeding intensity across infancy (descriptive)

Median relative abundance (with interquartile range, IQR) of selected genera is shown across scheduled sampling days, stratified by feeding category: exclusive breastfeeding, predominantly breastfed with water/supplements (High_BF), and low/no breastmilk intake (Low_no_BF). These plots provide an unadjusted descriptive overview of feeding-related taxon dynamics and highlight temporal patterns and variability across feeding strata.

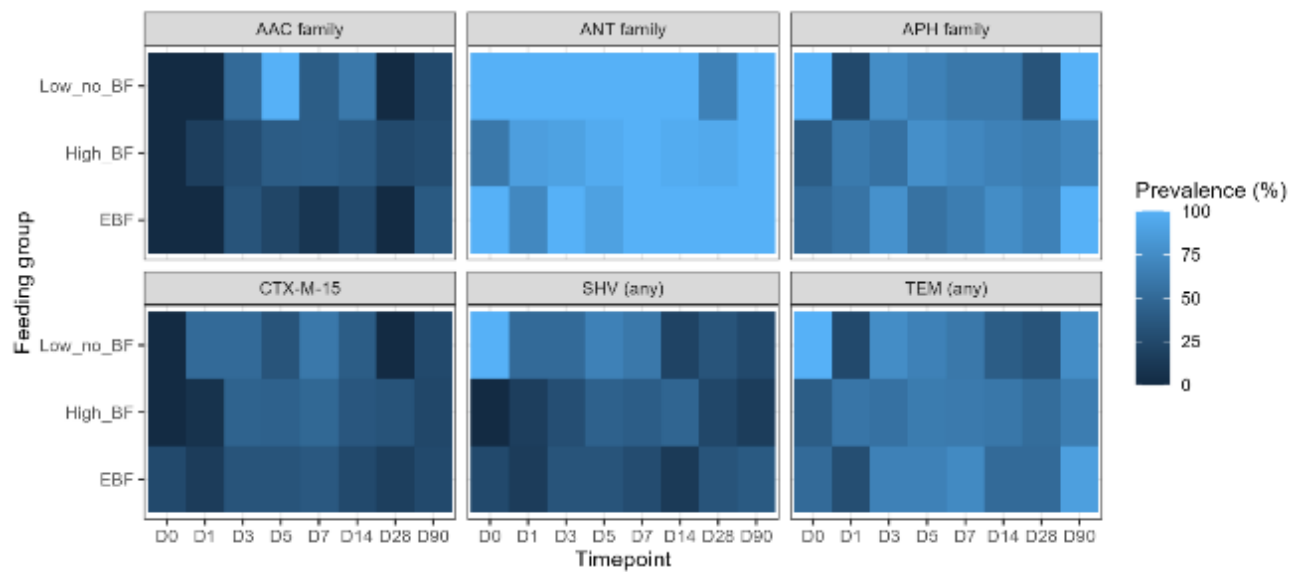


Figure 6.12: Prevalence of selected aminoglycoside-resistance gene families and β -lactamase markers across infancy by breastfeeding intensity (descriptive).

Heatmaps show the percentage of samples with non-zero relative abundance (“detected”) for aminoglycoside resistance gene families (AAC, ANT, APH) and β -lactamase markers (CTX-M-15, SHV, TEM), stratified by feeding category. Darker shading indicates lower prevalence within each feeding group $\times$ timepoint. Prevalence estimates are intended as descriptive summaries of temporal patterning and between-group differences.

6.4.4 Combined exposures and synergy

6.4.4.1 Multi-hit cumulative exposure and CTX-M-15 positive

To characterise how combined early-life exposures relate to acquisition of clinically relevant ESBL markers, a cumulative “multi-hit” exposure variable was defined as the count of three exposures: delivery by C-section (fixed), and timepoint-specific antibiotic exposure (any prior systemic antibiotics by the sampling timepoint) and feeding status (Low_no_BF at the sampling timepoint). The score was categorised as 0, 1, or ≥ 2 hits to preserve stability given sparse 2- and 3-hit strata. CTX-M-15 carriage was summarised across scheduled early visits (D1–D14) within each multi-hit group as n/N (%) and displayed as a prevalence trajectory

with denominators (Figure 6.13). The early timepoints are presented descriptively because combined-exposure regression at baseline and during the first week was undermined by sparse multi-hit strata and incomplete endpoint availability in some strata.

Across D1–D14, CTX-M-15 prevalence increased over time within each exposure stratum. Apparent differences between multi-hit groups were observed at several early timepoints; however, estimates for the ≥ 2 -hit group are based on small denominators at most visits and are therefore interpreted descriptively. The figure presents the underlying denominators at each timepoint, highlighting the limited sample sizes in the ≥ 2 -hit group across early visits.

Expanding the analysis to later timepoints, prevalence was summarised within each multi-hit group as n/N (%) at each timepoint and is presented alongside model denominators (Table 6.5). These later timepoints were selected for inferential analysis because endpoint availability was higher than at early visits and the exposure strata were sufficiently populated to support regression modelling. Observed *bla*_{CTX-M-15} prevalence differed across multi-hit strata at D28, whereas by D90 between-group differences were smaller and did not follow the same ordering (Table 6.5). In covariate-adjusted logistic regression models, the cumulative exposure score showed suggestive evidence of higher odds of *bla*_{CTX-M-15} carriage at D28 (aOR per step 2.55, 95% CI 0.97–6.70; $p = 0.058$), while no association was observed at D90 (aOR per step 0.95, 95% CI 0.37–2.47; $p = 0.917$) (Table 6.5; Figure 6.14). Estimates for the ≥ 2 -hit category should be interpreted cautiously due to the small number of infants contributing to this stratum.

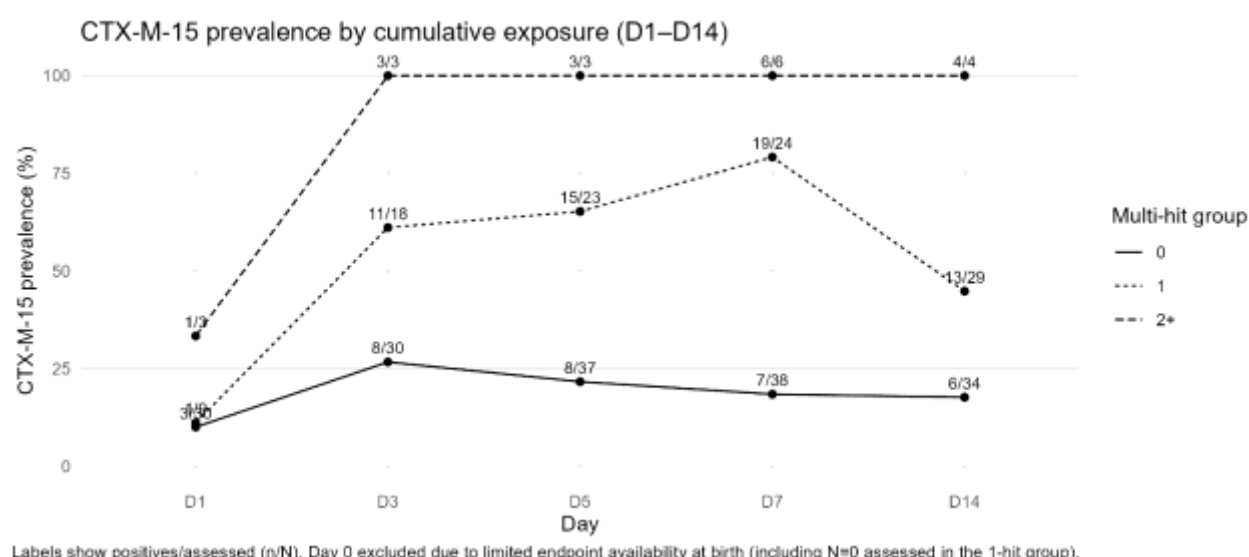


Figure 6.13: Early-life CTX-M-15 prevalence by cumulative multi-hit exposure group across D1–D14 (descriptive).

Prevalence of CTX-M-15 carriage is shown across early visits (D1, D3, D5, D7, D14) stratified by a cumulative multi-hit exposure score defined as the count of: C-section, prior antibiotic exposure by the sampling timepoint, and low/no breastfeeding status. The score was categorised as 0 hits, 1 hit, or ≥ 2 hits. Points show prevalence (%) within each group at each day, and labels indicate the underlying numerator and denominator (n/N) assessed per timepoint. Day 0 is excluded due to limited endpoint availability and non-comparable denominators across exposure strata at birth (including N=0 assessed in the 1-hit group).

Table 6.5: CTX-M-15 carriage by cumulative early-life exposure at D28 and D90

Day	0	hits	1	hit	≥ 2	hits	Adjusted OR per step	p-value	Model N	Events
	(n/N, %)		(n/N, %)		(n/N, %)		(95% CI)			(CTX-M-15)
D28	7/32		8/24		3/5		2.55 (0.97-6.70)	0.058	61	18
	(21.9%)		(33.3%)		(60.0%)					
D90	7/29		6/23		1/7		0.95 (0.37-2.47)	0.917	59	14
	(24.1%)		(26.1%)		(14.3%)					

Values are shown as n/N (%), where N is the number assessed for CTX-M-15 within each exposure stratum. “Hit” denotes C-section, infant antibiotic exposure, or low/no breastfeeding. Adjusted ORs are from logistic regression adjusted for sex, gestational age, and birth weight, and represent the per-step change across the collapsed cumulative exposure score (0/1/ ≥ 2 hits). Model N indicates complete cases included in the adjusted model.

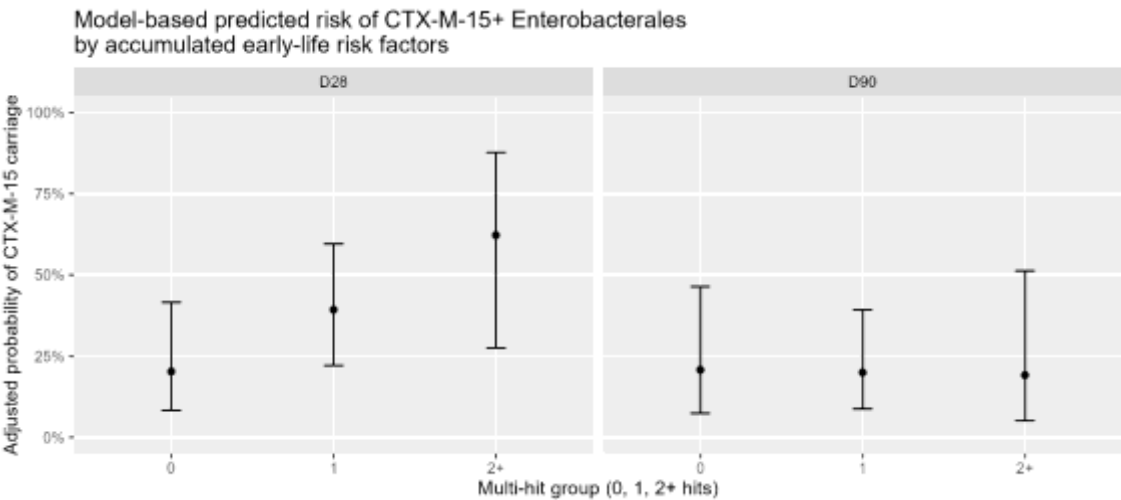


Figure 6.14: Model-based predicted probability of CTX-M-15 carriage at D28 and D90 by cumulative multi-hit exposure group.

Points show the adjusted predicted probability of CTX-M-15 carriage at D28 (left) and D90 (right) across cumulative multi-hit exposure categories (0, 1, ≥ 2 hits), derived from timepoint-specific logistic regression models. The multi-hit score was defined as the count of C-section, prior antibiotic exposure, and low/no breastfeeding status, collapsed to 0/1/ ≥ 2 to preserve stability.

6.4.4.2 Targeted effect modification at D28: antibiotics and delivery mode

To assess whether antibiotic exposure and delivery mode jointly patterned ESBL-related carriage at D28, prevalence was summarised in a two-way stratified table by delivery mode (C/S vs vaginal) and antibiotic exposure (Yes vs No). Results are presented descriptively as n/N (%) because antibiotic-exposed denominators were small within delivery-mode strata and formal interaction modelling would be unstable.

For CTX-M-15 carriage, prevalence was higher among C-section infants than vaginally delivered infants within both antibiotic strata (C/S: 7/12 without antibiotics and 3/5 with antibiotics; vaginal: 7/35 without antibiotics and 1/9 with antibiotics), although estimates in the antibiotic-exposed cells are based on small denominators. The same stratified comparison using the broader ESBL composite (Any ESBL marker) showed similarly higher carriage among C-section than vaginally delivered infants across antibiotic strata (C/S: 9/12 without antibiotics and 5/5 with antibiotics; vaginal: 20/35 without antibiotics and 4/9 with antibiotics) (Table 6.6).

Table 6.6: Prevalence of CTX-M-15 carriage and a composite ESBL marker outcome at Day 28, stratified by delivery mode and infant antibiotic exposure.

Values are n/N (%)

		Antibiotics:	
Outcome	Delivery_mode	No	Yes
CTX-M-15	C/S	7/12 (58.3%)	3/5 (60.0%)
CTX-M-15	SVD	7/35 (20.0%)	1/9 (11.1%)
Any ESBL marker	C/S	9/12 (75.0%)	5/5 (100.0%)
Any ESBL marker	SVD	20/35 (57.1%)	4/9 (44.4%)

Prevalence is presented as n/N (%) for CTX-M-15 and for any ESBL marker, stratified jointly by delivery mode (C-section [C/S] vs spontaneous vaginal delivery [SVD]) and prior antibiotic exposure by the D28 sampling timepoint (Yes/No). Results are reported descriptively to reflect sparse denominators in some strata, particularly among antibiotic-exposed infants within delivery-mode groups.

6.5 Discussion

Diet, antibiotic exposure, and mode of delivery are widely recognised as key modulators of early gut microbiome colonisation, with potential implications for immune priming and vulnerability to infection in infancy (Tamburini *et al.*, 2016; Bokulich *et al.*, 2016). A central unresolved question is whether such early perturbations are transient, restricted to the first weeks of life, or whether they persist into later infancy (Korpela *et al.*, 2016, 2020), particularly in settings with high background microbial and antimicrobial exposure. This chapter therefore examined how mode of delivery, infant antibiotic exposure, and breastfeeding intensity shape early gut microbiome and resistome development in Nigerian infants born in hospital, and whether these exposures combine to concentrate clinically important β -lactam resistance markers, including *bla*_{CTX-M-15}.

Across the first three months of life, these findings are characterised by (i) perinatal seeding and early selection of colonising microbial taxa and their associated ARGs during the first weeks of life, followed by (ii) rapid convergence of global microbiome and resistome metrics by around three months (Table 6.2; Figs 6.2–6.3; Fig 6.6). In this cohort, delivery mode, likely acting as a proxy for a bundle of perinatal exposures including microbial transfer pathways and hospital contact, was the primary early driver across α -diversity, community composition, and colonisation by selected taxa (Table 6.2; Figs 6.2–6.3; Table 6.3). By contrast, infant antibiotics and breastfeeding intensity showed weaker and less uniform associations with global endpoints, consistent with heterogeneity in exposure timing and with the possibility that the clinically relevant ARGs may be acquired very early and then persist (Table 6.2; Fig 6.5–6.6; Fig 6.8; Fig 6.10). However, the multi-hit analyses further suggest that high-priority ESBL markers disproportionately accumulate under stacked early-life exposures by the end of the neonatal period, even when individual factors show modest associations with global metrics (Fig 6.13–6.14; Tables 6.5–6.6).

6.5.1. Delivery mode: early signatures with rapid convergence

Mode of delivery was the primary early driver of gut community development in this cohort. In adjusted models at scheduled milestones, C-section was associated with lower Shannon diversity at D5 and higher total ARG burden at D5 and D28, with no significant difference observed by D90 (Section 6.4.1; Methods in Section 6.3.2). Delivery mode also explained a modest but detectable proportion of microbiome compositional variation at D5 ($R^2=0.062$, $p=0.001$) and D28 ($R^2=0.056$, $p=0.002$), with attenuation by D90 ($R^2=0.016$, $p=0.363$), while resistome composition showed the strongest and most consistent separation in the early neonatal period (Section 6.4.1.2). These community metrics were mirrored in taxon-level findings consistent with altered early seeding and succession: C-section was associated with reduced colonisation by obligate anaerobes implicated in early-life maturation (including *Bifidobacterium longum* and *Phocaeicola vulgatus* at D5), alongside transient enrichment of several opportunistic or hospital-associated taxa, although estimates for some taxa were imprecise (Section 6.4.1.3; Figure 6.4; Table 6.3.). Overall, the findings support the interpretation that delivery mode shapes the initial conditions of community assembly in the first postnatal week and contributes to early resistome burden and structure, with attenuation of delivery-mode differences in community metrics by around three months. The corresponding infant-level stacked-bar profiles (C/S vs SVD across all scheduled days) are provided in Appendix A.3, Figure S6.1.

Our findings mirror the 'stunted' early gut profiles identified in the UK Baby Biome Study (Shao et al., 2019), characterised by a loss of commensals, specifically *Bifidobacterium* and *Bacteroides/Phocaeicola*, and an enrichment of opportunistic taxa. Mechanistically, evidence suggests that a critical feature of the "caesarean signature" is reduced maternal gut/rectal strain transfer, often reflected in delayed or absent *Bacteroides/Phocaeicola*

acquisition, rather than vaginal canal exposure alone (Ferretti et al., 2018). The low prevalence of *Phocaeicola* and *Bacteroides* observed in this cohort is consistent with this evidence and reinforces the view that caesarean delivery disrupts early vertical transmission and succession. In interpreting the binary C-section contrast in the present cohort, an important nuance is that C-section delivery represents a composite exposure rather than a single mechanism. Elective and emergency caesarean pathways differ in labour exposure, membrane rupture, early contact patterns, and antibiotic practices, which can alter early microbial transfer (Mitchell et al., 2020). As a result, our single C-section indicator likely masks specific effects of elective versus emergency deliveries, contributing to imprecision in certain taxon-level findings.

From a clinical perspective, early depletion of obligate anaerobes alongside the transient enrichment of opportunistic taxa indicates a temporary “window of vulnerability”, during which colonisation resistance and community stability are reduced. This window coincides with the neonatal period, the time of highest risk for invasive bacterial infections, including sepsis (Okomo et al., 2019). In this cohort, the duration and impact of this vulnerability may be shaped by the surrounding environment. In the Nigerian setting, where constraints in WASH infrastructure and high microbial contact increase colonisation pressure (Edefo, 2025; WHO and UNICEF, 2024), environmental acquisition could act as a partial “leveller”. While this accelerates the convergence of overall metrics across delivery-mode, it likely reflects a high-exposure environment where persistent strain-level differences, such as clinically relevant markers, remain hidden beneath broad community-level metrics (Bäckhed et al., 2015; Shao et al., 2019).

6.5.2 Timing dependence and the neonatal window

A central pattern across the chapter is strong timing dependence. The associations with mode of delivery, antibiotic exposure, and feeding are most pronounced in the first weeks of life, with attenuation by the third month of life. This trajectory implies a narrow developmental window of peak sensitivity followed by rapid convergence in broad community measures. Such age-structured maturation is consistent with large-scale studies showing that early differences in "initial conditions" narrow as infants are colonised from broader maternal and household sources (Shao *et al.*, 2019; Stewart *et al.*, 2018).

However, the attenuation of community-level metrics, such as Shannon diversity or total ARG burden, does not necessarily indicate biological "recovery." These measures are aggregate summaries that can mask persistent differences in strain dominance or the genomic context of resistance genes (Bargheet *et al.*, 2025; X. Li *et al.*, 2024b). Convergence in diversity can coexist with meaningful differences in specific colonisers or prioritised resistance markers, which remain clinically relevant regardless of overall community structure.

6.5.3 Infant antibiotics: modifiers of an already-seeded resistome

Across the chapter, infant antibiotic exposure did not generate a uniform, cohort-wide step change in Shannon diversity or total ARG burden (Section 6.4.2). This pattern is best interpreted as biologically informative rather than a null result, driven by two plausible features of early-life microbial ecology. First, heterogeneity in timing and regimen intensity dilutes average effects in summary metrics. Antibiotic initiation occurred at variable postnatal ages with diverse spectra of activity; where perturbations are time-limited, characterised by acute disruption followed by partial recovery, pooling infants at different

points on this curve will naturally attenuate the mean contrast (Korpela et al., 2020). This interpretation is consistent with evidence from community cohorts in high-exposure settings, where antibiotic-associated shifts were found to be moderate and transient (Yassour et al., 2016) compared to intensive-care contexts (Gasparrini et al., 2019; Gibson et al., 2016).

Second, the windowed analyses suggest that, in this setting, antibiotics often act more as a modifier of an already-seeded resistome rather than as a singular "starting point" for AR acquisition. The observation that *bla*_{CTX-M-15} was frequently detectable prior to the first recorded antibiotic exposure (Section 6.4.2.3; Figure 6.8) implies that antibiotic pressure is selecting within a community that already contains clinically relevant markers. This aligns with the "pre-established resistome" hypothesis (X. Li et al., 2023; Yassour et al., 2016), demonstrating that infants in high-burden environments acquire resistance determinants extremely early via maternal and household reservoirs. It is also compatible with evidence that greater antibiotic intensity is required to suppress protective anaerobes and significantly alter community structure (Gibson et al., 2016).

A practical implication is that antibiotic stewardship remains necessary but may be insufficient to prevent early ESBL carriage if acquisition is primarily perinatal or environmental. In such contexts, stewardship must be paired with broader infection IPC and WASH-related interventions to reduce background colonisation pressure. This point is particularly relevant in the Nigerian setting, where high rates of peripartum maternal antibiotic use (84/87 mothers; Chapter 4) and constraints in hospital WASH infrastructure create sustained selection pressure during the critical neonatal window. Methodologically, these findings motivate the prioritisation of sensitive, high-resolution endpoints, such as specific *bla*_{CTX-M-15} carriage or gene-context measures, rather than relying solely on summary measures like Shannon diversity or total ARG burden to evaluate antibiotic-associated change.

6.5.4 Risk stacking: cumulative exposures and ESBL marker clustering

A further integrative signal from this chapter is that clinically prioritised β -lactamase marker carriage clusters under stacked early-life exposures, consistent with "risk stacking" rather than single-cause determinism (Section 6.4.3). Using a cumulative multi-hit score (caesarean delivery, prior infant antibiotics, and low/no breastfeeding), early *bla*_{CTX-M-15} trajectories suggested increasing separation by exposure burden across the first two postnatal weeks, although sparse denominators, particularly in the ≥ 2 -hit stratum, necessitate descriptive interpretation at these earliest visits (Figure 6.13). By D28, *bla*_{CTX-M-15} prevalence showed a graded pattern across 0-, 1-, and ≥ 2 -hit strata, and the adjusted per-step estimate was directionally consistent with higher odds of carriage, albeit with borderline statistical support and limited precision driven by the small ≥ 2 -hit group (Table 6.5). By D90, the gradient was not maintained, suggesting that stacking effects may be most detectable during the end-of-neonatal-period window rather than later infancy (Table 6.5; Figure 6.14).

Conceptually, the multi-hit score is best interpreted as a pragmatic proxy for convergent pathways: (i) disrupted vertical transmission and early hospital contact (delivery mode), (ii) selection pressure (antibiotics), and (iii) reduced successional support for protective anaerobes (feeding intensity). In a high-exposure setting where community-level metrics converge quickly, this combined-exposure framing can be more clinically informative than any single modulator analysed in isolation. This aligns with recent evidence demonstrating that the infant gut is shaped by a combination of overlapping maternal and environmental seeding routes, where the disruption of one pathway can be compounded by the absence of others (Bogaert et al., 2023). The approach employed here, treating risk factors as additive units to determine if a dose-response relationship exists between the number of exposures

and the likelihood of *bla*_{CTX-M-15} carriage, has been validated for assessing cumulative comorbidities in infectious disease risk and stratifying patient vulnerability.

Clinically, the implication is not that the present multi-hit score is implementation-ready, but that simple cumulative risk stratification could be developed and validated to target enhanced IPC, prioritised screening, and follow-up for infants most likely to concentrate high-risk β -lactamase markers during the neonatal window. This claim should be paired with a clear caution that small cells (particularly ≥ 2 hits) and correlated care pathways require replication and external validation before translation into policy. Nevertheless, the findings suggest that interventions intended to shift colonisation trajectories or reduce carriage of high-risk resistance markers are most plausibly time-sensitive. By around three months, summary measures are less discriminating; therefore, risk stratification and evaluation of intervention effects may need to focus on the neonatal window and prioritise higher-resolution endpoints, such as β -lactamase/ESBL-associated marker carriage (e.g., *bla*_{CTX-M-15}) and, where feasible, gene-context or host-association measures, rather than relying solely on summary diversity or aggregate ARG burden.

6.5.5 Synthesis and implications

This chapter demonstrates that early-life modulators, particularly delivery mode, shape initial gut microbiome and resistome assembly in Nigerian infants, but that these effects are most detectable during the neonatal period and attenuate in summary metrics by three months. The neonatal window emerges as a critical period of vulnerability, during which perturbations to vertical transmission, antibiotic selection pressure, and feeding patterns can concentrate clinically prioritised resistance markers. In this high-exposure setting, where universal early ARG seeding (Chapter 5) and high peripartum antibiotic use create sustained

colonisation pressure, interventions targeting the neonatal period, including enhanced IPC at the maternal-infant interface, WASH improvements, and rational antibiotic use, may offer the greatest opportunity to reduce carriage of therapy-relevant resistance determinants. These findings inform the intervention strategies and conceptual framework discussed in Chapter 7.

Chapter 7: General discussion

7.1 General Overview

The human microbiome is fundamental to health, serving as a critical interface between host physiology and the external environment. In early life, the developing gut microbiome plays an essential role in immune education, metabolic programming, and bacterial colonisation (Arrieta et al., 2014; Wampach et al., 2018). The first months of life represent a dynamic period of microbial assembly, during which pioneer colonisers establish communities that shape both immediate and long-term health outcomes (Bäckhed et al., 2015; Stewart et al., 2018). This process is not merely taxonomic succession but involves complex microbe-immune crosstalk that calibrates inflammatory responses, programs innate immune cells, and establishes tolerance pathways (Gensollen et al., 2016; Thaïss et al., 2016). Perturbations during this critical window, whether through antibiotic exposure, delivery mode, feeding, or environmental factors, can have lasting consequences for immune development and disease susceptibility (Tamburini et al., 2016).

In LMIC settings, the early-life microbiome development is shaped by distinct contextual realities. Higher infectious disease burdens, variable access to healthcare, diverse hygiene practices, and different patterns of antibiotic availability influence both microbial colonisation and the acquisition of AMR determinants (Gough et al., 2014; Yassour et al., 2016). Infants in these settings frequently carry clinically important ARGs from birth, even in the absence of direct antibiotic exposure (Carvalho et al., 2022; Shao et al., 2019). This early AMR colonisation reflects transmission from maternal, healthcare-associated, and environmental reservoirs, establishing a resistome that may persist and expand during subsequent microbial maturation (Gosalbes et al., 2016; K. Pärnänen et al., 2018). Yet despite the global health

importance of understanding these dynamics, longitudinal metagenomic data from African infant cohorts remain limited.

My thesis applied longitudinal metagenomic profiling to a Northern Nigerian infant cohort with daily sampling in the first week, and to our knowledge, is the first study comprising temporal resolution throughout the neonatal period, concurrently characterising the co-development of both the microbiome and resistome (Chapter 4). Analysis of the infant stool samples revealed dynamic temporal shifts in bacterial community structure where bacterial α -diversity was lowest at birth (Shannon median ~ 1.0 , often reflecting single-genus dominance) and increased progressively through the first postnatal week, albeit characterised by high inter-infant heterogeneity and no sustained linear increase through day 90 (Chapter 5, Figure 5.1). β -diversity patterns demonstrated that infant's age explained only a modest proportion of community variance ($R^2 = 0.122, p = 0.001$). While distinct differences in microbial composition based on delivery mode were detectable early (day 5: $R^2 = 0.062, p = 0.001$; day 28: $R^2 = 0.056, p = 0.002$), these disparities attenuated by day 90 ($R^2 = 0.016, p = 0.363$) (Chapter 6, Section 6.4.1). This temporal convergence of the infant gut microbiota aligns with observations by Matharu et al., (2022) (a subset of the HELMi birth cohort), who documented decreased β -diversity with age and convergence toward adult-like profiles; however, the Matharu *et al.* cohort showed persistent birth-mode signatures extending to 12 months, a longer duration than the attenuation of these differences observed here by D90. In contrast to studies from Western cohorts reporting delivery mode effects persisting for 2 years (Matharu et al., 2022; Patangia, Grimaud, O'Shea, et al., 2024b), my study suggests more rapid compositional convergence in the first three months.

Taxonomic composition showed early dominance by facultative anaerobes, with Proteobacteria (particularly *Escherichia* and *Klebsiella*) prominent in the first week, followed by increasing representation of obligate anaerobes over subsequent weeks (Chapter 5, Section 5.4.1.1). The

prominence of *Escherichia* and *Klebsiella* as primary hosts for clinically prioritised ESBL markers in this cohort, taxa that are also the leading causes of neonatal sepsis in West Africa (Meiring et al., 2025; Niyoyita et al., 2024), provides novel insight into the clinical relevance of early succession dynamics (Grant-Beurmann et al., 2022) in resource-limited settings where comparable longitudinal metagenomic data remains limited (Abdill et al., 2022).

Complementing these microbiome dynamics, resistome profiling revealed that ARGs relevant to first-line antimicrobial classes for the treatment of neonatal sepsis were detected almost immediately after birth, with dual β -lactam and aminoglycoside carriage ubiquitous by the end of the first postnatal week (Chapter 5, Section 5.4.4). ARG richness ($\log_{10}$ distinct ARG features) peaked in the first week (D0-D7) before declining modestly at later timepoints, whilst ARG Shannon diversity showed an inverse temporal pattern: highest at birth (mean approximately 3.3–3.6 at D0-D7), then declining significantly by day 28 (mean $\sim$ 2.7) and stabilising through day 90 (mean $\sim$ 2.8) (Chapter 5, Figure 5.4). Whilst previous work has documented age-associated declines in overall ARG burden and diversity across the first 14 months (Samarra et al., 2024), this study's focus on the 0-90 day window reveals a more nuanced pattern: ARG diversity peaks very early and then declines as the bacterial community structure matures, yet specific high-priority ESBL markers, particularly *bla*_{CTX-M-15} in *Escherichia* and *Klebsiella*, persist even as overall resistome diversity contracts. This finding has important implications for targeting early interventions, as it suggests that global resistome metrics may mask the persistence of clinically actionable threats.

Four distinct phases emerged from longitudinal analysis, providing a framework for understanding early-life AMR acquisition in this cohort:

- Phase 1: Immediate seeding (D0-D7). Characterised by detection of resistance determinants from birth, near-universal dual β -lactam and aminoglycoside resistance

carriage by day 7, high ARG richness and diversity, Proteobacteria dominance, significant delivery-mode divergence; specifically, Caesarean-born infants exhibited a distinct resistome signature with higher initial ARG richness compared to vaginally-born infants (resistome: $R^2 = 0.078$, $p = 0.001$ at day 1; $R^2 = 0.059$, $p = 0.001$ at day 5).

- Phase 2: Early selection (D7-D14). Marked by persistent delivery-mode differences which are compounded by early antibiotic signatures on resistome composition ($R^2 = 0.050$, $p = 0.003$ at day 7), establishment of multi-class resistance patterns in a large fraction of infants and beginning of taxonomic turnover towards obligate anaerobes.
- Phase 3: Succession and consolidation (D14-D28). During which ARG diversity declines significantly (Shannon drops from ~ 3.6 to ~ 2.7) and bacterial diversity begins to stabilise. Notably, delivery-mode effects on resistome composition persist and account for an increasing proportion of variance ($R^2 = 0.099$, $p = 0.002$ at day 28) even as microbiome composition between delivery groups begins to converge. Critically, the persistence of *bla*_{CTX-M-15} serves as a more clinically relevant predictor of sustained AMR carriage than the broad diversity metrics.
- Phase 4: Reservoir stabilisation (D28-D90). Where community structure stabilises, β -diversity convergence is nearly complete (no delivery-mode effect by day 90: $R^2 = 0.016$, $p = 0.363$), ARG diversity plateaus (Shannon ~ 2.8), but resistant strains, particularly *bla*_{CTX-M-15}-positive *Enterobacteriaceae*, remain detectable in a minority of infants.

This phased framework provides a finer temporal resolution during the critical first three months, when exposure-associated differences are most prominent and when empiric antibiotic treatment regimens are most likely to be deployed.

Critically, exposure-associated differences, including delivery mode, early antibiotic administration, and feeding practices, exerted their strongest observable effects during Phases 1-2, with progressive attenuation by Phase 4 as microbiome diversity converged (Chapter 6, Figures 6.2-6.3). However, whilst community-level diversity metrics showed increasing overlap across time, the persistence of specific high-priority AR markers provided a more sensitive indicator of sustained colonisation risk. At day 28, the consolidation point where ARG diversity has declined and community structure is stabilising, a cumulative multi-hit exposure score (including caesarean delivery, prior antibiotic exposure, and low/no breastfeeding) showed evidence of an association with *bla*_{CTX-M-15} carriage (adjusted odds ratio per step 2.55, 95% confidence interval 0.97-6.70; $p = 0.058$), with model-predicted carriage probabilities rising from 22% (0 hits) to 38% (1 hit) to 62% (≥ 2 hits), though this association attenuated by day 90 (Chapter 6, Table 6.5; Figure 6.14). This observation between overall diversity metrics and specific risk markers has rarely been explored in prior resistome studies. My study contributes to a crucial insight: in high-exposure settings where global diversity scores rapidly converge, analysing clinically prioritised AR markers offers more actionable risk information than relying on diversity indices alone.

This general discussion is organised to address three central questions arising from these findings. *First, how do contextual realities in LMIC settings structure the exposure landscape that drives early AMR acquisition?* (Section 7.2). Specifically, how does colonisation pressure (Section 7.2.1) and trajectories of microbiome development (Section 7.2.2) interact to produce the observed patterns? *Second, what mechanisms link specific exposures to observable patterns of resistome development, and where along this pathway can interventions be targeted?* (Section 7.3). This includes examining delivery pathways and the immediate postnatal environment (Section 7.3.1), selection pressure from antibiotics (Section 7.3.2), microbial maturation under realistic feeding and hygiene constraints (Section 7.3.3), and the

clinical utility of marker persistence signals (Section 7.3.4). *Third, what does persistence of clinically prioritised markers reveal about reservoir formation and future infection risk?* By synthesising findings across maternal carriage, delivery pathways, antibiotic exposure, and microbiome maturation, this discussion identifies high-leverage intervention points whilst maintaining a clear distinction between colonisation and clinical infection. The chapter concludes by proposing a data-driven conceptual framework (Section 7.6) that integrates these insights to guide future interventional research aimed at reducing early-life AMR burden in resource-limited settings, followed by key conclusions and take-home messages (Section 7.7).

7.2 How Exposure Shapes Early-Life AMR Acquisition

The concept of the 'exposome', defined by Wild, (2012) as the totality of environmental exposures from conception throughout life, provides a useful notion for understanding how contextual realities shape early-life microbial assembly. Ayeni et al. (2022) adapted this concept for African infant cohorts, noting that "data on microbiome-exposome interactions is lacking in African cohorts". This thesis cohort begins to address this gap by providing high-resolution temporal data on how the urban West African environment, characterised by high infectious disease burden, high post-partum antibiotic use, variable antibiotic access, and diverse cultural practices, shapes early-life microbial and resistome assembly.

7.2.1 Colonisation pressure in resource-limited settings

Most mechanistic descriptions of early-life microbiome and resistome development derive from HIC birth cohorts (Bäckhed et al., 2015; Shao et al., 2019; Stewart et al., 2018), where household sanitation standards constrain exposure intensity. In this present cohort, infants enter a high-contact microbial environment in which colonisation and selection pressures are

sustained from birth. When antibiotic exposure is intense and continuous, the marginal signal of any single modulator may appear weaker across time.

The intensity of colonisation pressure is quantified by time-to-first-detection curves (Chapter 5, Figure 5.8). Whilst longitudinal resistome studies have documented ARG presence at monthly or weekly intervals (Gibson et al., 2016; Samarra et al., 2024; Yassour et al., 2016), to our knowledge, no published cohort has used survival analysis methods to quantify acquisition rate with daily sampling resolution (D0, D1, D3, D5, D7) in the immediate neonatal period. Kaplan-Meier estimates reveal that 10% of infants carry β -lactam or aminoglycoside ARGs at birth, rising to 50% by day 3 and 90% by day 7. Dual β -lactam and aminoglycoside resistance carriage reaches 75% by day 7, plateauing near 95% by day 14. This dramatic early acquisition contrasts with the assumption in many HIC cohort studies that ARG accumulation is gradual and primarily driven by postnatal antibiotic exposure (Guo et al., 2020; Stearns et al., 2017; Wu et al., 2024). The 10% baseline carriage at birth and near-universal dual antibiotic resistance by day 14 strongly suggests that environmental and vertical transmission pathways dominate early resistome assembly in this high-exposure setting, with infant-level selection pressures acting on an already ARG colonised baseline.

Birth mode shapes the newborn's earliest contact network. In this cohort, delivery mode was associated with significant differences in both taxonomic and resistome composition during the first month. For the microbiome, mode of delivery explained 6.2% of between-infant variation at day 5 ($R^2 = 0.062$, $p = 0.001$), persisting at day 28 ($R^2 = 0.056$, $p = 0.002$), but with no evidence of effect by day 90 ($R^2 = 0.016$, $p = 0.363$). The resistome showed similar patterns: day 1 ($R^2 = 0.078$, $p = 0.001$), day 5 ($R^2 = 0.059$, $p = 0.001$), day 28 ($R^2 = 0.099$, $p = 0.002$), attenuating by day 90 ($R^2 = 0.024$, $p = 0.113$). Caesarean delivery was associated with lower Shannon diversity at day 5 (adjusted difference -0.665 , $p < 0.001$) and higher total ARG burden at days 5 and 28 (Chapter 6, Table 6.2).

This temporal population convergence contrasts with European cohorts where birth-mode signatures persist beyond the first year (Matharu et al., 2022; Patangia, Grimaud, Wang, et al., 2024). The more rapid convergence by day 90 potentially reflects high environmental microbial exposure, diverse feeding practices including prelacteal traditions, or overwhelming background AMR colonisation pressure that progressively dilutes the marginal contribution of birth-mode-associated seeding.

The healthcare facility environment shapes initial exposure, particularly for C-section infants who experience longer facility stays. Many healthcare facilities in low-resource settings lack consistent hand hygiene and reliable waste management (UNICEF & WHO, 2023, 2021). Studies from HIC settings demonstrate that hospital birth increases colonisation by Proteobacteria and opportunistic taxa (Guo et al., 2020; Stearns et al., 2017). The present cohort exhibits similar pattern of opportunistic pathogen enrichment but sustained by facility-level constraints and prolonged exposure to the hospital environment. C-section reflects not only surgical delivery, but also longer facility stays and greater cumulative contact with healthcare-associated microbial reservoirs (Shao et al., 2019), explaining why C-section differences are largest in the early time period (days 1-7) but attenuate as household and community-level microbial exposure dominates.

Antibiotic exposure may be embedded within the postpartum period. In this cohort, postpartum maternal antibiotic prescription was universal (87/87 mothers); however, actual exposure occurred in 84/87 cases, with the remaining three mothers unable to fulfil the prescription due to affordability constraint. Comparative data from HIC settings suggest maternal antibiotic use is typically 20-40% for C-section prophylaxis and less than 10% for uncomplicated vaginal delivery. In high-burden environments where postpartum infection risk and diagnostic uncertainty are salient, antibiotics function as routine risk-management tools, sustaining

selection pressure during the same neonatal window in which infant colonisation is being established.

Following discharge, colonisation pressure is heavily influenced by household and community exposure networks. Chapter 4 discusses the impact of dense household contact and frequent proximity to environmental contamination. WASH constraints are substantial: the Northwest zone has among the lowest access to improved sanitation nationally, with substantial open defecation prevalence and low per-capita water availability constraining hand hygiene (Eliana Drakopoulos, 2020). These constraints reflect structural inequities rather than individual choices. Recent evidence highlights strong associations between environmental pollution and ARG proliferation, with sewage-contaminated water, soil, and arthropod vectors acting as AMR reservoirs and carriers (Anand et al., 2021; Chatzigiannidou et al., 2025; Cook et al., 2025). Culturally patterned neonatal contact points, such as Qur'anic water administration and close-contact caregiving, cluster in the same neonatal period in which acquisition signatures are most detectable (Chapter 4). Collectively, these findings justify treating repeated exposure as a plausible default and explain why downstream stewardship alone may have limited leverage if the upstream environmental resistome remains unaddressed.

7.2.2 Trajectories of microbiome development

Early and regular sampling/observation provides the most reliable framework for interpreting early-life signals. Rather than a stable baseline interrupted by modulators, infancy is defined by rapid, age-structured succession. The TEDDY study (Stewart et al., 2018) followed 903 infants with monthly sampling from 3 to 46 months, documenting progressive diversification and declining Proteobacteria, with initial delivery-mode differences becoming progressively harder to distinguish as the microbiome matured. My study utilises these checkpoints to capture

these dynamics: D0-D7 (acquisition and divergence); D7-D14 (early selection); D28 (consolidation, end of neonatal period); and D90 (convergence). Early seeding events exert disproportionate influence on all subsequent trajectories (Bäckhed et al., 2015; Dominguez-Bello et al., 2010).

D0-D7: Acquisition and divergence. Differences driven by initial exposures are most apparent during the first postnatal week. Taxonomic composition is characterised by dominance of facultative anaerobic Proteobacteria, particularly *Escherichia* and *Klebsiella* (Chapter 5). In C-section infants, *Escherichia* comprised 50-100% of relative abundance at day 0, with substantial *Klebsiella* representation. Vaginally delivered infants also showed *Escherichia* dominance but with greater *Bacteroides* representation and, critically, the earlier presence of *Bifidobacterium*. A striking observation is the near-complete absence of *Bifidobacterium* in caesarean-section infants during days 0-7, with first appearance at day 14, resulting in an extended Proteobacteria-dominant phase.

This delayed *Bifidobacterium* colonisation is consistent with HIC observations: vaginally born infants show 21% *Bifidobacterium* carriage on day 1, rising to 64% by day 7, whilst C-section infants exhibit delayed expansion with catch-up by 8-12 weeks (Guo et al., 2020; Hill et al., 2017; Saturio et al., 2021). However, the near-complete absence at days 0-7 in C-section infants in this study suggests a more pronounced delay, potentially reflecting compounded effects of facility exposure, maternal antibiotics, and prelacteal feeding.

The clinical significance of these observations lies in *Bifidobacterium*'s role in colonisation resistance and immune education. The extended Proteobacteria-dominant phase (days 0-14) represents a vulnerability window during which *Enterobacteriaceae* carrying ESBL markers can establish stable carriage. This is supported by C-section infants having higher total ARG burden at day 5 and 28 (Chapter 6), with *Escherichia* and *Klebsiella*, the dominant genera

during this window, being primary hosts for *bla*_{CTX-M-15} and leading causes of neonatal sepsis in West Africa (Grant-Beurmann et al., 2022).

D7-D14: Early Selection and Persistence. The period from D7–D14 is marked by early selection pressure. The gut environment becomes less permissive due to competition from early anaerobes and declining oxygen tension. This pressure determines whether early colonisers persist into stable carriage states or decline. The delayed *Bifidobacterium* colonisation in C-section infants allows the initial ARGs acquired in the first week to consolidate their niche, reinforcing the interpretation that seeding precedes selection.

D28: Consolidation and Persistence. By four weeks, early colonisers either persist into stable carriage states or decline as the gut becomes less permissive. Day 28 is best interpreted as a consolidation checkpoint rather than a delayed acquisition start. The maturation trajectory, rising *Bifidobacterium* and reduced *Enterobacteriaceae*, provides biological evidence for why persistence, rather than acquisition, differentiates infants at this stage (Chapter 5). Biological mechanisms include declining gut oxygen tension favouring obligate anaerobes, breast milk oligosaccharides supporting *Bifidobacterium*, immune maturation strengthening colonisation resistance, and competition from anaerobes suppressing Proteobacteria (Milani et al., 2017).

The risk-stacking analysis, representing a cumulative risk of environmental and clinical exposures, shows a graded pattern of *bla*_{CTX-M-15} carriage by cumulative exposure at day 28 (adjusted odds ratio per step 2.55, 95% confidence interval 0.97–6.70; $p = 0.058$), with predicted probabilities rising from 22% (0 hits) to 62% (≥ 2 hits) (Chapter 6). This graded pattern suggests that whilst ARG diversity declines (Shannon drops from ~ 3.6 at days 0-7 to ~ 2.7 at day 28), specific high-priority markers persist in subsets of infants with cumulative exposure burdens. The TEDDY and INFANTMET cohorts documented consolidation phases at similar timelines (Hill et al., 2017; Stewart et al., 2018), but my cohort shows that resistome

compositional differences persist longer ($R^2 = 0.099$, $p = 0.002$ at day 28) than microbiome differences, suggesting that the resistome consolidation lags microbiome convergence. This comparison reveals a key finding: in settings with high environmental exposure, broad measures of microbiome structure become very similar across the cohort early in life. Because these diversity indices no longer show differences between individuals, they cannot be used to distinguish between high- and low-risk infants. Instead, detecting specific antibiotic resistance markers provides a clearer way to identify which infants are at risk, offering a more practical tool for clinical monitoring than general diversity scores.

D90: Attenuation and Convergence. By three months, exposure-linked contrasts are less distinct on community-wide metrics, with no evidence of delivery-mode effect on microbiome composition ($R^2 = 0.016$, $p = 0.363$). This attenuation does not invalidate earlier signals but reinforces that later infancy compresses between-group microbiome structures because of shared maturation constraints and repeated common exposures reduce the visibility of earlier differences. One interpretation is that reduced separability on diversity endpoints, not 'recovery' as suggested by earlier studies (e.g., Stearns et al., 2017), alongside marker-level carriage, remains informative for subsets of infants.

This faster convergence contrasts with European cohorts where birth-mode signatures persist to 12-24 months (Matharu et al., 2022; Patangia, Grimaud, Wang, et al., 2024). The more rapid convergence potentially reflects: (i) high environmental exposure overwhelming the signal of vertical transmission; (ii) diverse feeding practices introducing common exposures; (iii) background AMR pressure reducing visibility of modulator-specific effects; and (iv) shared biological maturation processes operating across all infants. Whilst diversity metrics converge, marker-level persistence continues. *bla*_{CTX-M-15} carriage remains detectable in a minority of infants at day 90, and the multi-hit exposure association attenuates (aOR 0.95, $p = 0.917$), suggesting a baseline colonisation pressure overwhelms the marginal signal of individual

modulators by three months. However, the subset maintaining high-ESBL carriage may represent a persistent reservoir with implications for subsequent infection risk. This divergence between overall community stability and specific marker persistence shows why it is important to track individual ARG, rather than relying on general diversity scores, in high-exposure settings.

7.3 From Mechanisms to Action: Identifying Pathways to Reduce AMR Colonisation

Early-life AMR colonisation is a clinically significant determinant of neonatal risk of invasive infection. Gut carriage of ESBL-producing *Enterobacteriaceae* establishes an endogenous reservoir from which invasive infections can arise. In a Malawian mother-infant cohort, neonates whose mothers carried ESBL-producing *Enterobacteriaceae* were significantly more likely to develop sepsis, with colonisation preceding invasive disease in numerous cases (Iqbal et al., 2024; Pearse et al., 2025; Schwartz et al., 2023). Studies from different settings report that ESBL carriage in early life is associated with increased risk of ESBL-producing Gram-negative bloodstream infections, with case-fatality rates exceeding 30% in some neonatal sepsis cohorts (Carvalho et al., 2022; Cornick et al., 2021; Labi et al., 2020; Musicha et al., 2017; Shao et al., 2019). While most colonised infants remain asymptomatic, the high prevalence of baseline carriage creates a substantial reservoir for potential outbreaks, particularly when augmented by healthcare-associated transmission.

In Northern Nigeria, the clinical risks are particularly high. Neonatal mortality rates remain high, with sepsis accounting for a larger proportion of deaths in the first month of life (National Population Commission (NPC) & ICF, 2019). Healthcare facilities face constraints in diagnostic capacity (both staff, instruments and consumables), and empiric treatment protocols are often based on outdated susceptibility patterns (K. C. Iregbu et al., 2021). When first-line

empiric therapy (ampicillin plus gentamicin) fails because the causative organism carries corresponding ARGs, the delay in appropriate treatment can be fatal.

The intervention goal ideally prioritises reducing the intensity and persistence of high-consequence AMR carriage during the neonatal window, acknowledging that total elimination is an unfeasible target in high-exposure settings. Specifically, interventions should aim to: (i) delay or prevent acquisition of clinically important antibiotic resistance markers (*bla*_{CTX-M-15}, carbapenemase genes, aminoglycoside resistance genes) during the first two weeks of life; (ii) support microbial maturation processes (particularly *Bifidobacterium* colonisation) that provide colonisation resistance to pathogens possessing relevant ARGs; and (iii) reduce selection pressure that amplifies antibiotic resistant clones during the consolidation phase (days 14-28). These harm-reduction intercessions acknowledge that complete prevention is not feasible but that targeted interventions during the critical window of neonatal development can shift population-level risk in clinically meaningful ways.

7.3.1 Delivery Pathways and Early Care Contacts

The delivery pathway and immediate neonatal care in this cohort acts as a concentrated, high-density exposure window that establishes the infant resistome baseline within the first postnatal week. The clear early separation by delivery mode described in Chapter 6 reflects how this brief period (days 0-7) concentrates acquisition opportunities. These patterns align with evidence from other LMIC settings where resistant bacteria are acquired rapidly through combination of maternal, healthcare-associated, and environmental transmission pathways (Carvalho et al., 2022; Musicha et al., 2017; Woerther et al., 2013). In the present cohort, the rapid transition from 10% carriage at birth to 90% by day 7 (Chapter 5, Figure 5.8) suggests that by the end of the first week, a high-resistance burden has already been consolidated.

This rapid seeding process is not confined to the delivery room. Infants are quickly moved from hospital to household environments, where initial acquisition is often reinforced. Studies examining household transmission demonstrate that resistant bacteria circulate within family units, with environmental reservoirs (water sources, food preparation surfaces) serving as persistent sources of re-exposure (Musicha et al., 2017; Ruppé & Schrenzel, 2018). In settings where open defecation persists and water availability is constrained (Section 7.2.1), preventing household-level exposure is extremely challenging.

The period immediately after birth is the most critical time to intervene. However, HIC strategies, such as maternal decolonisation or rigorous environmental screening (Cantey et al., 2013) are impractical for resource-constrained settings. In Northern Nigeria, infrastructure gaps and chronic staffing shortages render the sustainability of such high-intensity models improbable, if not impossible (K. C. Iregbu et al., 2021). My findings suggest that while hospital-level IPC may lower initial acquisition intensity, the negligible impact of delivery mode on the resistome by day 90 ($R^2 = 0.017$, $p = 0.288$) suggests that clinical efforts alone cannot eliminate exposure when household reservoirs are high. Therefore, the strategic goal should shift from total prevention to reduction of exposures. By prioritising the delay of acquisition, for instance, from day 0 to day 7, or reducing the burden of dual-antibiotic (β -lactam and aminoglycoside) resistance during the first two weeks, we create a tractable pathway to protect infants at their most vulnerable age.

7.3.2 Selection Pressure and Antibiotic Stewardship

As mentioned previously, postpartum maternal antibiotic exposure was ubiquitous in this cohort, with prophylactic over-prescription primarily involving metronidazole and co-amoxiclav. This is clinically important because such high exposure exerts multifaceted

selection pressure during the critical window of infant microbial assembly. Maternal antibiotics shape the infant gut resistome directly through the vertical transmission of AR strains (Patangia, Grimaud, Wang, et al., 2024) and indirectly through breast milk (Das et al., 2019). The near-ubiquitous exposure observed in my study contrasts with HIC settings, where maternal postpartum antibiotic use—beyond standard pre-operative prophylaxis—is typically limited to 20-40% following caesarean sections and under 10% for vaginal deliveries (Section 7.2.1). This disparity likely reflects a syndromic management approach for the mothers in this cohort who may have higher infection risk after delivery due to inadequate WASH infrastructure.

At the infant level, data presented in Chapter 6 supports the view that selection effects are most interpretable when aligned to timing of the antibiotic administration. In the antibiotic-window analysis, Shannon diversity is lowest in the 0-7 days after first antibiotic challenge and increases thereafter, reflecting an acute perturbation superimposed on a developing microbiome baseline (Chapter 6, Figure 6.6). Crucially, while diversity is suppressed, the total ARG burden shows only marginal shifts. This suggests that in this setting, antibiotics act upon a resistome that is already high and heterogeneous; the drugs select for existing resistant bacteria in the microbiome community rather than introducing them. This is consistent with the ZEBRA trial, which found that amoxicillin-clavulanic acid caused immediate resistome shifts and *Bifidobacterium* depletion, though these effects eventually normalised (Reyman et al., 2022). Similarly, Yassour *et al.* describes transient microbiome perturbations following infant antibiotic courses, with recovery trajectories dependent on antibiotic class, duration, and timing. These patterns support a restrained interpretation: antibiotics can shape short-term community structure without being the sole explanation for why resistance is present early.

Building on these observations, the implication is that antibiotic stewardship cannot be framed only as "reduce antibiotic use". In this setting, stewardship due diligence is dictated by complex

intersection of poor diagnostics and unregulated antibiotic access. In this cohort, maternal antibiotic acquisition was primarily by out-of-pocket expense managed through local pharmacies, where non-exposure was driven by affordability constraints (3/87 mothers). This reliance on the private sector introduces a critical and unmanageable variable: the uncertain quality and potency of antibiotics purchased outside regulated hospital pharmacies (A. A. Adamu et al., 2020; Auta et al., 2019; Gladys et al., 2025; Thomson et al., 2021). This creates a clinical paradox where infants face a double burden of selection pressure from poor-quality dosing, driven by financial necessity, and treatment failure when resistance renders standard empiric regimens ineffective.

Antibiotic stewardship is not aligned with upstream infection-risk reduction (IPC in facilities; feasible hygiene support post-discharge) and is likely to have limited traction, because it is asking clinicians and families to accept risk without changing the conditions that generate it. Effective antibiotic stewardship in LMIC neonatal settings requires a systems approach addressing diagnostic capacity, supply chain reliability, and economic access alongside traditional stewardship interventions (K. C. Iregbu et al., 2021; Laxminarayan et al., 2016; Okeke et al., 2024). This cohort's dataset, where antibiotic exposure produced detectable short-term effects on diversity, but overall ARG burden remained high across exposed infants, suggests that selection pressure operates against a backdrop of intense colonisation pressure, and reducing antibiotic selection alone is unlikely to eliminate high-antibiotic resistance carriage without concurrent reductions in acquisition intensity.

7.3.3 Microbial Maturation, Feeding, and Hygiene Constraints

Whilst colonisation and selection set the early stage, microbial succession and resilience determine whether the community transitions towards a more typical infant-associated

configuration. In my cohort, the dominant biological signal remains an age-structured succession: early Enterobacteriaceae prominence becomes less ubiquitous over time, whilst *Bifidobacterium* increases. This provides a practical definition of "resilience": not a return to a low-exposure baseline, but the capacity for maturational taxa to establish despite sustained exposure. The delayed *Bifidobacterium* colonisation in C-section infants, near-complete absence at days 0-7, represents a vulnerability window during which colonisation competition against pathogenic Enterobacteriaceae is compromised (Section 7.2.2). Supporting earlier *Bifidobacterium* establishment is therefore a plausible intervention strategy, as *Bifidobacterium* species acidify the gut environment, produce antimicrobial compounds, and competitively exclude facultative pathobionts (Guo et al., 2020; Samarra et al., 2025; Saturio et al., 2021).

Feeding-category contrasts in this cohort showed only modest differences in Shannon diversity and ARG burden (Chapter 6, Figure 6.9). This likely reflects the fact that breastfeeding is not binary in this setting; high breastfeeding prevalence is confounded by low exclusivity and water supplementation (Chapter 4). While HIC studies associate exclusive breastfeeding with pathogen suppression (Guo *et al.*, 2020), the introduction of water, often from contaminated sources, plausibly introduces competing exposure routes that overwhelm the protective qualities of breast milk (Mebrahtom et al., 2022; Ogbo et al., 2017).

In this context, breastfeeding support should include practical hygiene advice for any supplementary feeding and the equipment used. Interventions combining breastfeeding promotion with treated water, safe storage containers, and education on hygienic preparation of supplementary feeds may be more effective than breastfeeding promotion alone (Mebrahtom et al., 2022; Null et al., 2018). However, feasibility must be evaluated within local cultural contexts. In Northern Nigeria, prelacteal feeding with water (including Qur'anic water) is a culturally valued practice (Chapter 4), and interventions that ignore these practices are unlikely

to be adopted. This highlights the need for research into behavioural training that addresses cultural norms while promoting EBF. My ongoing work, supported by a Nestlé Foundation grant, seeks to evaluate whether EBF can effectively limit ARG carriage in this setting. By investigating the relationship between feeding exclusivity and the infant resistome, this research aims to determine if caregiver education, particularly regarding the 'thirst' misconception, can be leveraged as a pragmatic public health tool to delay environmental acquisition in high-burden environments.

7.3.4 Why Marker-Specific Carriage Is More Actionable Than Community-Wide Metrics

Community-wide metrics, such as Shannon diversity, total ARG burden, and ordination structure, capture overall ecological patterns but may obscure clinically relevant signals. In contrast, marker-specific endpoints (e.g., *bla*_{CTX-M-15} carriage in *Enterobacteriaceae*) track discrete, high-consequence carriage states that are more directly actionable for infection prevention and appropriate antibiotics. In both Chapters 5 and 6, community composition and diversity metrics are dominated by age-structured succession, and between-group separation attenuates by day 90. In that setting, marker-level endpoints remain informative because they are aligned with specific relevant antibiotic resistance markers than diffuse community-wide metrics.

The multi-hit analysis (Chapter 6, Figure 6.13) illustrates this distinction. When early-life exposures were represented cumulatively (caesarean delivery, antibiotic exposure, low breastfeeding), the predicted probability of *bla*_{CTX-M-15} at day 28 increased across strata: 22% (0 hits), 38% (1 hit), 62% (≥ 2 hits), showing a strong trend towards association (aOR per step 2.55, 95% CI 0.97–6.70; $p = 0.058$). This association attenuated by day 90, suggesting that day

28 functions as a consolidation checkpoint, where the 'memory' of early-life conditions is most detectable before environmental noise predominates.

Operationally, these findings support targeted screening for high-risk infants rather than universal surveillance. Identifying carriage at the Day 28 checkpoint provides a critical window for targeted infection control. This is especially vital where resistant pathogens challenge first-line antibiotic regimens and dominate cases of neonatal sepsis (Musicha et al., 2017; Thomson et al., 2021).

Section 7.6 uses these observations to propose a targeted prevention strategy. This framework focuses on the most effective interventions identified in the cohort, specifically improving hygiene during birth and using risk-stratified screening to manage infants who already carry resistant bacteria.

7.4 Strength and limitations of the study

A primary strength of this study is its contextual validity; by anchoring the findings to the distinct context of Northern Nigeria, this work provides a necessary comparison to HIC or low-environmental exposure context. While the age-structured microbial maturation aligns with global biological patterns, the magnitude of the early resistome burden and the rapid Day 90 convergence captured here reflect the real-world drivers, such as high environmental exposure and ubiquitous maternal antibiotic use, that define neonatal health in this region. The results offer a more accurate benchmark for other high-exposure settings, ensuring that future interventions are designed for the environments where they are most urgently needed.

Furthermore, the reliability and impact of these findings are inextricably linked to the context in which they were generated. This study transition from a pilot phase in a converted shipping container to the establishment of a permanent laboratory at MMSH allowed for the mitigation

of 'field-to-lab' risks, such as intermittent electricity and fragile cold chains (Chapter 3). This structural growth did more than just secure data integrity; it fostered local technical leadership. As detailed in my qualitative work on Nigerian research staff perspectives (Akpulu, Yakubu, et al., 2025), embedding these capabilities locally ensures that the expertise to monitor neonatal AMR remains a permanent asset of the community.

The interpretation of these findings is bounded by several technical and structural constraints. Primarily, metagenomic detection does not inherently equate to phenotypic resistance; while ARG detection confirms the genetic potential for resistance, it does not confirm its functional expression (Sommer et al., 2009). Although meta-transcriptomics would offer insight into active gene expression, such an approach remains technically and financially prohibitive for large-scale longitudinal surveillance (Heintz-Buschart & Wilmes, 2018). Therefore, this thesis frames early-life colonisation as a clinically significant risk indicator, measuring the potential of the endogenous reservoir, rather than a direct predictor of individual treatment failure.

This challenge of interpretation is compounded by limitations in genomic resolution. Without systematic reconstruction of plasmids or MGEs, bundled resistance patterns (e.g., dual β -lactam and aminoglycoside carriage) must be interpreted as taxonomic co-occurrence rather than confirmed physical linkage. Similarly, the lack of strain-level genomic linkage across timepoints precludes a definitive distinction between the persistence of a single clone and the repeated acquisition of new strains. These factors limit the ability to model HGT dynamics or prove specific transmission pathways between mothers, infants, and the environment.

From a statistical perspective, the stratified analyses often encountered the "small denominator" constraint typical of high-risk subgroups. Specifically, in the multi-hit *bla*_{CTX-M-15} analysis, the ≥ 2 -hit group at day 28 was small, leading to wide confidence intervals; by day 90, the attenuation of the gradient further suggests an exploratory trend rather than a definitive

population estimate. This lack of statistical power is matched by a lack of direct clinical outcome data. While the clinical relevance of early ARG carriage is established through regional literature linking ESBL colonisation to neonatal sepsis (Carvalho et al., 2022; Musicha et al., 2017), this study's primary contribution is as a foundational, exploratory analysis. As a result, it does not directly quantify infection risk or treatment failure rates within its own cohort. Instead, the unique prevalence data generated here, specifically the 62% carriage rate in the ≥ 2 -hit group, provides the critical prevalence estimates required for robust power calculations in subsequent, hypothesis-driven analysis designed to specifically measure infection risk.

7.5 Future Directions

7.5.1 Immediate Priority: Characterising Maternal and Household Reservoirs

An immediate research priority is the analysis of the archived maternal rectal swabs and environmental samples already collected and stored during this study (Chapter 4). While Chapters 5 and 6 established that the infant resistome is consolidated within the first week of life, it remains unclear whether these AR strains originate from maternal vertical seeding or horizontal acquisition from the immediate environment.

By analysing these archived samples, which include maternal rectal swabs and high-touch surfaces like delivery suite surfaces and household water sources, it will be possible to perform strain-level genomic linkage. This will provide creditable evidence needed to determine if the 90% carriage rate observed by day 7 (Chapter 5) is driven by maternal transmission or environmental contamination. Identifying the primary source of these early colonisers is a prerequisite for deciding whether local clinical policy should prioritise maternal decolonisation or structural improvements in facility and household hygiene.

7.5.2 Resolve Genetic Context and Mobile Genetic Elements for Clinically Prioritised Markers

A critical technical priority is to move beyond 'gene presence' to resolve the genomic architecture of high-priority determinants like *bla*_{CTX-M-15}. Utilising the long-read sequencing data already generated for this cohort, future work will move from taxonomic co-occurrence to establishing definitive physical linkage through specialised long-read assembly and plasmid reconstruction pipelines (e.g., MOB-suite). This structural resolution is essential for determining whether the 'bundled' resistance patterns observed in this cohort (e.g., co-occurring β -lactam and aminoglycoside resistance markers) reflect shared MGEs (Berglund et al., 2023). Distinguishing chromosomal from plasmid-borne resistance has direct implications for transmission risk and intervention design. While chromosomal integration can facilitate the clonal dissemination through vertical inheritance, highly mobile plasmids represent an additional threat by enabling horizontal gene transfer across diverse commensal and opportunistic species. Resolving these genomic locations is therefore critical to determining whether the observed resistance is driven by the spread of a single 'outbreak' strain or the circulation of a mobile genetic reservoir. Both routes can drive rapid spread in neonatal settings, but they imply different surveillance and intervention priorities (e.g., lineage-focused outbreak control versus containment of highly transmissible MGEs). Furthermore, as recent continental-scale surveys have identified over 1,000 novel bacterial species within the African gut (Maghini et al., 2025), there is significant potential for these long-read assemblies to uncover novel plasmids and MGEs that are currently absent from global databases.

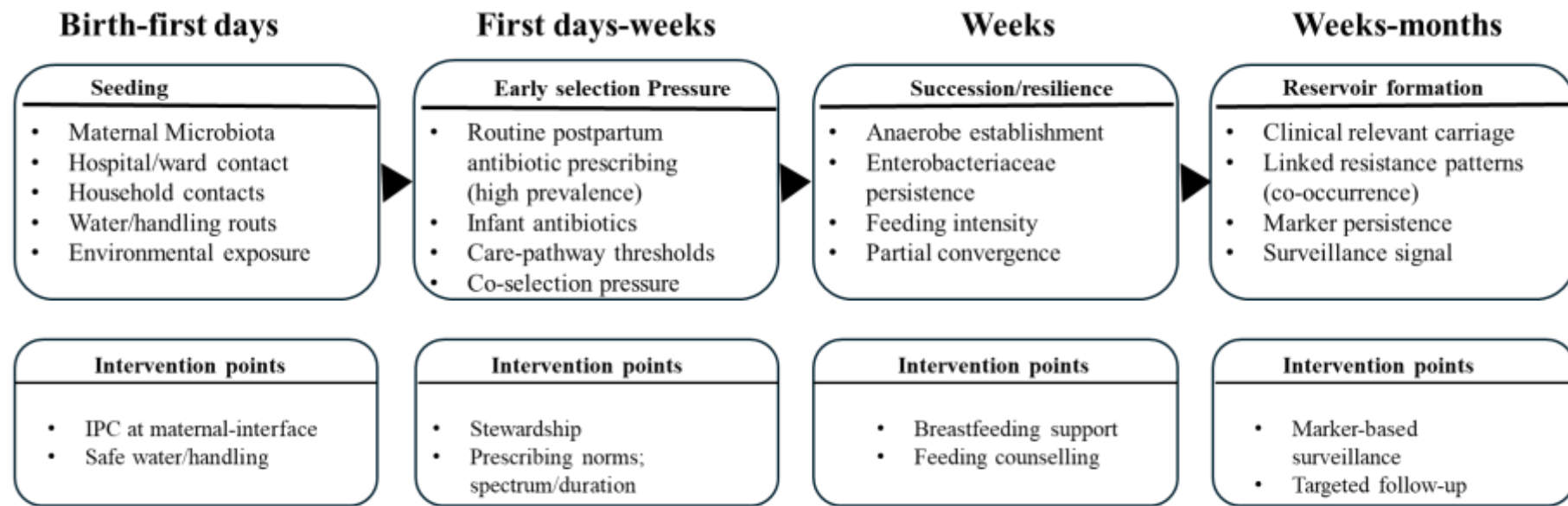
7.5.3 Longitudinal Stability and the Weaning Transition

While my thesis focused on the first 90 days of life, the study infrastructure included an extended collection phase. Stool samples through the weaning period (6-12 months) have

already been collected, and microbial DNA has been extracted and archived. Sequencing this extended longitudinal library is a high-priority next step. Weaning represents a secondary high-risk window in Northern Nigeria, where shifting dietary patterns and increased environmental exposure may reshape the resistome (Stewart et al., 2018). Analysing these archived samples will allow for a credible test of whether the Day 90 convergence observed in Chapter 6 represents a stable maturation or a transient phase disrupted by later-life factors. Such data will clarify if early-life 'risk markers' like *bla*_{CTX-M-15} persist as a permanent endogenous reservoir or fade as the gut microbiome matures.

7.5.4 Proposed Institutional Engagement and Protocol Review

A primary recommendation following this research is the formal dissemination of these findings to the study site's clinical leadership. The high prevalence resistome data generated here should serve as the evidentiary basis for a collaborative review of current postpartum antibiotic protocols. Specifically, the near-universal use of metronidazole and co-amoxiclav in most deliveries warrants re-evaluation to determine if a risk-stratified prescribing model could be implemented. By presenting the cohort's 90% day-7 carriage rate to hospital stewardship committees, this work can facilitate a shift away from routine prophylactic over-prescribing towards a more targeted therapy.



Thesis observed endpoints: Microbiome | resistome class burden | prioritised marker carriage | host-association (where feasible)

Figure 7.1: A time-structured framework for neonatal AMR risk management in Northern Nigeria

This conceptual framework maps four stages of the infant resistome (seeding, selection, succession, reservoir formation) to specific clinical windows identified in this cohort: the Day 0-7 Acquisition Window, the D7-D14 Early selection window, the Day 28 Consolidation Checkpoint, and the Day 90 Convergence. The figure highlights key biological processes, potential intervention points, and the primary endpoints used in this thesis (microbiome composition, resistome burden, and prioritised marker carriage) to validate each stage

7.6 A Conceptual Framework to Identify Future Interventional Strategies

My data reveal a time-structured process in which early-life AMR carriage emerges rapidly, is shaped by external drivers concentrated in the perinatal window and evolves through predictable maturational stages. Four key observations structure this interpretation.

First, colonisation occurs very early and from multiple sources. Ten percent of infants carry β -lactam and aminoglycoside resistance genes at birth, rising to 90% by day 7. This rapid early acquisition demonstrates that AR bacteria are introduced before most infant-level exposures are recorded. Delivery mode explains 6.2-7.8% of microbiome and resistome variation at days 1-5, reflecting how the immediate care environment, maternal microbiota, and facility contact concentrate acquisition opportunities. Infants then move quickly to household environments where WASH constraints, dense contact networks, and environmental reservoirs sustain colonisation.

Second, selection pressure overlaps temporally with early colonisation. Postpartum maternal antibiotic exposure was high (84/87 mothers), creating selection pressure when infant colonisation is being assembled. At the infant level, antibiotic exposure produced detectable short-term effects on Shannon diversity but weaker effects on total ARG burden, consistent with antibiotics acting on a resistome that is already high and heterogeneous. This supports the interpretation that selection amplifies AR strains acquired through early seeding, rather than initiating the resistome *de novo*.

Third, microbial succession proceeds despite sustained exposures, but ARG marker-level persistence better captures the lasting impact on the resistome. The dominant biological signal is age-structured succession: early Enterobacteriaceae prominence becomes less universal, whilst *Bifidobacterium* increases. By day 90, delivery-mode effects on microbiome composition attenuate, and community-wide metrics show increasing overlap. However, this

convergence does not imply elimination of clinically relevant carriage. The multi-hit analysis demonstrates that *bla*_{CTX-M-15} carriage at day 28 is associated with cumulative early-life exposures, with predicted probabilities rising from 22% (0 hits) to 62% (≥ 2 hits). This association attenuates by day 90, but the subset maintaining high ESBL carriage represents a persistent reservoir.

Fourth, timing is the most reliable organising principle. The thesis identifies critical time frames: D0-D7 (acquisition and divergence), D7-D14 (early selection), D28 (consolidation), and D90 (convergence). Exposure-linked signatures are most interpretable in the neonatal window. By day 28, early colonisers have either persisted or begun to decline. By day 90, shared maturation constraints compress between-group structure on diversity metrics, but marker-specific endpoints remain informative.

These findings suggest a conceptual framework (Figure 7.1) separating early-life AMR development into four linked stages, seeding, selection, succession, and reservoir formation, each aligned to specific time periods and intervention opportunities. In the seeding stage (birth to first days), maternal microbiota, hospital contact, household contacts, water routes, and environmental exposure introduce resistant organisms. Intervention points include IPC at the maternal-newborn interface and safe water/handling practices. In the early selection pressure stage (first days to weeks), routine postpartum antibiotic prescribing, infant antibiotics, and care-pathway thresholds amplify resistant clones. Intervention points include stewardship (prescribing norms, spectrum, duration). In the succession/resilience stage (weeks), anaerobe establishment, Enterobacteriaceae persistence, feeding intensity, and partial convergence shape community transitions. Intervention points include breastfeeding support and feeding counselling. In the reservoir formation stage (weeks to months), clinically relevant carriage, linked resistance patterns, marker persistence, and surveillance signals stabilise. Intervention points include marker-based surveillance and targeted follow-up. This framework identifies

potential intervention points based on timing, biological plausibility, and alignment with measurable endpoints.

By mapping microbiome composition, resistome class burden, and prioritised marker carriage onto each developmental stage, my work provides the necessary benchmarks to evaluate whether future interventions targeting specific nodes produce detectable shifts in colonisation potentials. The most effective strategies are those designed to limit the duration of ARG carriage during the neonatal window. This can be achieved by integrating targeted IPC within the neonatal window with antimicrobial stewardship and support for microbial maturation, tailored to the realities of the community (Sections 7.3.1-7.3.3).

References

- Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The Placenta Harbors a Unique Microbiome. *Science Translational Medicine*, 6(237). <https://doi.org/10.1126/scitranslmed.3008599>
- Aarts, C., Kylberg, E., Hörnell, A., Hofvander, Y., Gebre-Medhin, M., & Greiner, T. (2000). How exclusive is exclusive breastfeeding? A comparison of data since birth with current status data. *International Journal of Epidemiology*, 29(6), 1041–1046. <https://doi.org/10.1093/ije/29.6.1041>
- Abdill, R. J., Adamowicz, E. M., & Blekhman, R. (2022). Public human microbiome data are dominated by highly developed countries. *PLOS Biology*, 20(2), e3001536. <https://doi.org/10.1371/journal.pbio.3001536>
- Abramova, A., Karkman, A., & Bengtsson-Palme, J. (2024). Metagenomic assemblies tend to break around antibiotic resistance genes. *BMC Genomics*, 25(1), 959. <https://doi.org/10.1186/s12864-024-10876-0>
- Abubakar, R., Yohanna, S., & Zubairu, H. (2020). Influence of health decision making on maternal complications among women delivered at a general hospital, North-Western Nigeria. *Nigerian Journal of Clinical Practice*, 23(5), 675. https://doi.org/10.4103/njcp.njcp_659_18
- Adamu, A. A., Gadanya, M. A., Jalo, R. I., Uthman, O. A., & Wiysonge, C. S. (2020). Factors influencing non-prescription sales of antibiotics among patent and proprietary medicine vendors in Kano, Nigeria: a cross-sectional study. *Health Policy and Planning*, 35(7), 819–828. <https://doi.org/10.1093/heapol/czaa052>
- Adamu, A., Awosan, K. J., Mohammad, A. U., Muhammad, B. M., Muhammad, A., Bello, J. F., Ilah, G. B., Omenke, I. K., & Omoshalewa, U. M. (2021). Practice of Traditional Uvulectomy and Associated Factors Among Mothers of Under Five Years Children in Sokoto State, Nigeria. *Nigerian Journal of Basic and Applied Medical Sciences*, 1(1). <https://doi.org/10.53994/NJBAMS.202111.8>
- Adebola, S. O., Ogunkeyede, S. A., Obebe, F. A., Olaniyan, O. D., Fawole, O. B., & Salman, A. (2016). Profile of pediatric traditional uvulectomy in North-West Nigeria: The need for

caution and education. *International Journal of Pediatric Otorhinolaryngology*, 88, 194–198. <https://doi.org/10.1016/j.ijporl.2016.07.011>

Ajibola, O., Omisakin, O. A., Eze, A. A., & Omoleke, S. A. (2018). Self-Medication with Antibiotics, Attitude and Knowledge of Antibiotic Resistance among Community Residents and Undergraduate Students in Northwest Nigeria. *Diseases (Basel, Switzerland)*, 6(2). <https://doi.org/https://dx.doi.org/10.3390/diseases6020032>

Akpulu, C., Lankapalli, A. K., Toufiq, R., Cook, K., Portal, E. A., Khalid, R. Y., Mukkadas, A., Gambo, S., Aminu, A., Iregbu, K., Thomson, K., Walsh, T., & Sands, K. (2025). Comparative evaluation of DNA extraction protocols for neonatal gut microbiota profiling in a resource-limited setting. *The Microbe*, 7, 100398. <https://doi.org/10.1016/j.microb.2025.100398>

Akpulu, C., Yakubu, R., Tukur, F., Yusuf, A. H., Kassim, A., Mukaddas, A., Adeleye, A., Alkali, F., Maduekwe, A. O., Nsude, Q., Asunugwo, V., Okenu, M.-J., Yakubu, S., Audu, L., Amsa, M., Modibbo, F. I., Thomson, K., Achi, C. R., Iregbu, K. C., ... Sands, K. (2025). Beyond the samples: Nigerian research staff perspectives on a multicentre neonatal sepsis study's impact. *Nature Communications*, 16(1), 6745. <https://doi.org/10.1038/s41467-025-62102-4>

Al-lawama, M., AlZaatreh, A., Elrajabi, R., Abdelhamid, S., & Badran, E. (2019). Prolonged Rupture of Membranes, Neonatal Outcomes and Management Guidelines. *Journal of Clinical Medicine Research*, 11(5), 360–366. <https://doi.org/10.14740/jocmr3809>

Amarasekara, R., Jayasekara, R. W., Senanayake, H., & Dissanayake, V. H. W. (2015). Microbiome of the placenta in pre-eclampsia supports the role of bacteria in the multifactorial cause of pre-eclampsia. *Journal of Obstetrics and Gynaecology Research*, 41(5), 662–669. <https://doi.org/10.1111/jog.12619>

Amato, H. K., Loayza, F., Salinas, L., Paredes, D., Garcia, D., Sarzosa, S., Saraiva-Garcia, C., Johnson, T. J., Pickering, A. J., Riley, L. W., Trueba, G., & Graham, J. P. (2023). Risk factors for extended-spectrum beta-lactamase (ESBL)-producing *E. coli* carriage among children in a food animal-producing region of Ecuador: A repeated measures observational study. *PLOS Medicine*, 20(10), e1004299. <https://doi.org/10.1371/journal.pmed.1004299>

- Amenyogbe, N., Kollmann, T. R., & Ben-Othman, R. (2017). Early-Life Host–Microbiome Interphase: The Key Frontier for Immune Development. *Frontiers in Pediatrics*, 5. <https://doi.org/10.3389/fped.2017.00111>
- Aminu Kano Teaching Hospital (corporate author). (2025). *About AKTH*. Aminu Kano Teaching Hospital. <https://akth.gov.ng/>
- Anand, U., Reddy, B., Singh, V. K., Singh, A. K., Kesari, K. K., Tripathi, P., Kumar, P., Tripathi, V., & Simal-Gandara, J. (2021). Potential Environmental and Human Health Risks Caused by Antibiotic-Resistant Bacteria (ARB), Antibiotic Resistance Genes (ARGs) and Emerging Contaminants (ECs) from Municipal Solid Waste (MSW) Landfill. *Antibiotics*, 10(4), 374. <https://doi.org/10.3390/antibiotics10040374>
- Anderson, M., Ljungqvist, G., Van Kessel, R., Saint, V., & Mossialos, E. (2024). *The socioeconomic drivers and impacts of Antimicrobial Resistance (AMR) POLICY BRIEF 64 Implications for policy and research HEALTH SYSTEMS AND POLICY ANALYSIS*. www.euro.who.int
- Angell, B., Sanuade, O., Adetifa, I. M. O., Okeke, I. N., Adamu, A. L., Aliyu, M. H., Ameh, E. A., Kyari, F., Gadanya, M. A., Mabayoje, D. A., Yinka-Ogunleye, A., Oni, T., Jalo, R. I., Tsiga-Ahmed, F. I., DalGLISH, S. L., Abimbola, S., Colbourn, T., Onwujekwe, O., Owoaje, E. T., ... Abubakar, I. (2022). Population health outcomes in Nigeria compared with other west African countries, 1998–2019: a systematic analysis for the Global Burden of Disease Study. *The Lancet*, 399(10330), 1117–1129. [https://doi.org/10.1016/S0140-6736\(21\)02722-7](https://doi.org/10.1016/S0140-6736(21)02722-7)
- Arboleya, S., Sánchez, B., Milani, C., Duranti, S., Solís, G., Fernández, N., de los Reyes-Gavilán, C. G., Ventura, M., Margolles, A., & Gueimonde, M. (2015). Intestinal Microbiota Development in Preterm Neonates and Effect of Perinatal Antibiotics. *The Journal of Pediatrics*, 166(3), 538–544. <https://doi.org/10.1016/j.jpeds.2014.09.041>
- Arboleya, S., Saturio, S., & Gueimonde, M. (2022). Impact of intrapartum antibiotics on the developing microbiota: a review. *Microbiome Research Reports*, 1(3), 22. <https://doi.org/10.20517/mrr.2022.04>
- Arrieta, M.-C., Stiemsma, L. T., Amenyogbe, N., Brown, E. M., & Finlay, B. (2014). The Intestinal Microbiome in Early Life: Health and Disease. *Frontiers in Immunology*, 5. <https://doi.org/10.3389/fimmu.2014.00427>

- Arrieta, M.-C., Stiemsma, L. T., Dimitriu, P. A., Thorson, L., Russell, S., Yurist-Doutsch, S., Kuzeljevic, B., Gold, M. J., Britton, H. M., Lefebvre, D. L., Subbarao, P., Mandhane, P., Becker, A., McNagny, K. M., Sears, M. R., Kollmann, T., Mohn, W. W., Turvey, S. E., & Brett Finlay, B. (2015). Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Science Translational Medicine*, 7(307). <https://doi.org/10.1126/scitranslmed.aab2271>
- Aslam, B., Asghar, R., Muzammil, S., Shafique, M., Siddique, A. B., Khurshid, M., Ijaz, M., Rasool, M. H., Chaudhry, T. H., Aamir, A., & Baloch, Z. (2024). AMR and Sustainable Development Goals: at a crossroads. *Globalization and Health*, 20(1), 73. <https://doi.org/10.1186/s12992-024-01046-8>
- Audu, O., Onyemochi, A., Sonnyjohnbull, O., Umar, A. A., Ara, B. I., Raphael, A. E., Pius, E. O., & Polycarp, A. U. (2014). Preference for Health Provider's Gender amongst Women Attending Obstetrics/ Gynecology Clinic, ABUTH, Zaria, Northwestern Nigeria. *American Journal of Public Health Research*, 2(1), 21–26. <https://doi.org/10.12691/ajphr-2-1-5>
- Auta, A., Hadi, M. A., Oga, E., Adewuyi, E. O., Abdu-Aguye, S. N., Adeloye, D., Strickland-Hodge, B., & Morgan, D. J. (2019). Global access to antibiotics without prescription in community pharmacies: A systematic review and meta-analysis. *Journal of Infection*, 78(1), 8–18. <https://doi.org/10.1016/j.jinf.2018.07.001>
- Awoyemi, B. O., Makanju, A. A., Mpapalika, J., & Ekpeyo, R. S. (2023). A time series analysis of government expenditure and health outcomes in Nigeria. *Journal of Public Health in Africa*, 14(7), 6. <https://doi.org/10.4081/jphia.2023.1409>
- Ayeni, F. A., Biagi, E., Rampelli, S., Fiori, J., Soverini, M., Audu, H. J., Cristino, S., Caporali, L., Schnorr, S. L., Carelli, V., Brigidi, P., Candela, M., & Turrone, S. (2018). Infant and Adult Gut Microbiome and Metabolome in Rural Bassa and Urban Settlers from Nigeria. *Cell Reports*, 23(10), 3056–3067. <https://doi.org/10.1016/j.celrep.2018.05.018>
- Ayeni, K. I., Berry, D., Wisgrill, L., Warth, B., & Ezekiel, C. N. (2022). Early-life chemical exposome and gut microbiome development: African research perspectives within a global environmental health context. *Trends in Microbiology*, xx(xx). <https://doi.org/10.1016/j.tim.2022.05.008>

- Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., Khan, M. T., Zhang, J., Li, J., Xiao, L., Al-Aama, J., Zhang, D., Lee, Y. S., Kotowska, D., Colding, C., ... Wang, J. (2015). Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host & Microbe*, 17(5), 690–703. <https://doi.org/10.1016/j.chom.2015.04.004>
- Bajorek, S., Parker, L., Li, N., Winglee, K., Weaver, M., Johnson, J., Sioda, M., Gauthier, J., Lemas, D. J., Jobin, C., Lorca, G., Neu, J., & Fodor, A. A. (2019). Initial microbial community of the neonatal stomach immediately after birth. *Gut Microbes*, 10(3), 289–297. <https://doi.org/10.1080/19490976.2018.1520578>
- Baquero, F. (2004). From pieces to patterns: evolutionary engineering in bacterial pathogens. *Nature Reviews Microbiology*, 2(6), 510–518. <https://doi.org/10.1038/nrmicro909>
- Bargheet, A., Klingenberg, C., Esaiassen, E., Hjerde, E., Cavanagh, J. P., Bengtsson-Palme, J., & Pettersen, V. K. (2023). Development of early life gut resistome and mobilome across gestational ages and microbiota-modifying treatments. *EBioMedicine*, 92, 104613. <https://doi.org/10.1016/j.ebiom.2023.104613>
- Bargheet, A., Noordzij, H. T., Ponsero, A. J., Jian, C., Korpela, K., Valles-Colomer, M., Debelius, J., Kurilshikov, A., & Pettersen, V. K. (2025). Dynamics of gut resistome and mobilome in early life: a meta-analysis. *EBioMedicine*, 114, 105630. <https://doi.org/10.1016/j.ebiom.2025.105630>
- Baumann-Dudenhoeffer, A. M., D'Souza, A. W., Tarr, P. I., Warner, B. B., & Dantas, G. (2018). Infant diet and maternal gestational weight gain predict early metabolic maturation of gut microbiomes. *Nature Medicine*, 24(12), 1822–1829. <https://doi.org/10.1038/s41591-018-0216-2>
- Beaumont, A.-L., de Lauzanne, A., Criscuolo, A., Fabre, L., Rabenandrasana, M., Randriamanga, N., Bernabeu, S., Harimanana, A., Randremanana, R., Herindrainy, P., Collard, J.-M., Pring, L., Sr, N., Cheng, S., Borand, L., Kermorvant, E., Crucitti, T., Guillemot, D., Huynh, B.-T., & Group, B. S. (2025). *Early neonatal acquisition of extended-spectrum beta-lactamase (ESBL)-producing Enterobacterales in Madagascar and Cambodia: clonal and plasmid-level contributions of maternal gut carriage*. <https://doi.org/10.21203/rs.3.rs-6271227/v1>

- Beaumont, A.-L., Kermorvant-Duchemin, E., Breurec, S., & Huynh, B.-T. (2024). Neonatal Colonization With Antibiotic-Resistant Pathogens in Low- and Middle-Income Countries. *JAMA Network Open*, 7(11), e2441596. <https://doi.org/10.1001/jamanetworkopen.2024.41596>
- Belkaid, Y., & Hand, T. W. (2014). Role of the Microbiota in Immunity and Inflammation. *Cell*, 157(1), 121–141. <https://doi.org/10.1016/j.cell.2014.03.011>
- Benitez, A. J., Tanes, C., Friedman, E. S., Zackular, J. P., Ford, E., Gerber, J. S., DeRusso, P. A., Kelly, A., Li, H., Elovitz, M. A., Wu, G. D., Zemel, B., & Bittinger, K. (2025). Antibiotic exposure is associated with minimal gut microbiome perturbations in healthy term infants. *Microbiome*, 13(1), 21. <https://doi.org/10.1186/s40168-024-01999-3>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Berendes, D., Knee, J., Sumner, T., Capone, D., Lai, A., Wood, A., Patel, S., Nalá, R., Cumming, O., & Brown, J. (2019). Gut carriage of antimicrobial resistance genes among young children in urban Maputo, Mozambique: Associations with enteric pathogen carriage and environmental risk factors. *PLOS ONE*, 14(11), e0225464. <https://doi.org/10.1371/journal.pone.0225464>
- Berglund, F., Ebmeyer, S., Kristiansson, E., & Larsson, D. G. J. (2023). Evidence for wastewaters as environments where mobile antibiotic resistance genes emerge. *Communications Biology*, 6(1), 321. <https://doi.org/10.1038/s42003-023-04676-7>
- Bergström, A., Skov, T. H., Bahl, M. I., Roager, H. M., Christensen, L. B., Ejlerskov, K. T., Mølgaard, C., Michaelsen, K. F., & Licht, T. R. (2014). Establishment of Intestinal Microbiota during Early Life: a Longitudinal, Explorative Study of a Large Cohort of Danish Infants. *Applied and Environmental Microbiology*, 80(9), 2889–2900. <https://doi.org/10.1128/AEM.00342-14>
- Bethesda (MD): National Institute of Child Health and Human Development; 2006-. Metronidazole. (2024). *Drugs and Lactation Database (LactMed®)* [Internet]. <https://www.ncbi.nlm.nih.gov/books/NBK501315/>

- Betran, A. P., Ye, J., Moller, A.-B., Souza, J. P., & Zhang, J. (2021). Trends and projections of caesarean section rates: global and regional estimates. *BMJ Global Health*, 6(6), e005671. <https://doi.org/10.1136/bmjgh-2021-005671>
- Bevan, E. R., Jones, A. M., & Hawkey, P. M. (2017). Global epidemiology of CTX-M β -lactamases: temporal and geographical shifts in genotype. *Journal of Antimicrobial Chemotherapy*, 72(8), 2145–2155. <https://doi.org/10.1093/jac/dkx146>
- Bittinger, K., Zhao, C., Li, Y., Ford, E., Friedman, E. S., Ni, J., Kulkarni, C. V., Cai, J., Tian, Y., Liu, Q., Patterson, A. D., Sarkar, D., Chan, Siu. H. J., Maranas, C., Saha-Shah, A., Lund, P., Garcia, B. A., Mattei, L. M., Gerber, J. S., ... Wu, G. D. (2020). Bacterial colonization reprograms the neonatal gut metabolome. *Nature Microbiology*, 5(6), 838–847. <https://doi.org/10.1038/s41564-020-0694-0>
- Blair, J. M., Richmond, G. E., & Piddock, L. J. (2014). Multidrug Efflux Pumps in Gram-Negative Bacteria and Their Role in Antibiotic Resistance. *Future Microbiology*, 9(10), 1165–1177. <https://doi.org/10.2217/fmb.14.66>
- Bogaert, D., van Beveren, G. J., de Koff, E. M., Lusarreta Parga, P., Balcazar Lopez, C. E., Koppensteiner, L., Clerc, M., Hasrat, R., Arp, K., Chu, M. L. J. N., de Groot, P. C. M., Sanders, E. A. M., van Houten, M. A., & de Steenhuijsen Pijters, W. A. A. (2023). Mother-to-infant microbiota transmission and infant microbiota development across multiple body sites. *Cell Host & Microbe*, 31(3), 447-460.e6. <https://doi.org/10.1016/j.chom.2023.01.018>
- Boolchandani, M., D'Souza, A. W., & Dantas, G. (2019). Sequencing-based methods and resources to study antimicrobial resistance. *Nature Reviews Genetics*. <https://doi.org/10.1038/s41576-019-0108-4>
- Brigaglia, A. (2017). Fī Lawḥin Maḥfūz: Towards a Phenomenological Analysis of the Quranic Tablet. In *The Arts and Crafts of Literacy* (pp. 69–102). De Gruyter. <https://doi.org/10.1515/9783110541441-003>
- Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., Goulding, D., & Lawley, T. D. (2016). Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature*, 533(7604), 543–546. <https://doi.org/10.1038/nature17645>

- Bulabula, A. N. H., Dramowski, A., & Mehtar, S. (2017). Maternal colonization or infection with extended-spectrum beta-lactamase-producing Enterobacteriaceae in Africa: A systematic review and meta-analysis. *International Journal of Infectious Diseases*, 64, 58–66. <https://doi.org/10.1016/j.ijid.2017.08.015>
- Burgin, K. (1996). Defining terminology for improved breastfeeding research. In *Journal of nurse-midwifery* (Vol. 41, Number 3, pp. 193–194). [https://doi.org/10.1016/0091-2182\(96\)00017-1](https://doi.org/10.1016/0091-2182(96)00017-1)
- Busi, S. B., de Nies, L., Habier, J., Wampach, L., Fritz, J. V., Heintz-Buschart, A., May, P., Halder, R., de Beaufort, C., & Wilmes, P. (2021). Persistence of birth mode-dependent effects on gut microbiome composition, immune system stimulation and antimicrobial resistance during the first year of life. *ISME Communications*, 1(1). <https://doi.org/10.1038/s43705-021-00003-5>
- Cantey, J. B., Sreeramoju, P., Jaleel, M., Treviño, S., Gander, R., Hynan, L. S., Hill, J., Brown, C., Chung, W., Siegel, J. D., & Sánchez, P. J. (2013). Prompt Control of an Outbreak Caused by Extended-Spectrum β -Lactamase–Producing *Klebsiella pneumoniae* in a Neonatal Intensive Care Unit. *The Journal of Pediatrics*, 163(3), 672–679.e3. <https://doi.org/10.1016/j.jpeds.2013.03.001>
- Carvalho, M. J., Sands, K., Thomson, K., Portal, E., Mathias, J., Milton, R., Gillespie, D., Dyer, C., Akpulu, C., Boostrom, I., Hogan, P., Saif, H., Ferreira, A., Nieto, M., Hender, T., Hood, K., Andrews, R., Watkins, W. J., Hassan, B., ... Walsh, T. R. (2022). Antibiotic resistance genes in the gut microbiota of mothers and linked neonates with or without sepsis from low- and middle-income countries. *Nature Microbiology*, 7(9), 1337–1347. <https://doi.org/10.1038/s41564-022-01184-y>
- Centers for Disease Control and Prevention (CDC). (2023). *Estimated Healthcare Costs of Antimicrobial Resistance*. <https://www.cdc.gov/antimicrobial-resistance/stories/partner-estimates.html>
- Chatzigiannidou, I., Johansen, P. L., Dehli, R. K., Moll, J. M., Eriksen, C., Myers, P. N., Roager, H. M., Yang, L., Stokholm, J., Sørensen, S. J., Krogfelt, K. A., Laursen, M. F., Trivedi, U., Scheynius, A., Kristiansen, K., Mie, A., Alm, J., & Brix, S. (2025). Temporal dynamics and microbial interactions shaping the gut resistome in early infancy. *Nature Communications*, 16(1), 8139. <https://doi.org/10.1038/s41467-025-63401-6>

- Chen, P., Sun, Z., Wang, J., Liu, X., Bai, Y., Chen, J., Liu, A., Qiao, F., Chen, Y., Yuan, C., Sha, J., Zhang, J., Xu, L.-Q., & Li, J. (2023). Portable nanopore-sequencing technology: Trends in development and applications. *Frontiers in Microbiology*, 14. <https://doi.org/10.3389/fmicb.2023.1043967>
- Chiripanhura, B. M., & Niño-Zarazúa, M. (2016). The impacts of the food, fuel and financial crises on poor and vulnerable households in Nigeria: A retrospective approach to research inquiry. *Development Policy Review*, 34(6), 763–788. <https://doi.org/10.1111/dpr.12183>
- Cho, I., Yamanishi, S., Cox, L., Methé, B. A., Zavadil, J., Li, K., Gao, Z., Mahana, D., Raju, K., Teitler, I., Li, H., Alekseyenko, A. V., & Blaser, M. J. (2012). Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature*, 488(7413), 621–626. <https://doi.org/10.1038/nature11400>
- Choo, J. M., Leong, L. E., & Rogers, G. B. (2015). Sample storage conditions significantly influence faecal microbiome profiles. *Scientific Reports*, 5(1), 16350. <https://doi.org/10.1038/srep16350>
- Christensen, E. D., Hjelmsø, M. H., Thorsen, J., Shah, S., Redgwell, T., Poulsen, C. E., Trivedi, U., Russel, J., Gupta, S., Chawes, B. L., Bønnelykke, K., Sørensen, S. J., Rasmussen, M. A., Bisgaard, H., & Stokholm, J. (2022). The developing airway and gut microbiota in early life is influenced by age of older siblings. *Microbiome*, 10(1), 106. <https://doi.org/10.1186/s40168-022-01305-z>
- Christian, B. N., Christian, N. G., Keshinro, M. I., & Olutade-Babatunde, O. (2023). How to build bridges for Universal Health Coverage in Nigeria by linking formal and informal health providers. *BMJ Global Health*, 8(11), e014165. <https://doi.org/10.1136/bmjgh-2023-014165>
- Coenen, A. R., Hu, S. K., Luo, E., Muratore, D., & Weitz, J. S. (2020). A Primer for Microbiome Time-Series Analysis. *Frontiers in Genetics*, 11. <https://doi.org/10.3389/fgene.2020.00310>
- Collado, M. C., Rautava, S., Aakko, J., Isolauri, E., & Salminen, S. (2016). Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. *Scientific Reports*, 6(1), 23129. <https://doi.org/10.1038/srep23129>

- Collignon, P., Beggs, J. J., Walsh, T. R., Gandra, S., & Laxminarayan, R. (2018). Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. *The Lancet Planetary Health*, 2(9), e398–e405. [https://doi.org/10.1016/S2542-5196\(18\)30186-4](https://doi.org/10.1016/S2542-5196(18)30186-4)
- Committee Opinion No. 644. (2015). The Apgar Score. *Obstetrics & Gynecology*, 126(4), e52–e55. <https://doi.org/10.1097/AOG.0000000000001108>
- Cook, K., Premchand-Branker, S., Nieto-Rosado, M., Portal, E. A. R., Li, M., Rubio, C. O., Mathias, J., Aziz, J., Iregbu, K., Afegbua, S. L., Aliyu, A., Mohammed, Y., Nwafia, I., Oduyebo, O., Ibrahim, A., Tanko, Z., Walsh, T. R., Achi, C., & Sands, K. (2025). Flies as carriers of antimicrobial resistant (AMR) bacteria in Nigerian hospitals: A workflow for surveillance of AMR bacteria carried by arthropod pests in hospital settings. *Environment International*, 196, 109294. <https://doi.org/10.1016/j.envint.2025.109294>
- Cornick, J., Musicha, P., Peno, C., Seager, E., Iroh Tam, P.-Y., Bilima, S., Bennett, A., Kennedy, N., Feasey, N., Heinz, E., & Cain, A. K. (2021). Genomic investigation of a suspected *Klebsiella pneumoniae* outbreak in a neonatal care unit in sub-Saharan Africa. *Microbial Genomics*, 7(11). <https://doi.org/10.1099/mgen.0.000703>
- Costea, P. I., Zeller, G., Sunagawa, S., Pelletier, E., Alberti, A., Levenez, F., Tramontano, M., Driessen, M., Hercog, R., Jung, F.-E., Kultima, J. R., Hayward, M. R., Coelho, L. P., Allen-Vercoe, E., Bertrand, L., Blaut, M., Brown, J. R. M., Carton, T., Cools-Portier, S., ... Bork, P. (2017). Towards standards for human fecal sample processing in metagenomic studies. *Nature Biotechnology*, 35(11), 1069–1076. <https://doi.org/10.1038/nbt.3960>
- Cox, L. M., Yamanishi, S., Sohn, J., Alekseyenko, A. V., Leung, J. M., Cho, I., Rogers, A. B., Kim, S. G., Li, H., Gao, Z., Mahana, D., Za, J. G., Robine, N., Loke, P., & Blaser, M. J. (2014). *Altering the Intestinal Microbiota during a Critical Developmental Window Has Lasting Metabolic Consequences*. 705–721. <https://doi.org/10.1016/j.cell.2014.05.052>
- Cukrowska, B., Bierła, J. B., Zakrzewska, M., Klukowski, M., & Maciorkowska, E. (2020). The Relationship between the Infant Gut Microbiota and Allergy. The Role of *Bifidobacterium breve* and Prebiotic Oligosaccharides in the Activation of Anti-Allergic Mechanisms in Early Life. *Nutrients*, 12(4), 946. <https://doi.org/10.3390/nu12040946>

- Das, L., Virmani, R., Sharma, V., Rawat, D., & Singh, Y. (2019). Human Milk Microbiota: Transferring the Antibiotic Resistome to Infants. *Indian Journal of Microbiology*, 59(4), 410–416. <https://doi.org/10.1007/s12088-019-00824-y>
- Davies, J. E. (2007). *Origins, Acquisition and Dissemination of Antibiotic Resistance Determinants* (pp. 15–35). <https://doi.org/10.1002/9780470515358.ch2>
- D’Costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W. L., Schwarz, C., Froese, D., Zazula, G., Calmels, F., Debruyne, R., Golding, G. B., Poinar, H. N., & Wright, G. D. (2011). Antibiotic resistance is ancient. *Nature*, 477(7365), 457–461. <https://doi.org/10.1038/nature10388>
- D’Costa, V. M., McGrann, K. M., Hughes, D. W., & Wright, G. D. (2006). Sampling the Antibiotic Resistome. *Science*, 311(5759), 374–377. <https://doi.org/10.1126/science.1120800>
- Dhareel, D., Dhungana, R., Basnet, S., Gautam, S., Dhungana, A., Dudani, R., & Bhattarai, A. (2020). Breastfeeding practices within the first six months of age in mid-western and eastern regions of Nepal: a health facility-based cross-sectional study. *BMC Pregnancy and Childbirth*, 20(1), 59. <https://doi.org/10.1186/s12884-020-2754-0>
- Dierikx, T. H., Visser, D. H., Benninga, M. A., van Kaam, A. H. L. C., de Boer, N. K. H., de Vries, R., van Limbergen, J., & de Meij, T. G. J. (2020). The influence of prenatal and intrapartum antibiotics on intestinal microbiota colonisation in infants: A systematic review. *Journal of Infection*, 81(2), 190–204. <https://doi.org/10.1016/j.jinf.2020.05.002>
- DiGiulio, D. B., Romero, R., Kusanovic, J. P., Gómez, R., Kim, C. J., Seok, K. S., Gotsch, F., Mazaki-Tovi, S., Vaisbuch, E., Sanders, K., Bik, E. M., Chaiworapongsa, T., Oyarzún, E., & Relman, D. A. (2010). Prevalence and Diversity of Microbes in the Amniotic Fluid, the Fetal Inflammatory Response, and Pregnancy Outcome in Women with Preterm Pre-Labor Rupture of Membranes. *American Journal of Reproductive Immunology*. <https://doi.org/10.1111/j.1600-0897.2010.00830.x>
- Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences*, 107(26), 11971–11975. <https://doi.org/10.1073/pnas.1002601107>

- Dos Santos, S. J., Pakzad, Z., Elwood, C. N., Albert, A. Y. K., Gantt, S., Manges, A. R., Dumonceaux, T. J., Maan, E. J., Hill, J. E., & Money, D. M. (2021). Early Neonatal Meconium Does Not Have a Demonstrable Microbiota Determined through Use of Robust Negative Controls with *cpn* 60-Based Microbiome Profiling. *Microbiology Spectrum*, 9(2). <https://doi.org/10.1128/Spectrum.00067-21>
- Dramowski, A., Bolton, L., Fitzgerald, F., & Bekker, A. (2025). Neonatal Sepsis in Low- and Middle-income Countries: Where Are We Now? *Pediatric Infectious Disease Journal*, 44(6), e207–e210. <https://doi.org/10.1097/INF.00000000000004815>
- Edefo, J. W. (2025). Water and Sanitation Access in Nigeria. *Water Conservation Science and Engineering*, 10(2), 72. <https://doi.org/10.1007/s41101-025-00402-y>
- Edney, J. M., Kovats, S., Filippi, V., & Nakstad, B. (2022). A systematic review of hot weather impacts on infant feeding practices in low-and middle-income countries. *Frontiers in Pediatrics*, 10. <https://doi.org/10.3389/fped.2022.930348>
- Edwards, T., Williams, C. T., Olwala, M., Andang'o, P., Otieno, W., Nalwa, G. N., Akindolire, A., Cubas-Atienzar, A. I., Ross, T., Tongo, O. O., Adams, E. R., Nabwera, H., & Allen, S. (2023). Molecular surveillance reveals widespread colonisation by carbapenemase and extended spectrum beta-lactamase producing organisms in neonatal units in Kenya and Nigeria. *Antimicrobial Resistance & Infection Control*, 12(1), 14. <https://doi.org/10.1186/s13756-023-01216-0>
- Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., & Weyrich, L. S. (2019). Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends in Microbiology*, 27(2), 105–117. <https://doi.org/10.1016/j.tim.2018.11.003>
- Eliana Drakopoulos. (2020). *New survey reveals progress and gaps in Nigerians' access to water, sanitation and hygiene services.*
- Elliott, I., Batty, E. M., Ming, D., Robinson, M. T., Nawtaisong, P., de Cesare, M., Newton, P. N., & Bowden, R. (2020). Oxford Nanopore MinION Sequencing Enables Rapid Whole Genome Assembly of *Rickettsia typhi* in a Resource-Limited Setting. *The American Journal of Tropical Medicine and Hygiene*, 102(2), 408–414. <https://doi.org/10.4269/ajtmh.19-0383>

- El-Maradny, Y. A., Nortey, M. A., Hakayuwa, C. M., Anyamene, E. L., Mary, J. J. F., Engmann, S. T., Tsikata, C. Y., Ahmed, D. A., Onyeaghala, C., Heniedy, A. M., Gesaka, S. R., Saminu, A., & Edward, M. (2025). The impact of socioeconomic disparities climate factors and antimicrobial stewardship on antimicrobial resistance in Africa. *Discover Public Health*, 22(1), 247. <https://doi.org/10.1186/s12982-025-00631-5>
- Enav, H., Bäckhed, F., & Ley, R. E. (2022). The developing infant gut microbiome: A strain-level view. *Cell Host & Microbe*, 30(5), 627–638. <https://doi.org/10.1016/j.chom.2022.04.009>
- Fan, J., Huang, S., & Chorlton, S. D. (2021). BugSeq: a highly accurate cloud platform for long-read metagenomic analyses. *BMC Bioinformatics*, 22(1), 160. <https://doi.org/10.1186/s12859-021-04089-5>
- Federal Ministry of Water Resources (FMWR), G. of N. N. B. of S. (NBS) and U. (2022). *A Report of Findings Water Sanitation and Hygiene National Outcome Routine Mapping 2021 Federal Ministry of Water Resources*.
- Fernández-Pato, A., Sinha, T., Gacesa, R., Andreu-Sánchez, S., Gois, M. F. B., Gelderloos-Arends, J., Jansen, D. B. H., Kruk, M., Jaeger, M., Joosten, L. A. B., Netea, M. G., Weersma, R. K., Wijmenga, C., Harmsen, H. J. M., Fu, J., Zhernakova, A., & Kurilshikov, A. (2024). Choice of DNA extraction method affects stool microbiome recovery and subsequent phenotypic association analyses. *Scientific Reports*, 14(1), 3911. <https://doi.org/10.1038/s41598-024-54353-w>
- Ferretti, P., Allert, M., Johnson, K. E., Rossi, M., Heisel, T., Gonía, S., Knights, D., Fields, D. A., Albert, F. W., Demerath, E. W., Gale, C. A., & Blekhman, R. (2025). Assembly of the infant gut microbiome and resistome are linked to bacterial strains in mother's milk. *Nature Communications*. <https://doi.org/10.1038/s41467-025-66497-y>
- Ferretti, P., Pasolli, E., Tett, A., Asnicar, F., Gorfer, V., Fedi, S., Armanini, F., Truong, D. T., Manara, S., Zolfo, M., Beghini, F., Bertorelli, R., De Sanctis, V., Bariletti, I., Canto, R., Clementi, R., Cologna, M., Crifò, T., Cusumano, G., ... Segata, N. (2018). Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host & Microbe*, 24(1), 133-145.e5. <https://doi.org/10.1016/j.chom.2018.06.005>

- Fleischmann, C., Reichert, F., Cassini, A., Horner, R., Harder, T., Markwart, R., Tröndle, M., Savova, Y., Kisson, N., Schlattmann, P., Reinhart, K., Allegranzi, B., & Eckmanns, T. (2021). Global incidence and mortality of neonatal sepsis: a systematic review and meta-analysis. *Archives of Disease in Childhood*, 106(8), 745–752. <https://doi.org/10.1136/archdischild-2020-320217>
- Folgori, L., Bielicki, J., Heath, P. T., & Sharland, M. (2017). Antimicrobial-resistant Gram-negative infections in neonates: burden of disease and challenges in treatment. *Current Opinion in Infectious Diseases*, 30(3), 281–288. <https://doi.org/10.1097/QCO.0000000000000371>
- Fuchs, A., Bielicki, J., Mathur, S., Sharland, M., & Van Den Anker, J. N. (2018). Reviewing the WHO guidelines for antibiotic use for sepsis in neonates and children. *Paediatrics and International Child Health*, 38(sup1), S3–S15. <https://doi.org/10.1080/20469047.2017.1408738>
- Fuhrmeister, E. R., Harvey, A. P., Nadimpalli, M. L., Gallandat, K., Ambelu, A., Arnold, B. F., Brown, J., Cumming, O., Earl, A. M., Kang, G., Kariuki, S., Levy, K., Pinto Jimenez, C. E., Swarthout, J. M., Trueba, G., Tsukayama, P., Worby, C. J., & Pickering, A. J. (2023). Evaluating the relationship between community water and sanitation access and the global burden of antibiotic resistance: an ecological study. *The Lancet Microbe*, 4(8), e591–e600. [https://doi.org/10.1016/S2666-5247\(23\)00137-4](https://doi.org/10.1016/S2666-5247(23)00137-4)
- Gallo, D. M., Romero, R., Bosco, M., Gotsch, F., Jaiman, S., Jung, E., Suksai, M., Ramón y Cajal, C. L., Yoon, B. H., & Chaiworapongsa, T. (2023). Meconium-stained amniotic fluid. *American Journal of Obstetrics and Gynecology*, 228(5), S1158–S1178. <https://doi.org/10.1016/j.ajog.2022.11.1283>
- Gand, M., Bloemen, B., Vanneste, K., Roosens, N. H. C., & De Keersmaecker, S. C. J. (2023). Comparison of 6 DNA extraction methods for isolation of high yield of high molecular weight DNA suitable for shotgun metagenomics Nanopore sequencing to detect bacteria. *BMC Genomics*, 24(1), 438. <https://doi.org/10.1186/s12864-023-09537-5>
- Gasparri, A. J., Wang, B., Sun, X., Kennedy, E. A., Hernandez-Leyva, A., Ndao, I. M., Tarr, P. I., Warner, B. B., & Dantas, G. (2019). Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome. *Nature Microbiology*, 4(12), 2285–2297. <https://doi.org/10.1038/s41564-019-0550-2>

- Geneva: World Health Organization. (2018). *Guideline: Counselling of Women to Improve Breastfeeding Practices*. <https://www.ncbi.nlm.nih.gov/books/NBK539314/>
- Gensollen, T., Iyer, S. S., Kasper, D. L., & Blumberg, R. S. (2016). How colonization by microbiota in early life shapes the immune system. *Science*, 352(6285), 539–544. <https://doi.org/10.1126/science.aad9378>
- Gibson, M. K., Wang, B., Ahmadi, S., Burnham, C.-A. D., Tarr, P. I., Warner, B. B., & Dantas, G. (2016). Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nature Microbiology*, 1(4), 16024. <https://doi.org/10.1038/nmicrobiol.2016.24>
- Gladys, O. U., Gabriella, C. N., Lawal, F. O., Bukola, F. O., Amaechi, V. C., & Augustine, O. N. (2025). Antibiotics stewardship: prevalence, nature, and factors associated with dispensing of antibiotics without prescription among community pharmacists in Nigeria. *Journal of Pharmaceutical Policy and Practice*, 18(1). <https://doi.org/10.1080/20523211.2025.2498927>
- Gosalbes, M. J., Llop, S., Vallès, Y., Moya, A., Ballester, F., & Francino, M. P. (2013). Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clinical & Experimental Allergy*, 43(2), 198–211. <https://doi.org/10.1111/cea.12063>
- Gosalbes, M. J., Vallès, Y., Jiménez-Hernández, N., Balle, C., Riva, P., Miravet-Verde, S., de Vries, L. E., Llop, S., Agersø, Y., Sørensen, S. J., Ballester, F., & Francino, M. P. (2016). High frequencies of antibiotic resistance genes in infants' meconium and early fecal samples. *Journal of Developmental Origins of Health and Disease*, 7(1), 35–44. <https://doi.org/10.1017/S2040174415001506>
- Gough, E. K., Moodie, E. E. M., Prendergast, A. J., Johnson, S. M. A., Humphrey, J. H., Stoltzfus, R. J., Walker, A. S., Trehan, I., Gibb, D. M., Goto, R., Tahan, S., de Moraes, M. B., & Manges, A. R. (2014). The impact of antibiotics on growth in children in low and middle income countries: systematic review and meta-analysis of randomised controlled trials. *BMJ*, 348(apr15 6), g2267–g2267. <https://doi.org/10.1136/bmj.g2267>
- Grant-Beurmann, S., Jumare, J., Ndembi, N., Matthew, O., Shutt, A., Omoigberale, A., Martin, O. A., Fraser, C. M., & Charurat, M. (2022). Dynamics of the infant gut microbiota in the first 18 months of life: the impact of maternal HIV infection and breastfeeding. *Microbiome*, 10(1), 61. <https://doi.org/10.1186/s40168-022-01230-1>

- Grech, A., Collins, C. E., Holmes, A., Lal, R., Duncanson, K., Taylor, R., & Gordon, A. (2021). Maternal exposures and the infant gut microbiome: a systematic review with meta-analysis. *Gut Microbes*, *13*(1). <https://doi.org/10.1080/19490976.2021.1897210>
- Guitor, A. K., Yousuf, E. I., Raphenya, A. R., Hutton, E. K., Morrison, K. M., McArthur, A. G., Wright, G. D., & Stearns, J. C. (2022). Capturing the antibiotic resistome of preterm infants reveals new benefits of probiotic supplementation. *Microbiome*, *10*(1), 136. <https://doi.org/10.1186/s40168-022-01327-7>
- Guo, C., Zhou, Q., Li, M., Zhou, L., Xu, L., Zhang, Y., Li, D., Wang, Y., Dai, W., Li, S., & Zhang, L. (2020). Breastfeeding restored the gut microbiota in caesarean section infants and lowered the infection risk in early life. *BMC Pediatrics*, *20*(1), 532. <https://doi.org/10.1186/s12887-020-02433-x>
- Harrison, M. L., Dickson, B. F. R., Sharland, M., & Williams, P. C. M. (2024). Beyond Early- and Late-onset Neonatal Sepsis Definitions: What are the Current Causes of Neonatal Sepsis Globally? A Systematic Review and Meta-analysis of the Evidence. *Pediatric Infectious Disease Journal*, *43*(12), 1182–1190. <https://doi.org/10.1097/INF.0000000000004485>
- Hasrat, R., Kool, J., de Steenhuijsen Pijters, W. A. A., Chu, M. L. J. N., Kuiling, S., Groot, J. A., van Logchem, E. M., Fuentes, S., Franz, E., Bogaert, D., & Bosch, T. (2021). Benchmarking laboratory processes to characterise low-biomass respiratory microbiota. *Scientific Reports*, *11*(1), 17148. <https://doi.org/10.1038/s41598-021-96556-5>
- Heintz-Buschart, A., & Wilmes, P. (2018). Human Gut Microbiome: Function Matters. *Trends in Microbiology*, *26*(7), 563–574. <https://doi.org/10.1016/j.tim.2017.11.002>
- Hendriksen, R. S., Munk, P., Njage, P., van Bunnik, B., McNally, L., Lukjancenko, O., Röder, T., Nieuwenhuijse, D., Pedersen, S. K., Kjeldgaard, J., Kaas, R. S., Clausen, P. T. L. C., Vogt, J. K., Leekitcharoenphon, P., van de Schans, M. G. M., Zuidema, T., de Roda Husman, A. M., Rasmussen, S., Petersen, B., ... Aarestrup, F. M. (2019). Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nature Communications*, *10*(1), 1124. <https://doi.org/10.1038/s41467-019-08853-3>
- Henrick, B. M., Chew, S., Casaburi, G., Brown, H. K., Frese, S. A., Zhou, Y., Underwood, M. A., & Smilowitz, J. T. (2019). Colonization by *B. infantis* EVC001 modulates enteric

- inflammation in exclusively breastfed infants. *Pediatric Research*, 86(6), 749–757. <https://doi.org/10.1038/s41390-019-0533-2>
- Hidalgo-Cantabrana, C., Delgado, S., Ruiz, L., Ruas-Madiedo, P., Sánchez, B., & Margolles, A. (2017). Bifidobacteria and Their Health-Promoting Effects. *Microbiology Spectrum*, 5(3). <https://doi.org/10.1128/microbiolspec.BAD-0010-2016>
- Hill, C. J., Lynch, D. B., Murphy, K., Ulaszewska, M., Jeffery, I. B., O’Shea, C. A., Watkins, C., Dempsey, E., Mattivi, F., Tuohy, K., Ross, R. P., Ryan, C. A., O’ Toole, P. W., & Stanton, C. (2017). Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort. *Microbiome*, 5(1), 4. <https://doi.org/10.1186/s40168-016-0213-y>
- Holzhausen, E. A., Nikodemova, M., Deblois, C. L., Barnet, J. H., Peppard, P. E., Suen, G., & Malecki, K. M. (2021). Assessing the impact of storage time on the stability of stool microbiota richness, diversity, and composition. *Gut Pathogens*, 13(1), 75. <https://doi.org/10.1186/s13099-021-00470-0>
- Hsieh, Y.-H., Peterson, C. M., Raggio, A., Keenan, M. J., Martin, R. J., Ravussin, E., & Marco, M. L. (2016). Impact of Different Fecal Processing Methods on Assessments of Bacterial Diversity in the Human Intestine. *Frontiers in Microbiology*, 7. <https://doi.org/10.3389/fmicb.2016.01643>
- Huddleston, J. R. (2014). Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. *Infection and Drug Resistance*, 167. <https://doi.org/10.2147/IDR.S48820>
- Investigators of the Delhi Neonatal Infection Study (DeNIS) collaboration. (2016). Characterisation and antimicrobial resistance of sepsis pathogens in neonates born in tertiary care centres in Delhi, India: a cohort study. *The Lancet Global Health*, 4(10), e752–e760. [https://doi.org/10.1016/S2214-109X\(16\)30148-6](https://doi.org/10.1016/S2214-109X(16)30148-6)
- Iqbal, F., Siva, N., Shenoy, P. A., Lewis, L. E. S., Purkayastha, J., & Eshwara, V. K. (2024). Gut Pathogen Colonization: A Risk Factor to Bloodstream Infections in Preterm Neonates Admitted in the Neonatal Intensive Care Unit – A Prospective Cohort Study. *Neonatology*, 1–10. <https://doi.org/10.1159/000542335>
- Iregbu, K. C., Nwajiobi-Princewill, P. I., Medugu, N., Umeokonkwo, C. D., Uwaezuoke, N. S., Peter, Y. J., Nwafia, I. N., Elikwu, C., Shettima, S. A., Suleiman, M. R., Awopeju, A. T.

- O., Udoh, U., Adedosu, N., Mohammed, A., Oshun, P., Ekuma, A., Manga, M. M., Osaigbovo, I. I., Ejembi, C. J., ... Oduyebo, O. O. (2021). Antimicrobial Stewardship Implementation in Nigerian Hospitals: Gaps and Challenges. *African Journal of Clinical and Experimental Microbiology*, 22(1), 60–66. <https://doi.org/10.4314/ajcem.v22i1.8>
- Iregbu, K., Dramowski, A., Milton, R., Nsutebu, E., Howie, S. R. C., Chakraborty, M., Lavoie, P. M., Costelloe, C. E., & Ghazal, P. (2022). Global health systems' data science approach for precision diagnosis of sepsis in early life. *The Lancet Infectious Diseases*, 22(5), e143–e152. [https://doi.org/10.1016/S1473-3099\(21\)00645-9](https://doi.org/10.1016/S1473-3099(21)00645-9)
- J Olumade, T., A Adesanya, O., J Fred-Akintunwa, I., O Babalola, D., U Oguzie, J., A Ogunsanya, O., E George, U., D Akin-Ajani, O., & G Osasona, D. (2020). Infectious disease outbreak preparedness and response in Nigeria: history, limitations and recommendations for global health policy and practice. *AIMS Public Health*, 7(4), 736–757. <https://doi.org/10.3934/publichealth.2020057>
- Jakobsson, H. E., Abrahamsson, T. R., Jenmalm, M. C., Harris, K., Quince, C., Jernberg, C., Björkstén, B., Engstrand, L., & Andersson, A. F. (2014). Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by Caesarean section. *Gut*, 63(4), 559–566. <https://doi.org/10.1136/gutjnl-2012-303249>
- Jiménez, E., Fernández, L., Marín, M. L., Martín, R., Odriozola, J. M., Nueno-Palop, C., Narbad, A., Olivares, M., Xaus, J., & Rodríguez, J. M. (2005). Isolation of Commensal Bacteria from Umbilical Cord Blood of Healthy Neonates Born by Cesarean Section. *Current Microbiology*, 51(4), 270–274. <https://doi.org/10.1007/s00284-005-0020-3>
- Jokela, R., Pärnänen, K. M., Ponsero, A. J., Lahti, L., Kolho, K.-L., de Vos, W. M., & Salonen, A. (2024). A cohort study in family triads: impact of gut microbiota composition and early life exposures on intestinal resistome during the first two years of life. *Gut Microbes*, 16(1). <https://doi.org/10.1080/19490976.2024.2383746>
- Jonas, O. B. ;, Irwin, A., Berthe, F. C. J., Le Gall, F. G. ;, & Marquez, P. V. ; (2017). *DRUG-RESISTANT INFECTIONS A Threat to Our Economic Future*. <http://documents.worldbank.org/curated/en/323311493396993758>

- Jost, T., Lacroix, C., Braegger, C. P., Rochat, F., & Chassard, C. (2014). Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. *Environmental Microbiology*, 16(9), 2891–2904. <https://doi.org/10.1111/1462-2920.12238>
- Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282), 457–481. <https://doi.org/10.1080/01621459.1958.10501452>
- Karkman, A., Pärnänen, K., & Larsson, D. G. J. (2019). Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. *Nature Communications*, 10(1), 80. <https://doi.org/10.1038/s41467-018-07992-3>
- Kennedy, K. M., de Goffau, M. C., Perez-Muñoz, M. E., Arrieta, M.-C., Bäckhed, F., Bork, P., Braun, T., Bushman, F. D., Dore, J., de Vos, W. M., Earl, A. M., Eisen, J. A., Elovitz, M. A., Ganai-Vonarburg, S. C., Gänzle, M. G., Garrett, W. S., Hall, L. J., Hornef, M. W., Huttenhower, C., ... Walter, J. (2023). Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. *Nature*, 613(7945), 639–649. <https://doi.org/10.1038/s41586-022-05546-8>
- Kim, J. H., Jeon, J.-Y., Im, Y.-J., Ha, N., Kim, J.-K., Moon, S. J., & Kim, M.-G. (2023). Long-term taxonomic and functional stability of the gut microbiome from human fecal samples. *Scientific Reports*, 13(1), 114. <https://doi.org/10.1038/s41598-022-27033-w>
- Kleine Bardenhorst, S., Berger, T., Klawonn, F., Vital, M., Karch, A., & Rübsamen, N. (2021). Data Analysis Strategies for Microbiome Studies in Human Populations—a Systematic Review of Current Practice. *MSystems*, 6(1). <https://doi.org/10.1128/mSystems.01154-20>
- Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenent, L. T., & Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences*, 108(supplement_1), 4578–4585. <https://doi.org/10.1073/pnas.1000081107>
- Korpela, K., Salonen, A., Saxen, H., Nikkonen, A., Peltola, V., Jaakkola, T., de Vos, W., & Kolho, K.-L. (2020). Antibiotics in early life associate with specific gut microbiota signatures in a prospective longitudinal infant cohort. *Pediatric Research*, 88(3), 438–443. <https://doi.org/10.1038/s41390-020-0761-5>

- Korpela, K., Salonen, A., Virta, L. J., Kekkonen, R. A., Forslund, K., Bork, P., & de Vos, W. M. (2016). Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. *Nature Communications*, 7(1), 10410. <https://doi.org/10.1038/ncomms10410>
- Korpela, K., Zijlmans, M. A. C., Kuitunen, M., Kukkonen, K., Savilahti, E., Salonen, A., de Weerth, C., & de Vos, W. M. (2017). Childhood BMI in relation to microbiota in infancy and lifetime antibiotic use. *Microbiome*, 5(1), 26. <https://doi.org/10.1186/s40168-017-0245-y>
- Kothari, C., Gaiind, R., Singh, L. C., Sinha, A., Kumari, V., Arya, S., Chellani, H., Saxena, S., & Deb, M. (2013). Community acquisition of β -lactamase producing Enterobacteriaceae in neonatal gut. *BMC Microbiology*, 13(1), 136. <https://doi.org/10.1186/1471-2180-13-136>
- Kuehl, R., Morata, L., Boeing, C., Subirana, I., Seifert, H., Rieg, S., Kern, W. V, Kim, H. Bin, Kim, E. S., Liao, C.-H., Tilley, R., Lopez-Cortés, L. E., Llewelyn, M. J., Fowler, V. G., Thwaites, G., Cisneros, J. M., Scarborough, M., Nsutebu, E., Gurgui Ferrer, M., ... Price, J. (2020). Defining persistent Staphylococcus aureus bacteraemia: secondary analysis of a prospective cohort study. *The Lancet Infectious Diseases*, 20(12), 1409–1417. [https://doi.org/10.1016/S1473-3099\(20\)30447-3](https://doi.org/10.1016/S1473-3099(20)30447-3)
- Kumar, A., & Schweizer, H. (2005). Bacterial resistance to antibiotics: Active efflux and reduced uptake. *Advanced Drug Delivery Reviews*, 57(10), 1486–1513. <https://doi.org/10.1016/j.addr.2005.04.004>
- Kurowski, K. M., Marusinec, R., Amato, H. K., Saraiva-Garcia, C., Loayza, F., Salinas, L., Trueba, G., & Graham, J. P. (2021). Social and Environmental Determinants of Community-Acquired Antimicrobial-Resistant Escherichia coli in Children Living in Semirural Communities of Quito, Ecuador. *The American Journal of Tropical Medicine and Hygiene*, 105(3), 600–610. <https://doi.org/10.4269/ajtmh.20-0532>
- Labi, A.-K., Bjerrum, S., Enweronu-Laryea, C. C., Ayibor, P. K., Nielsen, K. L., Marvig, R. L., Newman, M. J., Andersen, L. P., & Kurtzhals, J. A. L. (2020). High Carriage Rates of Multidrug-Resistant Gram-Negative Bacteria in Neonatal Intensive Care Units From Ghana. *Open Forum Infectious Diseases*, 7(4). <https://doi.org/10.1093/ofid/ofaa109>

- Laforest-Lapointe, I., & Arrieta, M.-C. (2017). Patterns of Early-Life Gut Microbial Colonization during Human Immune Development: An Ecological Perspective. *Frontiers in Immunology*, 8. <https://doi.org/10.3389/fimmu.2017.00788>
- Lamberti, L. M., Fischer Walker, C. L., Noiman, A., Victora, C., & Black, R. E. (2011). Breastfeeding and the risk for diarrhea morbidity and mortality. *BMC Public Health*, 11(S3), S15. <https://doi.org/10.1186/1471-2458-11-S3-S15>
- Lane, A. A., McGuire, M. K., McGuire, M. A., Williams, J. E., Lackey, K. A., Hagen, E. H., Kaul, A., Gindola, D., Gebeyehu, D., Flores, K. E., Foster, J. A., Sellen, D. W., Kamau-Mbuthia, E. W., Kamundia, E. W., Mbugua, S., Moore, S. E., Prentice, A. M., Kvist, L. J., Otoo, G. E., ... Meehan, C. L. (2019). Household composition and the infant fecal microbiome: The INSPIRE study. *American Journal of Physical Anthropology*, 169(3), 526–539. <https://doi.org/10.1002/ajpa.23843>
- Laursen, M. F., Andersen, L. B. B., Michaelsen, K. F., Mølgaard, C., Trolle, E., Bahl, M. I., & Licht, T. R. (2016). Infant Gut Microbiota Development Is Driven by Transition to Family Foods Independent of Maternal Obesity. *MSphere*, 1(1). <https://doi.org/10.1128/mSphere.00069-15>
- Laursen, M. F., Sakanaka, M., von Burg, N., Mörbe, U., Andersen, D., Moll, J. M., Pekmez, C. T., Rivollier, A., Michaelsen, K. F., Mølgaard, C., Lind, M. V., Dragsted, L. O., Katayama, T., Frandsen, H. L., Vinggaard, A. M., Bahl, M. I., Brix, S., Agace, W., Licht, T. R., & Roager, H. M. (2021). Bifidobacterium species associated with breastfeeding produce aromatic lactic acids in the infant gut. *Nature Microbiology*, 6(11), 1367–1382. <https://doi.org/10.1038/s41564-021-00970-4>
- Laursen, M. F., Zachariassen, G., Bahl, M. I., Bergström, A., Høst, A., Michaelsen, K. F., & Licht, T. R. (2015). Having older siblings is associated with gut microbiota development during early childhood. *BMC Microbiology*, 15(1), 154. <https://doi.org/10.1186/s12866-015-0477-6>
- Lawal Bello, U., Haruna Ibrahim, A., Muhammad Ladan, A., Yunusa, U., & Alhaji Yaro, N. (2024). Pattern and Distribution of Open Defecation Practices in Kano State, Nigeria. *Journal of Institute of Africa Higher Education Research and Innovation (IAHERI)*, 1(1). <https://doi.org/10.59479/jiaheri.v1i001.40>

- Laxminarayan, R., Matsoso, P., Pant, S., Brower, C., Røttingen, J.-A., Klugman, K., & Davies, S. (2016). Access to effective antimicrobials: a worldwide challenge. *The Lancet*, 387(10014), 168–175. [https://doi.org/10.1016/S0140-6736\(15\)00474-2](https://doi.org/10.1016/S0140-6736(15)00474-2)
- Lebeaux, R. M., Coker, M. O., Dade, E. F., Palys, T. J., Morrison, H. G., Ross, B. D., Baker, E. R., Karagas, M. R., Madan, J. C., & Hoen, A. G. (2021). The infant gut resistome is associated with *E. coli* and early-life exposures. *BMC Microbiology*, 21(1), 201. <https://doi.org/10.1186/s12866-021-02129-x>
- Lebeaux, R. M., Madan, J. C., Nguyen, Q. P., Coker, M. O., Dade, E. F., Moroishi, Y., Palys, T. J., Ross, B. D., Pettigrew, M. M., Morrison, H. G., Karagas, M. R., & Hoen, A. G. (2022). Impact of antibiotics on off-target infant gut microbiota and resistance genes in cohort studies. *Pediatric Research*, 92(6), 1757–1766. <https://doi.org/10.1038/s41390-022-02104-w>
- Leo, S., Cetiner, O. F., Pittet, L. F., Messina, N. L., Jakob, W., Falquet, L., Curtis, N., & Zimmermann, P. (2023). Metagenomics analysis of the neonatal intestinal resistome. *Frontiers in Pediatrics*, 11. <https://doi.org/10.3389/fped.2023.1169651>
- Leong, K. S. W., McLay, J., Derraik, J. G. B., Gibb, S., Shackleton, N., Taylor, R. W., Glover, M., Audas, R., Taylor, B., Milne, B. J., & Cutfield, W. S. (2020). Associations of Prenatal and Childhood Antibiotic Exposure With Obesity at Age 4 Years. *JAMA Network Open*, 3(1), e1919681. <https://doi.org/10.1001/jamanetworkopen.2019.19681>
- Li, X., Brejnrod, A., Thorsen, J., Zachariasen, T., Trivedi, U., Russel, J., Vestergaard, G. A., Stokholm, J., Rasmussen, M. A., & Sørensen, S. J. (2023). Differential responses of the gut microbiome and resistome to antibiotic exposures in infants and adults. *Nature Communications*, 14(1), 8526. <https://doi.org/10.1038/s41467-023-44289-6>
- Li, X., Brejnrod, A., Trivedi, U., Russel, J., Thorsen, J., Shah, S. A., Vestergaard, G. A., Rasmussen, M. A., Nesme, J., Bisgaard, H., Stokholm, J., & Sørensen, S. J. (2024a). Co-localization of antibiotic resistance genes is widespread in the infant gut microbiome and associates with an immature gut microbial composition. *Microbiome*, 12(1), 87. <https://doi.org/10.1186/s40168-024-01800-5>
- Li, X., Brejnrod, A., Trivedi, U., Russel, J., Thorsen, J., Shah, S. A., Vestergaard, G. A., Rasmussen, M. A., Nesme, J., Bisgaard, H., Stokholm, J., & Sørensen, S. J. (2024b). Co-localization of antibiotic resistance genes is widespread in the infant gut microbiome and

- associates with an immature gut microbial composition. *Microbiome*, 12(1), 87. <https://doi.org/10.1186/s40168-024-01800-5>
- Li, Y., Xu, Z., Han, W., Cao, H., Umarov, R., Yan, A., Fan, M., Chen, H., Duarte, C. M., Li, L., Ho, P.-L., & Gao, X. (2021). HMD-ARG: hierarchical multi-task deep learning for annotating antibiotic resistance genes. *Microbiome*, 9(1), 40. <https://doi.org/10.1186/s40168-021-01002-3>
- Lim, M. Y., Hong, S., Kim, B.-M., Ahn, Y., Kim, H.-J., & Nam, Y.-D. (2020). Changes in microbiome and metabolomic profiles of fecal samples stored with stabilizing solution at room temperature: a pilot study. *Scientific Reports*, 10(1), 1789. <https://doi.org/10.1038/s41598-020-58719-8>
- Lim, M. Y., Song, E.-J., Kim, S. H., Lee, J., & Nam, Y.-D. (2018). Comparison of DNA extraction methods for human gut microbial community profiling. *Systematic and Applied Microbiology*, 41(2), 151–157. <https://doi.org/10.1016/j.syapm.2017.11.008>
- Lokossou, G. A. G., Kouakanou, L., Schumacher, A., & Zenclussen, A. C. (2022). Human Breast Milk: From Food to Active Immune Response With Disease Protection in Infants and Mothers. *Frontiers in Immunology*, 13. <https://doi.org/10.3389/fimmu.2022.849012>
- Lopatkin, A. J., Meredith, H. R., Srimani, J. K., Pfeiffer, C., Durrett, R., & You, L. (2017). Persistence and reversal of plasmid-mediated antibiotic resistance. *Nature Communications*, 8(1), 1689. <https://doi.org/10.1038/s41467-017-01532-1>
- Luchen, C. C., Chibuye, M., Spijker, R., Simuyandi, M., Chisenga, C., Bosomprah, S., Chilengi, R., Schultz, C., Mende, D. R., & Harris, V. C. (2023). Impact of antibiotics on gut microbiome composition and resistome in the first years of life in low- to middle-income countries: A systematic review. *PLOS Medicine*, 20(6), e1004235. <https://doi.org/10.1371/journal.pmed.1004235>
- Lyons, K. E., Ryan, C. A., Dempsey, E. M., Ross, R. P., & Stanton, C. (2020). Breast Milk, a Source of Beneficial Microbes and Associated Benefits for Infant Health. *Nutrients*, 12(4), 1039. <https://doi.org/10.3390/nu12041039>
- Ma, J., Li, Z., Zhang, W., Zhang, C., Zhang, Y., Mei, H., Zhuo, N., Wang, H., Wang, L., & Wu, D. (2020). Comparison of gut microbiota in exclusively breast-fed and formula-fed

- babies: a study of 91 term infants. *Scientific Reports*, 10(1), 15792. <https://doi.org/10.1038/s41598-020-72635-x>
- Maghini, D. G., Moss, E. L., Vance, S. E., & Bhatt, A. S. (2021). Improved high-molecular-weight DNA extraction, nanopore sequencing and metagenomic assembly from the human gut microbiome. *Nature Protocols*, 16(1), 458–471. <https://doi.org/10.1038/s41596-020-00424-x>
- Maghini, D. G., Oduaran, O. H., Olubayo, L. A. I., Cook, J. A., Smyth, N., Mathema, T., Belger, C. W., Agongo, G., Boua, P. R., Choma, S. S. R., Gómez-Olivé, F. X., Kisiangani, I., Mashaba, G. R., Micklesfield, L., Mohamed, S. F., Nonterah, E. A., Norris, S., Sorgho, H., Tollman, S., ... Hazelhurst, S. (2025). Expanding the human gut microbiome atlas of Africa. *Nature*, 638(8051), 718–728. <https://doi.org/10.1038/s41586-024-08485-8>
- Magiorakos, A. P., Srinivasan, A., Carey, R. B., Carmeli, Y., Falagas, M. E., Giske, C. G., Harbarth, S., Hindler, J. F., Kahlmeter, G., Olsson-Liljequist, B., Paterson, D. L., Rice, L. B., Stelling, J., Struelens, M. J., Vatopoulos, A., Weber, J. T., & Monnet, D. L. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance. *Clinical Microbiology and Infection*, 18(3), 268–281. <https://doi.org/10.1111/j.1469-0691.2011.03570.x>
- Magnúsdóttir, S., Ravcheev, D., de Crécy-Lagard, V., & Thiele, I. (2015). Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Frontiers in Genetics*, 6. <https://doi.org/10.3389/fgene.2015.00148>
- Mahtab, S., Madewell, Z. J., Baillie, V., Dangor, Z., Lala, S. G., Assefa, N., Berihun, M., Madrid, L., Regassa, L. D., Scott, J. A. G., Ameh, S., Bangura, J. S., Ita, O., Kaluma, E., Ogbuanu, I. U., Gaume, B., Kotloff, K. L., Sow, S. O., Tapia, M. D., ... Bassat, Q. (2024). Etiologies and comorbidities of meningitis deaths in children under 5 years in high-mortality settings: Insights from the CHAMPS Network in the post-pneumococcal vaccine era. *Journal of Infection*, 89(6), 106341. <https://doi.org/10.1016/j.jinf.2024.106341>
- Marcobal, A., Barboza, M., Froehlich, J. W., Block, D. E., German, J. B., Lebrilla, C. B., & Mills, D. A. (2010). Consumption of Human Milk Oligosaccharides by Gut-Related

- Microbes. *Journal of Agricultural and Food Chemistry*, 58(9), 5334–5340. <https://doi.org/10.1021/jf9044205>
- Martinez, J. L., & Baquero, F. (2000). Mutation Frequencies and Antibiotic Resistance. *Antimicrobial Agents and Chemotherapy*, 44(7), 1771–1777. <https://doi.org/10.1128/AAC.44.7.1771-1777.2000>
- Mashau, R. C., Meiring, S. T., Dramowski, A., Magobo, R. E., Quan, V. C., Perovic, O., von Gottberg, A., Cohen, C., Velaphi, S., van Schalkwyk, E., & Govender, N. P. (2022). Culture-confirmed neonatal bloodstream infections and meningitis in South Africa, 2014–19: a cross-sectional study. *The Lancet Global Health*, 10(8), e1170–e1178. [https://doi.org/10.1016/S2214-109X\(22\)00246-7](https://doi.org/10.1016/S2214-109X(22)00246-7)
- Matharu, D., Ponsero, A. J., Dikareva, E., Korpela, K., Kolho, K.-L., de Vos, W. M., & Salonen, A. (2022). Bacteroides abundance drives birth mode dependent infant gut microbiota developmental trajectories. *Frontiers in Microbiology*, 13. <https://doi.org/10.3389/fmicb.2022.953475>
- Mebrahtom, S., Worku, A., & Gage, D. J. (2022). The risk of water, sanitation and hygiene on diarrhea-related infant mortality in eastern Ethiopia: a population-based nested case-control. *BMC Public Health*, 22(1), 343. <https://doi.org/10.1186/s12889-022-12735-7>
- Meiring, S., Quan, V., Mashau, R., Perovic, O., Magobo, R., Smith, M., Mpembe, R., von Gottberg, A., de Gouveia, L., Walaza, S., Cohen, C., Kapongo, C., Mackay, C., Mailula, M. T., Mekgoe, O., Motjale, L., Phayane, R., Dramowski, A., Govender, N. P., ... Pearce, V. (2025). Pathogen aetiology and risk factors for death among neonates with bloodstream infections at lower-tier South African hospitals: a cross-sectional study. *The Lancet Microbe*, 6(5), 100989. <https://doi.org/10.1016/j.lanmic.2024.100989>
- Milani, C., Duranti, S., Bottacini, F., Casey, E., Turrone, F., Mahony, J., Belzer, C., Delgado Palacio, S., Arbolea Montes, S., Mancabelli, L., Lugli, G. A., Rodriguez, J. M., Bode, L., de Vos, W., Gueimonde, M., Margolles, A., van Sinderen, D., & Ventura, M. (2017). The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiology and Molecular Biology Reviews*, 81(4). <https://doi.org/10.1128/MMBR.00036-17>
- Mills, M., Lee, S., Piperata, B. A., Garabed, R., Choi, B., & Lee, J. (2023). Household environment and animal fecal contamination are critical modifiers of the gut microbiome

- and resistome in young children from rural Nicaragua. *Microbiome*, 11(1), 207. <https://doi.org/10.1186/s40168-023-01636-5>
- Milton, R., Alkali, F. I., Modibbo, F., Sanders, J., Mukaddas, A. S., Kassim, A., Sa'ad, F. H., Tukur, F. M., Pell, B., Hood, K., Ghazal, P., & Iregbu, K. C. (2021). A qualitative focus group study concerning perceptions and experiences of Nigerian mothers on stillbirths. *BMC Pregnancy and Childbirth*, 21(1), 830. <https://doi.org/10.1186/s12884-021-04207-4>
- Milton, R., Gillespie, D., Dyer, C., Taiyari, K., Carvalho, M. J., Thomson, K., Sands, K., Portal, E. A. R., Hood, K., Ferreira, A., Hender, T., Kirby, N., Mathias, J., Nieto, M., Watkins, W. J., Bekele, D., Abayneh, M., Solomon, S., Basu, S., ... Paterson, L. (2022). Neonatal sepsis and mortality in low-income and middle-income countries from a facility-based birth cohort: an international multisite prospective observational study. *The Lancet Global Health*, 10(5), e661–e672. [https://doi.org/10.1016/S2214-109X\(22\)00043-2](https://doi.org/10.1016/S2214-109X(22)00043-2)
- Milton, R., Modibbo, F., Gillespie, D., Alkali, F. I., Mukaddas, A. S., Kassim, A., Sa'ad, F. H., Tukur, F. M., Khalid, R. Y., Muhammad, M. Y., Bello, M., Edwin, C. P., Ogudo, E., Iregbu, K. C., Jones, L., Hood, K., Ghazal, P., Sanders, J., Hassan, B., ... Walsh, T. R. (2022). Incidence and sociodemographic, living environment and maternal health associations with stillbirth in a tertiary healthcare setting in Kano, Northern Nigeria. *BMC Pregnancy and Childbirth*, 22(1), 692. <https://doi.org/10.1186/s12884-022-04971-x>
- Mitchell, C. M., Mazzoni, C., Hogstrom, L., Bryant, A., Bergerat, A., Cher, A., Pochan, S., Herman, P., Carrigan, M., Sharp, K., Huttenhower, C., Lander, E. S., Vlamakis, H., Xavier, R. J., & Yassour, M. (2020). Delivery Mode Affects Stability of Early Infant Gut Microbiota. *Cell Reports Medicine*, 1(9), 100156. <https://doi.org/10.1016/j.xcrm.2020.100156>
- Moore, A. M., Ahmadi, S., Patel, S., Gibson, M. K., Wang, B., Ndao, I. M., Deych, E., Shannon, W., Tarr, P. I., Warner, B. B., & Dantas, G. (2015). Gut resistome development in healthy twin pairs in the first year of life. *Microbiome*, 3(1), 27. <https://doi.org/10.1186/s40168-015-0090-9>
- Moss, E. L., Maghini, D. G., & Bhatt, A. S. (2020). Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nature Biotechnology*, 38(6), 701–707. <https://doi.org/10.1038/s41587-020-0422-6>

- Mukesh Dherani, Daniel Pope, Maya Mascarenhas, Kirk R. Smith, Martin Weber, & Nigel Bruce. (2008). Indoor air pollution from unprocessed solid fuel use and pneumonia risk in children aged under five years: a systematic review and meta-analysis. *Bulletin of the World Health Organization*, 86(5), 390–398. <https://doi.org/10.2471/BLT.07.044529>
- Munk, P., Brinch, C., Møller, F. D., Petersen, T. N., Hendriksen, R. S., Seyfarth, A. M., Kjeldgaard, J. S., Svendsen, C. A., van Bunnik, B., Berglund, F., Bego, A., Power, P., Rees, C., Lambrinidis, D., Neilson, E. H. J., Gibb, K., Coventry, K., Collignon, P., Cassar, S., ... Aarestrup, F. M. (2022). Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance. *Nature Communications*, 13(1), 7251. <https://doi.org/10.1038/s41467-022-34312-7>
- Murray, C. J., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., ... Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*, 399(10325), 629–655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)
- Musa, S. (2023). Traditional and cultural newborn care practices in the Dakace and Tsibiri communities of Nigeria: an ethnographic study. *Annals of Clinical and Biomedical Research*, 4(1). <https://doi.org/10.4081/acbr.2023.304>
- Musicha, P., Cornick, J. E., Bar-Zeev, N., French, N., Masesa, C., Denis, B., Kennedy, N., Mallewa, J., Gordon, M. A., Msefula, C. L., Heyderman, R. S., Everett, D. B., & Feasey, N. A. (2017). Trends in antimicrobial resistance in bloodstream infection isolates at a large urban hospital in Malawi (1998–2016): a surveillance study. *The Lancet Infectious Diseases*, 17(10), 1042–1052. [https://doi.org/10.1016/S1473-3099\(17\)30394-8](https://doi.org/10.1016/S1473-3099(17)30394-8)
- Nadimpalli, M. L., Rojas Salvatierra, L., Chakraborty, S., Swarthout, J. M., Cabrera, L. Z., Pickering, A. J., Calderon, M., Saito, M., Gilman, R. H., & Pajuelo, M. J. (2024). Effects of breastfeeding on children's gut colonization with multidrug-resistant Enterobacterales in peri-urban Lima, Peru. *Gut Microbes*, 16(1). <https://doi.org/10.1080/19490976.2024.2309681>
- Naghavi, M., Vollset, S. E., Ikuta, K. S., Swetschinski, L. R., Gray, A. P., Wool, E. E., Robles Aguilar, G., Mestrovic, T., Smith, G., Han, C., Hsu, R. L., Chalek, J., Araki, D. T., Chung,

- E., Raggi, C., Gershberg Hayoon, A., Davis Weaver, N., Lindstedt, P. A., Smith, A. E., ... Murray, C. J. L. (2024). Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *The Lancet*, 404(10459), 1199–1226. [https://doi.org/10.1016/S0140-6736\(24\)01867-1](https://doi.org/10.1016/S0140-6736(24)01867-1)
- Naha, S., Basak, P., Sands, K., Milton, R., Carvalho, M. J., Mitra, S., Bhattacharjee, A., Sinha, A., Mukherjee, S., Saha, B., Chattopadhyay, P., Chakravorty, P. S., Nandy, R. K., Dutta, S., Walsh, T. R., & Basu, S. (2023). Carriage and within-host diversity of *mcr-I.1*-harbouring *Escherichia coli* from pregnant mothers: inter- and intra-mother transmission dynamics of *mcr-I.1*. *Emerging Microbes & Infections*, 12(2). <https://doi.org/10.1080/22221751.2023.2278899>
- Nassereddine, Z. N., Opara, S. D., Coutinho, O. A., Qyteti, F., Book, R., Heinicke, M. P., Napieralski, J., & Tiquia-Arashiro, S. M. (2024). Critical perspectives on advancing antibiotic resistant gene (ARG) detection technologies in aquatic ecosystems. *Science of The Total Environment*, 957, 177775. <https://doi.org/10.1016/j.scitotenv.2024.177775>
- National Institute for Health and Care Excellence (NICE). (2021). *Neonatal infection: antibiotics for prevention and treatment NICE guideline*. www.nice.org.uk/guidance/ng195
- National Population Commission & National Bureau of Statistics. (2023). *Demographic Statistics Bulletin 2022*.
- National Population Commission (NPC), & ICF. (2019). *The Federal Republic of Nigeria Nigeria Demographic and Health Survey 2018 National Population Commission Abuja, Nigeria*. www.DHSprogram.com.
- Navab-Daneshmand, T., Friedrich, M. N. D., Gächter, M., Montealegre, M. C., Mlambo, L. S., Nhiwatiwa, T., Mosler, H.-J., & Julian, T. R. (2018). *Escherichia coli* Contamination across Multiple Environmental Compartments (Soil, Hands, Drinking Water, and Handwashing Water) in Urban Harare: Correlations and Risk Factors. *The American Journal of Tropical Medicine and Hygiene*, 98(3), 803–813. <https://doi.org/10.4269/ajtmh.17-0521>
- Neemann, K., Olateju, E. K., Izevbigie, N., Akaba, G., Olanipekun, G. M., Richard, J. C., Duru, C. I., Kocmich, N. J., Samson, K. K., Rezac-Elgohary, A., Anigilaje, E. A., Yunusa, T., Megafu, C. O. N., Ajose, T. O., Medugu, N., Meza, J., & Obaro, S. (2020). Neonatal

- outcomes associated with maternal recto-vaginal colonization with extended-spectrum β -lactamase producing Enterobacteriaceae in Nigeria: a prospective, cross-sectional study. *Clinical Microbiology and Infection*, 26(4), 463–469. <https://doi.org/10.1016/j.cmi.2019.07.013>
- Nel Van Zyl, K., Whitelaw, A. C., & Newton-Foot, M. (2020). The effect of storage conditions on microbial communities in stool. *PLOS ONE*, 15(1), e0227486. <https://doi.org/10.1371/journal.pone.0227486>
- Nieto-Rosado, M., Sands, K., Portal, E. A. R., Thomson, K. M., Carvalho, M. J., Mathias, J., Milton, R., Dyer, C., Akpulu, C., Boostrom, I., Hogan, P., Saif, H., Sanches Ferreira, A. D., Hender, T., Portal, B., Andrews, R., Watkins, W. J., Zahra, R., Shirazi, H., ... Walsh, T. R. (2024). Colonisation of hospital surfaces from low- and middle-income countries by extended spectrum β -lactamase- and carbapenemase-producing bacteria. *Nature Communications*, 15(1), 2758. <https://doi.org/10.1038/s41467-024-46684-z>
- Nigerian National Bureau of Statistics, & World Bank Living Standards Measurement Study. (2024). *Nigeria General Household Survey-Panel (Wave 5) 2023/2024 Tracking Nigerian Households to Understand Their Resilience Over Time*.
- Niyoyita, J. C., Ndayisenga, J., Omolo, J., Niyompano, H., Bimenyimana, P. C., Dzinamarira, T., Nsekuye, O., Chavez, I., & Hakizayezu, F. (2024). Factors associated with neonatal sepsis among neonates admitted in Kibungo Referral Hospital, Rwanda. *Scientific Reports*, 14(1), 15961. <https://doi.org/10.1038/s41598-024-66818-z>
- Nogacka, A. M., Salazar, N., Arbolea, S., Suárez, M., Fernández, N., Solís, G., de los Reyes-Gavilán, C. G., & Gueimonde, M. (2018). Early microbiota, antibiotics and health. *Cellular and Molecular Life Sciences*, 75(1), 83–91. <https://doi.org/10.1007/s00018-017-2670-2>
- Null, C., Stewart, C. P., Pickering, A. J., Dentz, H. N., Arnold, B. F., Arnold, C. D., Benjamin-Chung, J., Clasen, T., Dewey, K. G., Fernald, L. C. H., Hubbard, A. E., Kariger, P., Lin, A., Luby, S. P., Mertens, A., Njenga, S. M., Nyambane, G., Ram, P. K., & Colford, J. M. (2018). Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: a cluster-randomised controlled trial. *The Lancet Global Health*, 6(3), e316–e329. [https://doi.org/10.1016/S2214-109X\(18\)30005-6](https://doi.org/10.1016/S2214-109X(18)30005-6)

- Obayiuwana, A., & Ibekwe, A. M. (2020). Antibiotic Resistance Genes Occurrence in Wastewaters from Selected Pharmaceutical Facilities in Nigeria. *Water*, 12(7), 1897. <https://doi.org/10.3390/w12071897>
- O'Brien, C. E., Meier, A. K., Cernioglo, K., Mitchell, R. D., Casaburi, G., Frese, S. A., Henrick, B. M., Underwood, M. A., & Smilowitz, J. T. (2022). Early probiotic supplementation with *B. infantis* in breastfed infants leads to persistent colonization at 1 year. *Pediatric Research*, 91(3), 627–636. <https://doi.org/10.1038/s41390-020-01350-0>
- Ogbo, F. A., Agho, K., Ogeleka, P., Woolfenden, S., Page, A., & Eastwood, J. (2017). Infant feeding practices and diarrhoea in sub-Saharan African countries with high diarrhoea mortality. *PLOS ONE*, 12(2), e0171792. <https://doi.org/10.1371/journal.pone.0171792>
- Ogunkunle, T. O., Bello, S. O., Ibrahim, T. L., Sanni, U. A., Aliu, R., Sanusi, I., Katibi, S. O., & Ibraheem, R. (2024). Antibiotics use practices for hospitalized asphyxiated neonates: a cross-sectional survey across selected Nigerian newborn units. *Journal of Xiangya Medicine*, 9, 16–16. <https://doi.org/10.21037/jxym-24-7>
- Okeke, I. N., de Kraker, M. E. A., Van Boeckel, T. P., Kumar, C. K., Schmitt, H., Gales, A. C., Bertagnolio, S., Sharland, M., & Laxminarayan, R. (2024). The scope of the antimicrobial resistance challenge. *The Lancet*, 403(10442), 2426–2438. [https://doi.org/10.1016/S0140-6736\(24\)00876-6](https://doi.org/10.1016/S0140-6736(24)00876-6)
- Okomo, U., Akpalu, E. N. K., Le Doare, K., Roca, A., Cousens, S., Jarde, A., Sharland, M., Kampmann, B., & Lawn, J. E. (2019). Aetiology of invasive bacterial infection and antimicrobial resistance in neonates in sub-Saharan Africa: a systematic review and meta-analysis in line with the STROBE-NI reporting guidelines. *The Lancet Infectious Diseases*, 19(11), 1219–1234. [https://doi.org/10.1016/S1473-3099\(19\)30414-1](https://doi.org/10.1016/S1473-3099(19)30414-1)
- Olin, A., Henckel, E., Chen, Y., Lakshmikanth, T., Pou, C., Mikes, J., Gustafsson, A., Bernhardsson, A. K., Zhang, C., Bohlin, K., & Brodin, P. (2018). Stereotypic Immune System Development in Newborn Children. *Cell*, 174(5), 1277-1292.e14. <https://doi.org/10.1016/j.cell.2018.06.045>
- Olsen, N. S., & Riber, L. (2025). Metagenomics as a Transformative Tool for Antibiotic Resistance Surveillance: Highlighting the Impact of Mobile Genetic Elements with a Focus on the Complex Role of Phages. *Antibiotics*, 14(3), 296. <https://doi.org/10.3390/antibiotics14030296>

- Onderdonk, A. B., Hecht, J. L., McElrath, T. F., Delaney, M. L., Allred, E. N., & Leviton, A. (2008). Colonization of second-trimester placenta parenchyma. *American Journal of Obstetrics and Gynecology*, 199(1), 52.e1-52.e10. <https://doi.org/10.1016/j.ajog.2007.11.068>
- O'Neill, J. (2016). *TACKLING DRUG-RESISTANT INFECTIONS GLOBALLY: FINAL REPORT AND RECOMMENDATIONS THE REVIEW ON ANTIMICROBIAL RESISTANCE CHAIRED BY JIM O'NEILL*. <https://amr-review.org/Publications.html>
- Organisation for Economic Co-operation and Development (OECD). (2023). *Fighting Antimicrobial Resistance in EU and EEA countries EMBRACING A ONE HEALTH APPROACH FOCUS ON*.
- Oyedemi, O. T., Shaw, S., Martin, J. C., Ayeni, F. A., & Scott, K. P. (2022). Changes in the gut microbiota of Nigerian infants within the first year of life. *PLOS ONE*, 17(3), e0265123. <https://doi.org/10.1371/journal.pone.0265123>
- Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A., & Brown, P. O. (2007). Development of the Human Infant Intestinal Microbiota. *PLoS Biology*, 5(7), e177. <https://doi.org/10.1371/journal.pbio.0050177>
- Pärnänen, K., Karkman, A., Hultman, J., Lyra, C., Bengtsson-Palme, J., Larsson, D. G. J., Rautava, S., Isolauri, E., Salminen, S., Kumar, H., Satokari, R., & Virta, M. (2018). Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. *Nature Communications*, 9(1), 3891. <https://doi.org/10.1038/s41467-018-06393-w>
- Pärnänen, K. M., Hultman, J., Markkanen, M., Satokari, R., Rautava, S., Lamendella, R., Wright, J., McLimans, C. J., Kelleher, S. L., & Virta, M. P. (2022). Early-life formula feeding is associated with infant gut microbiota alterations and an increased antibiotic resistance load. *The American Journal of Clinical Nutrition*, 115(2), 407–421. <https://doi.org/10.1093/ajcn/nqab353>
- Partridge, S. R., Kwong, S. M., Firth, N., & Jensen, S. O. (2018). Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clinical Microbiology Reviews*, 31(4). <https://doi.org/10.1128/CMR.00088-17>

- Patangia, D. V., Grimaud, G., O'Shea, C.-A., Ryan, C. A., Dempsey, E., Stanton, C., & Ross, R. P. (2024a). Early life exposure of infants to benzylpenicillin and gentamicin is associated with a persistent amplification of the gut resistome. *Microbiome*, 12(1), 19. <https://doi.org/10.1186/s40168-023-01732-6>
- Patangia, D. V., Grimaud, G., O'Shea, C.-A., Ryan, C. A., Dempsey, E., Stanton, C., & Ross, R. P. (2024b). Early life exposure of infants to benzylpenicillin and gentamicin is associated with a persistent amplification of the gut resistome. *Microbiome*, 12(1), 19. <https://doi.org/10.1186/s40168-023-01732-6>
- Patangia, D. V., Grimaud, G., Wang, S., Ross, R. P., & Stanton, C. (2024). Influence of age, socioeconomic status, and location on the infant gut resistome across populations. *Gut Microbes*, 16(1). <https://doi.org/10.1080/19490976.2023.2297837>
- Pearse, O., Lester, R., Zuza, A., Mangochi, H., Siyabu, P., Tewesa, E., Edwards, T., Thomson, N. R., Kawaza, K., Musicha, P., Cornick, J., Heinz, E., Feasey, N., & Jewell, C. (2025). *Extended-Spectrum Beta-Lactamase Klebsiella pneumoniae on a Malawian neonatal unit is amplified by neonates and transmitted by maternal hands, cots and ward surfaces*. <https://doi.org/10.1101/2025.08.13.25333346>
- Pehrsson, E. C., Tsukayama, P., Patel, S., Mejía-Bautista, M., Sosa-Soto, G., Navarrete, K. M., Calderon, M., Cabrera, L., Hoyos-Arango, W., Bertoli, M. T., Berg, D. E., Gilman, R. H., & Dantas, G. (2016). Interconnected microbiomes and resistomes in low-income human habitats. *Nature*, 533(7602), 212–216. <https://doi.org/10.1038/nature17672>
- Peng, Z., Mao, Y., Zhang, N., Zhang, L., Wang, Z., & Han, M. (2021). Utilizing Metagenomic Data and Bioinformatic Tools for Elucidating Antibiotic Resistance Genes in Environment. *Frontiers in Environmental Science*, 9. <https://doi.org/10.3389/fenvs.2021.757365>
- Perry, J., Waglechner, N., & Wright, G. (2016). The Prehistory of Antibiotic Resistance. *Cold Spring Harbor Perspectives in Medicine*, 6(6), a025197. <https://doi.org/10.1101/cshperspect.a025197>
- Pickard, J. M., Zeng, M. Y., Caruso, R., & Núñez, G. (2017). Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. *Immunological Reviews*, 279(1), 70–89. <https://doi.org/10.1111/imr.12567>

- Pickering, A. J., Null, C., Winch, P. J., Mangwadu, G., Arnold, B. F., Prendergast, A. J., Njenga, S. M., Rahman, M., Ntozini, R., Benjamin-Chung, J., Stewart, C. P., Huda, T. M. N., Moulton, L. H., Colford, J. M., Luby, S. P., & Humphrey, J. H. (2019). The WASH Benefits and SHINE trials: interpretation of WASH intervention effects on linear growth and diarrhoea. *The Lancet Global Health*, 7(8), e1139–e1146. [https://doi.org/10.1016/S2214-109X\(19\)30268-2](https://doi.org/10.1016/S2214-109X(19)30268-2)
- Portal, E. A. R., Sands, K., Farley, C., Boostrom, I., Jones, E., Barrell, M., Carvalho, M. J., Milton, R., Iregbu, K., Modibbo, F., Uwaezuoke, S., Akpulu, C., Audu, L., Edwin, C., Yusuf, A. H., Adeleye, A., Mukkadas, A. S., Maduekwe, D., Gambo, S., ... Spiller, O. B. (2024). Characterisation of colistin resistance in Gram-negative microbiota of pregnant women and neonates in Nigeria. *Nature Communications*, 15(1), 2302. <https://doi.org/10.1038/s41467-024-45673-6>
- Portik, D. M., Brown, C. T., & Pierce-Ward, N. T. (2022). Evaluation of taxonomic classification and profiling methods for long-read shotgun metagenomic sequencing datasets. *BMC Bioinformatics*, 23(1), 541. <https://doi.org/10.1186/s12859-022-05103-0>
- Preywisch, R., Løvlie, R., & Eidens, M. (2025). Stabilization of human RNA and DNA in stool samples at room temperature. *BioTechniques*, 1–9. <https://doi.org/10.1080/07366205.2025.2546762>
- Prince, A. L., Ma, J., Kannan, P. S., Alvarez, M., Gisslen, T., Harris, R. A., Sweeney, E. L., Knox, C. L., Lambers, D. S., Jobe, A. H., Chougnet, C. A., Kallapur, S. G., & Aagaard, K. M. (2016). The placental membrane microbiome is altered among subjects with spontaneous preterm birth with and without chorioamnionitis. *American Journal of Obstetrics and Gynecology*, 214(5), 627.e1-627.e16. <https://doi.org/10.1016/j.ajog.2016.01.193>
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9), 833–844. <https://doi.org/10.1038/nbt.3935>
- Rajar, P., Dhariwal, A., Salvadori, G., Junges, R., Åmdal, H. A., Berild, D., Fugelseth, D., Saugstad, O. D., Lausten-Thomsen, U., Greisen, G., Haaland, K., & Petersen, F. C. (2022). Microbial DNA extraction of high-host content and low biomass samples: Optimized

- protocol for nasopharynx metagenomic studies. *Frontiers in Microbiology*, 13. <https://doi.org/10.3389/fmicb.2022.1038120>
- Ranjeva, S. L., Warf, B. C., & Schiff, S. J. (2018). Economic burden of neonatal sepsis in sub-Saharan Africa. *BMJ Global Health*, 3(1), e000347. <https://doi.org/10.1136/bmjgh-2017-000347>
- Ravi, A., Avershina, E., Foley, S. L., Ludvigsen, J., Storrø, O., Øien, T., Johnsen, R., McCartney, A. L., L'Abée-Lund, T. M., & Rudi, K. (2015). The commensal infant gut meta-mobilome as a potential reservoir for persistent multidrug resistance integrons. *Scientific Reports*, 5(1), 15317. <https://doi.org/10.1038/srep15317>
- Reyman, M., van Houten, M. A., van Baarle, D., Bosch, A. A. T. M., Man, W. H., Chu, M. L. J. N., Arp, K., Watson, R. L., Sanders, E. A. M., Fuentes, S., & Bogaert, D. (2019). Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life. *Nature Communications*, 10(1), 4997. <https://doi.org/10.1038/s41467-019-13014-7>
- Reyman, M., van Houten, M. A., Watson, R. L., Chu, M. L. J. N., Arp, K., de Waal, W. J., Schiering, I., Plötz, F. B., Willems, R. J. L., van Schaik, W., Sanders, E. A. M., & Bogaert, D. (2022). Effects of early-life antibiotics on the developing infant gut microbiome and resistome: a randomized trial. *Nature Communications*, 13(1), 1–12. <https://doi.org/10.1038/s41467-022-28525-z>
- Rice, L. B. (2008). Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE. *The Journal of Infectious Diseases*, 197(8), 1079–1081. <https://doi.org/10.1086/533452>
- Roswall, J., Olsson, L. M., Kovatcheva-Datchary, P., Nilsson, S., Tremaroli, V., Simon, M.-C., Kiilerich, P., Akrami, R., Krämer, M., Uhlén, M., Gummesson, A., Kristiansen, K., Dahlgren, J., & Bäckhed, F. (2021). Developmental trajectory of the healthy human gut microbiota during the first 5 years of life. *Cell Host & Microbe*, 29(5), 765–776.e3. <https://doi.org/10.1016/j.chom.2021.02.021>
- Ruppé, E., & Schrenzel, J. (2018). Messages from the second International Conference on Clinical Metagenomics (ICCMg2). *Microbes and Infection*, 20(4), 222–227. <https://doi.org/10.1016/j.micinf.2018.02.005>

- Russell, N. J., Stöhr, W., Plakkal, N., Cook, A., Berkley, J. A., Adhisivam, B., Agarwal, R., Ahmed, N. U., Balasegaram, M., Ballot, D., Bekker, A., Berezin, E. N., Bilardi, D., Boonkasidecha, S., Carnevalheiro, C. G., Chami, N., Chaurasia, S., Chiurchiu, S., Colas, V. R. F., ... Sharland, M. (2023). Patterns of antibiotic use, pathogens, and prediction of mortality in hospitalized neonates and young infants with sepsis: A global neonatal sepsis observational cohort study (NeoOBS). *PLOS Medicine*, 20(6), e1004179. <https://doi.org/10.1371/journal.pmed.1004179>
- Ryan, F. J., Clarke, M., Lynn, M. A., Benson, S. C., McAlister, S., Giles, L. C., Choo, J. M., Rossouw, C., Ng, Y. Y., Semchenko, E. A., Richard, A., Leong, L. E. X., Taylor, S. L., Blake, S. J., Mugabushaka, J. I., Walker, M., Wesselingh, S. L., Licciardi, P. V., Seib, K. L., ... Lynn, D. J. (2025). Bifidobacteria support optimal infant vaccine responses. *Nature*, 641(8062), 456–464. <https://doi.org/10.1038/s41586-025-08796-4>
- Saak, C. C., Dinh, C. B., & Dutton, R. J. (2020). Experimental approaches to tracking mobile genetic elements in microbial communities. *FEMS Microbiology Reviews*, 44(5), 606–630. <https://doi.org/10.1093/femsre/fuaa025>
- Saeed, M. H., & Hall, L. J. (2023). Early-life antibiotic usage and impact on the gut microbiota, including emergence of antimicrobial resistant *Enterococcus*. *Microbiota and Host*, 1(1). <https://doi.org/10.1530/MAH-23-0002>
- Salonen, A., Nikkilä, J., Jalanka-Tuovinen, J., Immonen, O., Rajilić-Stojanović, M., Kekkonen, R. A., Palva, A., & de Vos, W. M. (2010). Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: Effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *Journal of Microbiological Methods*, 81(2), 127–134. <https://doi.org/10.1016/j.mimet.2010.02.007>
- Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., & Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, 12(1), 87. <https://doi.org/10.1186/s12915-014-0087-z>
- Samarra, A., Alcañiz, A. J., Martínez-Costa, C., Marina, A., Comas, I., Segata, N., Quijada, N. M., & Collado, M. C. (2025). Breastfeeding and early Bifidobacterium-driven microbial colonization shape the infant gut resistome. *Nature Communications*, 16(1), 6099. <https://doi.org/10.1038/s41467-025-61154-w>

- Samarra, A., Cabrera-Rubio, R., Martínez-Costa, C., & Collado, M. C. (2024). Unravelling the evolutionary dynamics of antibiotic resistance genes in the infant gut microbiota during the first four months of life. *Annals of Clinical Microbiology and Antimicrobials*, 23(1), 72. <https://doi.org/10.1186/s12941-024-00725-z>
- Sands, K., Carvalho, M. J., Portal, E., Thomson, K., Dyer, C., Akpulu, C., Andrews, R., Ferreira, A., Gillespie, D., Hender, T., Hood, K., Mathias, J., Milton, R., Nieto, M., Taiyari, K., Chan, G. J., Bekele, D., Solomon, S., Basu, S., ... Walsh, T. R. (2021). Characterization of antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low- and middle-income countries. *Nature Microbiology*, 6(4), 512–523. <https://doi.org/10.1038/s41564-021-00870-7>
- Sands, K., Carvalho, M. J., Spiller, O. B., Portal, E. A. R., Thomson, K., Watkins, W. J., Mathias, J., Dyer, C., Akpulu, C., Andrews, R., Ferreira, A., Hender, T., Milton, R., Nieto, M., Zahra, R., Shirazi, H., Muhammad, A., Akif, S., Jan, M. H., ... Walsh, T. R. (2022). Characterisation of Staphylococci species from neonatal blood cultures in low- and middle-income countries. *BMC Infectious Diseases*, 22(1), 593. <https://doi.org/10.1186/s12879-022-07541-w>
- Sands, K., Spiller, O. B., Thomson, K., Portal, E. A., Iregbu, K. C., & Walsh, T. R. (2022). Early-Onset Neonatal Sepsis in Low- and Middle-Income Countries: Current Challenges and Future Opportunities. *Infection and Drug Resistance*, Volume 15, 933–946. <https://doi.org/10.2147/IDR.S294156>
- Saturio, S., Nogacka, A. M., Suárez, M., Fernández, N., Mantecón, L., Mancabelli, L., Milani, C., Ventura, M., de los Reyes-Gavilán, C. G., Solís, G., Arbolea, S., & Gueimonde, M. (2021). Early-Life Development of the Bifidobacterial Community in the Infant Gut. *International Journal of Molecular Sciences*, 22(7), 3382. <https://doi.org/10.3390/ijms22073382>
- Saturio, S., Rey, A., Samarra, A., Collado, M. C., Suárez, M., Mantecón, L., Solís, G., Gueimonde, M., & Arbolea, S. (2023). Old Folks, Bad Boon: Antimicrobial Resistance in the Infant Gut Microbiome. *Microorganisms*, 11(8), 1907. <https://doi.org/10.3390/microorganisms11081907>
- Schwartz, D. J., Shalon, N., Wardenburg, K., DeVaux, A., Wallace, M. A., Hall-Moore, C., Ndao, I. M., Sullivan, J. E., Radmacher, P., Escobedo, M., Burnham, C.-A. D., Warner, B.

- B., Tarr, P. I., & Dantas, G. (2023). Gut pathogen colonization precedes bloodstream infection in the neonatal intensive care unit. *Science Translational Medicine*, 15(694). <https://doi.org/10.1126/scitranslmed.adg5562>
- Sedova, D. A., Golovin, S. N., Shebeko, S. K., & Ermakov, A. M. (2025). Microbiological Diversity, Formation, Ecological Role and Research Methods of the Pig Gut Microbiota: Review. *Ecological Genetics*, 23(2), 173–184. <https://doi.org/10.17816/ecogen642085>
- Seki, D., Mayer, M., Hausmann, B., Pjevac, P., Giordano, V., Goeral, K., Unterasinger, L., Klebermaß-Schrehof, K., De Paepe, K., Van de Wiele, T., Spittler, A., Kasprian, G., Warth, B., Berger, A., Berry, D., & Wisgrill, L. (2021). Aberrant gut-microbiota-immune-brain axis development in premature neonates with brain damage. *Cell Host & Microbe*, 29(10), 1558-1572.e6. <https://doi.org/10.1016/j.chom.2021.08.004>
- Shannon, H., Allen, C., Clarke, M., Dávila, D., Fletcher-Wood, L., Gupta, S., Keck, K., Lang, S., Ludolph, R., & Kahangire, D. A. (2018). *Web Annex A Report of the systematic review on the effect of household crowding on health*. <http://apps.who.int/bookorders>.
- Shao, Y., Forster, S. C., Tsaliki, E., Vervier, K., Strang, A., Simpson, N., Kumar, N., Stares, M. D., Rodger, A., Brocklehurst, P., Field, N., & Lawley, T. D. (2019). Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature*, 574(7776), 117–121. <https://doi.org/10.1038/s41586-019-1560-1>
- Sharma, V., Leight, J., Giroux, N., AbdulAziz, F., & Nyqvist, M. B. (2019). “That’s a woman’s problem”: a qualitative analysis to understand male involvement in maternal and newborn health in Jigawa state, northern Nigeria. *Reproductive Health*, 16(1), 143. <https://doi.org/10.1186/s12978-019-0808-4>
- Simane-Netshisaulu, K. G., Lebeso, R. T., & Tulelo, P. M. (2022). Provision of neonatal care: An African indigenous perspective. In *Working with indigenous knowledge: Strategies for health professionals* (pp. 107–125). AOSIS Books. <https://doi.org/10.4102/aosis.2022.BK296.07>
- Soler, N., & Forterre, P. (2020). Vesiduction: the fourth way of <scp>HGT</scp>. *Environmental Microbiology*, 22(7), 2457–2460. <https://doi.org/10.1111/1462-2920.15056>

- Sommer, M. O. A., Dantas, G., & Church, G. M. (2009). Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora. *Science*, 325(5944), 1128–1131. <https://doi.org/10.1126/science.1176950>
- Stearns, J. C., Simioni, J., Gunn, E., McDonald, H., Holloway, A. C., Thabane, L., Mousseau, A., Schertzer, J. D., Ratcliffe, E. M., Rossi, L., Surette, M. G., Morrison, K. M., & Hutton, E. K. (2017). Intrapartum antibiotics for GBS prophylaxis alter colonization patterns in the early infant gut microbiome of low risk infants. *Scientific Reports*, 7(1), 16527. <https://doi.org/10.1038/s41598-017-16606-9>
- Stewart, C. J., Ajami, N. J., O'Brien, J. L., Hutchinson, D. S., Smith, D. P., Wong, M. C., Ross, M. C., Lloyd, R. E., Doddapaneni, H., Metcalf, G. A., Muzny, D., Gibbs, R. A., Vatanen, T., Huttenhower, C., Xavier, R. J., Rewers, M., Hagopian, W., Toppari, J., Ziegler, A.-G., ... Petrosino, J. F. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature*, 562(7728), 583–588. <https://doi.org/10.1038/s41586-018-0617-x>
- Stinson, L. F., Keelan, J. A., & Payne, M. S. (2018). Comparison of Meconium DNA Extraction Methods for Use in Microbiome Studies. *Frontiers in Microbiology*, 9. <https://doi.org/10.3389/fmicb.2018.00270>
- Street, R., & Ababa, A. (2021). *AFRICAN ANTIBIOTIC TREATMENT GUIDELINES FOR COMMON BACTERIAL INFECTIONS AND SYNDROMES*. www.africacdc.organdwww.cddep.org.
- Sudjai, A., Suriye, K., Kruasuwan, W., Jenjaroenpun, P., Foongladda, S., & Wongsurawat, T. (2022). Long-Read 16S rRNA Amplicon and Metagenomic Data of Swine Feed-Additive Probiotics Product. *Microbiology Resource Announcements*, 11(9). <https://doi.org/10.1128/mra.00397-22>
- Tamburini, S., Shen, N., Wu, H. C., & Clemente, J. C. (2016). The microbiome in early life: implications for health outcomes. *Nature Medicine*, 22(7), 713–722. <https://doi.org/10.1038/nm.4142>
- Tanaka, M., & Nakayama, J. (2017). Development of the gut microbiota in infancy and its impact on health in later life. *Allergology International*, 66(4), 515–522. <https://doi.org/10.1016/j.alit.2017.07.010>

- Tesema, G. A., & Seifu, B. L. (2023). Factors associated with mother's healthcare-seeking behavior for symptoms of acute respiratory infection in under-five children in sub-Saharan Africa: a multilevel robust Poisson regression modelling. *BMC Health Services Research*, 23(1), 1061. <https://doi.org/10.1186/s12913-023-10065-x>
- Thaiss, C. A., Zmora, N., Levy, M., & Elinav, E. (2016). The microbiome and innate immunity. *Nature*, 535(7610), 65–74. <https://doi.org/10.1038/nature18847>
- Thänert, R., Sawhney, S. S., Schwartz, D. J., & Dantas, G. (2022). The resistance within: Antibiotic disruption of the gut microbiome and resistome dynamics in infancy. *Cell Host & Microbe*, 30(5), 675–683. <https://doi.org/10.1016/j.chom.2022.03.013>
- Theophilus, R. J., & Taft, D. H. (2023). Antimicrobial Resistance Genes (ARGs), the Gut Microbiome, and Infant Nutrition. *Nutrients*, 15(14), 3177. <https://doi.org/10.3390/nu15143177>
- Thomas, S. P., Denizer, E., Zuffa, S., Best, B. M., Bode, L., Chambers, C. D., Dorrestein, P. C., Liu, G. Y., Momper, J. D., Nizet, V., Tsunoda, S. M., & Tremoulet, A. H. (2023). Transfer of antibiotics and their metabolites in human milk: Implications for infant health and microbiota. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, 43(5), 442–451. <https://doi.org/10.1002/phar.2732>
- Thomson, K. M., Dyer, C., Liu, F., Sands, K., Portal, E., Carvalho, M. J., Barrell, M., Boostrom, I., Dunachie, S., Farzana, R., Ferreira, A., Frayne, F., Hassan, B., Jones, E., Jones, L., Mathias, J., Milton, R., Rees, J., Chan, G. J., ... Walsh, T. R. (2021). Effects of antibiotic resistance, drug target attainment, bacterial pathogenicity and virulence, and antibiotic access and affordability on outcomes in neonatal sepsis: an international microbiology and drug evaluation prospective substudy (BARNARDS). *The Lancet. Infectious Diseases*, 21(12), 1677–1688. [https://doi.org/10.1016/S1473-3099\(21\)00050-5](https://doi.org/10.1016/S1473-3099(21)00050-5)
- Tigoi, C., Bourdon, C., Ngari, M., Musyimi, R., Timbwa, M., Mwaringa, S., Ngao, N., Maronga, C., Mburu, M., Ndirangu, A., Arif, F., Kazi, Z., Ejaz, M. S., Saleem, A. F., Singa, B. O., Mupere, E., Shahid, A. S. M. S. Bin, Khan, A. F., Chisti, M. J., ... Berkley, J. A. (2025). Antimicrobial Usage Among Acutely Ill Hospitalized Children Aged 2–23 Months in Sub-Saharan Africa and South Asia. *Open Forum Infectious Diseases*, 12(9). <https://doi.org/10.1093/ofid/ofaf487>

- Trosvik, P., Noordzij, H. T., & de Muinck, E. J. (2024). Antibiotic resistance gene dynamics in the commensal infant gut microbiome over the first year of life. *Scientific Reports*, 14(1), 18701. <https://doi.org/10.1038/s41598-024-69275-w>
- Tun, H. M., Konya, T., Takaro, T. K., Brook, J. R., Chari, R., Field, C. J., Guttman, D. S., Becker, A. B., Mandhane, P. J., Turvey, S. E., Subbarao, P., Sears, M. R., Scott, J. A., & Kozyrskyj, A. L. (2017). Exposure to household furry pets influences the gut microbiota of infants at 3–4 months following various birth scenarios. *Microbiome*, 5(1), 40. <https://doi.org/10.1186/s40168-017-0254-x>
- Turunen, J., Tejesvi, M. V., Paalanne, N., Pokka, T., Amatya, S. B., Mishra, S., Kaisanlahti, A., Reunanen, J., & Tapiainen, T. (2024). Investigating prenatal and perinatal factors on meconium microbiota: a systematic review and cohort study. *Pediatric Research*, 95(1), 135–145. <https://doi.org/10.1038/s41390-023-02783-z>
- United Nations Children’s Fund (UNICEF), & World Health Organization (WHO). (2023). *WHO/UNICEF JOINT MONITORING PROGRAMME FOR WATER SUPPLY, SANITATION AND HYGIENE*. <https://washdata.org>
- Van Uffelen, A., Posadas, A., Roosens, N. H. C., Marchal, K., De Keersmaecker, S. C. J., & Vanneste, K. (2024). Benchmarking bacterial taxonomic classification using nanopore metagenomics data of several mock communities. *Scientific Data*, 11(1), 864. <https://doi.org/10.1038/s41597-024-03672-8>
- Victor, C., Kupoluyi, J. A., Oyinlola, F. F., & Sule, V. O. (2025). Prevalence and factors associated with water, sanitation and hygiene (WASH) facilities deprivation among children in Nigeria. *BMC Pediatrics*, 25(1), 102. <https://doi.org/10.1186/s12887-025-05463-5>
- von Wintersdorff, C. J., Wolffs, P. F., Savelkoul, P. H., Nijssen, R. R., Lau, S., Gerhold, K., Hamelmann, E., & Penders, J. (2016). The gut resistome is highly dynamic during the first months of life. *Future Microbiology*, 11(4), 501–510. <https://doi.org/10.2217/fmb.15.154>
- Wampach, L., Heintz-Buschart, A., Fritz, J. V., Ramiro-Garcia, J., Habier, J., Herold, M., Narayanasamy, S., Kaysen, A., Hogan, A. H., Bindl, L., Bottu, J., Halder, R., Sjöqvist, C., May, P., Andersson, A. F., de Beaufort, C., & Wilmes, P. (2018). Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential. *Nature Communications*, 9(1), 5091. <https://doi.org/10.1038/s41467-018-07631-x>

- Wasswa, F. B., Kassaza, K., Nielsen, K., & Bazira, J. (2022). MinION Whole-Genome Sequencing in Resource-Limited Settings: Challenges and Opportunities. *Current Clinical Microbiology Reports*, 9(4), 52–59. <https://doi.org/10.1007/s40588-022-00183-1>
- Wesolowska-Andersen, A., Bahl, M. I., Carvalho, V., Kristiansen, K., Sicheritz-Pontén, T., Gupta, R., & Licht, T. R. (2014). Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome*, 2(1), 19. <https://doi.org/10.1186/2049-2618-2-19>
- WHO. (2013). *Pocket book of hospital care for children: guidelines for the management of common illnesses with limited resources* (2nd ed.). Geneva.
- WHO. (2021). *2021 AWaRe classification*.
- WHO, & UNICEF. (2024). *Global breastfeeding scorecard | Global Breastfeeding Collective*. <https://www.globalbreastfeedingcollective.org/global-breastfeeding-scorecard>
- Wild, C. P. (2012). The exposome: from concept to utility. *International Journal of Epidemiology*, 41(1), 24–32. <https://doi.org/10.1093/ije/dyr236>
- Wiley, K. S., Gregg, A. M., Fox, M. M., Lagishetty, V., Sandman, C. A., Jacobs, J. P., & Glynn, L. M. (2024). Contact with caregivers is associated with composition of the infant gastrointestinal microbiome in the first 6 months of life. *American Journal of Biological Anthropology*, 183(4). <https://doi.org/10.1002/ajpa.24858>
- Woerther, P.-L., Angebault, C., Jacquier, H., Clermont, O., El Mniai, A., Moreau, B., Djossou, F., Peroz, G., Catzefflis, F., Denamur, E., & Andremont, A. (2013). Characterization of Fecal Extended-Spectrum- β -Lactamase-Producing *Escherichia coli* in a Remote Community during a Long Time Period. *Antimicrobial Agents and Chemotherapy*, 57(10), 5060–5066. <https://doi.org/10.1128/AAC.00848-13>
- Wongsurawat, T., Nakagawa, M., Atiq, O., Coleman, H. N., Jenjaroenpun, P., Allred, J. I., Trammel, A., Puengrang, P., Ussery, D. W., & Nookaew, I. (2019). An assessment of Oxford Nanopore sequencing for human gut metagenome profiling: A pilot study of head and neck cancer patients. *Journal of Microbiological Methods*, 166, 105739. <https://doi.org/10.1016/j.mimet.2019.105739>

- World Health Organization (WHO), & the United Nations Children's Fund (UNICEF). (2024). *Water, sanitation, hygiene, environmental cleaning and waste management in health care facilities: 2023 data update and special focus on primary health care*.
- World Bank. (2024). *Out-of-pocket expenditure (% of current health expenditure) – Nigeria*.
World Development Indicators.
<https://data.worldbank.org/indicator/SH.XPD.OOPC.CH.ZS?locations=NG>
- World Health Organization. (2013). *WHO recommendations on Postnatal care of the mother and newborn 2013*.
- World Health Organization. (2023). *Antimicrobial Resistance: Fact Sheet*.
<https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>
- World Health Organization (WHO), & the United Nations Children's Fund (UNICEF). (2021). *Progress on Household Drinking Water, Sanitation and Hygiene 2000-2020 : Five Years into the SDGs*. World Health Organization.
- Wright, G. D. (2007). The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Reviews Microbiology*, 5(3), 175–186. <https://doi.org/10.1038/nrmicro1614>
- Wu, X., Guo, R., Fan, Y., Chen, S., Zheng, W., Shu, X., Chen, B., Li, X., Xu, T., Shi, L., Chen, L., Shan, L., Zhu, Z., Tao, E., & Jiang, M. (2024). Dynamic impact of delivery modes on gut microbiota in preterm infants hospitalized during the initial 4 weeks of life. *International Journal of Medical Microbiology*, 315, 151621. <https://doi.org/10.1016/j.ijmm.2024.151621>
- Xiao, L., & Zhao, F. (2023). Microbial transmission, colonisation and succession: from pregnancy to infancy. *Gut*, 72(4), 772–786. <https://doi.org/10.1136/gutjnl-2022-328970>
- Yamin, D., Uskoković, V., Wakil, A., Goni, M., Shamsuddin, S., Mustafa, F., Alfouzan, W., Alissa, M., Alshengeti, A., Almaghrabi, R., Fares, M., Garout, M., Al Kaabi, N., Alshehri, A., Ali, H., Rabaan, A., Aldubisi, F., Yean, C., & Yusof, N. (2023). Current and Future Technologies for the Detection of Antibiotic-Resistant Bacteria. *Diagnostics*, 13(20), 3246. <https://doi.org/10.3390/diagnostics13203246>
- Yassour, M., Vatanen, T., Siljander, H., Hämäläinen, A.-M., Härkönen, T., Ryhänen, S. J., Franzosa, E. A., Vlamakis, H., Huttenhower, C., Gevers, D., Lander, E. S., Knip, M., & Xavier, R. J. (2016). Natural history of the infant gut microbiome and impact of antibiotic

treatment on bacterial strain diversity and stability. *Science Translational Medicine*, 8(343). <https://doi.org/10.1126/scitranslmed.aad0917>

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

Zaidi, A. K., Huskins, W. C., Thaver, D., Bhutta, Z. A., Abbas, Z., & Goldmann, D. A. (2005). Hospital-acquired neonatal infections in developing countries. *The Lancet*, 365(9465), 1175–1188. [https://doi.org/10.1016/S0140-6736\(05\)71881-X](https://doi.org/10.1016/S0140-6736(05)71881-X)

Zewdie, S., Kassa, A. A., Yayehrad, A. T., Bizuneh, M. M., Ayenew, W., Zewudie, M., Mulat, S., Bitew, B. E., Zewudie, S., Meharie, B. G., Tesfaye, T. C., Zerga, A. A., Ayele, F. Y., Toleha, H. N., Workineh, B. D., & Bayked, E. M. (2024). Non-prescription antibiotic use and its predictors among children in low- and middle-income countries: a systematic review and meta-analysis. *Italian Journal of Pediatrics*, 50(1), 260. <https://doi.org/10.1186/s13052-024-01808-5>

Appendix A.1:

[illegible]

Supplementary Figure 2.1: Take-home household event grid used to support caregiver recall between study visits.

Caregivers recorded dates of key infant events and exposures using predefined codes. The grid was reviewed at each follow-up visit and reconciled with CRF responses.

Supplementary data 1:

[GitHub - ldenies007/Long_reads_microbiome_AMR: Pipeline to analyze ONT long read data \(Taxonomy, AMR\)](#)

Supplementary data 2: https://github.com/Chyakpulu/PhD_Thesis

Appendix A. 2

Supplementary data 1: Author contributions

Chinenye Akpulu: Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Aisha Mukkadas: Writing – review & editing, Project administration, Methodology, Investigation.

Khalid Rashida: Writing – review & editing, Project administration, Methodology, Investigation.

Portal Edward: Writing – review & editing, Validation, Methodology.

Lankapalli Aditya: Writing – review & editing, Visualization, Validation, Software, Resources, Formal analysis, Data curation.

Kate Cook: Writing – review & editing, Validation, Methodology.

Rida Toufiq: Writing – review & editing, Validation, Methodology.

Aliyu Aminu: Writing – review & editing, Visualization, Investigation.

Safiya Gambo: Writing – review & editing, Validation, Investigation.

Kenneth Iregbu: Writing – review & editing, Validation, Investigation.

Kathryn Thomson: Writing – review & editing, Validation, Supervision, Investigation. Timothy

Walsh: Writing – review & editing, Validation, Supervision, Investigation, Funding acquisition.

Kirsty Sands: Writing – review & editing, Validation, Supervision, Methodology, Investigation, Conceptualization.

Appendix A. 2

Supplementary Table 3.1: Per-condition pairwise comparisons of DNA extraction kit yield (ng) across preservation and storage conditions

Condition	Friedman χ^2 (p)	PS–SM: HL ng (95% CI), p (adj)	ZB–SM: HL ng (95% CI), p (adj)	PS–ZB: HL ng (95% CI), p (adj)
Day 0 (no preserv.)	7.90 (0.0193)	1256.0 (512.0–2000.0), 0.0625 (0.1875)	1097.7 (195.3–2000.0), 0.0625 (0.1875)	228.8 (–1488.0–1714.3), 0.6250 (1.000)
Day 1 (Shield –80 °C)	7.30 (0.0260)	48.5 (2.0–517.3), 0.0625 (0.1875)	129.4 (0.0–263.3), 0.1250 (0.3750)	8.0 (–252.7–291.3), 0.4375 (1.000)
Day 1 (Shield 4 °C)	7.60 (0.0224)	75.7 (2.1–490.7), 0.0625 (0.1875)	114.5 (2.3–226.7), 0.0625 (0.1875)	17.6 (–212.4–302.3), 1.000 (1.000)
Day 1 (Shield RT)	6.30 (0.0429)	93.3 (0.0–618.7), 0.1250 (0.3750)	110.9 (2.1–237.0), 0.0625 (0.1875)	17.0 (–225.2–459.7), 1.000 (1.000)
Day 2 (Shield RT)	5.20 (0.0743)	33.2 (0.0–91.2), 0.2500 (0.7500)	102.5 (2.1–203.0), 0.0625 (0.1875)	–84.6 (–182.9–8.0), 0.6250 (1.000)
Week 6 (Swab 4 °C)	8.40 (0.0150)	675.3 (56.8–1690.0), 0.0625 (0.1875)	588.5 (73.0–1486.7), 0.0625 (0.1875)	93.6 (–16.2–203.3), 0.1250 (0.3750)

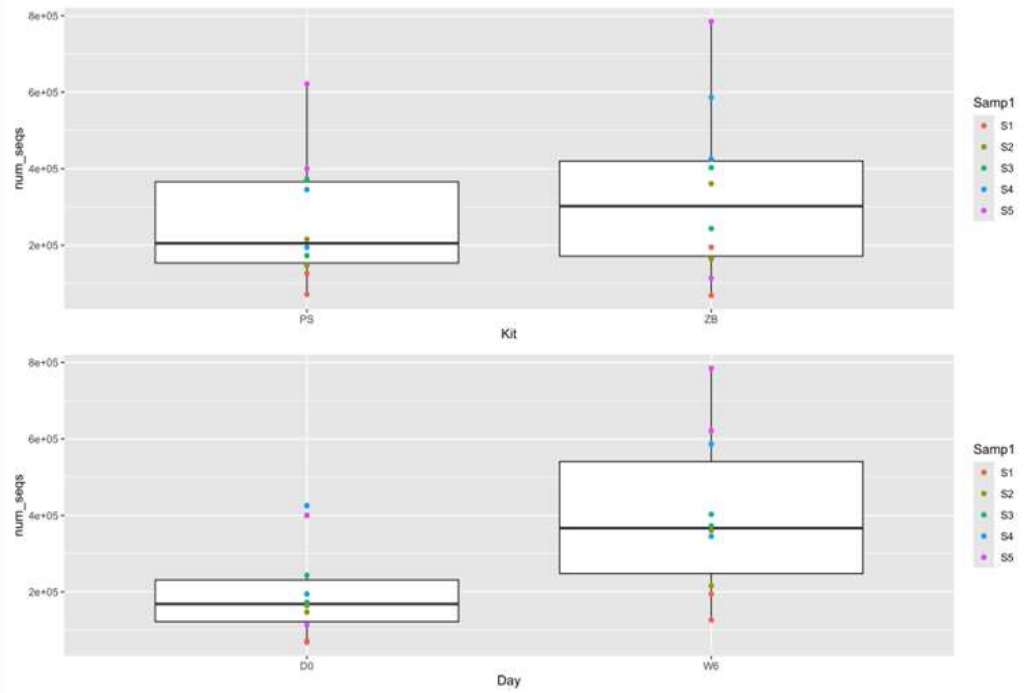
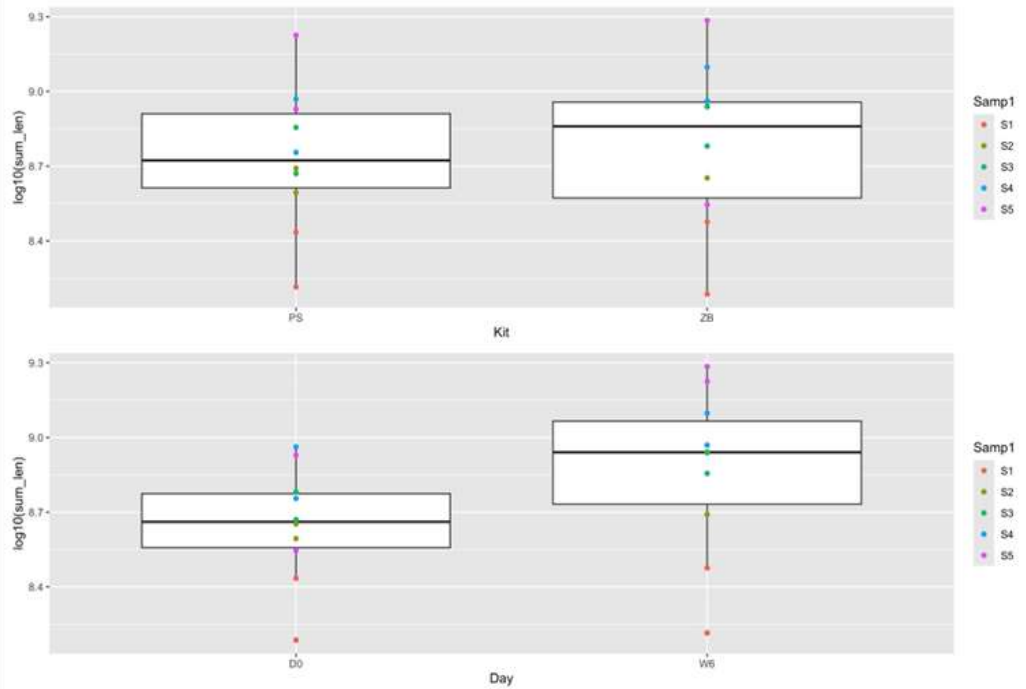
For each condition, an overall Friedman test assessed within-infant differences in yield between kits (PS, ZB, SM), followed by paired Wilcoxon signed-rank contrasts with Holmadjusted p-values. Pairwise effects are reported as Hodges–Lehmann (HL) median differences in yield (ng) with 95% confidence intervals for PS–SM, ZB–SM, and PS–ZB

Appendix A. 2

Supplementary Table 3.2: DNA yield by extraction kit across preservation and storage conditions

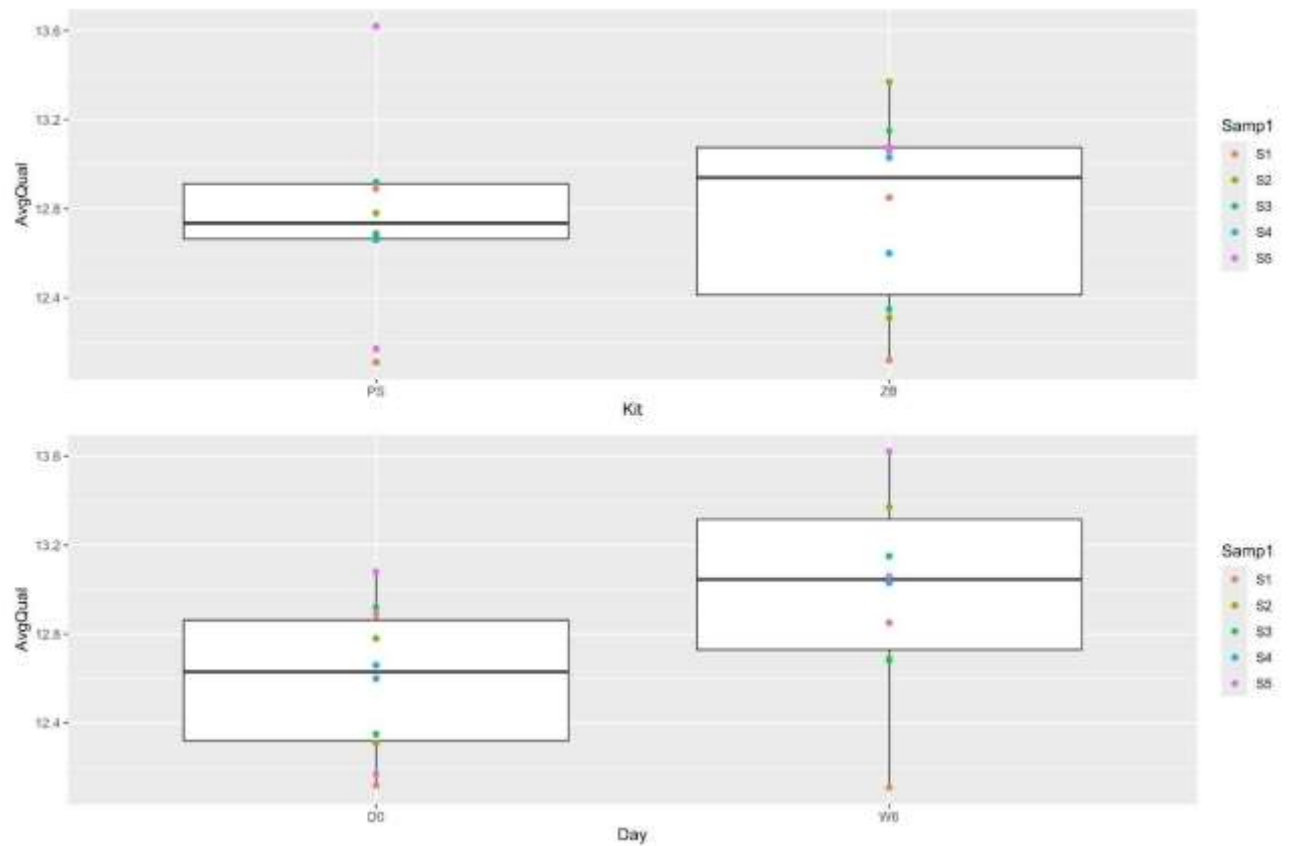
Condition	PS	SM	ZB
Day 0 (no preserv.)	742.7 (561.3-2000.0)	0.0 (0.0-0.0)	285.7 (285.0-2000.0)
Day 1 (Shield 4°C)	75.7 (14.2-123.0)	0.0 (0.0-0.0)	133.3 (30.2-188.3)
Day 1 (Shield RT)	45.6 (10.7-86.3)	0.0 (0.0-0.0)	78.3 (32.7-226.0)
Day 1 (Shield -80°C)	93.3 (11.8-100.0)	0.0 (0.0-0.0)	110.9 (48.5-159.0)
Day 2 (Shield RT)	20.1 (0.0-66.4)	0.0 (0.0-0.0)	83.2 (62.9-172.7)
Week 6 (swab 4°C)	426.3 (370.7-980.0)	0.0 (0.0-0.0)	351.7 (311.3-865.7)

Values are reported as median yield (ng) with interquartile range (IQR) for each kit (PS, SM, ZB) within each sampling condition/time point (Day 0 no preservative; Day 1 Shield stored at 4°C, room temperature, or -80°C; Day 2 Shield at room temperature; Week 6 swab stored at 4°C).

A**B**

Supplementary Figure 3.1: Sequencing throughput summaries for the sequenced libraries (PS and ZB extracts at Day 0 and Week 6).

Boxplots show (A) the number of reads generated per library (read count; num_seqs) and (B) total sequencing yield expressed as log10 total bases (log10(sum_len)), each summarised by extraction kit (PS vs ZB) and by time point (D0 vs W6). Coloured points denote individual infants (S1–S5).



Supplementary Figure 3.2: Mean read quality (Q score) for the sequenced libraries (PS and ZB extracts at Day 0 and Week 6).

Boxplots summarise per-library average read quality (AvgQual) by extraction kit (top; PS vs ZB) and by time point (bottom; D0 vs W6). Coloured points denote individual infants (S1–S5)

Appendix A. 3

Supplementary data 1: Data readiness and QC outcomes

This section summarises sequencing throughput and QC outcomes for infant stool metagenomes and controls, to justify the analysis-ready dataset used for microbiome and resistome trajectory analyses.

Read quality (before/after fastp cleaning)

Raw sequencing output showed the expected contrast between infant stool samples, positive controls, and negative controls. Infant stool libraries typically yielded high numbers of long reads, whereas extraction/library negative controls produced low-yield, short-fragment profiles consistent with background signal (Figure S5.1). After fastp cleaning, infant stool samples retained the great majority of reads, with only ~1–2% removed per sample on average. In contrast, negative controls lost a larger fraction of reads (around ~12% on average), consistent with higher proportions of short, low-complexity, or poor-quality fragments.

Read-length structure was preserved in infant stool samples

Read-length N50 remained approximately ~5 kb before and after fastp, and median average read length was essentially unchanged (Figure S5.2). Base-level quality improved modestly after trimming, consistent with removal of low-quality tails. Negative controls remained dominated by very short reads (N50 on the order of ~100 bp) and remained clearly separable from stool samples after cleaning. Libraries showing heavy loss during cleaning (>20% reads removed) were rare and were confined to negative controls.

Host-DNA burden (before/after host removal)

Host-DNA burden was generally low in infant stool libraries. The median fraction of reads removed as host was approximately 2.5%, and three-quarters of samples were below ~7% host removal. A subset of libraries (57/557; 10.2%) met a “high-host” criterion of $\geq 20\%$ host removal, distributed across 39 of 90 infants. A very small number of libraries (6/557; ~1%) were extremely host-dominated ($\geq 70\%$ host reads removed). These extreme cases occurred across both early and later timepoints and are consistent with intermittent collection of very low-biomass material (e.g., mucus/epithelial-rich specimens rather than formed stool).

Despite host removal, most infant stool libraries retained substantial microbial sequencing depth. The median number of non-host reads retained was on the order of $4\text{--}5 \times 10^5$ reads per library, and three-quarters of samples retained more than $\sim 2 \times 10^5$ non-host reads. Only a minority fell below 10^5 retained reads ($\sim 9\%$), and only 2 of 557 fell below 10^4 . The lowest yield library retained after host removal still contained 5,508 non-host reads. These patterns are summarised in Figures S5.3A-B.

Negatives-informed background assessment and decontamination

Reagent and handling background was assessed using extraction and library blanks. In principle, decontamination can be performed per flowcell or extraction batch using that batch's own negative control; however, several sequencing runs did not include a blank. Restricting subtraction only to runs with a blank would therefore have treated some infants' libraries differently from others, introducing an artificial batch effect. To avoid this, all available negatives across runs were pooled to define a composite background profile used to identify low-level background signal across libraries.

Three complementary approaches were applied to this pooled background: (i) microDecon (proportional subtraction under non-negativity constraints); (ii) decontam in prevalence mode (taxa more frequent in negatives than in true samples); and (iii) a prevalence/blank-dominance screen. Overlap between approaches is shown in Figure S5.4. Taxa consistently identified as background were predominantly environmental and biologically implausible as neonatal gut microbiota (e.g., plant-, soil-, and aquatic-associated genera, plus low-abundance fungi/yeasts). A small number of clinically plausible gut-associated taxa flagged by prevalence criteria were reviewed and retained rather than automatically removed, to avoid removing genuine early-life colonisation signal.

To minimise over-correction, the microDecon-adjusted (contaminant-subtracted) table was used as the primary correction output for downstream analyses, while decontam and rule-based screening contributed contaminant flags for annotation and sensitivity checks. The impact of decontamination on within-sample diversity was assessed using Shannon diversity. Within flowcells, Shannon values typically decreased only slightly after decontamination (on the order of $\sim 0.1\text{--}0.2$ units), and the relative ordering of samples by diversity remained stable (Figure S5.5). No sequencing run showed a large downward shift consistent with loss of dominant gut taxa.

Conclusion

Overall, these outcomes indicate that the infant stool samples retained enough high-quality microbial data for detailed analysis. The controls performed as expected, and the removal of background noise produced minor, logical adjustments. This confirms that the dataset is robust and suitable for the longitudinal analysis of the microbiome and resistome presented in this thesis.

The credibility of any downstream microbiome and resistome inference depend not just on which exposures infants experienced, but on whether the metagenomic data generated under these field conditions are analytically usable. The quality check in section 4.4.9 indicates that they are, with a few borderline samples. After removal of human reads, most infant stool libraries retained substantial microbial reads to support downstream analysis. This pattern is consistent with what is known about neonatal stool biology: the earliest infant stool can be extremely low biomass (Fierer et al., 2025; Kennedy et al., 2023), dominated by host cells and mucus, and technically fragile (Gallo et al., 2023; Turunen et al., 2024). Notably, high-host libraries were not confined to a single infant or an isolated early time point. Instead, they recurred across multiple infants and were also observed at later timepoints. This suggests that intermittent, low-biomass, host-rich samples are an authentic biological feature of the cohort rather than a one-off technical failure. Accordingly, such libraries can be explicitly identified, flagged, and included in sensitivity analyses, rather than silently influencing inference.

Background contamination was systematically characterised and controlled in accordance with published recommendations for low-biomass metagenomic studies (Eisenhofer et al., 2019; Salter et al., 2014). Salter et al. (2014) and others have shown that, when inputs are low, kit- and handling-derived taxa (often environmental or skin-associated) can appear at low abundance in true samples and even dominate apparent signal (Eisenhofer et al., 2019; Salter et al., 2014). In this cohort, extraction and library blanks were used to model that background, and a conservative subtraction strategy was applied. The taxa flagged as background were not typical neonatal gut microbiota but rather plant-, soil-, and aquatic-associated bacteria, plus low-abundance filamentous fungi and yeasts, which is in line with reagent and handling noise (Eisenhofer et al., 2019; Salter et al., 2014). Clinically relevant gut organisms such as *Enterococcus faecalis* and *Salmonella enterica* were not discarded solely because low-level signal was also detectable in negative controls. Instead, these taxa were retained in the analytical dataset but explicitly flagged, reflecting both their plausibility as early-life gut

colonisers and the recognised potential for low-level barcode crosstalk in multiplexed sequencing runs (Kennedy et al., 2023). Following background subtraction, within-sample diversity decreased only slightly (on the order of ~0.1–0.2 Shannon units), and the relative ordering of sequencing runs by diversity remained stable. This pattern suggests that the procedure primarily removed batch-associated background noise rather than artificially reshaping the underlying microbial community. Consequently, the downstream taxonomic and resistome analyses presented in later chapters can be interpreted with confidence, even though these metagenomes were generated in a setting of intermittent power, variable available stool volume, and incomplete negative controls per run, practical constraints that are seldom reported in high-income neonatal cohorts but are intrinsic to microbiome work in low-resource hospitals.

In summary, this cohort combines dense early-life sampling with detailed exposure data in a high-contact, high-environmental-load urban setting, under real operational constraints. The QC analysis in the second half of the chapter indicates that despite host-rich, low-biomass material in a subset of samples and variable low-level background across flowcells, the majority of libraries retain sufficient non-host depth after host removal, and contaminant subtraction behaves in a controlled, biologically plausible way. This supports the use of these metagenomic data, including downstream taxonomic and resistome analyses, in subsequent chapters.

References

- Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., & Weyrich, L. S. (2019). Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends in Microbiology*, 27(2), 105–117. <https://doi.org/10.1016/j.tim.2018.11.003>
- Fierer, N., Leung, P. M., Lappan, R., Eisenhofer, R., Ricci, F., Holland, S. I., Dragone, N., Blackall, L. L., Dong, X., Dorador, C., Ferrari, B. C., Goordial, J., Holmes, S. P., Inagaki, F., Korem, T., Li, S. S., Makhalanyane, T. P., Metcalf, J. L., Nagarajan, N., ... Greening, C. (2025). Guidelines for preventing and reporting contamination in lowbiomass microbiome studies. *Nature Microbiology*, 10(7), 1570–1580. <https://doi.org/10.1038/s41564-025-02035-2>
- Gallo, D. M., Romero, R., Bosco, M., Gotsch, F., Jaiman, S., Jung, E., Suksai, M., Ramón y Cajal, C. L., Yoon, B. H., & Chaiworapongsa, T. (2023). Meconium-stained amniotic fluid. *American Journal of Obstetrics and Gynecology*, 228(5), S1158–S1178. <https://doi.org/10.1016/j.ajog.2022.11.1283>
- Kennedy, K. M., de Goffau, M. C., Perez-Muñoz, M. E., Arrieta, M.-C., Bäckhed, F., Bork, P., Braun, T., Bushman, F. D., Dore, J., de Vos, W. M., Earl, A. M., Eisen, J. A., Elovitz,

- M. A., Ganal-Vonarburg, S. C., Gänzle, M. G., Garrett, W. S., Hall, L. J., Hornef, M. W., Huttenhower, C., ... Walter, J. (2023). Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. *Nature*, *613*(7945), 639–649. <https://doi.org/10.1038/s41586-022-05546-8>
- Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., & Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*(1), 87. <https://doi.org/10.1186/s12915-014-0087-z>
- Turunen, J., Tejesvi, M. V., Paalanne, N., Pokka, T., Amatya, S. B., Mishra, S., Kaisanlahti, A., Reunanen, J., & Tapiainen, T. (2024). Investigating prenatal and perinatal factors on meconium microbiota: a systematic review and cohort study. *Pediatric Research*, *95*(1), 135–145. <https://doi.org/10.1038/s41390-023-02783-z>

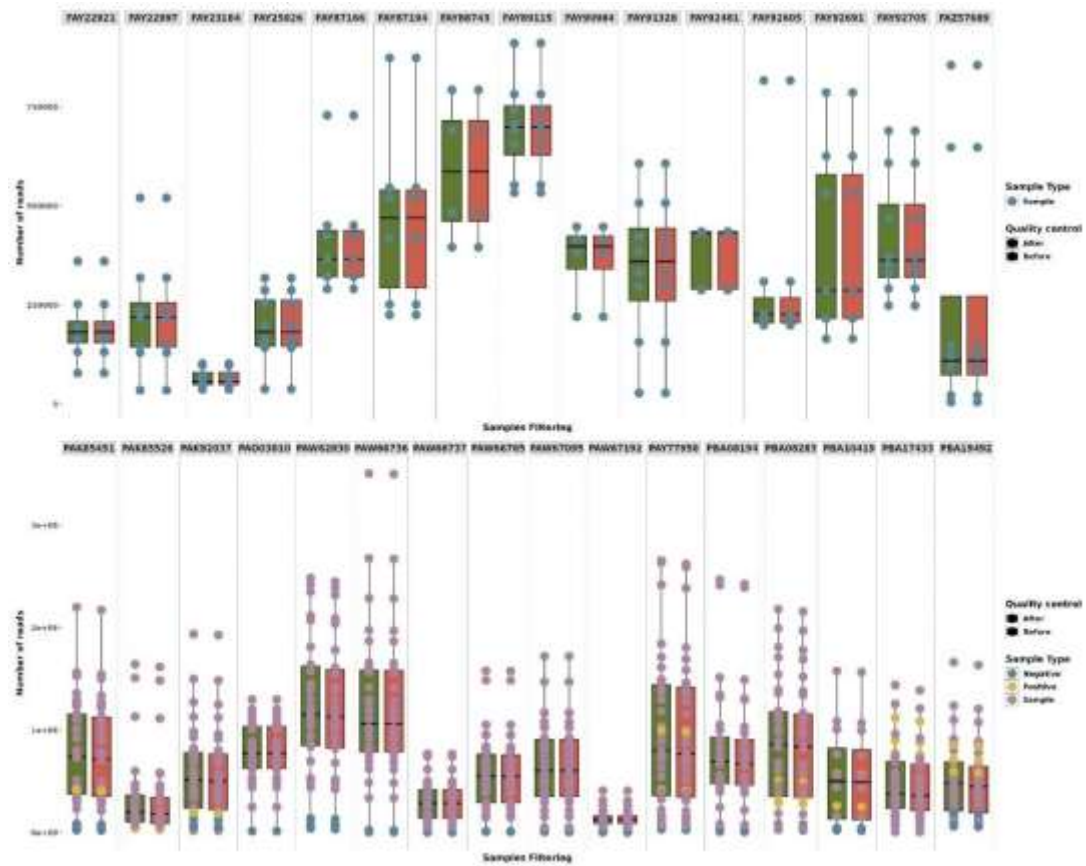


Figure S5.1: Read-count distributions before and after quality control across infant stool libraries and sequencing controls

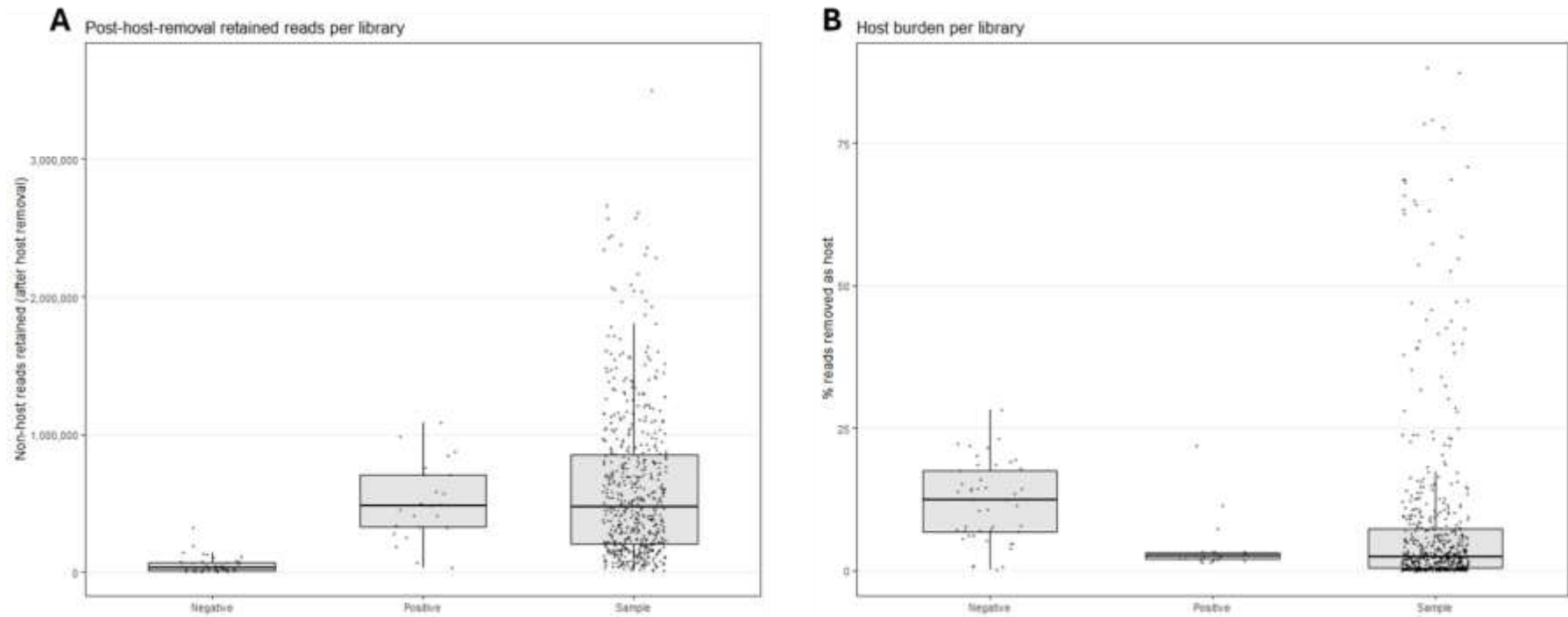


Figure S5.3: Host-read removal and retained non-host sequencing depth across infant stool libraries and controls.

(A) Number of non-host reads retained per library after host-DNA removal, shown for negative controls, positive controls, and infant stool samples.
(B) Host-DNA burden per library, expressed as the percentage of reads removed as host.

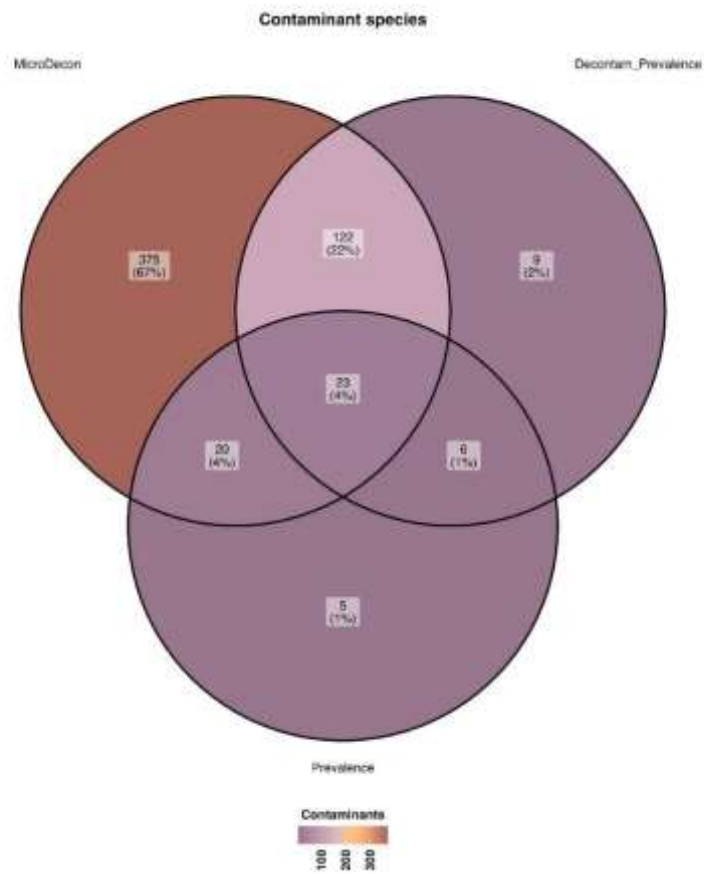


Figure S5.4: Overlap of contaminant taxa identified by three background detection approaches

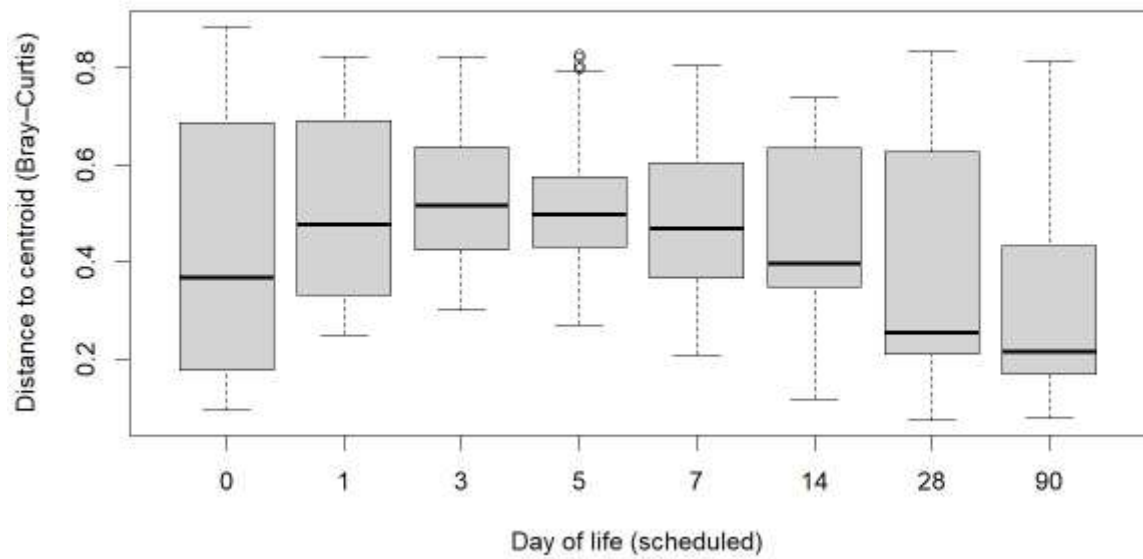


Figure S5.6: Within-timepoint community dispersion across scheduled sampling days

Boxplots show the distribution of each sample's Bray-Curtis distance to the group centroid (i.e., multivariate dispersion) at each scheduled day of life (D0, D1, D3, D5, D7, D14, D28, D90). Lower distances indicate tighter clustering (greater within-day compositional similarity), whereas higher distances indicate greater heterogeneity in community composition among infants sampled at that time point.

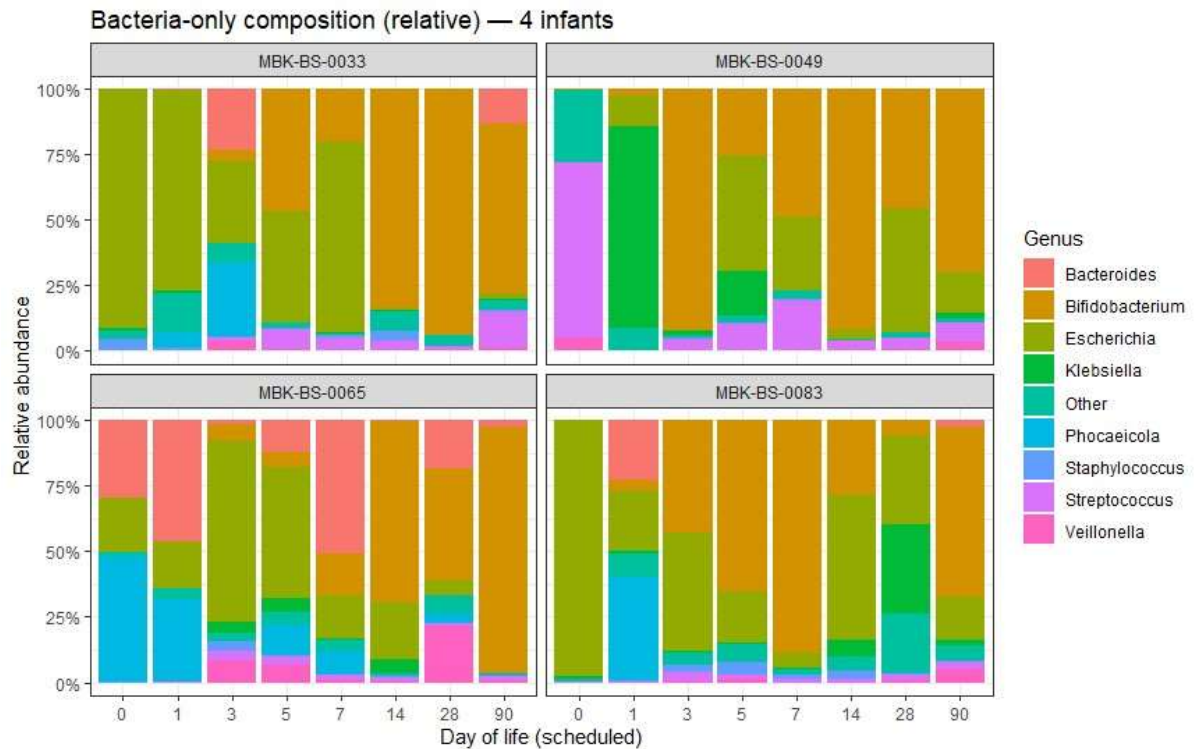
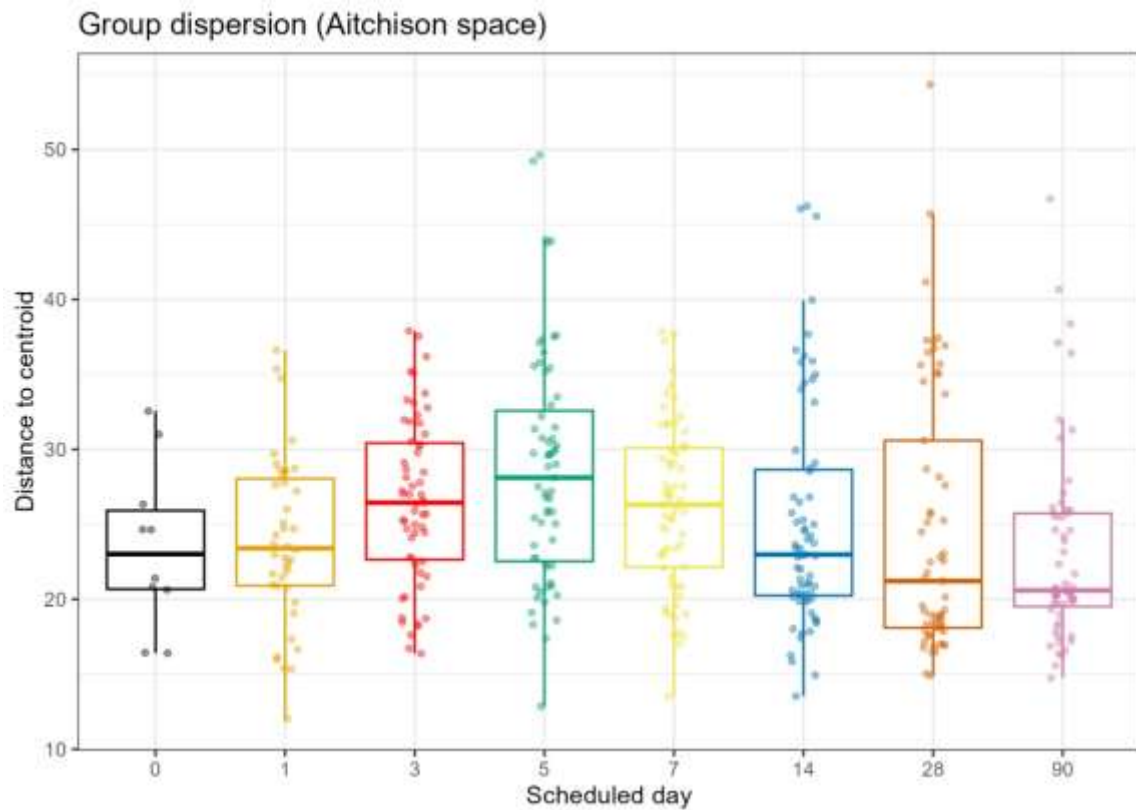


Figure S5.7: Longitudinal gut bacterial genus composition in four representative infants across scheduled sampling days.

Stacked bar charts show relative abundance of dominant bacterial for four infants with complete samples at all time points at Day 0, 1, 3, 5, 7, 14, 28, and 90. Profiles illustrate pronounced early-life inter-individual variability and temporal succession, with early dominance by Enterobacterales-associated genera (e.g., *Escherichia*, *Klebsiella*) in some infants and increasing abundance of beneficial taxa (e.g., *Bifidobacterium*) later in follow-up; remaining low-abundance taxa are grouped as “Other.”



Supplementary Figure S5.8: Within-timepoint community dispersion across scheduled sampling days in Aitchison space.

Boxplots show each sample's distance to the day-specific centroid computed in Aitchison (compositional) space for scheduled days D0, D1, D3, D5, D7, D14, D28, and D90. Points represent individual infant samples. Lower distances indicate tighter clustering (greater within-day compositional similarity), whereas higher distances indicate greater heterogeneity in community composition among infants at that time point.

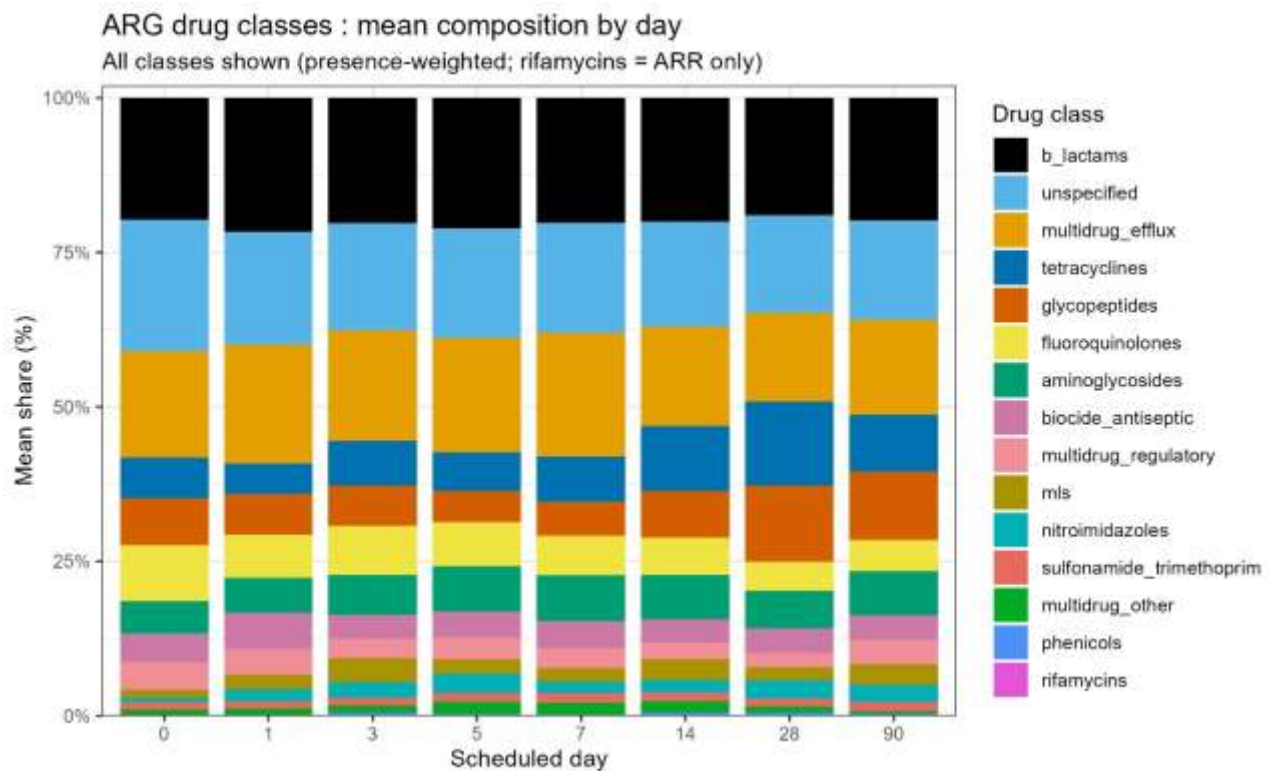


Figure S5.9: Mean antimicrobial resistance gene (ARG) composition by drug class across scheduled sampling days.

Stacked bars show the mean proportional contribution (%) of ARG drug classes at each scheduled day (D0, D1, D3, D5, D7, D14, D28, D90). Class contributions are calculated using a presence-weighted approach so that all detected classes are represented; rifamycin class assignments reflect ARR-family determinants only. The plot summarises how the relative profile of therapy-relevant resistance classes (e.g., β -lactams, aminoglycosides, tetracyclines, fluoroquinolones, and multidrug efflux/regulatory categories) varies over early infancy.

Table S5. 1: Day-specific co-carriage enrichment for pairs of therapy-relevant ARG drug classes.

Day of life	Antibioticclass pair	N	obs	expected	log2_enrich	p	p_adj
0	beta & ag	10	0.77	0.67	0.21	0.20	0.50
1	beta & ag	42	0.83	0.82	0.01	1.00	1.00
3	beta & ag	59	0.84	0.78	0.11	0.00	0.00
5	beta & ag	63	0.93	0.91	0.03	0.06	0.16
7	beta & ag	68	0.95	0.90	0.08	0.00	0.00
14	beta & ag	67	0.82	0.75	0.12	0.00	0.00
28	beta & ag	61	0.72	0.63	0.19	0.00	0.00
90	beta & ag	59	0.84	0.81	0.06	0.05	0.05
0	beta & fq	10	0.86	0.75	0.21	0.10	0.50
1	beta & fq	42	0.87	0.86	0.01	1.00	1.00
3	beta & fq	59	0.89	0.86	0.06	0.02	0.03
5	beta & fq	63	0.98	0.97	0.01	1.00	1.00
7	beta & fq	68	0.92	0.89	0.05	0.01	0.01
14	beta & fq	67	0.80	0.73	0.14	0.00	0.00
28	beta & fq	61	0.75	0.66	0.19	0.00	0.00
90	beta & fq	59	0.83	0.78	0.09	0.00	0.01
0	beta & tet	10	0.59	0.59	0.01	1.00	1.00
1	beta & tet	42	0.85	0.82	0.05	0.14	0.20
3	beta & tet	59	0.91	0.87	0.06	0.01	0.02
5	beta & tet	63	0.98	0.95	0.03	0.02	0.08
7	beta & tet	68	0.93	0.91	0.03	0.09	0.10
14	beta & tet	67	0.86	0.83	0.05	0.04	0.04
28	beta & tet	61	0.83	0.80	0.05	0.10	0.12
90	beta & tet	59	0.93	0.89	0.06	0.01	0.01
0	beta & sxt	10	0.50	0.43	0.21	1.00	1.00
1	beta & sxt	42	0.59	0.57	0.05	0.40	0.51
3	beta & sxt	59	0.66	0.61	0.11	0.01	0.02
5	beta & sxt	63	0.76	0.74	0.03	0.24	0.40
7	beta & sxt	68	0.79	0.75	0.08	0.01	0.01
14	beta & sxt	67	0.67	0.59	0.19	0.00	0.00
28	beta & sxt	61	0.57	0.50	0.19	0.00	0.00
90	beta & sxt	59	0.69	0.65	0.09	0.03	0.03

0	ag & fq	10	0.77	0.67	0.21	0.20	0.50
1	ag & fq	42	0.85	0.76	0.16	0.00	0.00
3	ag & fq	59	0.83	0.78	0.08	0.01	0.02
5	ag & fq	63	0.93	0.92	0.01	1.00	1.00
7	ag & fq	68	0.92	0.89	0.05	0.01	0.01
14	ag & fq	67	0.80	0.71	0.17	0.00	0.00
28	ag & fq	61	0.70	0.54	0.38	0.00	0.00
90	ag & fq	59	0.79	0.71	0.16	0.00	0.00
0	ag & tet	10	0.59	0.53	0.17	1.00	1.00
1	ag & tet	42	0.78	0.72	0.11	0.03	0.06
3	ag & tet	59	0.83	0.79	0.06	0.06	0.06
5	ag & tet	63	0.93	0.91	0.03	0.06	0.16
7	ag & tet	68	0.93	0.91	0.03	0.09	0.10
14	ag & tet	67	0.85	0.82	0.05	0.05	0.05
28	ag & tet	61	0.69	0.65	0.07	0.13	0.14
90	ag & tet	59	0.84	0.81	0.06	0.05	0.05
0	ag & sxt	10	0.50	0.39	0.37	0.44	0.89
1	ag & sxt	42	0.59	0.50	0.24	0.00	0.01
3	ag & sxt	59	0.66	0.55	0.25	0.00	0.00
5	ag & sxt	63	0.76	0.70	0.11	0.00	0.02
7	ag & sxt	68	0.79	0.75	0.08	0.01	0.01
14	ag & sxt	67	0.67	0.58	0.22	0.00	0.00
28	ag & sxt	61	0.57	0.41	0.48	0.00	0.00
90	ag & sxt	59	0.69	0.59	0.22	0.00	0.00
0	fq & tet	10	0.59	0.59	0.01	1.00	1.00
1	fq & tet	42	0.80	0.76	0.08	0.09	0.15
3	fq & tet	59	0.89	0.87	0.03	0.19	0.19
5	fq & tet	63	0.98	0.97	0.01	1.00	1.00
7	fq & tet	68	0.93	0.90	0.05	0.00	0.01
14	fq & tet	67	0.83	0.79	0.08	0.00	0.00
28	fq & tet	61	0.72	0.68	0.07	0.09	0.12
90	fq & tet	59	0.83	0.78	0.09	0.00	0.01
0	fq & sxt	10	0.50	0.43	0.21	1.00	1.00
1	fq & sxt	42	0.59	0.53	0.16	0.02	0.06
3	fq & sxt	59	0.66	0.61	0.11	0.01	0.02
5	fq & sxt	63	0.76	0.75	0.01	1.00	1.00

7	fq & sxt	68	0.78	0.74	0.07	0.03	0.04
14	fq & sxt	67	0.67	0.56	0.27	0.00	0.00
28	fq & sxt	61	0.57	0.43	0.42	0.00	0.00
90	fq & sxt	59	0.68	0.57	0.24	0.00	0.00
0	tet & sxt	10	0.50	0.34	0.55	0.17	0.50
1	tet & sxt	42	0.57	0.50	0.18	0.03	0.06
3	tet & sxt	59	0.66	0.62	0.09	0.04	0.04
5	tet & sxt	63	0.76	0.74	0.03	0.24	0.40
7	tet & sxt	68	0.78	0.76	0.03	0.37	0.37
14	tet & sxt	67	0.67	0.63	0.08	0.03	0.04
28	tet & sxt	61	0.56	0.52	0.09	0.15	0.15
90	tet & sxt	59	0.69	0.65	0.09	0.03	0.03

For each scheduled day of life, the table summarises observed co-occurrence frequency (obs) of each antibiotic-class pair (β -lactams [beta], aminoglycosides [ag], fluoroquinolones [fq], tetracyclines [tet], and sulfonamide/trimethoprim [sxt]) relative to the expected frequency under independence (expected) among samples with complete class calls (N). Enrichment is reported as $\log_2(\text{observed}/\text{expected})$ ($\log_2_enrich$). Statistical significance was assessed using a day-stratified cooccurrence test; p values are reported alongside multiple-testing adjusted values (p_adj)

Table S5.2: Longitudinal transition summary of antibiotic resistance genes (ARGs)

	endpoint	interval	N pos. at start	p11	p10
Transition Summary	Aminoglycosides	D0-D1	6	1	0
Transition Summary	Aminoglycosides	D0-D3	1	1	0
Transition Summary	Aminoglycosides	D0-D5	1	1	0
Transition Summary	Aminoglycosides	D14-D28	43	0.67	0.33
Transition Summary	Aminoglycosides	D14-D90	9	0.89	0.11
Transition Summary	Aminoglycosides	D1-D14	2	1	0
Transition Summary	Aminoglycosides	D1-D3	24	0.83	0.17
Transition Summary	Aminoglycosides	D1-D5	9	1	0
Transition Summary	Aminoglycosides	D1-D7	1	1	0
Transition Summary	Aminoglycosides	D28-D90	34	0.76	0.24
Transition Summary	Aminoglycosides	D3-D14	2	1	0
Transition Summary	Aminoglycosides	D3-D28	1	1	0
Transition Summary	Aminoglycosides	D3-D5	40	0.9	0.1
Transition Summary	Aminoglycosides	D3-D7	6	1	0
Transition Summary	Aminoglycosides	D5-D14	8	0.75	0.25
Transition Summary	Aminoglycosides	D5-D7	51	0.94	0.06
Transition Summary	Aminoglycosides	D7-D14	52	0.87	0.13
Transition Summary	Aminoglycosides	D7-D28	8	1	0
Transition Summary	Aminoglycosides	D7-D90	1	1	0
Transition Summary	Beta-lactams	D0-D1	6	1	0
Transition Summary	Beta-lactams	D0-D3	2	1	0
Transition Summary	Beta-lactams	D0-D5	1	1	0
Transition Summary	Beta-lactams	D14-D28	43	0.91	0.09
Transition Summary	Beta-lactams	D14-D90	9	1	0
Transition Summary	Beta-lactams	D1-D14	2	1	0
Transition Summary	Beta-lactams	D1-D3	30	0.93	0.07

Transition Summary	Beta-lactams	D1-D5	8	1	0
Transition Summary	Beta-lactams	D1-D7	1	1	0
Transition Summary	Beta-lactams	D28-D90	43	0.88	0.12
Transition Summary	Beta-lactams	D3-D14	2	1	0
Transition Summary	Beta-lactams	D3-D28	1	1	0
Transition Summary	Beta-lactams	D3-D5	44	0.98	0.02
Transition Summary	Beta-lactams	D3-D7	7	1	0
Transition Summary	Beta-lactams	D5-D14	10	0.9	0.1
Transition Summary	Beta-lactams	D5-D7	52	0.94	0.06
Transition Summary	Beta-lactams	D7-D14	52	0.85	0.15
Transition Summary	Beta-lactams	D7-D28	8	1	0
Transition Summary	Beta-lactams	D7-D90	1	1	0
Transition Summary	Both	D0-D1	6	1	0
Transition Summary	Both	D0-D3	1	1	0
Transition Summary	Both	D0-D5	1	1	0
Transition Summary	Both	D14-D28	40	0.7	0.3
Transition Summary	Both	D14-D90	9	0.89	0.11
Transition Summary	Both	D1-D14	2	1	0
Transition Summary	Both	D1-D3	24	0.83	0.17
Transition Summary	Both	D1-D5	8	1	0
Transition Summary	Both	D1-D7	1	1	0
Transition Summary	Both	D28-D90	34	0.76	0.24
Transition Summary	Both	D3-D14	2	1	0
Transition Summary	Both	D3-D28	1	1	0
Transition Summary	Both	D3-D5	40	0.9	0.1
Transition Summary	Both	D3-D7	6	1	0
Transition Summary	Both	D5-D14	8	0.75	0.25
Transition Summary	Both	D5-D7	51	0.94	0.06

Transition Summary	Both	D7-D14	52	0.81	0.19
Transition Summary	Both	D7-D28	8	1	0
Transition Summary	Both	D7-D90	1	1	0

(Aminoglycosides, β -lactams, and concurrent carriage in the infant gut microbiome across various timepoints. The table displays the number of infants positive at the start of an interval (N pos. at start) and the proportions of those infants who remained positive (p11) or became negative (p10) by the end of the interval, highlighting a general trend of high short-term persistence and lower long-term retention of ARGs.

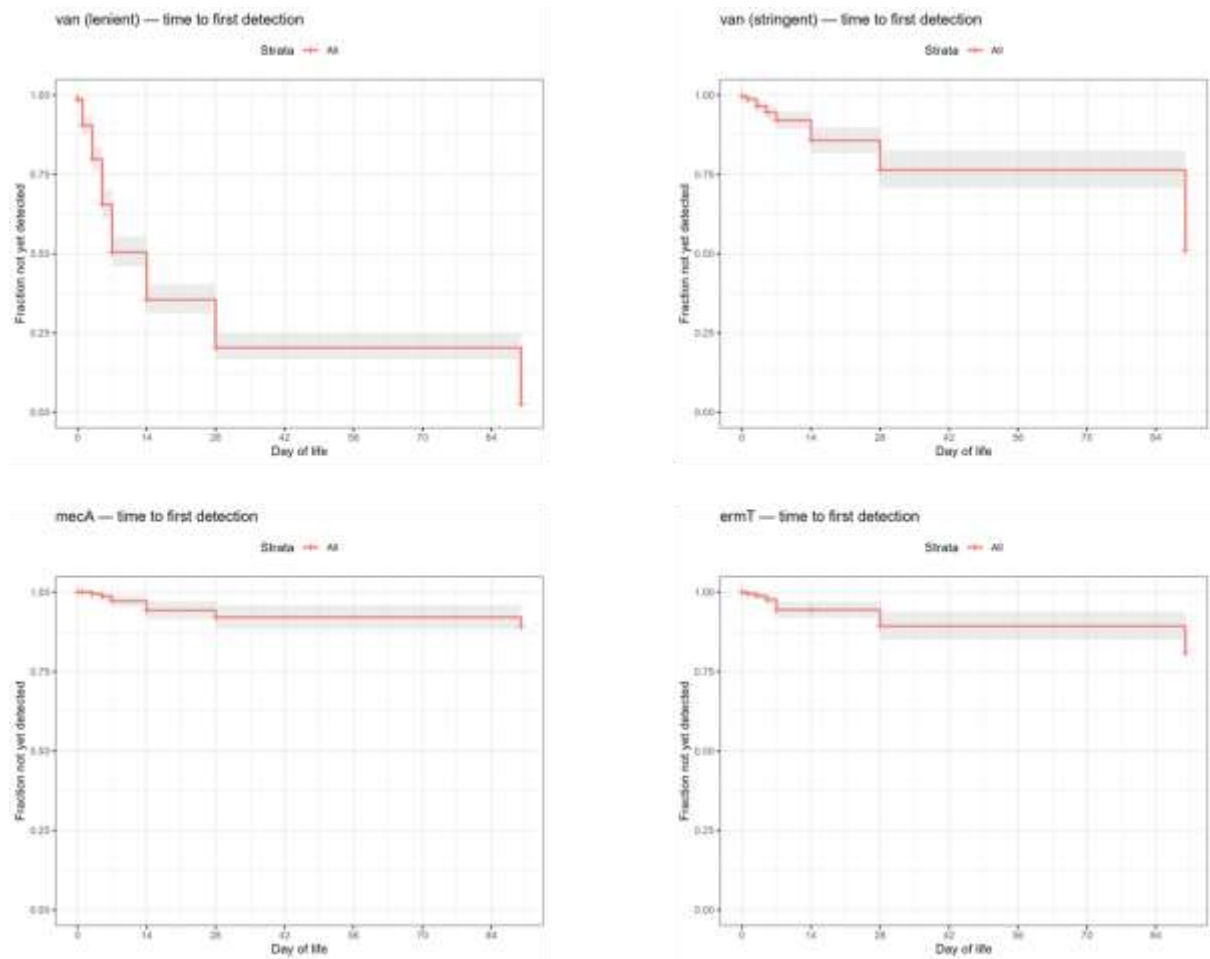


Figure S5.10: Time to first detection of selected rare but high-risk ARGs across the first three months of life.

KaplanMeier curves show the fraction of infants not yet detected as carrying each determinant over scheduled follow-up (D0-D90): *van* (lenient definition), *van* (stringent definition), *mecA*, and *ermT*. Step decreases indicate first detection events; shaded bands denote 95% confidence intervals. The comparison of lenient versus stringent *van* definitions illustrates sensitivity of detection timing to calling thresholds, while the largely flat curves for rarer markers reflect infrequent acquisition over follow-up

Appendix A. 3

Supplementary data 2

COHORT STRUCTURE Enrolment
and Sample Availability:

- 90 infants enrolled (total recruitment)
- 87 infants provided at least one stool sample for analysis: used for taxonomic classification (microbiome analysis)
- 85 infants had sufficient sequencing depth for ARG analysis (resistome)

Missingness Across Timepoints: Not all infants have samples at all timepoints. Sample sizes vary by scheduled day:

- D0: n=10
- D1: n=42
- D3: n=59
- D5: n=63
- D7: n=68
- D14: n=69
- D28: n=63
- D90: n=61

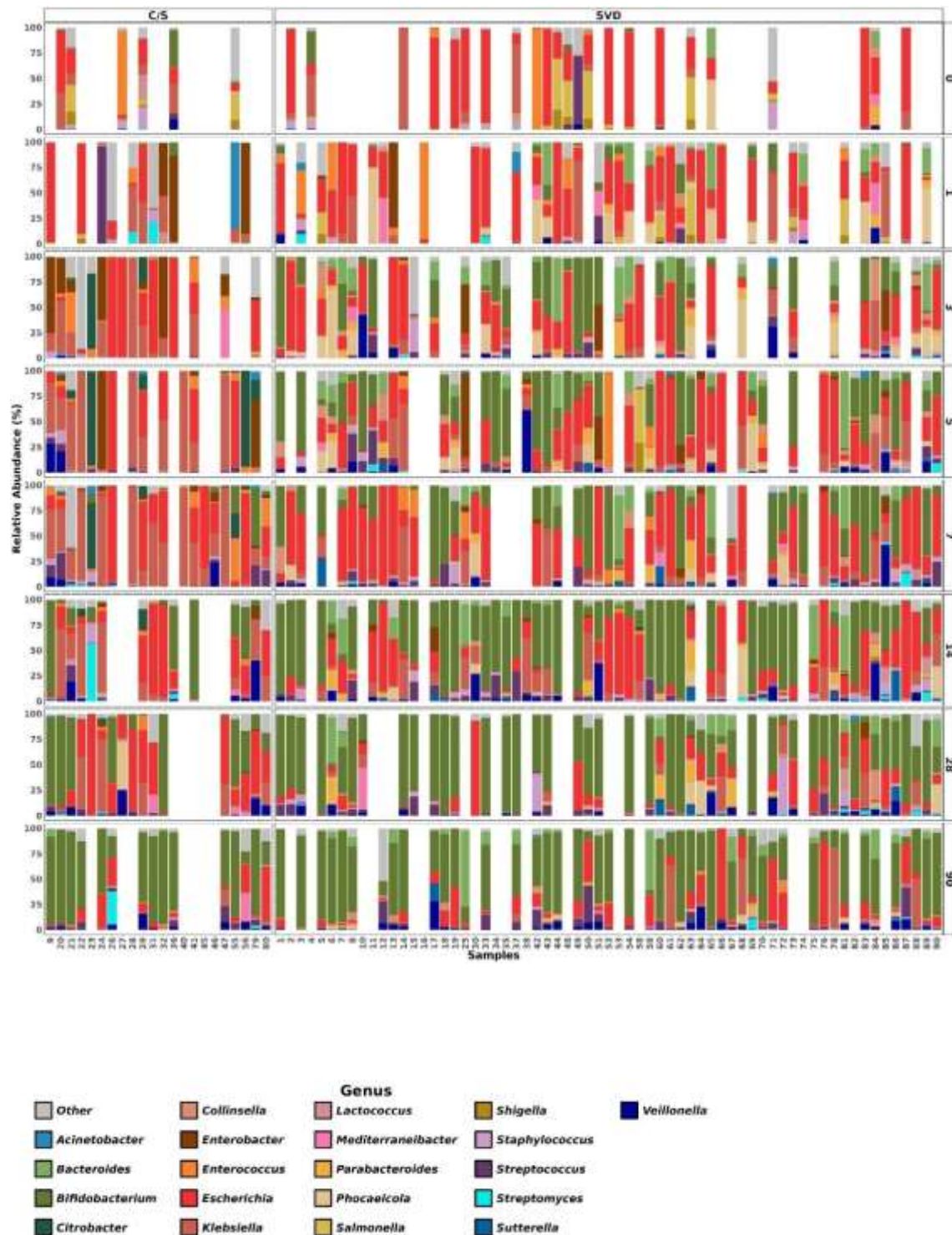


Figure S6.1: Infant-level longitudinal genus composition stratified by mode of delivery across scheduled sampling days.

Stacked bar plots show relative abundance (%) of dominant bacterial genera for each infant sample across the scheduled day-of-life visits (D0, D1, D3, D5, D7, D14, D28, D90). Panels are stratified by delivery mode (C/S vs SVD), with each bar representing an individual sample and colours denoting genera (legend shown below; low-abundance taxa grouped as “Other”).

Figure plotted by Aditiya Lankapalli