

https://dspace.org/

Snapshot of the gut microbial diversity of an urban South African population in relation to diet, lifestyle, health and socioeconomic status

Nkera-Gutabara, Claudine Kanyana

2025
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/49958
Nkera-Gutabara, Claudine Kanyana . (2025). The role of design houses [PhD thesis, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/49958

Downloaded from a DSpace repository. https://dspace.org/

Snapshot of the gut microbial diversity of an urban South African population in relation to diet, lifestyle, health and socioeconomic status

Claudine Kanyana Nkera-Gutabara

734113

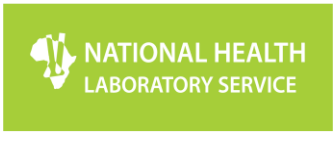
A thesis submitted to the Faculty of Health Science, University of the Witwatersrand, in fulfilment of the requirements for the degree of Doctor of Philosophy

Johannesburg, September 2025









I DECLARATION

I, Claudine K. Nkera-Gutabara, declare that this Thesis is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

_____

(Signature of candidate)

_____day of_____20_____in_____

II DEDICATION

This work is dedicated to my late Uncle Habib, in loving memory of him

1944 – 2024

To my loving husband, Benjamin Ndimurukundo, my pillar of strength and support throughout this journey

To my friends, extended family, siblings and parents, most especially my praying mother, Jeannette Nyiratunga – thank you for your encouragement every step of the way

To God, the Almighty, for being my guiding light

"…but we also boast in our affliction, knowing that affliction brings about endurance, and endurance brings about proven character, and proven character brings about hope." – Romans 5:3-4

III PRESENTATIONS AND PUBLICATIONS

Publications

Perspective Commentary – “Microbiomics: The next pillar of precision medicine and its role in the future of African health”; Authors: Nkera-Gutabara, CK., Kerr, R., Scholefield, J., Hazelhurst, S. and Naidoo, J.; Journal: Frontiers in Genetics (April 2022). [https://www.frontiersin.org/articles/10.3389/fgene.2022.869610/full].

Book chapter – “Chapter 5.4 Functional applications of human microbiome diversity studies”; Authors: Nkera-Gutabara, CK*., Hurrell, T*., Naidoo, J.; in Sadeepa, D., Sirisena, K. and Manage, P. (April 2024). “Microbial Diversity in the Genomic Era: Functional Diversity and Community Analysis”. **SECOND EDITION** from Academic Press, Elsevier. [https://www.researchgate.net/publication/379484897]. *: equal contribution.

Presentations

Co-hosted the South African Society of Human Genetics Young Researchers’ Online Symposium (SASHG YROS, August 2021).

Oral presentation at CSIR Workshop - “Probiotics – Opportunities for Technology Localisation” (July 2022).

Poster presentation at the Cold Spring Harbor “Microbiome Meeting” in New York, USA (October 2022).

Poster presentation at the 14th International Congress of Human Genetics in Cape Town, SA (February 2023).

Poster presentation - African Microbiome Day (Virtual, 11th September 2023).

YEGG Committee member/Local Organising Committee member (Program coordinator) and Poster presentation - Young and Early Career Geneticists Group (YEGG) Symposium (School of Public Health, Wits, 26th October 2024)

Poster presentation – South African Society for Human Genetics (SASHG) 2024 Conference (Sun City, 27th – 29th October 2024).

IV ABSTRACT

The human gut microbiome, comprising symbiotic, pathogenic, and commensal organisms, generally maintains a stable composition in adults but is influenced by diet, lifestyle, and environment. South Africa's (SA) shift toward urbanization, with associated increases in sedentary lifestyles and dietary changes, may profoundly impact health, yet understanding of these effects on SA's gut microbiome remains limited. This study details the aim of a snapshot characterization of the gut microbiome in a population of *n* = 220 urban professional South Africans recruited from the CSIR campus (Pretoria) and the University of the Witwatersrand (Johannesburg). Utilizing comprehensive demographic, lifestyle and dietary surveys (initial 123 variables refined to 101 informative variables) and a stool collection for microbiome profiling, alongside PacBio 16S long-read sequencing and complementary short-read shotgun metagenomics, this research further aims to reveal significant associations between microbial community structure and host demographics, socioeconomic status (SES), diet, lifestyle, and health-related factors.

Analyses demonstrated that microbiome composition significantly varies by age, gender, BMI, and ancestry. Younger adults exhibited higher microbial diversity with enriched beneficial taxa such as *Bifidobacterium longum* and *Roseburia faecis*, whereas older adults showed lower diversity coupled with increased susceptibility pathways, including *Staphylococcus aureus* infections. Gender-specific differences emerged, with males enriched in *Prevotella stercorea* and phenylpropanoid biosynthesis pathways, and females characterized by *Bacteroides dorei* and distinct lipid metabolism processes. Ancestry-related microbiome differences correlated with dietary adaptations, including *Prevotella* dominance in African ancestry groups and *Bacteroides* enrichment in Asian/Multi-ancestry groups. BMI analysis indicated microbial taxa and metabolic pathways uniquely associated with underweight, normal, overweight, and obese groups, highlighting metabolic flexibility and potential dysbiotic states. Notably, despite substantial inter-individual taxonomic variability and limited species-level core microbiome persistence over time, the functional profile remained remarkably stable (82% of functional pathways), emphasizing metabolic redundancy and resilience within urban South African gut ecosystems. Methodological comparisons between 16S long-read and shotgun sequencing highlighted platform-specific biases, underscoring the importance of integrated multi-omics approaches for accurate functional interpretation.

This study contributes to the field by characterizing a novel, understudied South African population, challenging Eurocentric microbiome paradigms, and emphasizing urbanization's dual role as disruptor and adapter of gut microbial communities. It advocates for locally tailored microbiome interventions and highlights microbial signatures linked to modifiable urban stressors such as diet and environmental exposures. Future research should prioritize

longitudinal multi-omics studies, causal inference analyses, and subpopulation-specific validations to inform targeted public health strategies in rapidly urbanizing African contexts.

V ACKNOWLEDGEMENTS

I would like to thank my supervisors Prof Scott Hazelhurst, Dr Jerolen Naidoo and Dr Robyn Kerr for their consistent guidance and encouragement and support throughout the years of this PhD project and the thesis writing process.

A special thanks to Dr Dylan Maghini, Dr Jakob Wirbel and Dr David Twesigomwe for significant assistance with the bioinformatics analyses and Caitlin Wheeler for logistical and moral support.

Thank you to the University of the Witwatersrand, Sydney Brenner Institute for Molecular Bioscience (SBIMB), and the Council for Scientific and Industrial Research (CSIR) for provision of facilities and funding necessary to carry out the research and presentation of the research at scientific conferences. Thank you as well to colleagues at the Division of Human Genetics, SBIMB and CSIR BIGG laboratory for support.

A big thank you to the participants of the CSIR Microbiome Mapping Initiative for consenting to be a part of the study – without whom none of this would be possible.

This research was funded by an NRF Thuthuka Grant awarded to Dr Jerolen Naidoo, Wits Postgraduate Merit Award and by a DSI-CSIR PhD Bursary.

VI Table of Contents

1. Introduction ---- 23
1.1. Background and rationale ---- 24
1.1.1. Gut dysbiosis and health outcomes ---- 25
1.1.2. Dietary habits and lifestyle factors ---- 25
1.1.3. Gut microbiome analysis methods ---- 26
Cost-benefit analysis ---- 27
1.2. Research aim and objectives ---- 29
1.3. Significance of the study ---- 30
1.4. Thesis structure ---- 30
2. Literature review ---- 30
2.1. Overview of the gut microbiome ---- 30
2.1.1. The establishment of the gut microbiome ---- 31
2.1.2. Composition and function of the human gut microbiome ---- 31
2.1.3. Intrinsic and extrinsic factors influencing gut microbiome diversity ---- 32
2.2. Gut microbiome and human health ---- 34
2.2.1. Gut-organ axes ---- 34
The gut-brain axis and neurological health ---- 35
The gut-liver axis and cellular models ---- 36
The gut-lung axis and respiratory health ---- 37
2.2.2. Gut microbiome pathologies ---- 38
Obesity ---- 38
Infectious diseases – HIV and TB ---- 39
Cancer ---- 40
2.2.3. Gut microbiome and therapeutic interventions ---- 41
Fermented foods ---- 42
Faecal Microbiota Transplantation (FMT) ---- 43
Probiotics ---- 43
Postbiotics ---- 44

2.3. Gut microbiome studies in South African populations ---- 45
2.3.1. Previous research findings ---- 45
2.3.2. Gaps in knowledge ---- 50
3. Materials and methods ---- 52
3.1. Participant recruitment ---- 53
3.2. Participant data collection and analysis ---- 54
3.2.1. Ethical considerations and consent procedures ---- 54
3.2.2. Anthropometric and health data ---- 55
3.2.3. Descriptive analysis of survey data ---- 55
3.3. Stool sample collection and processing ---- 56
3.4. Microbiome sequencing techniques ---- 58
3.4.1. PacBio 16S rRNA long-read sequencing ---- 58
3.4.2. Illumina shotgun metagenomics ---- 59
3.5. Bioinformatics Analysis ---- 60
I. 16S LONG-READ BIOINFORMATICS ANALYSIS ---- 61
3.5.1. QIIME2 workflow for 16S data ---- 61
i. Feature-Table and Feature-Data ---- 63
ii. Phylogenetic Tree generation ---- 63
3.5.2. Normalization through rarefaction depth analysis ---- 63
3.5.3. Diversity analysis ---- 64
i. Alpha diversity analysis ---- 65
ii. Beta diversity analysis ---- 65
3.5.4. Taxonomic classification ---- 66
3.5.5. Differential abundance analysis ---- 66
3.5.6. Predicted functional profiling analysis ---- 67
II. SHOTGUN BIOINFORMATICS ANALYSIS ---- 69
3.5.7. MAG workflow for shotgun data ---- 69
i. Preprocessing ---- 71
ii. Classification ---- 71

iii. Assembly ---- 72
iv. Binning ---- 73
v. Post-binning – Genome Taxonomy Database ---- 74
vi. Observed functional profiling analysis ---- 74
3.6. Long-read and shotgun metagenomic dataset concordance ---- 76
3.6.1. Taxonomic concordance assessment ---- 76
3.6.2. Distribution and abundance analysis ---- 77
3.6.3. Method agreement evaluation ---- 77
3.6.4. Functional analysis comparison ---- 77
i. Differential abundance analysis ---- 77
ii. Comparative functional analysis ---- 78
iii. Metadata integration analysis ---- 78
3.7. Core versus temporal microbiome profile: Longitudinal analysis ---- 78
3.7.1. Compositional stability assessment ---- 78
3.7.2. Diversity and taxonomic analysis ---- 79
3.7.3. Functional stability analysis ---- 79
3.7.4. Integration with participant metadata ---- 79
3.7.5. Core and variable component analysis ---- 79
3.8. Comparative analysis to other cohort populations ---- 80
3.8.1. Data processing and standardization ---- 80
3.8.2. Diversity analysis and taxonomic analysis ---- 80
3.8.3. Functional pathway analysis ---- 81
3.9. Exploratory analysis of community clusters ---- 82
4. Results ---- 82
4.1. Descriptive analysis ---- 82
4.1.1. Study cohort characteristics ---- 82
4.1.2. Data collection and sample processing overview ---- 85
4.1.3. Data cleaning and variable selection ---- 85
4.1.4. Significant demographic associations ---- 85

4.2. Sequencing data quality control (QC)---- 86
4.2.1. 16S rRNA gene sequencing (long-read data)---- 86
4.2.2. Shotgun metagenomic sequencing (short-read data)---- 86
4.3. Gut microbiome composition in the urban South African population ---- 86
4.3.1. Microbial community richness estimates and differences using long-read data 87
i. Alpha diversity across demographic variables---- 87
ii. Beta diversity across demographic variables ---- 90
iii. Diversity across response variables---- 93
4.3.2. Differential taxonomic abundance analyses with LEfSe---- 98
i. Differential abundance across demographic variables ---- 99
ii. Differential abundance across response variables ---- 104
SES---- 104
Diet---- 105
Lifestyle---- 108
Health---- 111
4.3.3. Predicted functional profiling analyses with PICRUSt and *ggpicrust2*------- 112
4.4. Investigation of community stratification in the SAN population ---- 114
4.4.1. Identification and validation of community clusters---- 115
4.4.2. Characterization of cluster-specific community structures---- 116
4.4.3. Association of clusters with participant characteristics ---- 117
4.5. Long-read and short-read dataset concordance---- 121
4.5.1. Taxonomic resolution and classification---- 121
4.5.2. Microbial community composition---- 121
4.5.3. Diversity metrics comparison---- 121
4.5.4. Method agreement and bias analysis---- 121
4.5.5. Functional profiling concordance---- 122
4.6. Core versus temporal dynamics of the gut microbiome---- 122
4.6.1. Diversity Analysis---- 123
i. Taxonomic analysis---- 123

ii. Functional analysis---------- 128
4.6.2. Taxonomic composition analysis ---------- 131
4.6.3. Core microbiome analysis---------- 141
i. Taxonomic analysis---------- 141
ii. Functional analysis---------- 148
4.7. Comparative analysis with other cohort populations ---------- 151
4.7.1. MAG quality and novelty analysis ---------- 152
i. Quality analysis ---------- 152
ii. Taxonomic analysis---------- 153
iii. Novelty analysis ---------- 158
4.7.2. Diversity patterns across cohorts---------- 161
i. Alpha diversity---------- 161
ii. Beta diversity ---------- 163
4.7.3. Taxonomic and differential abundance ---------- 167
4.7.4. Functional diversity across cohorts ---------- 174
5. Discussion ---------- 182
5.1. Overview of findings---------- 182
5.2. Methodological considerations and limitations ---------- 183
Technical considerations---------- 184
Study design ---------- 184
Analytical challenges---------- 184
5.3. Demographic drivers of microbiome composition---------- 185
i. Age-related patterns ---------- 185
Diversity ---------- 186
Differential taxonomy---------- 187
Functional pathways ---------- 187
ii. Gender-related patterns ---------- 188
Diversity ---------- 189
Differential taxonomy---------- 189

Functional pathways ---------- 191
iii. BMI group-related patterns ---------- 191
Diversity and differential taxonomy ---------- 192
Functional pathways ---------- 194
iv. Ancestry-related patterns ---------- 194
Diversity and differential taxonomy ---------- 195
Functional pathways ---------- 196
5.4. Natural community stratification suggests underlying biological differences ------ 198
5.5. Implications of diet, lifestyle, SES and health factors ---------- 199
i. SES variables ---------- 199
ii. Diet variables ---------- 200
iii. Lifestyle variables ---------- 201
iv. Health variables ---------- 203
5.6. Longitudinal analysis reveals temporal dynamics of a core yet variable microbiome 204
5.7. Revealing the broader landscape ---------- 207
Diversity across resolution and taxonomic divergence ---------- 207
MAG analysis to bridge platforms ---------- 209
Functional profiling and concordance ---------- 210
5.8. Future directions for research ---------- 211
Population-specific investigations ---------- 211
Clinical applications ---------- 212
Methodological advancements ---------- 213
6. Conclusion ---------- 215
6.1. Summary of findings ---------- 215
6.2. Contributions to the field ---------- 216
References ---------- 218

VII List of abbreviations / Nomenclature

Actinobacteria = Actinomycetota

Bacteroidetes = Bacteroidota

Firmicutes = Bacillota

Proteobacteria = Pseudomonadota

Verrucomicrobia = Verrucomicrobiota

ACE2 - Angiotensin-Converting Enzyme 2

ADHD - Attention Deficit Hyperactivity Disorder

AF - Alignment Fraction

AGP - American Gut Project

ALDEx2 - ANOVA-Like Differential Expression tool

ALS - Amyotrophic Lateral Sclerosis

ANI - Average Nucleotide Identity

ARV - Antiretroviral

ASD - Autism Spectrum Disorder

ASV - Amplicon Sequence Variant

AUD - Alcohol Use Disorder

AWI - AWIGEN1 (Phase 1) microbiome cohort

AWI-Gen1 - Africa Wits-INDEPTH Partnership for Genomic Research Phase 1

BA - Blueberry Anthocyanins

BCAA - Branched-Chain Amino Acid

BLAST - Basic Local Alignment Search Tool

BMI - Body Mass Index

BWA - Burrows-Wheeler Aligner

CAZymes - Carbohydrate-Active Enzymes

CD - Celiac Disease

CDI - Clostridioides difficile Infection

CFR - Code of Federal Regulations

CGA - Chlorogenic Acid

CheckM - Check Metagenome

Concoct - Clustering contigs on coverage and composition

CRC - Colorectal Cancer

CSIR - Council for Scientific and Industrial Research

CVD - Cardiovascular Disease

DADA2 - Divisive Amplicon Denoising Algorithm 2

DASTool - Dereplication, Aggregation and Scoring Tool

EC - Enzyme Commission

FastANI - Fast Alignment-free Sequence-based Average Nucleotide Identity

FDA - Food and Drug Administration

FDR - False Discovery Rate

FISMA - Federal Information Security Management Act

FMT - Fecal Microbiota Transplantation

GABA - Gamma-Aminobutyric Acid

GDPR - General Data Protection Regulation

ggPICRUSt2 - R package for PICRUSt2 analysis and visualization

GOS - Galacto-Oligosaccharides

GTDB-Tk - Genome Taxonomy Database Toolkit

HIPAA - Health Insurance Portability and Accountability Act

HPCSA - Health Professions Council of South Africa

HUMAnN3 - HMP Unified Metabolic Analysis Network version 3.9

IBD - Inflammatory Bowel Disease

IBS - Irritable Bowel Syndrome

IBS-C - Constipation-Predominant Irritable Bowel Syndrome

ICC - Intraclass Correlation Coefficient

KDE - Kernel Density Estimation

KEGG - Kyoto Encyclopaedia of Genes and Genomes

KO - KEGG Ortholog

KW - Kruskal-Wallis

LDA - Linear Discriminant Analysis

LEfSe - Linear Discriminant Analysis Effect Size

LI - Lactose Intolerance

LPG - Liquified Petroleum Gas

LPS - Lipopolysaccharides

M. tb - Mycobacterium tuberculosis

MaAsLin2 - Multivariate Association with Linear Models 2

MAG - Metagenome-Assembled Genome

MaxBin - automated binning of assembled metagenomic scaffolds using an expectation-maximization algorithm

Megahit - Metagenome Assembly with Graph-based Hierarchical Iterative Treatment

MetaBAT - Metagenome Binning with Abundance and Tetra-nucleotide frequencies

MetaPhlAn - Metagenomic Phylogenetic Analysis

MIAME - Minimum Information About a Microarray Experiment

NAFLD - Non-Alcoholic Fatty Liver Disease

NCBI - National Centre for Biotechnology Information

NCD - Non-Communicable Disease

NSTI - Nearest Sequenced Taxon Index

PACS - Post-Acute Sequelae

PCoA - Principal Coordinates Analysis

PCOS - Polycystic Ovary Syndrome

PERMANOVA - Permutational Multivariate Analysis of Variance

PICRUSt2 - Phylogenetic Investigation of Communities by Reconstruction of Unobserved States 2

PLWH - People Living With HIV

PoPIA - Protection of Personal Information Act

Prodigal - Prokaryotic Dynamic Programming Genefinding Algorithm

QIIME2 - Quantitative Insights Into Microbial Ecology 2

Quast - Quality assessment tool

RED - Relative Evolutionary Divergence

REDCap - Research Electronic Data Capture

SAN - South African (urban professional) microbiome population

SCFA - Short-Chain Fatty Acid

SES - Socioeconomic Status

SILVA - Small Subunit Ribosomal RNA Gene Database

SL - Sphingolipid

SNP - Single-Nucleotide Polymorphism

STC - Slow Transit Constipation

STORMS - Standardized Metadata Reporting for Microbiome Studies

TB - Tuberculosis

Wits - University of the Witwatersrand

VIII List of Figures

Chapter 2

Figure 2.1 Overview of the gut-organ axes and their implications for various health conditions within the African context.

Figure 2.2 Overview of various gut microbiome therapeutic approaches targeting the gut microbiome to enhance health outcomes.

Chapter 3

Figure 3.1 Overview of the study design and methodology.

Figure 3.2 Descriptive analysis workflow.

Figure 3.3 Sample collection timeline.

Figure 3.4 Comparative schematic of microbial DNA Sequencing.

Figure 3.5 Overview of PacBio SMRT Sequencing:

Figure 3.6 Overview of the QIIME2 pipeline for 16S rRNA long-read sequencing data.

Figure 3.7 Summary information of the demultiplexed single-end 16S rRNA long-reads obtained from QIIME2.

Figure 3.8 PICRUSt2 computational workflow.

Figure 3.9 PICRUSt2 workflow diagram for metagenome analysis.

Figure 3.10 Overview of the bioinformatics pipeline for shotgun sequencing reads using the Metagenome-assembled genome (MAG) workflow.

Figure 3.11 MAG preprocessing workflow for shotgun sequencing data.

Figure 3.12 MAG workflow for assembly, binning, and post-binning analysis of shotgun metagenomic data.

Figure 3.13 Bioinformatics workflow for shotgun metagenomic analysis utilizing HUMAnN3 (with Mamba environment) pipeline.

Chapter 4

Figure 4.1 Comparison of alpha diversity measures across demographic variables.

Figure 4.2 Beta diversity principal coordinates analysis (PCoA) plots with Bray-Curtis Dissimilarity measure.

Figure 4.3 Comparison of Shannon diversity indices between individuals who use public transport ("Yes/Checked") and those who do not ("No/Unchecked").

Figure 4.4 Comparison of Shannon diversity indices between individuals who exclude dairy ("Yes/Checked") and those who do not ("No/Unchecked").

Figure 4.5 Comparison of Shannon diversity indices between individuals who use gas heating ("Yes/Checked") and those who do not ("No/Unchecked").

Figure 4.6 Comparison of Shannon diversity indices between individuals who consume multivitamins ("Yes") and those who do not ("No").

Figure 4.7 Principal Coordinates Analysis (PCoA) plot of Weighted UniFrac Distances comparing microbial community composition between individuals who use herbal supplements ('Yes') and those who do not ('No').

Figure 4.8 Principal Coordinates Analysis (PCoA) plot of Weighted UniFrac Distances comparing microbial community composition between individuals who do not have food allergies ('Checked') and those who do ('Unchecked').

Figure 4.9 Differential abundance of microbial taxa in "age group" within the SAN population.

Figure 4.10 Differential abundance of microbial taxa in "gender" groups within the SAN population.

Figure 4.11 Differential abundance of microbial taxa in "ancestry" groups within the SAN population.

Figure 4.12 Differential abundance of microbial taxa in "BMI group" within the SAN population.

Figure 4.13 Differential abundance of microbial taxa for "public transport" use within the SAN population.

Figure 4.14 Differential abundance of microbial taxa for "herbal supplements" use within the SAN population.

Figure 4.15 Differential abundance of microbial taxa for "gas heating" use within the SAN population.

Figure 4.16 Differential abundance of microbial taxa for "multivitamins" use within the SAN population.

Figure 4.17 Differential abundance of microbial taxa for "food allergies (none)" within the SAN population.

Figure 4.18 Distribution of the top 10% of KEGG (MetaCyc) pathways as identified from PICRUSt functional analysis of 16S long-read data.

Figure 4.19 Principal Coordinates Analysis (PCoA) plot showing the distribution of SAN population samples clustered using k-means clustering (k = 2).

Figure 4.20 Distribution of relative abundances for each bacterial phylum across the two identified clusters in the SAN population.

Figure 4.21 Distribution of metadata variables significantly associated with microbial community clusters ($p < 5.0 \times 10\text{-}2$).

Figure 4.22 Bacterial phyla showing significant associations with lifestyle variables in the SAN population clusters.

Figure 4.23 Comparison of intra- and inter-individual gut microbiome variation.

Figure 4.24 Shannon Diversity Index comparison between baseline (cross-sectional) and follow-up (longitudinal) timepoints.

Figure 4.25 Individual changes in Shannon Diversity between timepoints.

Figure 4.26 Principal Coordinates Analysis (PCoA) of stool microbiome samples based on Bray-Curtis dissimilarities.

Figure 4.27 Total variance of KEGG ortholog (KO) pathway abundances plotted against their log10-transformed relative abundances.

Figure 4.28 Intra- versus inter-individual variance components of KO pathway abundances.

Figure 4.29 Distribution of bacterial phyla relative abundance across collection times.

Figure 4.30 Subject-specific mean relative abundance of bacterial phyla across collection times.

Figure 4.31 Distribution of the 14 most abundant bacterial classes across collection timepoints.

Figure 4.32 Class-level bacterial abundance changes by subject.

Figure 4.33 Distribution of the 14 most abundant bacterial orders across collection timepoints.

Figure 4.34 Order-level bacterial abundance changes by subject.

Figure 4.35 Genus-level bacterial differences between timepoints.

Figure 4.36 LEfSe analysis results displaying differentially abundant genera between timepoints.

Figure 4.37 Genus abundance changes by subject.

Figure 4.38 LEfSe analysis results displaying differentially abundant species between timepoints.

Figure 4.39 Species abundance changes by subject.

Figure 4.40 Distribution of core taxa across taxonomic levels identified using two prevalence thresholds.

Figure 4.41 Mean relative abundance versus prevalence of core taxa across taxonomic levels.

Figure 4.42 Taxonomic composition of core microbiome members at different prevalence thresholds.

Figure 4.43 Relationship between prevalence and mean relative abundance of core taxa at different taxonomic levels.

Figure 4.44 Proportion of stable and variable taxa across taxonomic levels.

Figure 4.45 Volcano plot depicting individual-level temporal dynamics of microbial taxa between baseline and follow-up samples (n = 22 subjects).

Figure 4.46 Distribution of core KO functions by prevalence and abundance.

Figure 4.47 Volcano plot of KO functional temporal dynamics.

Figure 4.48 Distribution of metagenome-assembled genome (MAG) quality across three human gut microbiome populations (AGP, AWI, and SAN).

Figure 4.49 Taxonomic composition at phylum level for metagenome-assembled genomes (MAGs) recovered from three human gut microbiome cohorts -- American Gut Project (AGP), AWIGEN1 (AWI) and the urban professional South African population (SAN).

Figure 4.50 Taxonomic composition at genus level for metagenome-assembled genomes (MAGs) recovered from three human gut microbiome populations (AGP, AWI, and SAN).

Figure 4.51 Distribution of the five most abundant taxa at each taxonomic level (phylum to species) across three human gut microbiome populations (AGP, AWI, and SAN).

Figure 4.52 Novelty assessment of bacterial MAGs against GTDB reference genomes.

Figure 4.53 Novelty assessment of archaeal MAGs against GTDB reference genomes.

Figure 4.54 Shannon diversity comparison across three human gut microbiome populations at the phylum level.

Figure 4.55 Shannon diversity comparison across three human gut microbiome populations at the genus level.

Figure 4.56 Shannon diversity comparison across three human gut microbiome populations at the species level.

Figure 4.57 Beta diversity analysis using log-Euclidean distances visualized through Principal Coordinate Analysis (PCoA) across three human gut microbiome populations at the phylum level.

Figure 4.58 Beta diversity analysis using log-Euclidean distances visualized through Principal Coordinate Analysis (PCoA) across three human gut microbiome populations at the genus level.

Figure 4.59 Beta diversity analysis using log-Euclidean distances visualized through Principal Coordinate Analysis (PCoA) across three human gut microbiome populations at the species level.

Figure 4.60 Mean relative abundance of bacterial phyla across three populations.

Figure 4.61 Differential abundance analysis comparing phyla between SAN and other cohorts (AGP, AWIGEN1).

Figure 4.62 Mean relative abundance of bacterial genera across three populations.

Figure 4.63 Differential abundance analysis comparing genera between SAN and other cohorts (AGP, AWIGEN1).

Figure 4.64 Mean relative abundance of bacterial species across three populations.

Figure 4.65 Differential abundance analysis comparing species between SAN and other cohorts (AGP, AWIGEN1).

Figure 4.66 Correlation of predicted metabolic pathway abundances between shotgun metagenomics and 16S approaches.

Figure 4.67 Distribution of abundance differences between prediction methods.

Figure 4.68 Top 20 pathways showing largest abundance differences between methods.

Figure 4.69 Top 20 pathways showing largest abundance differences between methods.

Figure 4.70 Distribution of pathway detection across functional categories.

Figure 4.71 Distribution of abundance differences across functional categories.

Figure 4.72 Summary of key findings from the urban professional South African gut microbiome study.

Chapter 5

Figure 5.1 Age-related gut microbiome findings showing key differences in microbial diversity, taxonomic composition, and functional pathways between age groups.

Figure 5.2 Gender-related findings revealing differences in gut microbiome composition and functionality between males and females.

Figure 5.3 BMI-associated variations in gut microbiome characteristics showing relationships between body mass index and microbial communities.

Figure 5.4 Ancestry-related patterns in gut microbiome composition demonstrating differences across various ancestral groups.

Figure 5.5 Key gut microbiome findings and associations of taxa with diet, lifestyle, socioeconomic status and health variables.

Figure 5.6 Microbiome profile spectrum and influential factors across global and local South African populationss.

IX List of Tables

Chapter 1

Table 1.1 Comparative summary of the cost, quality, read length and yield of short- and long-read sequencing platforms.

Chapter 2

Table 2.1 Summary of gut microbiome studies conducted in South African populations.

Chapter 4

Table 4.1 Descriptive characteristics of the urban professional South African study population

Table 4.2 Differentially abundant LEfSe taxa biomarkers uniquely identified in a response variable.

Table 4.3 Survey responses in subjects with largest Bray-Curtis Dissimilarity temporal shifts.

Table 4.4 High-abundance KO functions exhibiting substantial temporal variation.

Table 4.5 Statistical comparison of MAG quality distributions between populations.

Table 4.6 Top 5 taxa distribution across taxonomic ranks in three human gut microbiome populations.

1. Introduction

The microbiome is a dynamic ecosystem comprising diverse microorganisms that interact with each other and their environment. These interactions can lead to functional stability under certain conditions, often described as homeostasis (Berg et al., 2020). The human microbiome includes communities of commensal, symbiotic, and pathogenic microorganisms inhabiting distinct biogeochemical niches across the body. While homeostasis is a common framework for understanding microbiome dynamics, it is not universally consistent; disruptions (e.g., dysbiosis) i.e. transient states of imbalance, frequently occur due to external stressors or competitive interactions. These microbial communities – comprising bacteria, viruses, archaea, and fungi – engage in cooperative, neutral, and antagonistic (including competitive) interactions, which may shift in response to environmental pressures (Lebba et al., 2016).

Biogeochemical niches across the body include sites like the gastrointestinal tract (i.e. “gut”), urogenital, and respiratory tracts, as well as organs like the liver, skin, lungs, and kidneys. Each of these sites has a unique microenvironment that shapes its local microbial communities (Reynoso-García et al., 2022). The skin and respiratory tract, for example, are dominated by strict aerobic species, while the gastrointestinal tract houses both facultative aerobes and anaerobic species (Dekaboruah et al., 2020). Bacteria account for more than 90% of the human microbiome with the average human body hosting approximately 39 trillion bacterial cells, compared to 30 ± 8 trillion human cells, majority of which reside in the gut (Hatton et al., 2023; Sender & Milo, 2021).

The human gut, specifically the colon, is recognized as the richest source of microbial density in the human body and is a source of microbes that colonize other anatomical sites (Reynoso-García et al., 2022; Sender et al., 2016). With more than 23 million genes in total, the gut microbiome encodes 50- to 100-fold more genes than the human genome (Sender et al., 2016; Tierney et al., 2019). The gut microbiome's composition and diversity influence various physiological processes, including digestion, metabolism, and immunity, and is in turn influenced by various factors such as diet, lifestyle, and environmental exposures (Zimmermann et al., 2019). For instance, the production of short-chain fatty acids (SCFAs) by gut bacteria during the fermentation of dietary fibres plays a significant role in energy metabolism and immune regulation, highlighting the microbiome's impact on host physiology (Fujisaka et al., 2023). Thus, the human microbiome plays a crucial role in the overall function of the human body and is an essential component of human health.

Geographic location is a major factor in shaping the environmental landscape and contributes up to 20% of the variance observed in an individual's gut microbiome (Rothschild et al., 2018; Vangay et al., 2018). In line with the global trend, South Africa (SA) has seen an increase in

non-communicable diseases (NCDs) prevalence, with a 27% probability of dying between the ages of 30 and 70 years as a result of cardiovascular diseases (CVDs), diabetes, cancers, or chronic respiratory conditions (Johnson et al., 2021; World Health Organization, 2023). The increase in NCDs prevalence in SA is a cause for concern and these conditions are associated with imbalances of the gut microbiome, which may contribute to their onset and progression (Qin et al., 2012; Ferreira et al., 2018; Helliwell et al., 2019; Oduaran et al., 2020). Recent seminal efforts have sought to address the historical lack of representation for South African cohorts in global microbiome datasets (Allali et al., 2021; Maghini et al., 2025). However further studies are needed to fully capture the diverse diet, lifestyle and socioeconomic factors linked to microbiome diversity within SA. The characterisation of gut microbiome profiles across diverse local population groups will support efforts to more comprehensively illuminate the complex interplay between the environment, human health, and the gut microbiome, particularly in the context of South Africa's unique ecological and cultural landscape.

1.1. Background and rationale

South Africa’s cultural diversity, encompassing unique lifestyles, diets, and social practices, significantly shapes the gut microbial profiles of its population (van Zyl et al., 2020). Recent urbanization trends have driven shifts toward processed diets and sedentary behaviours, contributing to rising rates of metabolic and mental health conditions such as hyperglycaemia, obesity, and depression (Millington, 2019; Mutyambizi et al., 2020; Ramaboli et al., 2024). While lack of exercise is not inherently a pathology, it is a critical modifiable risk factor that exacerbates these conditions and influences gut microbiota composition (Wegierska et al., 2022). For instance, sedentary lifestyles correlate with reduced microbial diversity and altered SCFA production, which may mediate their metabolic consequences (Bressa et al., 2017). These transitions underscore the "microbiota-dependent" nature of diseases linked to under- and over-nutrition, which remain a public health burden across South Africa and the African continent (Blanton et al., 2016; Gómez-Gallego et al., 2019; J. L. Sonnenburg & Bäckhed, 2016).

Despite the growing interest in gut microbiome research in Africa, studies have primarily focused on gut dysbiosis associated with malnutrition and its treatment, with an emphasis on the "rural to urban" transition that contributes to increasing disease risks in African populations (Alou et al., 2016). However, there is a significant gap in research on the gut microbial profiles of adults in urban South Africa (Brewster et al., 2019). This study aims to address this gap by characterizing the gut microbial profile of a novel and diverse urban professional South African niche group to investigate links between microbial community composition and individual phenotypes.

1.1.1. Gut dysbiosis and health outcomes

Gut dysbiosis is a term used to describe an imbalance in the gut microbial communities that can lead to negative health outcomes (Humphreys, 2020). Dysbiosis can be caused by a range of factors, including antibiotic use, dietary changes, and lifestyle factors (Messer & Chang, 2018). Studies have shown that gut dysbiosis is associated with a range of health conditions, including metabolic disorders, inflammatory bowel disease, and mental health disorders (Brown et al., 2012; Kho & Lal, 2018). The gut microbiome plays a crucial role in regulating metabolic processes, including glucose and lipid metabolism, which is involved in diseases such as type 2 diabetes and obesity (Qin et al., 2012; Schoeler & Caesar, 2019). In addition, inflammatory bowel disease, which includes Crohn's disease and ulcerative colitis, has also been linked to alterations in the gut microbiome, with dysbiosis potentially contributing to the onset of these conditions (Zhang et al., 2022). Mental health and neurocognitive disorders, including anxiety, depression and Alzheimer's Disease, have also been directly correlated to alterations in the gut microbiome (Tremlett et al., 2017; Yang et al., 2022).

In South Africa, persistent high rates of obesity are linked to increased risk of mortality, especially in young women. Obesity prevalence in adults is significantly higher in SA than most low- and middle-income countries, affecting approximately 39% and 10% of women and men, respectively (Kruger et al., 2022; NCDRisC, 2016; UNICEF, 2022). This may be due to changes in dietary habits and sedentary lifestyles, leading to gut dysbiosis (Wandai et al., 2020). Therefore, investigating the association between gut dysbiosis, dietary habits, lifestyle factors and other variables in a diverse urban South African population is crucial to improving future understanding around the potential role of gut dysbiosis in the development of these health conditions. The findings of this study may also provide insights into potential interventions to prevent or treat these conditions by improving the gut microbiome through personalised medicine and other strategies.

1.1.2. Dietary habits and lifestyle factors

Dietary habits and lifestyle factors can influence gut microbial composition and diversity. In South Africa, a range of dietary habits and lifestyle factors may contribute to gut dysbiosis. For example, the consumption of a high-fat, high-sugar diet has been linked to reduced microbial diversity and increased levels of bacteria, like *Bacteroides* and *Prevotella*, in the gut (Hildebrandt et al., 2009; Shon et al., 2023). Additionally, physical inactivity has been associated with similar adverse alterations in diversity (van Zyl et al., 2020).

The process of urbanization in South Africa has brought about significant changes in the food landscape, with a shift from subsistence farming and local produce to a greater reliance on packaged, processed foods available at supermarkets and street vendors (Pereira, 2014).

This transition has been accompanied by a pervasive influence of media and advertising on consumer behaviour, influencing the purchasing and consumption decisions made by people across the country. Unfortunately, many consumers struggle to make healthy food choices, as they are often unable to decipher food labels (Yenerall et al., 2015). As a result, the dietary patterns of South Africans have changed dramatically in recent years, with an increased intake of energy-dense, nutrient-poor foods being a major contributor to the high prevalence of chronic diseases in the country (Kruger et al., 2022).

This study aims to investigate the association between gut microbiome status and factors such as those introduced that influence gut microbial composition in a diverse urban South African population. This will be achieved by collecting data on dietary habits, physical activity levels, and other lifestyle factors using validated questionnaires and a relatively novel microbiome sequencing technique, to characterize and unveil South African microbial diversity.

1.1.3. Gut microbiome analysis methods

Second-generation short-read sequencing (e.g., Illumina) has revolutionized microbiome research through cost-effective, high-throughput profiling of microbial communities, primarily via targeted 16S rRNA gene sequencing of hypervariable regions (e.g., V3 – V4) (Tedersoo et al., 2018; Xu et al., 2019). However, this approach suffers from accuracy and reproducibility limitations: taxonomic resolution is constrained by variable phylogenetic signals across hypervariable regions, with cases of bacterial genera being indistinguishable at the species level due to > 97% sequence similarity (Caudill & Brayton, 2022; Větrovský & Baldrian, 2013); and primer bias introduces variability, as demonstrated by researchers (Biol-Aquino et al., 2019), where V3 – V4 primers inflated *Bifidobacteria* abundance in infant gut samples compared to V4 – V5 primers. Short-read lengths (< 500 bp) further hinder assembly of structural variations critical for strain-level analysis, especially in underrepresented populations (Zeevi et al., 2019).

Third-generation long-read sequencing (e.g., Pacific Biosciences (PacBio) HiFi, Oxford Nanopore Technologies (ONT)) addresses these limitations by capturing full-length 16S rRNA genes (~ 1,500 bp) in a single read, eliminating primer bias and standardizing taxonomic classification (Ho et al., 2020). Full-length or long-read 16S enhances reproducibility by ensuring consensus across studies and improves accuracy through species-level discrimination (Lin et al., 2024). Recent advances have also drastically reduced error rates where long-reads reach > 99.9% accuracy (Q30), surpassing early long-read platforms (11 – 14% errors) and rivalling short-read accuracy (Earl et al., 2018; Gehrig et al., 2022; J. S. Johnson et al., 2019; D. Lang et al., 2020; Roberts et al., 2013).

Shotgun metagenomics, applicable to both short- and long-read platforms, bypasses PCR amplification by sequencing all DNA in a sample, enabling whole-genome reconstruction and functional insights (Gand et al., 2024). While short-read shotgun sequencing dominates for cost, long-read shotgun excels in resolving complex genomic regions (e.g., mobile elements, repetitive sequences) and linking antimicrobial resistance genes to hosts (Gand et al., 2024). However, taxonomic assignment challenges persist due to database inaccuracies and lack of appropriate geographical representation (Campos et al., 2022). Public repositories (e.g., SILVA, Greengenes) contain misclassified sequences (0.2 – 2.5%) and gaps in unclassified taxa (Campos et al., 2022; Kozlov et al., 2016). Custom databases improve resolution (Myer et al., 2020), but emerging tools leveraging FL 16S data now reduce misclassification by resolving ambiguous regions (Cabezas et al., 2024). Therefore, while short-read technologies remain practical for large-scale 16S surveys, long-read sequencing – whether FL 16S or shotgun – offers superior reproducibility, accuracy, and genomic context, enhancing standards for microbiome analysis.

Advances in sequencing technologies, such as ONT and PacBio, have provided powerful tools for resolving metagenomic complexity, yet their adoption in African microbiome studies remains limited compared to shotgun sequencing. Despite their advantages, challenges such as cost, infrastructure requirements, and technical expertise have hindered widespread use in Africa. For instance, the AWI-Gen 2 Microbiome Project primarily employed Illumina-based shotgun sequencing to maximize data consistency and accessibility across diverse African sites, while highlighting the potential for long-read sequencing to expand reference genome collections in future studies (Maghini et al., 2025). Similarly, short-read sequencing dominated earlier studies in South Africa, such as the analysis of urban and rural gut microbiomes, which noted the limitations of incomplete genomic resolution (Tamburini et al., 2022). Increased adoption of long-read sequencing platforms in Africa could significantly enhance our understanding of the continent's unique microbiome diversity and its implications for health.

Cost-benefit analysis

Second-generation sequencing (NGS) technologies have become indispensable tools for exploring the gut microbiome, but the choice between short-read and long-read approaches involves careful consideration of cost, accuracy, and the specific research question. While second-generation short-read sequencing initially dominated due to its high throughput and lower per-base cost, third-generation long-read sequencing is rapidly becoming a competitive and often advantageous alternative, especially for comprehensive 16S rRNA gene characterization (Ciuffreda et al., 2021).

The cost of genome sequencing has dramatically decreased, driven by technological advancements and market competition. The cost to sequence a human genome has fallen significantly in recent years, reaching levels as low as $600 in early 2025 and with ongoing efforts to achieve the "$100 genome" (Giani et al., 2020; 3billion, 2023; Complete Genomics, 2024; GlobeNewswire, 2024, 2024). However, for microbiome studies, the cost per gigabase (Gb) and overall yield are more relevant metrics for platform comparison. Illumina platforms offer a range of options, from the more affordable iSeq series to the high-throughput NovaSeq series. While older data (Vilella, 2019) suggested a substantial cost difference between short-read and long-read platforms (Table 1.1), recent pricing indicates a narrowing gap, particularly when considering the benefits of long-read sequencing for 16S rRNA gene analysis.

Table 1.1 Comparative summary of the cost, quality, read length and yield of short- and long-read sequencing platforms

Platform*	Cost per Gb*	Quality	Read length*	Yield*
Illumina iSeq100	$521	≥ 89% bases ≥ Q30**	2x 150bp paired-end	1.2Gb
Illumina NovaSeq S4	$4.80	≥ 85% bases ≥ Q30	2x 150bp paired-end	6000Gb
Illumina NovaSeq X Plus	~ $3- $7	≥ 90% bases ≥ Q30	2x 150bp paired-end	Up to 16,000Gb
PacBio RSII	$200	Q50 (>99.999%)°	20Kbp	12Gb
PacBio Revio	$10 - $20	HiFi >Q30 (>99.9%)	Up to 75Kbp	Up to 1,500Gb

*Data adapted from: Dr A Vilella, 2019; Illumina, 2021; Macrogen, 2020; Illumina, 2024; PacBio, 2024; ONT, 2024. Jumina, 2017; °Macrogen, 2020. Cost per Gb and Yield ranges are approximate and depend on kit configuration and sequencing run parameters. Quality metrics are representative and may vary.

While circular consensus sequencing (CCS) in PacBio and 2D sequencing or high-accuracy base calling in ONT were initially associated with higher costs compared to short-read sequencing (Giani et al., 2020), this is no longer universally true. PacBio HiFi sequencing, utilizing CCS, now achieves high accuracy (> 99.9% or Q30) with costs becoming increasingly competitive (PacBio, 2024). ONT's MinION, known for its affordability and portability, particularly regarding instrument cost, can be especially attractive in resource-limited settings (Adewale, 2020; Kraft & Kurth, 2019). The MinION Mk1C, for instance, offers a relatively low entry barrier and can be deployed in field settings, reducing sample transportation costs and turnaround time (ONT, 2024). Furthermore, for targeted 16S rRNA gene sequencing, the long reads generated by PacBio and ONT eliminate the need for assembly and primer bias associated with short-read approaches, potentially reducing computational costs and improving the accuracy and reproducibility of taxonomic profiling (Ciuffreda et al., 2021).

While Illumina short-read platforms still offer unmatched throughput for applications requiring massive sequencing depth, for many gut microbiome studies, particularly those focused on comprehensive 16S rRNA gene characterization and in settings with budget constraints, long-read sequencing platforms are emerging as a highly viable and increasingly cost-effective option (Cha et al., 2023; Ciuffreda et al., 2021). The choice ultimately depends on the specific research question, desired depth of sequencing, and available resources, but the advantages

of long-read sequencing in terms of accuracy, full-gene coverage, and potential for cost savings in specific applications warrant careful consideration, especially in resource-limited environments. Utilizing full-length or third-generation sequencing technologies enables more accurate resolution of novel microbial variants in understudied regions, such as the gut microbiome of South African urban professional populations, thereby advancing our aim to characterize unique microbial diversity and its functional relevance.

1.2. Research aim and objectives

The overarching aim of this study is to generate a snapshot of the gut microbial diversity within a niche, urban South African population group. The primary aims and objectives nested in this research are:

1. To produce a novel full-length 16S sequencing microbiome dataset of urban professional South Africans (SAN) and optimize bioinformatics pipelines for their analysis. This will be achieved through the following objectives:
    a. To recruit 250 participants from two urban locations (Council for Scientific and Industrial Research (CSIR), Pretoria; University of the Witwatersrand, Johannesburg) using a convenience sampling approach.
    b. To collect electronic anthropological participant data using REDCap following informed consent.
    c. To collect biological samples (stool) from all participants at a single time point, as well as a second sample as a longitudinal population subset.
    d. To obtain raw DNA sequencing data extracted from stool samples followed by long-read sequencing and shotgun metagenomics using the PacBio and Illumina platforms, respectively.
    e. To assess and validate optimized bioinformatics pipelines on the long-read metagenomic data and compare with shotgun sequencing.
2. To conduct an exploratory analysis to characterize the microbial diversity to inform associations with diet, lifestyle, socioeconomic status and health in the South African context. The research objectives of this aim are:
    a. To investigate the alpha and beta diversity and taxonomic composition of the SAN study population across different groups derived from the anthropometric survey.
    b. To assess the differential abundance between groups and the predicted functional profiles of the gut microbiome.
    c. To assess the taxonomic and observed functional profile concordance between the long-read and the subset short-read dataset.
    d. To assess the compositional and functional stability of a longitudinal subset of the SAN population gut microbiome over two timepoints.

e. To contextualize the SAN study population microbial profile with comparative African (AWIGEN1) and global (American Gut Project) cohorts.

1.3. Significance of the study

Urbanization drives profound shifts in human lifestyles, yet its impacts on gut microbial ecology in transitional African populations remain poorly understood. South Africa's urban professionals – a demographic navigating rapid dietary, environmental, and socioeconomic change – represent a critical but understudied cohort for investigating how modern stressors reshape microbiome stability and health linkages. Current microbiome research disproportionately focuses on people living in Western countries, perpetuating biases in biomarker discovery and health interventions. This study addresses these gaps by employing third-generation PacBio long-read 16S sequencing and shotgun metagenomics to characterize gut microbial composition and function in urban South Africans. By resolving taxonomic complexity at species-level resolution and contextualizing findings within local dietary, lifestyle, and environmental exposures, this work aims to clarify how microbial communities adapt to – or destabilize under – urbanization's pressures. The methodological integration of long-read sequencing with metagenomics may further expose limitations in functional inference from 16S data, advancing best practices for studying underrepresented populations. Ultimately, this research seeks to initiate the infrastructure platform to inform equitable public health strategies tailored to South Africa's unique urbanization challenges, where non-communicable diseases disproportionately burden transitioning communities.

1.4. Thesis structure

The thesis will follow a 'block' format, which is the conventional monograph format. The thesis will be divided into eight chapters, including Introduction, Literature review, Methods, Results, Discussion, Conclusion, References and Appendices. All the results will be grouped together into one 'block', while the references will be presented as the final item of the thesis, followed by relevant appendices. Additional data and figures will also be included as supplementary files and referenced accordingly. This format allows for a clear and organized presentation of the research, with a focus on the results and their interpretation.

2. Literature review

2.1. Overview of the gut microbiome

The human gut, specifically the colon, harbours the richest microbial density in the human body and is a source of microbes that colonize other body sites like the liver and mammary glands (Reynoso-García et al., 2022; Sender et al., 2016). The gut microbiome contains over 23 million genes, exceeding the human genome's gene count by 50- to 100-fold (Sender et

al., 2016; Tierney et al., 2019). The gut microbiome functions as a "second genome" for humans, encoding enzymes, proteins, and signalling molecules that regulate host physiology from energy production to inflammation across the body (Grice & Segre, 2012; Tierney et al., 2019). The establishment of this intricate microbial community begins at birth and develops rapidly during infancy.

2.1.1. The establishment of the gut microbiome

Although the establishment of the gut microbiome is still an active area of research (Aagaard, 2021), the current scientific consensus is that the gut is sterile in utero but becomes rapidly colonized by microbes at birth and during early infancy. Colonization involves four phases: 1) a sterile prenatal phase, though microbial DNA in placental tissues suggests possible in utero seeding; 2) acquisition from maternal and environmental microbes during delivery; 3) breastfeeding or formula feeding which differently diversify the microbiome; and 4) introduction of solids leading to an adult-like microbiota by 3 years old (Aagaard et al., 2014; Jiménez et al., 2008). Vaginal delivery seeds the newborn gut with maternal vaginal microbes, while C-section delivers skin microbes, originally thought to durably impact microbiome composition (Dominguez-Bello et al., 2010). However, after 6 weeks, delivery mode effects seem to subside (Chu et al., 2017). Generally, breastfed infants harbour more *Bifidobacteria*, which regulate immunity and early community structure, whilst formula feeding promotes more diverse gut microbes with lower *Bifidobacteria* (Laursen, 2021; Lay et al., 2005; Ruiz et al., 2017).

The first colonizers in infancy include *Bacteroides*, *Lactobacillus*, *Bifidobacterium*, *Escherichia*, and *Klebsiella* (Bäckhed et al., 2015; Chu et al., 2017; Koenig et al., 2011). They enrich metabolic pathways that utilize glucose and synthesize amino acids, cell wall components, and B vitamins to support growth, metabolism, and neurodevelopment in a critical developmental window (Chu et al., 2017; Kennedy, 2016). Though many studies focus on non-African populations, research across diverse and African-centric cohorts is needed to fully understand microbiome variability worldwide, which informs its composition and functional properties in geo-specific contexts across the individual lifespan.

2.1.2. Composition and function of the human gut microbiome

The composition and function of the human gut microbiome are shaped by a dynamic interplay of diet, lifestyle, genetics, and environmental exposures such as medications and hygiene practices (Gomez et al., 2016; Wu et al., 2011). While much of this understanding derives from Western populations, emerging research highlights distinct microbial signatures in African cohorts, reflecting unique dietary, socioeconomic, and environmental contexts (Brewster et al., 2019). For instance, African gut microbiomes often exhibit inverse gradients of *Bacteroides*

and *Prevotella* abundance, driven by protein / fat-rich versus plant-based diets – a pattern starkly contrasting with Western populations (Brewster et al., 2019). These differences underscore the critical need for region-specific studies, particularly in Africa, where factors like antibiotic overuse and variable sanitation practices disproportionately disrupt microbial equilibrium, favouring pathogenic taxa and xenobiotic metabolism (Hansen et al., 2019).

Globally, gut microbiomes cluster into three functional "enterotypes" dominated by *Bacteroides* (enterotype 1), *Prevotella* (enterotype 2), or *Ruminococcus* (enterotype 3), which are independent of geography, ancestry, or demographics (Arumugam et al., 2011). These enterotypes reflect specialized metabolic roles: *Bacteroides*-enriched communities prioritize carbohydrate and protein fermentation, while *Prevotella* and *Ruminococcus* dominate systems adept at degrading mucin glycoproteins (Derrien et al., 2004). Interactions between these taxa drive ecological equilibria, balancing competition and cooperation to sustain community structure (Mobeen et al., 2018). All enterotypes contribute to vitamin synthesis (e.g., biotin, riboflavin), though their relative metabolic strengths differ, emphasizing functional redundancy as a hallmark of microbiome resilience (Arumugam et al., 2011).

Functionally, the gut microbiome produces metabolites like short-chain fatty acids (SCFAs) and lipopolysaccharides, which regulate inflammation, immunity, and cross-organ communication (Louis et al., 2014). Despite taxonomic variability – dominated by Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria – functional stability often persists, suggesting metabolic pathways are conserved even as species turnover occurs (Sender et al., 2016). However, reduced microbial diversity, a hallmark of dysbiosis, is consistently linked to disease (Gacesa et al., 2022), underscoring the need to preserve microbial richness through context-specific interventions. In Africa, this requires reconciling urbanization’s dietary shifts with traditional practices to mitigate risks of non-communicable diseases. Therefore, understanding the factors that contribute to maintaining a diverse and balanced gut microbiome is crucial for promoting optimal health outcomes.

2.1.3. Intrinsic and extrinsic factors influencing gut microbiome diversity

The human gut microbiome is a complex and dynamic ecosystem, shaped by a multitude of interrelated environmental, dietary, and host-associated factors. Geography and location have an important influence, with significant differences observed between rural and urban areas (Yatsunenko et al., 2012; Oduaran et al., 2020; Maghini et al., 2025), as well as between different countries and regions globally (Schnorr et al., 2014). These geographical differences may be driven by variations in diet, cultural practices, sanitation, and other environmental exposures.

Diet is considered one of the most impactful factors modulating gut microbial diversity and function (O'Hara & Shanahan, 2006). High-fat, high-sugar Western diets associate with decreased diversity, reduced levels of beneficial microbes, increased pathogens, and higher risk of dysbiosis (Gacesa et al., 2022). In contrast, diets rich in fibre and plant polysaccharides promote microbial diversity, elevate beneficial microbes like *Bifidobacterium* and *Lactobacillus*, and support gut health (Aagaard et al., 2014). The dietary components provide selective metabolic nutrients favouring the growth of distinct microbial communities. Age is another key host factor affecting the developmental trajectory of the gut microbiome over the human lifespan. Infants harbour a relatively simple microbiome which diversifies over the first 2-3 years of life (Laursen, 2021; Yatsunenko et al., 2012). Aging leads to reduced diversity and altered composition, including decreased levels of health-promoting *Faecalibacterium* (Claesson et al., 2010).

Beyond intrinsic host factors, a variety of external exposures and interventions can disrupt the gut microbial ecosystem. The administration of antibiotics significantly depletes microbial diversity and alters community structure, with effects sometimes lasting months or years after treatment cessation (Dethlefsen & Relman, 2011). Non-antibiotic medications like proton pump inhibitors also impact the relative abundance of certain microbial taxa and promote dysbiosis (Reveles et al., 2018). Environmental contaminants such as heavy metals and pesticides have been linked to changes in the gut microbiome as well, including decreased diversity and richness (Vallès & Francino, 2018). Chronic psychological stress may also contribute to dysbiosis through effects on gastrointestinal physiology and immune function (Doroszkiewicz et al., 2021).

Lifestyle factors additionally intersect with microbial community structure and activity. Smoking, alcohol consumption, physical inactivity, disturbed circadian rhythms, and other unhealthy lifestyle choices have been associated with dysbiotic patterns in the gut microbiome (Khaledi et al., 2024; Wang et al., 2021). Chronic sleep deprivation may reduce microbial richness and diversity, while exercise appears to increase microbial diversity and elevate SCFA-producing bacteria (Wang et al., 2021; Yu et al., 2019). The multidimensional interplay between diet, medication, stress, toxins, lifestyle behaviours, and the gut microbiome remains incompletely characterized but highlights potential avenues for modulating the microbiome through lifestyle and environmental modifications.

The role of host genetics in shaping the gut microbiome remains an active area of investigation with evolving perspectives. Rothschild et al. (2018) conducted a landmark study of 1,046 healthy individuals with both metagenomic and genetic data, finding that the microbiome was not significantly associated with genetic ancestry and estimating overall microbiome

heritability at only 1.9%. Their analysis demonstrated that environmental factors explained approximately 20% of microbiome β-diversity variance, while genetics contributed minimally. Similarly, He et al. (2018) examined over 7,000 individuals across different regions in China and found that host location rather than genetics was the strongest determinant of microbiome variation, with microbiome-based disease models failing when applied across different geographical locations.

More recent work from the Dutch Microbiome Project (Gacesa et al., 2022) confirmed these patterns while reporting slightly higher heritability rates – about 6.6% of taxa showed significant heritability, though 48.6% of taxa variance was explained by cohabitation rather than genetics. These studies collectively employed different methodological approaches including twin studies, family structures, and population-level analyses, yet consistently demonstrate that while certain specific taxa show heritable patterns, the overall microbiome composition is predominantly shaped by environmental factors, diet, medication use, and shared household environments.

Despite progress, the relative contributions of host genetics versus environmental drivers remain poorly resolved. Targeted studies in underrepresented populations are critical to disentangle these interactions and advance personalized strategies for optimizing gut ecosystems. Prioritizing diet and lifestyle modifications offers a pragmatic approach to mitigating dysbiosis and its associated disease burden.

2.2. Gut microbiome and human health

The human gut microbiome plays a critical role in health. Dysbiosis, or imbalance of gut microbes, associates with diverse pathologies including immune disorders like asthma and inflammatory bowel disease, metabolic disorders like obesity and diabetes, and neurological conditions like Parkinson's disease (Frati et al., 2018; Musso et al., 2010; Scheperjans et al., 2015; Xavier & Podolsky, 2007). Dysbiosis also links to major African health challenges including infectious diseases, cardiovascular disease, cancer, and mental illness (Bhattacharjee & Lukiw, 2013; Feng et al., 2016; Koay et al., 2018; Liu et al., 2021). While further research is needed to establish causality, the gut microbiome shows potential as a biomarker or therapeutic target for major diseases in Africa and globally.

2.2.1. Gut-organ axes

The gut is not only responsible for the digestion and absorption of nutrients but also plays a vital role in the regulation of other organ systems in the body. Emerging evidence suggests that the gut microbiome interacts with multiple organ systems through neural, immune, endocrine, and metabolic pathways, termed gut-organ axes (Figure 2.1). The bi-directional communication along these axes allows gut microbes to influence physiology and disease at

both intestinal and extra-intestinal sites. Dysbiosis can disrupt gut-organ communication, contributing to systemic effects and disorders in distant tissues and organs, which will be explored further in the following sections.

The gut-brain axis and neurological health

The gut-brain axis integrates neurological, immunological, endocrine, and metabolic pathways to enable bi-directional communication between the gut microbiome and the brain. Research in germ-free mice demonstrates the microbiota's role in neurodevelopment, brain function, neurotransmitters, stress response, anxiety, and cognition (Luczynski et al., 2016). Western diets high in fat and sugar can disrupt gut microbes, increase intestinal permeability, drive inflammation, and compromise the blood-brain barrier, negatively impacting the gut-brain axis (Noble et al., 2017). Along with diet – stress, infections, and antibiotics also promote dysbiosis, contributing to neurodegenerative disorders (Doroszkiewicz et al., 2021).

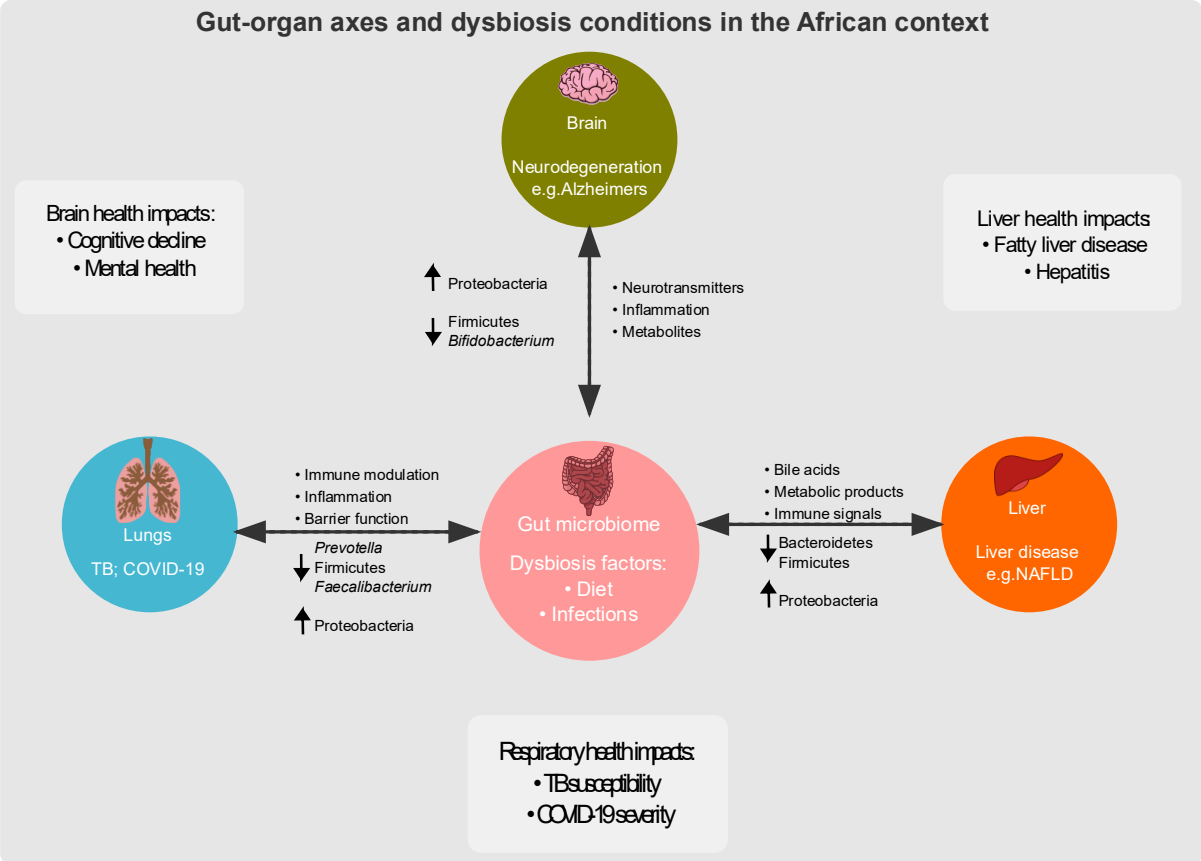


Figure 2.1 Overview of the gut-organ axes and their implications for various health conditions within the African context. This diagram shows relationships between the gut microbiome and the brain, lungs, and liver, highlighting how diet and infections influence conditions like Alzheimer's, TB, COVID-19, and liver diseases. Key microbial taxa linked to each health condition are also shown.

Links between the gut microbiome and psychiatric and neurological conditions like Alzheimer's, Parkinson's, depression, autism, and multiple sclerosis continue to emerge

(Doroszkiewicz et al., 2021; Sherwin et al., 2018; Westfall et al., 2017). Neurodegenerative diseases involve progressive loss of neurons, leading to long-term cognitive and motor impairments influenced by genetic and environmental factors. Alterations in the gut microbiome may impact neurodegenerative disease risk by modulating nervous system development and function (Fung et al., 2017).

In Alzheimer's disease, amyloid beta plaques may originate from the gut microbiome before translocating to the brain, suggesting a gut-brain axis role in pathogenesis (Sun et al., 2020). Alzheimer's patients exhibit decreased gut microbial diversity, including fewer Firmicutes and more Bacteroidetes (Vogt et al., 2017). Transferring healthy gut microbes to Alzheimer's model mice alleviates pathology and symptoms, indicating microbiome-based treatment potential (Kim et al., 2020). In Parkinson's disease, a prevalent neurodegenerative disorder, dual hits in the gut and nasal cavity may precede disease onset in the brain. Parkinson's patients frequently have earlier gastrointestinal dysfunction and distinct gut microbiota profiles, supporting a gut-brain connection (Fang et al., 2020; Petrov et al., 2017).

Human cell-based models like gut and brain organoids show promise for elucidating mechanisms of gut-brain crosstalk absent in animal models. With increased life expectancy in Africa to ~65 years (Macrotrends, 2025), the burden of age-related neurodegenerative disorders is rising. Although gut microbial signatures associate with amyotrophic lateral sclerosis, Parkinson's, and Alzheimer's (Fang et al., 2020; Peng et al., 2018; Petrov et al., 2017), transplanting beneficial microbes ameliorates neurodegenerative symptoms in animal models, highlighting the gut microbiome's therapeutic potential (Kim et al., 2020; Sun et al., 2018). This warrants exploration to address associated mental health disorders like anxiety and depression prevalent in Africa (*SADAG*, 2019).

The gut-liver axis and cellular models

The gut-liver axis enables critical bidirectional communication between the intestinal microbiome and the liver. The gut microbiome regulates liver function through production of metabolites like short-chain fatty acids that modulate hepatic metabolism and inflammation (Racanelli & Rehermann, 2006). Dysbiosis of the gut microbiome associates with chronic liver diseases including non-alcoholic fatty liver disease (NAFLD), alcoholic liver disease, cirrhosis, hepatitis C virus infection, and liver cancer (Inoue et al., 2022; Schneider et al., 2022; Wang et al., 2021).

Non-alcoholic fatty liver disease represents a major public health issue, affecting 30% of adults globally (Younossi et al., 2023). It encompasses progressive liver pathology from benign steatosis to cirrhosis and hepatocellular carcinoma. NAFLD links to altered gut microbial composition independent of obesity and insulin resistance, with depleted Bacteroidetes and

Firmicutes and increased faecal propionate and isobutyric acid (Da Silva et al., 2018). Obese individuals exhibit a Proteobacteria-enriched liver microbiome signature correlating with NAFLD severity, suggesting the hepatic microbiome as an early NAFLD risk factor (Suppli et al., 2021). The gut microbiome also correlates with the intrahepatic immune repertoire, both of which reduce in diversity during liver fibrosis progression (Liang et al., 2020). Distinct gut microbial and metabolic profiles mark NAFLD-cirrhosis patients developing hepatocellular carcinoma, highlighting the gut-liver axis in disease progression (Behary et al., 2021).

In vitro gut-liver models utilizing organoids or tissue engineering can capture inter-organ interactions beyond isolated cell models. Gut-liver micro-physiological systems emulate lipid accumulation in the liver following absorption by the gut epithelium (Lee & Sung, 2018), and interconnected gut-liver microfluidic devices model early NAFLD pathogenesis (Yang et al., 2020). Integrated gut-liver-immune-brain models will further elucidate mechanisms of multi-organ crosstalk relevant to health in Africa and globally (Hawkins et al., 2020; Trapecar et al., 2021). The microbiome-gut-liver axis represents a promising target for innovative diagnostics, preventatives, and therapeutics to address the increasing burden of chronic liver disease (Vento et al., 2018).

The gut-lung axis and respiratory health

The gut-lung axis describes bidirectional communication between the gut microbiome and respiratory system, mediated by immune modulation, microbial metabolites, and neuroendocrine pathways (Barcik et al., 2020; Dang & Marsland, 2019; Price & O'Toole, 2021). This axis is particularly relevant to respiratory diseases prevalent in Africa, including tuberculosis (TB) and coronavirus disease (COVID-19), where gut dysbiosis may exacerbate systemic inflammation or impair host immunity.

Tuberculosis (TB) is caused by *Mycobacterium tuberculosis* (*M. tb*), but emerging evidence suggests gut microbiome composition influences disease susceptibility and progression. For instance, mice infected with *M. tb* exhibit rapid gut microbial depletion (e.g., reduced *Bacteroidetes* and *Firmicutes*), which correlates with impaired pulmonary immune responses and increased bacterial burden (Winglee et al., 2014). In humans, TB-associated gut dysbiosis may indirectly affect lung outcomes: severe pulmonary inflammation alters the respiratory microenvironment, leading to sputum microbiome shifts dominated by Proteobacteria and Bacteroidetes (Cheung et al., 2013; Cui et al., 2012). These changes likely reflect secondary inflammatory damage rather than direct gut-lung interaction but highlight the systemic nature of TB pathology.

The SARS-CoV2 disease further highlights the gut-lung axis's clinical importance. Gut microbiome composition predicts disease severity and post-acute sequelae (PACS), with

reduced diversity and depletion of beneficial taxa (e.g., *Faecalibacterium*) correlating with prolonged symptoms (Lau et al., 2022; Liu et al., 2021; Yeoh et al., 2021). Mechanistically, gut bacteria like *Firmicutes* may enhance intestinal ACE2 expression, a receptor for SARS-CoV-2, potentially facilitating viral entry while modulating immune responses (Zuo et al., 2020). Bacterial metabolites, such as secondary bile acids, also exhibit antiviral activity against SARS-CoV-2 *in vitro*, suggesting therapeutic potential (Piscotta et al., 2021). However, long-term impacts of pandemic measures (e.g., sanitation, antibiotic overuse) on early-life microbiome establishment require urgent study (Campbell et al., 2021; Chevallier et al., 2021).

Innovative models integrating gut-lung interactions are critical for unravelling these mechanisms. Lung cell lines, air-liquid interface cultures, and organoids replicate respiratory anatomy and physiology (Baptista et al., 2022; Varghese et al., 2022), while co-culture systems with gut microbiota could elucidate how microbial metabolites (e.g., short-chain fatty acids) regulate lung immune cells. Such models are vital for studying African populations, where genetic, dietary, and environmental factors shape unique gut-lung dynamics.

2.2.2. Gut microbiome pathologies

Research continues to uncover links between dysbiosis or alterations in the microbiome and a wide range of pathological conditions. As alluded to earlier, the gut microbiome in particular has been implicated in common NCDs like obesity, diabetes, cancer, as well as infectious diseases such as HIV. This section will examine some of the key associations found between gut microbiome dysbiosis and pathologies highly relevant to African populations. The involvement of the gut microbiome in such diverse conditions further supports the need to develop therapeutic interventions targeting the microbiome in precision medicine approaches, as will be discussed in the following section.

Obesity

The prevalence of obesity in African populations doubled between 1990 and 2016, affecting over 10 million individuals (Biadgilign et al., 2017; World Health Organization, 2016). Obesity has been directly linked to various non-communicable diseases (NCDs) including type II diabetes, which has increased by over 120% in Africa since 1980 and is predicted to cost $60 billion United States Dollars (USD) in sub-Saharan Africa alone by 2030 (Jaffar & Gill, 2017). Based on recent estimates, overweight and obesity cases are costing South Africa's health system R33 billion (USD 1.9bn) a year, which represents 15.38% of government health expenditure and is equivalent to 0.67% of gross domestic product (GDP) (Boachie et al., 2022).

Clinical and experimental studies in human cohorts have established correlations between gut microbiota and obesity, with baseline microbial composition directly impacting insulin

sensitivity, halting progression of type I diabetes, and other outcomes (de Groot et al., 2021; Kootte et al., 2017). Longitudinal studies in the UK and USA have also shown an increased risk of neurocognitive deficits and autism spectrum disorders (ASDs) associated with maternal obesity (Basatemur et al., 2013; Pugh et al., 2015). Murine models revealed maternal high-fat diets (mHFD) induced long-term cognitive deficits across generations (Sarker & Peleg-Raibstein, 2018). This was confirmed by evidence that maternal obesity disrupted offspring gut microbiomes, strongly associated with cognitive and social dysfunctions (Liu et al., 2021). Given over 20% of women above 20 years old in South Africa are obese (Lobstein & Brinsden, 2017), this could pose a serious long-term and generational health challenge.

Microbiome-linked therapies have provided promising interventions, using high-fibre diets to produce microbiota-derived SCFAs that alleviate behavioural deficits via the gut-brain axis (Liu et al., 2021). Mouse models have demonstrated *A. muciniphila's* inverse correlation with obesity, restoring damaged gut structure and function while improving hepatic function, insulin resistance, and glucose homeostasis in high fat-diet (HFD) mice (de Vos, 2017; Everard et al., 2013; Yang et al., 2020). In addition, its use as a probiotic in humans with metabolic disorders has potential as an accessible precision medicine therapy (Depommier et al., 2019).

Infectious diseases – HIV and TB

Human Immunodeficiency Virus (HIV) remains a global crisis, with ~39.9 million infected in 2023 and more than 30% untreated (UNAIDs, 2023). While vaccines are tested, the gut's role in HIV creates opportunities to improve treatment, as infection associates with chronic gut inflammation and microbiota changes (Brenchley & Douek, 2008; Leggat et al., 2022; Lozupone, Li, et al., 2013). Recent large-scale African cohort studies, such as Maghini et al. (2025), confirm that HIV infection significantly reduces gut microbiome alpha-diversity and alters beta-diversity, with 131 prokaryotic taxa differentially abundant in people living with HIV (PLWH). Notably, *Faecalibacterium prausnitzii* – a butyrate producer linked to reduced inflammation – is depleted in PLWH, alongside novel taxa like *Dysosmobacter welbionis* and *Enterocloster sp*., which may reflect region-specific microbial adaptations. Delineating differences in microbiomes of people on antiretrovirals (ARVs) versus healthy individuals, such as increased *Fusobacterium* (contrasting with Maghini et al.'s findings of *Fusobacterium* depletion in PLWH, potentially due to ART or regional variability), *Prevotella*, and *Megasphaera*, could support microbiome-directed therapies (Meng et al., 2022). Machine learning models further demonstrate the gut microbiome's potential as an HIV biomarker, accurately distinguishing PLWH across diverse African cohorts (Maghini et al., 2025).

Infectious diseases have a significant burden in Africa, with HIV/AIDS and tuberculosis (TB) as top mortality causes (Statista, 2021). Despite promising vaccine evidence like the RV144

Thailand trial, contrasting results were seen in South Africa's HVTN 702 trial (Gray et al., 2021; Rerks-Ngarm et al., 2009). However, modulating microbiomes to enhance immune responses, for example via probiotics shown to decrease HIV-targeted cells in macaques (Klatt et al., 2021), or restoring taxa like *F. prausnitzii*, remains a promising therapeutic avenue.

Tuberculosis (TB) is the leading cause of death in South Africa, with 56,000 deaths in 2023, including 31,000 HIV-co-infected individuals (World Health Organization, 2024). While studies find contrasting sputum microbiomes in TB patients versus healthy individuals (Cheung et al., 2013; Cui et al., 2012), limited consistency of a TB-specific microbial signature has been shown in African populations. For example, the most recent and largest African study by Maghini et al. (2025) observed no significant gut microbiome associations with TB co-infection in PLWH, highlighting the need for further research into TB-specific microbial signatures in African populations. Contrastingly, research by Naidoo et al., 2021 showed TB-specific microbial related pro-inflammatory responses in pulmonary TB patients. Other prior work links *M. tuberculosis* infection to gut microbiota alterations in mice (Winglee et al., 2014), and metabolomics studies support the microbiome's role in TB treatment outcomes (Das et al., 2015; Luies et al., 2017; Yu et al., 2023). These findings strongly support continued pharmacomicrobiomics research in Africa to address the dual burden of NCDs and infectious diseases.

Cancer

Metagenomics continues to revolutionize cancer research by targeting microbial diversity (Banerjee et al., 2015). However, the microbiome's precise role in cancer aetiology and treatment is complex and disputed (Lee et al., 2021; Sepich-Poore et al., 2021). The gut microbiome associates with numerous cancers, especially colorectal cancer (CRC) (Lee et al., 2021). Comparisons of African stool microbiomes between healthy individuals and colorectal cancer (CRC) patients in Morocco revealed significant differences in bacterial profiles and abundance (Allali et al., 2018). Similar research found more tumour load and bacterial composition/abundance in African versus non-African cancer specimens, suggesting microbiome-driven carcinogenesis may worsen prognoses for Africans (Feng et al., 2019). Differences between African-Americans and European-ancestry individuals also suggest microbial dysbiosis may contribute to CRC racial disparities (Farhana et al., 2018). These initial African studies highlight important host-microbiome relationships to incorporate in cancer therapies. Thus, understanding the immuno-oncology-microbiome axis will require large, multi-site, longitudinal cohorts to identify cancer-associated microbiota (Sepich-Poore et al., 2021).

While research on microbiome-based therapeutics is still an emerging field, current evidence supports the microbiome's role in many diseases and the need to further develop precision interventions tailored to an individual's microbiome profile. The following section will provide an overview of current and potential microbiome-based therapeutic interventions relevant to African contexts.

2.2.3. Gut microbiome and therapeutic interventions

The gut microbiome's profound influence on human health – spanning metabolic disorders, infectious diseases, and cancer – has positioned it as a critical therapeutic target. Dysbiosis is increasingly implicated in disease pathogenesis, but emerging research highlights opportunities to harness microbiome-directed therapies, such as probiotics, postbiotics, and faecal microbiota transplants (FMTs), to restore microbial equilibrium and improve clinical outcomes (Figure 2.2). Beyond treatment, the microbiome's role in drug metabolism and immune modulation highlights its potential to shape personalized diagnostics and precision medicine (Wang et al., 2024; Zhao et al., 2023).

These insights into the microbiome's clinical impact are further substantiated by groundbreaking research on microbiome-drug interactions. Zimmermann et al. (2019) comprehensively mapped how gut bacteria metabolize medications, demonstrating that approximately two-thirds of tested drugs were modified by at least one bacterial strain and identifying specific microbial gene products responsible for these transformations. This metabolic capacity extends beyond direct drug modification to profound immunomodulatory effects, as evidenced by Viaud et al. (2013), who discovered that cyclophosphamide's anticancer efficacy partially depends on specific gut bacteria that enhance antitumor immune responses by promoting T helper 17 cell development. Similarly, Gopalakrishnan et al. (2018) revealed that melanoma patients with diverse gut microbiomes, particularly those enriched in *Faecalibacterium* species, showed significantly improved responses to anti-PD-1 immunotherapy compared to non-responders whose microbiomes were dominated by Bacteroidales.

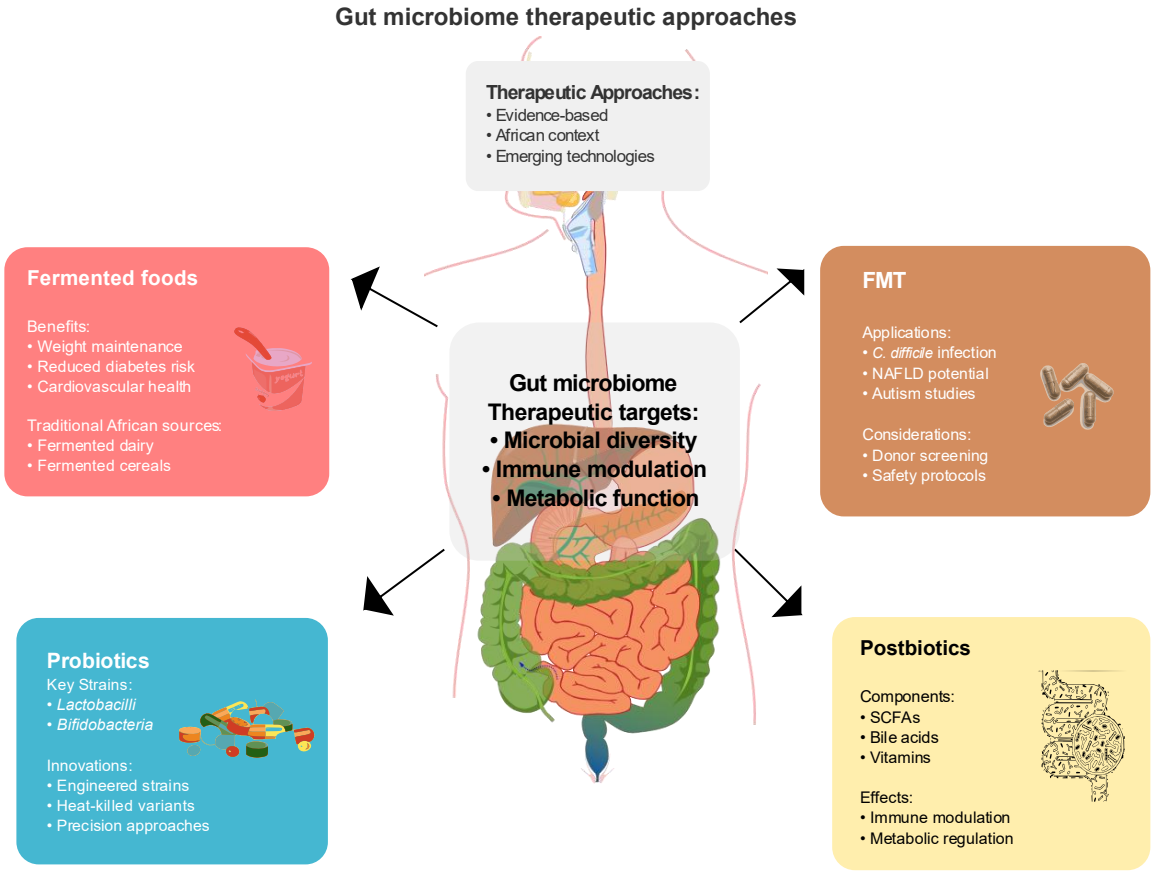


Figure 2.2 Overview of various gut microbiome therapeutic approaches targeting the gut microbiome to enhance health outcomes. This figure depicts key interventions – fermented foods, probiotics, faecal microbiota transplantation (FMT), and postbiotics – and their benefits, applications, considerations and effects.

Collectively, these studies highlight the gut microbiome's dual capacity to directly transform therapeutic compounds and fundamentally shape immune responses to treatment, with genetic profiling of key bacterial species offering potential predictive biomarkers of drug efficacy. However, translating this promise into practice requires deeper insights into disease-specific microbiome signatures, mechanisms of host-microbe interaction, and the drivers of inter-individual variation. In this context, we evaluate therapeutic strategies that aim to leverage the microbiome's dual role as a biomarker and modulator of health.

Fermented foods

Fermented foods, staples in many African diets, include popular South African foods such as sour milk and maize porridge (FACTS, 2019). These have gained popularity through research highlighting their potential health benefits including weight maintenance and decreased risk of diabetes, cancer, and CVD in animal models and humans. In addition, fermented foods may

act as modulators of the human microbiome-immune axis, promoting resistance to NCDs (Díaz-López et al., 2016; Dimidi et al., 2019; Gille et al., 2018; Villarreal-Soto et al., 2018). The SARS-CoV2 (COVID-19) pandemic has brought renewed interest in these foods as possible sources of beneficial microbes. Specifically, they could provide compensatory exposure to safe foodborne microbes eliminated by industrialization and sanitizing the industrial environment (Carter et al., 2023). In this way, fermented foods may act as precision microbiome modulators, enabling resistance to chronic disease through their impact on the microbiome-immune system axis (Wastyk et al., 2021).

Faecal Microbiota Transplantation (FMT)

Faecal Microbiota Transplantation (FMT), a precise strategy for treating gut-associated infections and dysbiosis, involves the transfer of faecal matter from a healthy individual to a patient with a less diverse or depleted microbiota (Khanna et al., 2017; Weingarden et al., 2015). Such a treatment strategy paved the way for the development of defined bacterial cocktails and has shown promising results in the treatment of recurrent *Clostridioides difficile* infection (CDI) and may hold potential for treating other diseases (Lam et al., 2019; Munoz et al., 2016).

In November 2022, Rebyota, an FMT-based therapy, became the first of its kind to gain federal approval in the USA for treating CDI (FDA, 2022). Although the specific mechanism of action of FMT is not fully understood, evidence suggests that it may restore a patient's gut microbiome profile to a functionally balanced and healthy state (Bajaj et al., 2022). Furthermore, FMT has been studied for its potential in treating NAFLD and improving the prognosis of alcohol-associated hepatitis (Craven et al., 2020; Philips et al., 2017, 2018). In addition, FMT has shown promise in reducing autism spectrum disorder symptoms and gastrointestinal issues in children (Kang et al., 2019; N. Li et al., 2021). However, FMT is not without risks, such as the potential transmission of disease and adverse effects, necessitating stringent donor screening protocols (Bajaj et al., 2022; Lang & Schnabl, 2020).

Probiotics

Probiotics, live microorganisms that can alter the gut's ecological profile to promote health, are considered functional food products (Kuntz & Gilbert, 2017; Sanders et al., 2018). Traditional probiotic strains, such as *Lactobacilli* and *Bifidobacteria*, have been safely used for long periods to regulate intestinal flora, maintain the epithelial barrier, inhibit pathogens, and modulate and mature the immune system (Wang et al., 2021; Wieërs et al., 2019). However, the emerging biotechnological landscape aims to incorporate the rational design of bacterial strains with tailored probiotic properties (Douillard & de Vos, 2019). Gene editing technologies

further enable the development of engineered probiotics, which are have ongoing clinical trials for various diseases (Ma et al., 2022).

Due to some safety concerns regarding live probiotics, scientists tested and discovered advantages in using heat-killed bacteria, which retain key probiotic effects with favourable properties observed in animal models, *in vitro*, and clinical trials (Yang et al., 2020). Scientists have proposed using multi-omics frameworks and machine learning algorithms to identify correlations and causal evidence of disease through targeted addition or removal of microorganisms as probiotics. This precision medicine approach could achieve optimal health by predicting individualized responsiveness to dietary components (Mills et al., 2019).

Innovative use of probiotics in humans with metabolic disorders and obesity has shown potential as an accessible precision medicine therapy (Aoun et al., 2020). Promising animal models have also shown microbiota transplantation and probiotic treatment can ameliorate neurodegenerative disease pathology and behaviour (Bonfili et al., 2017; Kim et al., 2020a). As microbiomics and gut-organ axes are further elucidated, opportunities will continue expanding for health-promoting bacterial species.

Postbiotics

Postbiotics refer to the metabolites and cell components produced by microorganisms that confer health benefits to the host. (Agus et al., 2021; Krautkramer et al., 2021). As probiotics are living microorganisms that promote gut health when consumed, postbiotics are the non-viable cellular products and metabolites remaining after probiotic fermentation. Of interest are bile acids, tri-methylamine oxide, tryptophan derivatives, SCFAs, vitamins, enzymes, and cell wall fragments (Agus et al., 2021). Postbiotics exhibit pleiotropic properties including immunomodulation, infection prevention, and mucous production stimulation (Mosca et al., 2022; Żółkiewicz et al., 2020). The pleiotropic effects arise from the diversity of these postbiotic origins and components (Mosca et al., 2022).

Increased postbiotic intake can be achieved through greater consumption of fermented foods like kefir, tempeh, and kimchi. The diverse components promote a healthy gut microbiome, thereby benefiting overall health (Golen, 2021). In animal models, *Lactobacillus paracasei*-derived postbiotics slowed weight gain, suppressed insulin resistance, and decreased hepatic steatosis and inflammation, suggesting potential for managing liver disease progression (Pan et al., 2022). Through microbiome-host interactions, postbiotics influence immunity, metabolic function, and pathogen resistance (Krautkramer et al., 2021).

Clinical trials have provided evidence for microbial postbiotic therapeutics in treating gastrointestinal, respiratory, immune, stress, and metabolic disorders (Mosca et al., 2022). As research continues elucidating mechanisms, postbiotic applications will likely expand. Overall,

interventions that leverage the gut microbiome hold significant promise as therapeutic strategies in the African context and globally. However, these techniques require further research to ensure their safety, efficacy, and customization to specific populations and disease states.

2.3. Gut microbiome studies in South African populations

While previous research has provided initial insights into the gut microbiome in South African populations, critical knowledge gaps remain regarding the microbial diversity and its determinants within specific groups. Much of the existing work has focused on rural inhabitants or people living with HIV (Claassen-Weitz et al., 2018; Hansen et al., 2019; van Zyl et al., 2020). However, few studies have characterized the gut microbiome and its drivers in urban South African populations (Ramaboli et al., 2024). Diet, lifestyle, and socioeconomics shape the microbial community, yet their impacts remain underexplored in this key demographic. This highlights the need for research examining the gut microbiome and its associations with these factors in urban South Africans. The following sections will summarize findings from previous microbiome research in South Africa and highlight critical gaps, motivating this PhD research project to further address unanswered questions regarding the gut microbiome and its determinants in this understudied cohort.

2.3.1. Previous research findings

Various studies have been conducted to understand gut microbial composition and function in different populations, including South Africa, revealing fascinating findings that contribute to the global understanding of the gut microbiome. These studies have been identified and referenced in Table 2, highlighting the focus of the research, the type of sample and methodology used, the number of participants, and key findings. This table was adapted from (Allali et al., 2021) which thoroughly reviews additional studies conducted on the broader African continent.

The AWI-Gen 2 Microbiome Project, recently reported by Maghini et al., (2025), represents a landmark advancement in African microbiome research. This cross-sectional study of 1,801 women from six research centres across four African countries (including three South African sites: Agincourt, DIMAMO, and Soweto) constitutes the largest population-representative survey of African gut microbiomes to date. Using shotgun metagenomic sequencing, the researchers identified significant taxonomic differences between rural, peri-urban, and urban populations, with geography emerging as the strongest factor influencing microbiome composition. The study's most remarkable contribution is its extensive catalogue of novel microbial diversity: 1,005 previously unknown bacterial metagenome-assembled genomes (MAGs) and over 40,000 new viral genomes, demonstrating that African populations harbour

substantial microbiome diversity, absent from existing reference collections. The researchers also observed distinctive HIV-associated microbiome signatures that differed from those reported in high-income countries, including associations with bacterial species not previously linked to HIV infection. Importantly, the study established that microbiome sampling in these populations had not reached saturation, indicating that further research will continue to uncover new microbial diversity. This work not only significantly expands our understanding of the South African gut microbiome but also establishes a framework for equitable, community-engaged microbiome research that can inform future studies across the continent.

Overall, these studies highlight the complex and multifaceted role of the gut microbiome in health and disease in South African populations. They emphasize the need for continued research in this area, considering the unique environmental, dietary, and cultural factors that characterize these populations.

Table 2.1 Summary of gut microbiome studies conducted in South African populations

SAMPLE ORIGIN (COUNTRY)	AFFILIATION (COUNTRY)	FOCUS	SAMPLE TYPE	METHODOLOGY	NUMBER OF PARTICIPANTS	FINDINGS	REFERENCE
SOUTH AFRICA	Germany, South Africa	Respiratory, gastrointestinal, and other diseases	Stool	16S rRNA, V3–V4 regions, Illumina Miseq	34	Preliminary evidence suggests faecal dysbiosis in ill young children.	(Krishnamoorthy et al., 2020)
SOUTH AFRICA	South Africa, USA	Atopic dermatitis	Stool	16S rRNA, V4 region, Illumina MiSeq	38	No significant microbial diversity differences were observed between children with atopic dermatitis (AD) and control children.	(Mahdavinia et al., 2017)
SOUTH AFRICA, USA	Germany, South Africa, UK, USA	Colorectal cancer	Stool	16S rRNA, V4 region, Illumina MiSeq	21	Alaskan Native diet, carcinogen exposure associated with tumour-promoting colonic milieu.	(Ocvirk et al., 2020)
SOUTH AFRICA, USA	South Africa, The Netherlands, USA	Colon cancer risk	Stool	16S rRNA, 454 pyrosequencing	12	Results suggest gut microbiota's metabolite balance affects colon cancer risk.	(Ou et al., 2013)
SOUTH AFRICA	South Africa, USA	HIV	Stool	16S rRNA, V4 region, Illumina MiSeq	155	Non-exclusive breastfeeding modifies gut microbiota, potentially affecting HIV transmission.	(Wood et al., 2018)
SOUTH AFRICA	South Africa, USA	None	Stool	16S rRNA, V4 region, Illumina MiSeq	197	Infants' meconium exhibited elevated levels of Proteobacteria, particularly within the Enterobacteriaceae family.	(Claassen-Weitz et al., 2018)
UGANDA	The Netherlands, Norway, South Africa, Uganda	Malnutrition	Stool	16S rRNA	147	Maternal education intervention positively impacted child development and growth; unaffected gut microbiota.	(Atukunda et al., 2019)

ZIMBABWE	Australia, Norway, South Africa, UK, Zimbabwe	HIV	Rectal swabs	16S rRNA, V4 region, Illumina MiSeq	280	HIV-infected children exhibit altered gut microbiota, potentially restored by antiretroviral therapy.	(Flygel et al., 2020)
BURKINA FASO	Burkina Faso, Germany, Niger, South Africa, USA	Antibiotics	Rectal swabs	16S rRNA, V3–V4 regions, Illumina	62	No evidence of antibiotic impact on cohabiting children was found.	(Oldenburg et al., 2019)
BURKINA FASO	Burkina Faso, Germany, South Africa, USA	Antibiotics	Rectal swabs	16S rRNA, V3–V4 regions, Illumina	115	Azithromycin alters paediatric intestinal microbiome composition, while impact of amoxicillin and cotrimoxazole is uncertain.	(Oldenburg et al., 2018)
KENYA	Kenya, South Africa, Switzerland, The Netherlands	None	Stool	16S rRNA, V3–V6 regions, 454 pyrosequencing	115	Iron-containing Microencapsulated Nutrient Powders (MNPs) given to weaning infants disrupt the gut microbiome, promoting pathogens and inflammation.	(Jaeggi et al., 2015)
SOUTH AFRICA	South Africa, USA	None	Stool	Short and long-read WGS, HiSeq4000, Illumina, Nanopore sequencing, ONT	169	The gut microbiome of South Africans does not conform to a simple western-nonwestern axis.	(Tamburini et al., 2022)
SOUTH AFRICA	South Africa, USA	Obesity	Stool	16S rRNA, V3-V4 regions, Illumina MiSeq	170	The overall gut microbiome of our cohorts was reflective of their ongoing epidemiological transition.	(Oduaran et al., 2020)
SOUTH AFRICA	South Africa	None	Stool	16S rRNA, V3-V4 regions, Illumina MiSeq, ITS region, Qiagen	100	Geographic location and lifestyle factors are major drivers of gut mycobiota in rural and urban African populations.	(Kabwe et al., 2020)
SOUTH AFRICA	South Africa	Infant gut colonization	Stool, meconium, breastmilk	16S rRNA, V4–V5 regions, Illumina MiSeq	20 dyads	Breastmilk vital for seeding infant gut microbiome, distinct communities in South Africa.	(Wallenborn et al., 2022)

GHANA, SOUTH AFRICA, SEYCHELLES, JAMAICA, USA	USA, South African, Ghana, Jamaica, Seychelles	Toxic metals exposure and cardiometabolic risk	Stool	16S rRNA, V4 region, Illumina NovaSeq 6000	178	Lead and arsenic exposure associated with significant beta diversity shifts; numerous taxa linked to high metals also associated with obesity/T2DM markers; porphyrin metabolism enriched among metal-associated taxa.	(Jorgenson et al., 2024 - preprint)
GHANA, SOUTH AFRICA, SEYCHELLES, JAMAICA, USA	USA, South African, Ghana, Jamaica, Seychelles	Faecal SCFAs and adiposity across the urban epidemiological transition	Stool	16S rRNA, V4 region, Illumina NovaSeq 6000	1,867	Gut diversity and SCFA profiles varied by site and adiposity; lower microbial co-occurrence stability and altered SCFA patterns correlated with higher adiposity across the transition.	(Ecklu-Mensah et al., 2023)
GHANA, SOUTH AFRICA, JAMAICA, USA	USA, South Africa, Ghana, Jamaica, Switzerland	Sleep duration	Stool	16S rRNA, V4 region, Illumina MiSeq	655	Short sleepers showed lower diversity and enrichment of lysine-pathway butyrate genes; distinct taxa and network structures in short and long sleepers versus normal.	(Fei et al., 2021)
GHANA, SOUTH AFRICA, JAMAICA, USA	USA, South Africa, Ghana, Jamaica, Switzerland	Cardiometabolic (CM) risk across the urban epidemiological transition	Stool, saliva	16S rRNA, V4 region, Illumina MiSeq	655	Higher CM risk linked to lower gut alpha diversity and distinct beta diversity; enrichment of Lachnospiraceae and oral *Streptococcus/Prevotella/Veillonella* with CM risk; predicted butyrate amino acid pathways and LPS genes enriched with higher CM risk; geography modified associations.	(Fei et al., 2019)
SOUTH AFRICA	South Africa, USA, Germany	Diet	Stool, food, hand	16S rRNA, V3-V4 regions, Illumina MiSeq	44	Urban Xhosa gut microbiome less diverse; shifts towards Western-diet associated taxa increases colorectal cancer risk; elevated deoxycholic acid in faeces; distinct faecal virome and microbiota differences between rural and urban environments.	(Ramaboli et al., 2024)
BURKINA FASO, GHANA, KENYA, SOUTH AFRICA	USA, South Africa, Ghana, Burkina Faso, Kenya	Composition and diversity across a range of rural, peri-urban, and urban African populations; HIV	Stool	Short-read WGS, NovaSeq 6000 platform, Illumina	1,801	Geographic origin primarily shapes microbiome composition; identified 1,005 new prokaryotic African-related MAGs; *Treponema succinifaciens* varies with urbanization and antibiotic use; potential of microbiome-based HIV biomarkers.	(Maghini et al., 2025)

2.3.2. Gaps in knowledge

The gut microbiome has been extensively studied in various populations worldwide, including South African populations. However, there are still gaps in knowledge that need to be addressed to better understand the gut microbiome's role in human health and disease. One area that requires further investigation is the relationship between the gut microbiome and clearly delineated, high-priority conditions in adults such as cardiometabolic risk factors (obesity, insulin resistance, hypertension, dyslipidaemia), gastrointestinal functional disorders and diet-related intolerances (e.g., IBS, lactose and gluten intolerance), allergies, and colorectal cancer risk pathways.

Research on the gut microbiome's relationship to disease has yielded inconsistent findings, highlighting these knowledge gaps. For instance, studies in South African populations have suggested associations between gut microbiome composition and respiratory and gastrointestinal diseases in children in align with global research (Krishnamoorthy et al., 2020). However, when examining other conditions like atopic dermatitis, researchers found no significant differences in microbial diversity between affected and healthy children, despite correlative findings in other populations (Kim & Kim, 2019; Mahdavinia et al., 2017). Similarly, while evidence suggests potential links between gut microbiome composition and colon cancer risk (Ocvirk et al., 2020; Ou et al., 2013), these relationships may vary across populations. Further research, particularly in adult South African populations, is needed to fully characterize the gut microbiome's role the above cardiometabolic, gastrointestinal, allergic, and colorectal neoplasia-related pathogenesis and to determine whether these associations could inform prevention and treatment strategies.

Another area that requires further investigation is the impact of antibiotics and dietary supplements on the gut microbiome. While studies have shown that azithromycin alters the paediatric intestinal microbiome composition (Oldenburg et al., 2018, 2019), the impact of other antibiotics like amoxicillin and cotrimoxazole in adults uncertain. Similarly, dietary supplements like iron-containing Microencapsulated Nutrient Powders (MNPs) have been found to disrupt the gut microbiome, promoting pathogens and inflammation, yet are a necessity to combat malnutrition (Jaeggi et al., 2015). Further research is needed to understand the long-term effects of these interventions, and the potential of alternative safer approaches, on the gut microbiome and human health. In urban professional settings, additional common exposures (e.g., supplement use, alcohol consumption, and household energy sources such as gas heating) also warrant systematic evaluation for their links to xenobiotic metabolism functional pathways.

Furthermore, while studies have explored the gut microbiome's role in diseases such as TB, COVID-19 and HIV transmission and progression (Flygel et al., 2020; Nagata et al., 2022; Wood et al., 2018), the impact of other infectious diseases on the gut microbiome is still unclear. For adult urban cohorts, however, the most immediately relevant disease domains remain noncommunicable and diet-linked

gastrointestinal conditions as outlined above. Additionally, while the use of relatively novel long-read sequencing technology has been applied to an urban South African population to explore the gut microbiome's composition and function (Tamburini et al., 2022), further studies are needed to compare the results obtained from this technology with those obtained from alternatives (e.g. PacBio long-read sequencing) and traditional 16S rRNA sequencing. Moreover, while studies have explored the infant gut microbiome's initial colonization and development (Claassen-Weitz et al., 2018; Wallenborn et al., 2022), the long-term impact of environmental factors such as geography, lifestyle, and education levels on the gut microbiome is still unclear within context of the South African landscape. Further research is needed to better understand the gut microbiome's role in shaping human health across South African and other diverse populations.

In addition to the gaps in knowledge mentioned above, it is also important to note the context of this research project. Urbanization in Africa is accelerating, yet its impacts on gut microbial ecology in transitional populations – particularly urban professionals navigating hybrid lifestyles – remain poorly characterized. While rural African microbiomes have been studied extensively, urban cohorts like South Africa’s professionals represent a critical gap, as their microbial communities lie at the intersection of traditional diets, processed food consumption, and socioeconomic stressors. The mechanisms by which urbanization drives taxonomic volatility while maintaining functional resilience, or how demographic factors like ancestry mediate microbial adaptation to dietary transitions, are unresolved. Furthermore, socioeconomic inequalities, deeply entrenched in South Africa’s history, likely modulate microbiome-host interactions, but these dynamics remain unexplored in populations with overlapping access to urban resources and persistent environmental exposures. Accordingly, urban professional South African populations offer a tractable setting to develop hypothesis-driven studies that move beyond correlation toward causal frameworks (e.g., prospective designs and multi-omics) within the disease areas specified above.

Current microbiome research suffers from geographic and methodological inequities. Reference databases disproportionately represent Western populations, limiting the identification of regionally relevant taxa or biomarkers tied to African health contexts (Maghini et al., 2025). Methodologically, while short-read sequencing has dominated microbiome studies, its current limited capacity to resolve strain-level diversity or distinguish misclassified taxa in such underrepresented groups underscores the need for integrative approaches combining long-read and shotgun metagenomics. Additionally, the functional implications of intra-individual microbiome variability – a hallmark of transitional populations – are poorly understood, hindering the translation of microbial ecology into actionable public health strategies.

This study aims to address these gaps by focusing on urban professionals, a cohort characteristic of Africa’s rapid urbanization. By examining how demographic factors interact with dietary, occupational,

health and lifestyle factors, this work contributes to characterizing and understanding microbial adaptation in transitional societies. Crucially, it challenges the extrapolation of Western- and Euro-centric microbial norms to African populations, advocating for equitable, context-driven research paradigms. Ultimately, this work seeks to contribute to our understanding of gut microbial diversity within an understudied population group within South Africa.

3. Materials and methods

The methods outlined in this study describe a thorough scientific analysis which aimed to characterize relationships between various anthropometric and health parameters and the gut microbiome, in a South African urban professional cohort (SAN). The methodological framework detailed here in Chapter 3 is summarized in Figure 3.1 below. This encompassed comprehensive participant recruitment strategies (Chapter 3.1) and in-depth data collection surveying including anthropometric, dietary, Socioeconomic status (SES) and health assessments. All research activities were in line with the institutional research ethics committee approval (Chapter 3.2) and were compliant with the requirements of the Protection of Personal Information Act (PoPIA). These initial steps set the foundation for the subsequent collection and processing of stool samples (Chapter 3.3), which were analysed using gold standard and third generation microbiome sequencing techniques; specifically, Illumina shotgun metagenomic and PacBio 16S long-read sequencing (Chapter 3.4).

The bioinformatics pipeline (Chapter 3.5), optimised to ensure high-fidelity data interpretation, included the Quantitative Insights into Microbial Ecology 2 (QIIME2) workflow for 16S rRNA gene sequencing data and a Metagenome-Assembled Genomes (MAG) workflow for metagenomic data, complemented by a comparative analysis which validated and integrated findings from both sequencing approaches (Chapter 3.6). The profiles derived from these analyses were compared with a longitudinal subset of samples to investigate the temporal variability of the gut microbiome (Chapter 3.7). This was followed by comparing these overall community patterns to established global population cohorts, to contextualize the SAN population within broader human gut microbiome diversity patterns (Chapter 3.8) and exploring the uncovered intra-cohort microbial patterns (Chapter 3.9).

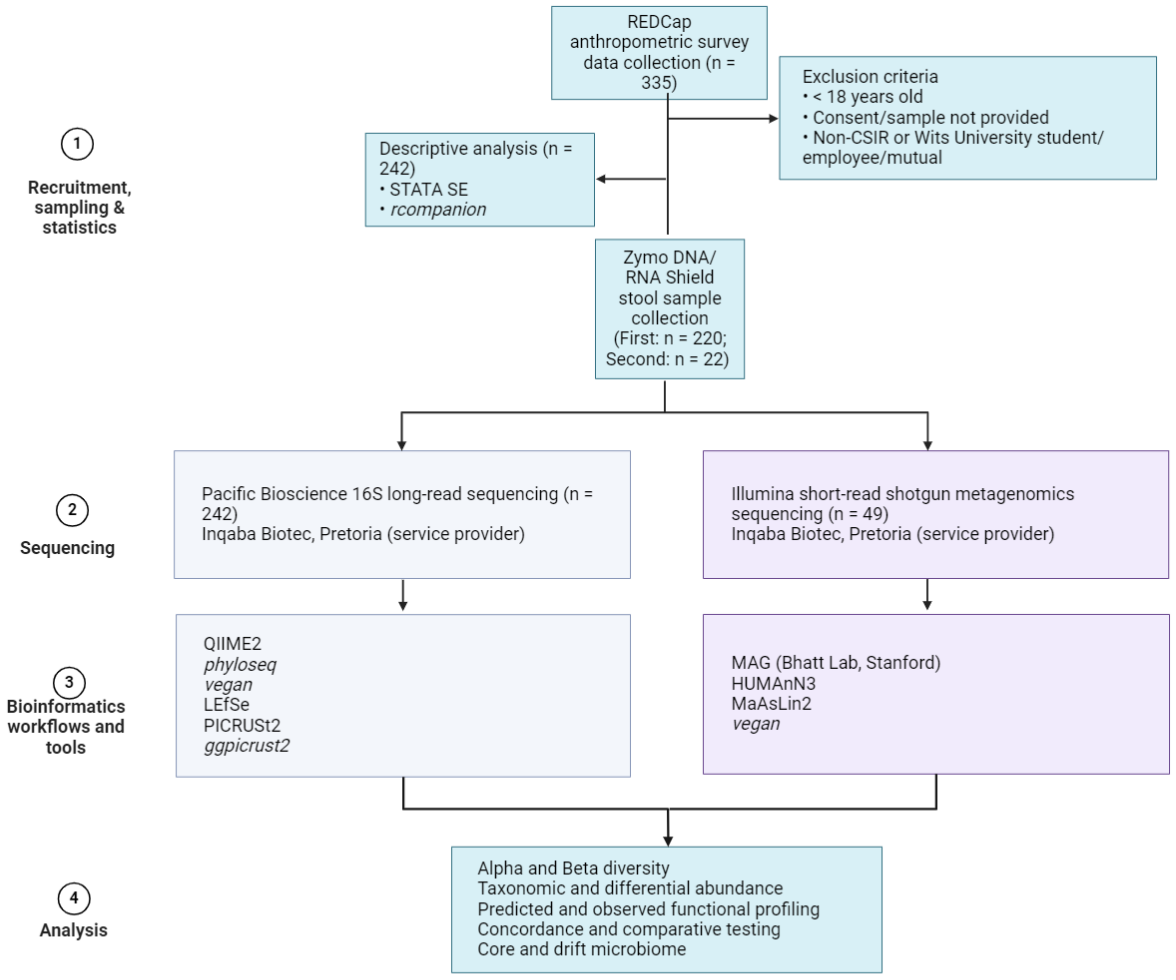


Figure 3.1 Overview of the study design and methodology. Tools written in italics indicate R software packages.

3.1. Participant recruitment

The study recruited participants from the Council for Scientific Industrial Research (CSIR) Pretoria campus and the University of the Witwatersrand (Wits, Johannesburg) to establish a novel South African urban professional cohort. Recruitment used both online methods (CSIR website https://cmmi.csir.co.za, email campaigns) and offline approaches (flyers, seminars). The target sample size was approximately 250 individuals, including CSIR employees, Wits students and faculty, and their household members.

The only exclusion criterion was a minimum age requirement of 18 years for participants. This cohort was selected as it represents South Africa's urban professional population diversity, spanning various occupational levels from highly skilled researchers to unskilled labour categories (CSIR, 2019; University of the Witwatersrand, 2022). This enabled analysis of gut microbiome composition in a South African urban professional setting, allowing investigation of potential intra-population diversity influences (Nieuwoudt et al., 2018).

3.2. Participant data collection and analysis

Following recruitment and informed consent, participants completed an online questionnaire using the Research Electronic Data Capture (REDCap) platform (Harris et al., 2019). REDCap is a secure web application for managing online surveys and databases, designed to maintain data integrity and compliance with key regulations including 21 Code of Federal Regulations (CFR) Part 11, Federal Information Security Management Act (FISMA), Health Insurance Portability and Accountability Act (HIPAA), and General Data Protection Regulation (GDPR), making it ideal for handling sensitive research data (Harris et al., 2019; Nichols, 2023).

The REDCap questionnaire contained 123 compulsory data fields covering participants' health, diet, environment, occupation, fitness, location and lifestyle (Supplementary Material S3.1). Designed to take approximately 20 minutes, participants could complete it in multiple sessions. The questionnaire was available in English, Afrikaans, and isiZulu, selected based on their prevalence in the target population according to a General Household Survey (Statistics South Africa, 2019). Each participant received a unique identifier code to ensure anonymity and link metadata with corresponding gut microbiome sample and sequencing data.

3.2.1. Ethical considerations and consent procedures

The ethics procedures and consent for this research strictly adhered to the guidelines set forth by the Wits Human Research Ethics Committee (HREC Medical) and the CSIR Research Ethics Committee (CSIR REC) – certificate *M210240* and the CSIR (reference number *292/2019*). Prior to participating in the survey and sample collection, all participants were provided with detailed information about the study, its purpose and procedures (Supplementary Material S3.2) and a PoPIA Privacy Statement (Supplementary Material S3.4). Along with this, ethics approval at CSIR (reference number *292/2019*). These measures ensured that participant rights, privacy, and confidentiality were protected throughout the study. REDCap was used to administer the consenting process. The survey's design, providing multiple examples and selectable options, aimed to enhance the accuracy and ease of reporting by helping participants recall and report their daily choices and medical history more effectively (Supplementary Material S3.1).

3.2.2. Anthropometric and health data

Anthropometric data (*n* = 123 fields) were collected through the REDCap online questionnaire adapted from the American Gut Project (AGP) (McDonald et al., 2018), facilitating convenient and non-intrusive data collection. Participants provided their weight in kilograms and height in centimetres, methods commonly utilized in large-scale studies for their reasonable accuracy in population-level analyses (Stommel & Schoenborn, 2009), and BMI (kg/m^2) was derived from that. Despite potential individual variations and reporting biases, self-reporting remains a practical and economical approach for anthropometric data collection in larger cohort studies (Aminianfar et al., 2021; Tuomela et al., 2019).

Other sections included dietary, health and lifestyle assessment to analyse participants' eating habits, physical activity levels, medication use, recent travel and clinical history. The dietary component covered a comprehensive range of food groups and consumption frequency, to understand dietary patterns and the importance of various foods in participants' diets. Lifestyle factors were also recorded, including exercise frequency and type, which are significant for understanding health outcomes (Z. Wang et al., 2022). SES was assessed to explore how income levels, access to amenities, and other socio-economic factors influence health within the urban South African context.

3.2.3. Descriptive analysis of survey data

Descriptive statistical analyses were conducted using STATA SE version 18.0 and R package *rcompanion* version 2.4.35 on the de-identified survey data. Data cleaning excluded variables with homogeneous responses (e.g., Autism Spectrum Disorder diagnosis) and merged subcategories with fewer than twenty responses (e.g., medical conditions diagnosed by different practitioners were simplified to "Yes / No" categories like for Irritable Bowel Syndrome). Afrikaans surveys were translated to English for consistency. Numeric variables including birth dates and body mass index (BMI) were converted to categorical groups, and frequency data were simplified to binary variables. This yielded 89 variables for analysis: 7 for SES, 22 for diet, 27 for lifestyle, and 33 health-related variables (Supplementary Table S4.2).

Fisher's exact test assessed associations between demographic variables and responses, with Monte Carlo simulation (1000 iterations) for sparse exercise-related data (Dexter et al., 2017). The Bonferroni correction method was applied to address multiple testing, and standardized residuals identified significant pattern deviations. Results included raw and adjusted p-values plus Cramér's V (φc) statistics to measure association strength: weak (0.10 – 0.29), moderate (0.30 – 0.49), or strong (0.50 – 1) (H.-Y. Kim, 2017). Variables with Bonferroni-adjusted p-value $< 5.0 \times 10^{-2}$ and $\varphi c > 0.5$ were prioritized for downstream analyses (Supplementary Data S4.1).

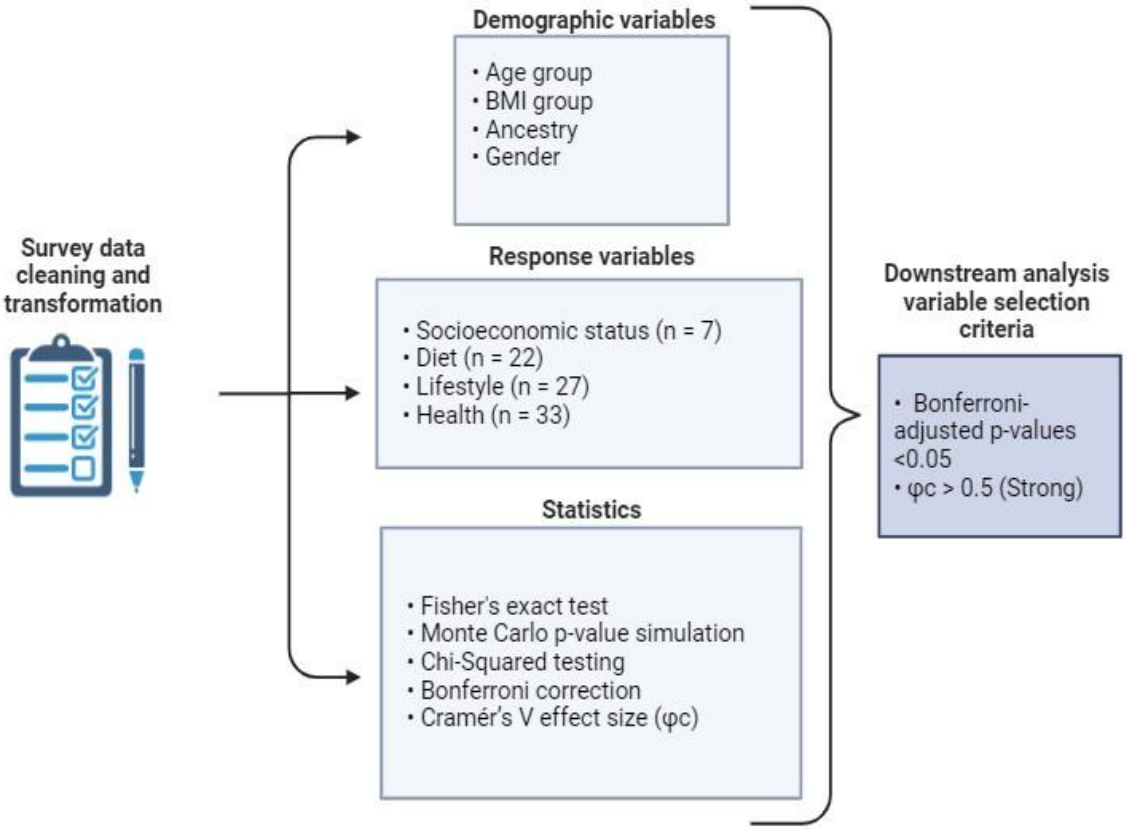


Figure 3.2 Descriptive analysis workflow. Survey data is extracted from the REDCap database, followed by data cleaning and transformation of variables. Demographic groups are compared to response variables with the statistical methods, resulting in variables prioritized for downstream analysis that meet the selection criteria.

3.3. Stool sample collection and processing

A single stool sample was collected from all participants for a recruitment and collection period from September 2021 to September 2023 (Figure 3.3). A longitudinal sample was collected from a subset of *n* = 22 individuals for a second collection period from July 2023 to September 2023. Time gaps between the two samples occurred approximately 7 – 24 months after initial sampling. Participants indicated their preferred collection site from the six available locations, which included sites at both the CSIR (one site) and Wits University (five sites). Upon completion of their survey, each customized kit labelled with their name and anonymous identification (ID) number that linked to their survey, was provided in an envelope for sample collection and return.



Figure 3.3 Sample collection timeline. From September 2021 to September 2023, four recruitment drives and five sample collections occurred. The cohort expanded from CSIR only in 2021 to Wits University and household members from 2022, with ethical approval. The fifth batch was a second longitudinal sample collection.

The DNA/RNA Shield™ Fecal Collection kit (Zymo Research Corp, USA) was used to ensure standardized and efficient sample collection. The Fecal Collection Tubes were designed for the preservation of nucleic acids from stool samples, by facilitating robust DNA recovery and were compatible with various downstream sequencing workflows (Hughes et al., 2017). This enabled a microbial snapshot of a sample while inactivating viruses and infectious agents making samples safe and ready for transport. Significant advantage included (a) their stabilization of samples at room temperature (DNA > 1 year), and (b) superior performance in comparison to other kits in the preservation of the taxonomic and functional profile of the gut microbiome over a long-term period (Kim et al., 2023). Participants were able to easily collect a spoonful of stool into the collection tubes prefilled with DNA/RNA Shield™ stabilization buffer as a preservative and shake vigorously to ensure proper stabilization.

Participants returned the collection tubes containing their sample to the designated drop-off sites. The samples were then transported to the CSIR Bioengineering and Integrative Genomic Research Group laboratory (BIGG) for cataloguing and storage at 4°C, until they were batched and prepared for DNA extraction and sequencing. The subsequent steps involved outsourcing the batching, DNA extraction and sequencing processes to the Health Professions Council of South Africa (HPCSA) certified laboratory at Inqaba Biotec™ in Pretoria. DNA extraction and sequencing were performed under strict quality control measures discussed in section 3.4 below.

3.4. Microbiome sequencing techniques

Two different sequencing techniques were used to analyse the microbiome composition in this study. For all samples (*n* = 242), PacBio 16S rRNA long-read sequencing was utilized.

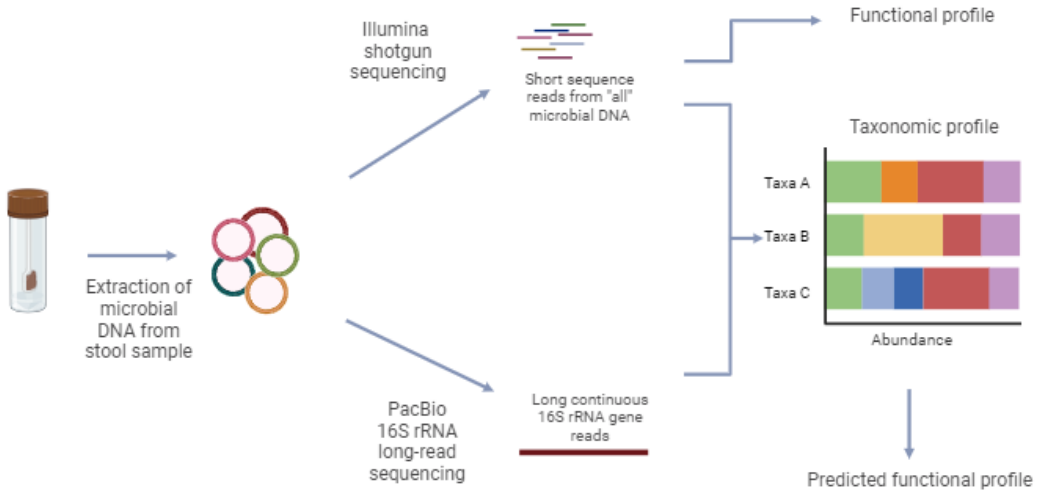


Figure 3.4 Comparative schematic of microbial DNA Sequencing. Illumina shotgun sequencing yields short sequence reads from the entire microbial DNA, facilitating a taxonomic and functional profile, while PacBio 16S rRNA long-read sequencing produces long continuous reads of the 16S rRNA gene, enabling taxonomic profiling and prediction of functional capabilities from this profile.

This technique provided high-resolution analysis of microbial communities by generating full-length 16S rRNA gene sequences (Figure 3.4). Additionally, a subset of samples (*n* = 49) underwent Illumina shotgun metagenomics sequencing, which offered a comprehensive analysis of the entire genomic content, including taxonomic composition, functional annotations, gene predictions, and metabolic pathways. Combining these sequencing approaches allowed for a thorough understanding of microbial diversity, taxonomic structure, and functional potential of the gut microbiota of the cohort.

3.4.1. PacBio 16S rRNA long-read sequencing

The DNA extraction from the refrigerated stool samples was carried out by a laboratory technician at the Inqaba Biotec™ laboratory. A ZymoBIOMICS DNA Miniprep extraction kit was utilized for this purpose (Zymo Research, 2023). This kit was specifically designed to purify DNA for downstream microbiome and metagenome analysis. It employed a lysis system that effectively addresses bias associated with unequal lyzing efficiencies of different organisms (Zymo Research, 2023).

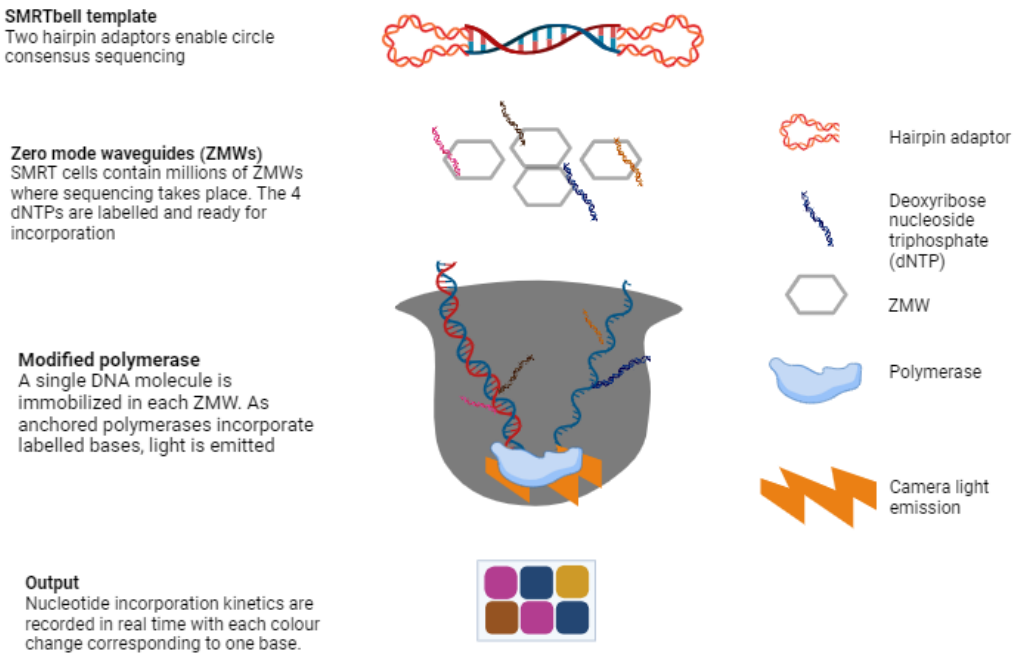


Figure 3.5 Overview of PacBio SMRT Sequencing: The process begins with a DNA template prepared with hairpin adapters, allowing for circular consensus sequencing. Within the Zero Mode Waveguides (ZMWs) of the SMRT cell, a single DNA molecule is immobilized and sequenced as modified polymerases incorporate fluorescently labelled deoxynucleotide triphosphates (dNTPs). The real-time detection of nucleotide incorporation is achieved through emitted light, with each colour representing a different base, allowing for continuous long-read sequencing.

To amplify the V1-V9 region of the bacterial 16S rRNA gene, PCR was performed using the 27F and 1492R universal primer pair. These primers were modified to include PacBio M13 adapter sequences at the 5' end of the forward and reverse sequences, allowing for barcoding of each amplicon. Subsequently, the barcoded amplicons underwent limited cycle PCR to introduce PacBio M13 barcodes for multiplexing (Figure 3.5). After quantification, the amplicons were equimolarly pooled and purified using Ampure PB beads. Library preparation for this study was performed using Single Molecule Real Time (SMRT) technology by PacBio (Pacific Bioscience, 2023). To prepare the library for high-quality sequencing on the PacBio Sequel IIe system, the steps of sequencing primer annealing and polymerase binding were carried out following the manufacturers protocol (Pacific Bioscience, 2023). Proper disposal of excess samples and waste adhered to the regulations set forth by the HPCSA.

3.4.2. Illumina shotgun metagenomics

Following a similar strategy to the American Gut Project, which performed shotgun metagenomic analysis on approximately 10% of samples (McDonald *et al*., 2018), *n* = 50 samples (20.7%) were selected for whole-genome sequencing. The shotgun metagenomic sequencing was performed by Inqaba Biotec™ using the Illumina MiSeq platform (Illumina Inc, USA). Following the DNA extraction

procedure described above, this comprehensive sequencing approach sampled all genes of the organisms present in these selected samples, with multiple samples combined in a single sequencing run to achieve high sequence coverage per sample (300 Cycles).

This technique involved Nextera library preparation and PCR amplification, followed by quantification using the PicoGreen dsDNA assay kit and library pooling (McDonald *et al.*, 2018). The libraries were then sequenced on the MiSeq platform, which offered reagents capable of producing up to 30 giga-bases (Gb) of output with 25 x 10^6 sequencing reads and 2 × 150 base pair (bp) read lengths.

3.5. Bioinformatics Analysis

The bioinformatic analysis of the gut microbiome was performed for the characterization of gross taxonomic classifications of microbiota present. The raw 16S long-read and shotgun data received in fastq.gz format was analysed using the using the “Quantitative Insights Into Microbial Ecology 2” (QIIME2, www.qiime2.org – version 2023.7) analysis pipeline (Bolyen et al., 2019) and a Metagenome-Assembled Genome (MAG) workflow from the AWI-Gen 2 Microbiome GitHub repository, respectively (Bhatt Lab, 2023). The QIIME2 command-line interface, AWI-Gen2 repository and plug-ins were accessed for this analysis using the Wits Core Research Cluster.

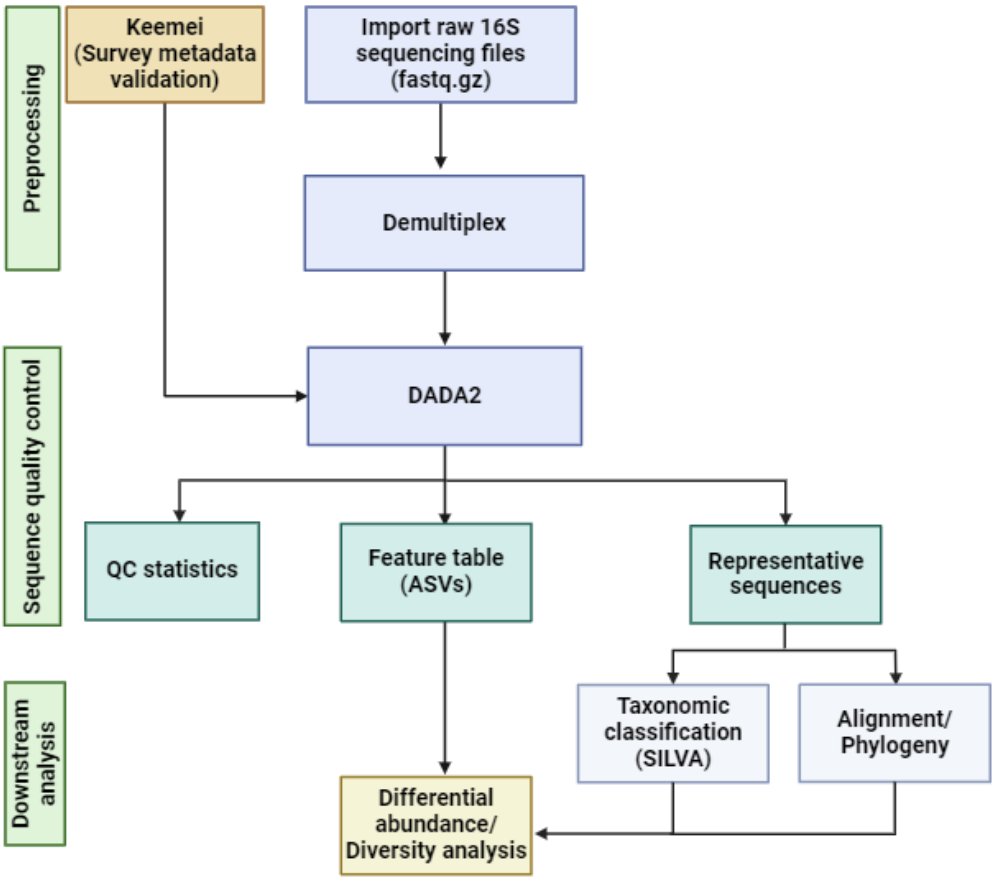


Figure 3.6 Overview of the QIIME2 pipeline for 16S rRNA long-read sequencing data. This figure summarises the core processes involved in microbial community analysis using the QIIME 2 platform. This is followed by a demultiplexing step to assign sequences to individual samples. The sequence quality control (QC) phase is managed by the DADA2 algorithm, which also includes the generation of QC statistics. The output of this process is a feature table containing amplicon sequence variants (ASVs) and a set of representative sequences. These outputs serve as the foundation for the downstream analysis, which encompasses diversity analysis and differential abundance assessments, alongside taxonomic classification using the SILVA reference database. In addition, alignment and phylogenetic analysis of the representative sequences is performed to infer evolutionary relationships key to diversity analysis.

I. 16S LONG-READ BIOINFORMATICS ANALYSIS

3.5.1. QIIME2 workflow for 16S data

To process the 16S sequences and test for associations to diet, lifestyle, SES and health, a metadata mapping file containing the relevant information was created for the study (Supplementary Data S3.1). This tabular file was validated for use in QIIME2 using the “Keemei” Google Sheets tool (Rideout et al., 2016). The raw 16S sequences were imported to the Wits Core Research Cluster for use in QIIME2 (Figure 3.6).

Two full-length 16S rRNA sequence data sets from gut microbiome samples were used to provide geographical, gender, and health status comparisons for our primary SAN population, particularly

in terms of diversity metrics and functional pathway analyses. The sample group sizes were small and thus used for observations of patterns and extrapolated differences were interpreted with caution. The groups were:

- Weinstock et al., (2019) sequenced samples from four healthy male American athletes PacBio RS II platform using an amplicon-based metagenomic approach with PCR selection (SRA accession ID PRJNA552603 (Weinstock *et al*., 2019). This sequencing strategy was designed to enable species and strain-level microbiome analysis to identify and measure the abundance of bacteria present in stool samples from athletes.
- Yang et al., (2023) sequenced eight gut microbiome samples from Chinese diarrhoeal patients using full-length 16S rRNA gene sequencing on a PacBio Sequel II platform using an amplicon-based approach with PCR selection (SRA PRJNA968050). This sequencing strategy was designed to analyse the gut microbiota composition in diarrhoeal patients.

Using the SingleEndFastqManifestPhred33V2 input format, the fastq.gz files of raw sequence data were imported and demultiplexed into a QIIME2 environment as an "artifact" – this contained information about the type and source of the data (Bolyen *et al*., 2019). The summary information obtained from the demultiplexed single-end reads determined how many sequences were obtained per sample, as well as the sequence quality at each base position (Figure 3.7) and visualised using QIIME2View (https://view.qiime2.org).

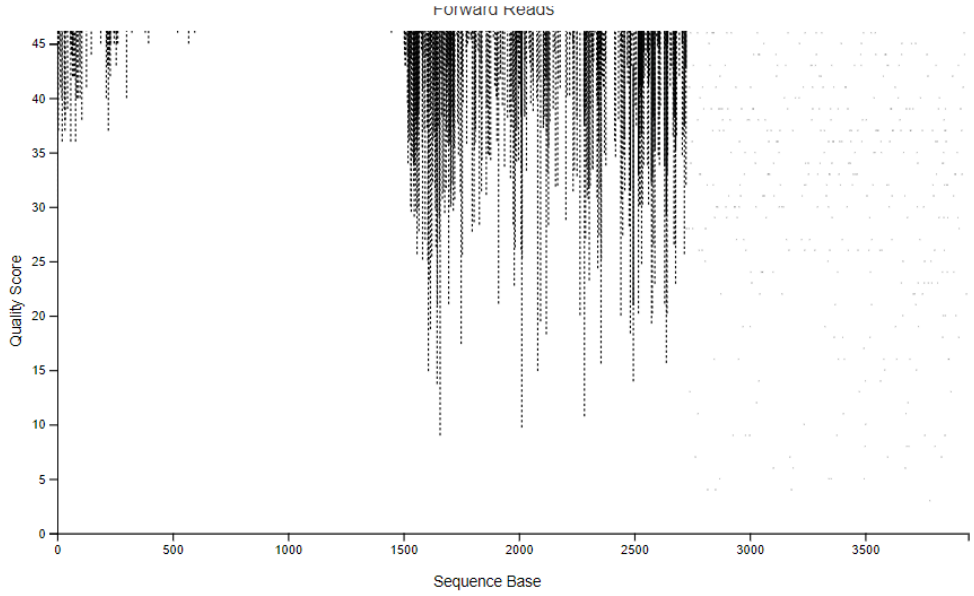


Figure 3.7 Summary information of the demultiplexed single-end 16S rRNA long-reads obtained from QIIME2. This provides a visualisation in QIIME2View of the number of sequences per sample as well as the average quality scores across all base positions.

To produce high quality sequence variant data, the Divisive Amplicon Denoising Algorithm 2 (DADA2) was used (Callahan et al., 2016). This involved importing the demultiplexed reads and trimming them

to specified lengths that correspond to the typical range of the 16S rRNA gene, using the parameter thresholds of --p-min-len 1000 and --p-max-len 1600. The data output from this process were artifacts of representative sequences, a feature table, and related summary statistics of the retained reads.

i. Feature-Table and Feature-Data

Following the quality filtering procedure using DADA2, visualizations were generated to examine the representative sequences and create a feature table of amplicon sequence variants (ASVs). These visualizations provided insights into the distribution of sequences associated with each sample and feature, incorporating information from the metadata mapping file. Histograms depicting these distributions, along with relevant summary statistics, were also generated. Additionally, this process facilitated the mapping of feature identities (IDs) to sequences and provided links for convenient utilization of the Basic Local Alignment Search Tool (BLAST). With this tool, each representative sequence could be compared against the National Centre for Biotechnology Information (NCBI) nucleotide database (Altschul et al., 1990) to gather further information.

ii. Phylogenetic Tree generation

To uncover meaningful patterns in microbial diversity during downstream analyses, a phylogenetic tree was constructed. This process involved several steps. First, the pipeline aligned multiple sequences in the Feature Data [Sequence] using the *mafft* program (Katoh, 2002) to generate the Feature Data [Aligned Sequence]. Next, the alignment was masked to eliminate highly variable positions that could introduce noise into the tree. The *FastTree* tool (M. N. Price et al., 2010) was then applied to generate the tree from the masked alignment. Finally, the unrooted tree obtained was subjected to midpoint rooting. This phylogenetic tree serves as a crucial component for analysing various diversity metrics presented in this research.

3.5.2. Normalization through rarefaction depth analysis

A detailed rarefaction analysis was performed in R software using the *phyloseq* package (version 1.46.0) for normalizing the feature-table ASV outputs of the 16S long-read sequencing data from QIIME2. The purpose of this analysis was to determine the optimal sequencing depth that would maintain a robust sample size while capturing the full complexity of microbial diversity present in the samples, as described in Weiss et al. (2017). Rarefaction depths of 90, 215, 300, and 450 were systematically evaluated. The selection of rarefaction depths was crucial for balancing sample representation and the reliability of diversity estimates:

- **90 sequences**: Chosen slightly above the first quartile (88 sequences), this depth maximizes sample retention while minimizing sampling error from the lowest-count samples. At this depth, 60 samples were excluded, and 1,628 features retained across 173 samples.

- **215 sequences**: Matching the median sample sum, this depth balances substantial sample retention with robust diversity estimates. Here, 116 samples were excluded, with 1,954 features retained across 117 samples.
- **300 sequences**: A moderate increase from the median, aimed at enhancing diversity reliability while still maintaining a significant sample size. At this depth, 140 samples were excluded, and 1,933 features were retained across 93 samples.
- **450 sequences**: Positioned between the median and the third quartile to enrich sequence coverage while accepting a reduction in sample size for higher reliability in the diversity estimates. This depth led to the exclusion of 168 samples, retaining 1,798 features in 65 samples.

These depths were strategically selected to ensure comprehensive gut microbiome analysis in relation to various factors (Weiss et al., 2017). Each depth balanced sample retention with reliable diversity estimates. To identify significant diversity differences between depths, non-parametric tests were performed: Kruskal-Wallis test, followed by pairwise Wilcoxon rank-sum tests with Benjamini-Hochberg procedure for multiple comparison correction. Statistical analysis across five sequencing batches investigated potential batch effects on diversity. Assessment led to selecting a rarefaction depth of 90, optimized for the South African (SAN) cohort to retain substantial samples (173) while providing sufficient microbiome diversity detail. This larger sample size was essential for robust correlation analyses, improving power to detect true microbiome-variable associations.

Principal Coordinates Analysis (PCoA) measured the chosen depth's adequacy, supported by permutational multivariate analysis of variance (PERMANOVA), to investigate community composition differences across sequencing depths. Findings aimed to confirm microbial community stability across tested depths and rarefaction process reliability. With rarefaction depth at 90, read distribution normalization was completed for further analysis. Normalized read density distribution was plotted across demographic group variables, using Kernel density estimation (KDE) for smoothing (noting that resulting negative values are KDE visual artifacts, not actual read counts).

3.5.3. Diversity analysis

Within the framework of the *phyloseq* package in R, a comprehensive approach was adopted for analysing microbial diversity across the cohort. The *phyloseq* package facilitated the integration of the QIIME2 outputs: the phylogenetic tree with survey metadata, and abundance data from the feature table, enabling the calculation of alpha and beta diversity indices. Alpha and beta diversity, as defined by (Whittaker, 1960), quantifies the local biodiversity within a sample, and the regional biodiversity differences between samples, respectively. For each sample, these indices were computed across the key sampling depths for demographic variables. Following the application of the specified

rarefaction depth of 90, the analysis was extended to include lifestyle, diet, health, and SES variables, selected based on criteria of Fisher's Bonferroni-adjusted $p \leq 5.0 \times 10^{-2}$ and φc effect size ≥ 0.5 from descriptive statistics outputs to balance between data retention and statistical robustness. As iterated above, this rarefaction process was crucial for enabling equitable comparisons across samples with varying original sequence counts.

Subsequent to calculating these diversity indices, statistical tests were performed to explore associations between microbial diversity and group variables. The analyses utilized the non-parametric Wilcoxon rank-sum test to handle the non-normal distribution of microbial diversity data effectively. The p-values were adjusted using the Benjamini-Hochberg method to control the false discovery rate, ensuring that the statistical inferences made were reliable. This methodological approach, executed within the *phyloseq* environment, provided a powerful and integrated toolset for the comprehensive analysis of microbial community structure and its association with South African demographic and anthropometric factors.

i. Alpha diversity analysis

For each demographic variable, diversity indices including Shannon, Simpson, and Observed OTUs were calculated to quantify the richness and evenness of microbial communities within the samples. The Shannon index, a measure of community diversity that accounts for both abundance and evenness of the species present (Shannon, 1948), and the Simpson index, which reflects the probability of two individuals randomly selected from a sample belonging to the same species (Simpson, 1949), were utilized alongside Observed OTUs. Notably, the term "Observed OTUs" refers to the count of unique Amplicon Sequence Variants (ASVs) identified, a direct output from DADA2 processing in QIIME2, representing species richness (Callahan *et al*., 2016). The demographic variables were treated as categorical factors, and the levels of these factors were meticulously verified to ensure accurate grouping in subsequent analyses. Statistical comparisons were conducted using the Wilcoxon rank-sum test to assess differences in diversity indices between groups within each demographic category. The p-values from these tests were adjusted using the Benjamini-Hochberg method to control for false discovery rates, ensuring the reliability of the findings (Benjamini & Hochberg, 1995).

ii. Beta diversity analysis

Beta diversity was examined, focusing on the Bray-Curtis dissimilarity and Weighted UniFrac distances to explore microbial community differences across demographic variables like ancestry. The Bray-Curtis dissimilarity is a non-phylogenetic measure that quantifies the compositional dissimilarity between two communities, based on counts of shared and unique species (Bray & Curtis, 1957). In addition, the Weighted UniFrac distance incorporates phylogenetic relationships between

microbial taxa and also accounts for the relative abundance of taxa (C. Lozupone & Knight, 2005), providing a more nuanced view of community structure.

For each distance measure, PCoA was executed to visualize the microbial community differences. The PCoA plots were enhanced with ellipses representing the 95% confidence intervals around group centroids and annotated with the percentage of variation explained by the principal coordinates. The association between microbial community composition and demographic variables was statistically tested using PERMANOVA (Anderson, 2001). This method partitions the variance in microbial community data explained by group categories and tests the significance of these partitions using 999 permutations. Results from the PERMANOVA were summarized with p-values indicating the statistical significance of the associations observed. These various beta diversity metrics were employed to facilitate the analyses described above, enabling the examination of dissimilarity and similarity patterns in gut microbial community composition among samples.

3.5.4. Taxonomic classification

The taxonomic analysis used QIIME2 with VSEARCH tool (Rognes et al., 2016) –cluster-features-de-novo method for clustering 16S long-read sequences. Using 99% identity threshold, this generated *de novo* tabulated and representative sequence files, which were assigned taxonomic labels using the vsearch classifier. The classifier, trained on Naïve Bayes algorithm, aligned sequences against the "SILVA 132 99% Operational Taxonomic Unit (OTU) marker gene reference database" (Bokulich et al., 2018). Taxonomy visualization used the metadata tabulate command, while taxa barplot command incorporated metadata to produce interactive summaries of gut microbial composition's relative abundance per sample. The tabulated data was exported to Biological Observation Matrix (BIOM) format for downstream analyses in RStudio version 2023.12.1.402 (Posit team, 2024).

Further analyses used package in R to explore phylum-level taxonomic abundance with data rarefied to 90. The taxonomy table was filtered to exclude PacBio American athlete and Chinese diarrhoeal samples (reserved for functional profile analyses with internal normalization process) and samples with missing metadata (*age group, BMI group, ancestry, gender*). Taxonomic data was cleaned by removing non-matching prefixes and standardizing unclear labels to "Unclassified" before updating the *phyloseq* object. The analysis revealed bacterial phyla distribution across the cohort, showing relative abundance and prevalence of each phylum.

3.5.5. Differential abundance analysis

In the context of differential abundance analysis of 16S long-read sequencing data and metadata, the Linear Discriminant Analysis (LDA) Effect Size (LEfSe) pipeline (Segata et al., 2011) was used to validate representative taxa and perform additional analyses in RStudio, with the R library package *lefser* version 1.12.1. This method was used to identify significant differences in microbial taxa across the various demographic variables including *age group, gender, ancestry,* and *BMI group*.

The analysis was conducted separately at both genus and species taxonomic levels using the run_lefse function, implemented in the R package, based on the methodology described by Segata *et al*. (2011). For each analysis, taxa abundances were normalized to copies per million (CPM) to account for different sequencing depths. A non-parametric Kruskal-Wallis rank-sum test was first applied to identify taxa with differential abundance across multiple categories with a significance threshold of p-value ≤ 5.0×10^{-2}. For taxa meeting this criterion, pairwise comparisons were subsequently performed using the Wilcoxon rank-sum test when more than two groups were compared, maintaining the same p-value cutoff (Mann & Whitney, 1947). Taxa that remained significant in these tests were subjected to Linear Discriminant Analysis (LDA) to estimate the effect size of differences, with an LDA score threshold of 2.0 used to identify meaningful biological differences.

Results were visualized using dot plots, where the size of each dot corresponded to the negative logarithm of the p-value, indicating the statistical significance of the abundance difference for each taxon. Additionally, the analysis was extended to include lifestyle, diet, health, and SES variables, selected based on criteria of Fisher's Bonferroni-adjusted p-value ≤ 5.0×10^{-2} and φc effect size ≥ 0.5 from descriptive statistics outputs.

3.5.6. Predicted functional profiling analysis

To elucidate the functional potential of microbial communities, a comprehensive analysis was performed using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) (Douglas et al., 2018), and additional statistical analysis and visualization with the R package *ggPICRUSt2* version 1.7.3. The initial input data for PICRUSt2 consisted of ASV tables in BIOM format, and representative sequences in FASTA format generated from QIIME2 analysis of the 16S long-read sequences. This analysis was facilitated by utilizing the feature-table and feature-classifier plugins for sequence classification and table generation.

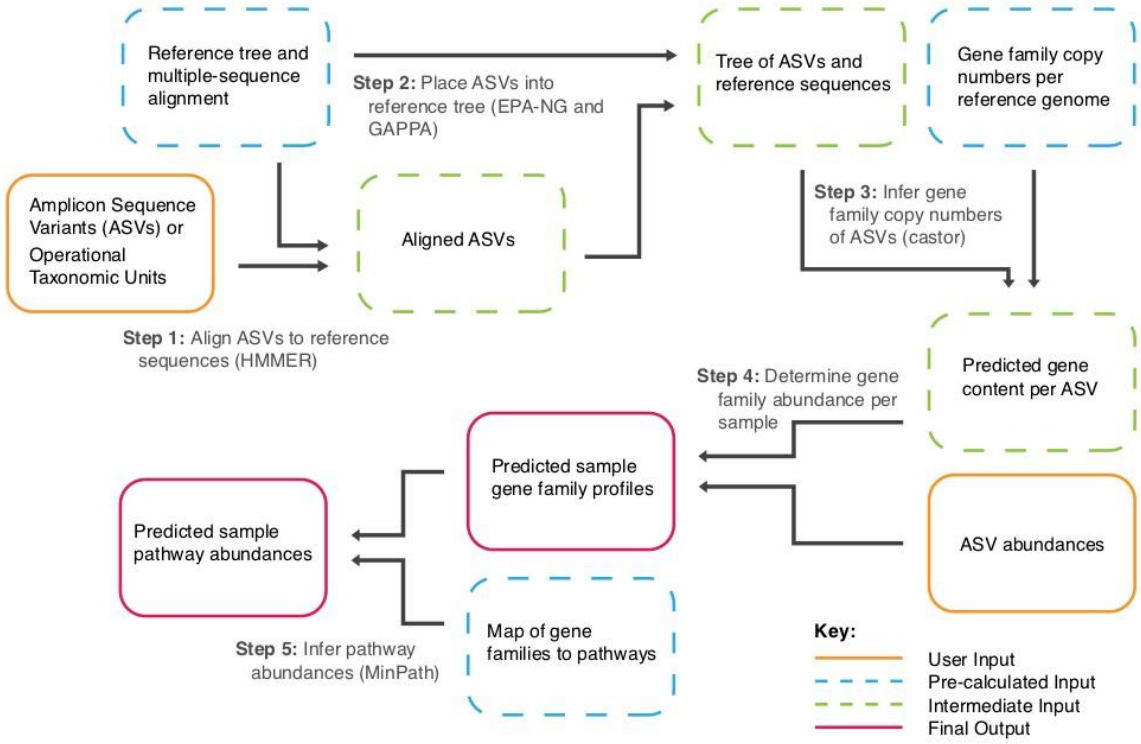


Figure 3.8 PICRUSt2 computational workflow. PICRUSt2 is a bioinformatics tool used for predicting functional profiles of microbial communities from the 16S long-read ASV data. The workflow begins with the alignment of ASVs to reference sequences using HMMER (Hidden Markov Model ER), progresses through the phylogenetic placement of aligned ASVs into a reference tree using EPA-NG (Evolutionary Placement Algorithm - Next Generation) and GAPPA (Genetic Algorithm for Placement and Phylogenetic Assignment), and then infers gene family copy numbers with the *castor* R package. Subsequently, the abundance of gene families per sample is calculated, which leads to the final step of inferring pathway abundances via MinPath (Minimal set of Pathways). (Douglas *et al*., 2020)

The PICRUSt2 pipeline was executed using the picrust2_pipeline.py script (Douglas, 2021), incorporating the standardized steps such as sequence alignment, placement against a reference phylogenetic tree, hidden-state prediction, metagenome prediction through the inference of gene family abundances, and subsequent pathway-level predictions (Figure 3.8). The latter predictions capitalized on the reconstruction of Enzyme Commission (EC) numbers and Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) groups, along with the calculation of Nearest Sequenced Taxon Index (NSTI) scores, which serve as a measure of the predictive accuracy based on the relatedness of input sequences to those in the reference database.

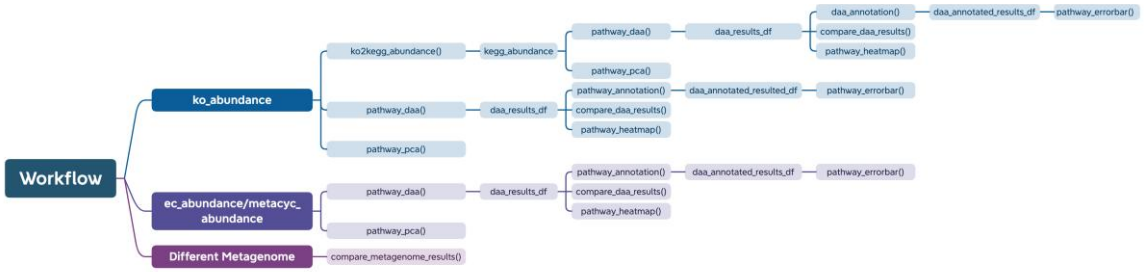


Figure 3.9 PICRUSt2 workflow diagram for metagenome analysis. This diagram illustrates the step-by-step process for analysing metagenomic data using the PICRUSt2 pipeline, highlighting the sequential functions and data outputs such as KO (KEGG Orthology) and EC (Enzyme Commission) abundance, pathway predictions, and comparative analysis results. (Yang *et al.*, 2023)

Statistical analysis and visualization of predicted functional profiles used *ggPICRUSt2* in R. As shown in Figure 3.9, the workflow began with ko_abundance or ec_abundance / metacyc_abundance input, followed by ko2kegg_abundance() function for annotated Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway abundances. With ko_to_kegg = "true", output included KEGG pathway annotations (name, description, class, map), enabling nuanced interpretation. With ko_to_kegg = "false", output focused on enzyme descriptions for each KEGG Orthology (KO) group or broader MetaCyc metabolic pathway predictions.

The pathway_daa() function enabled differential abundance analysis incorporating variables meeting specific selection criteria. “Analysis Of Differential Abundance Taking Sample Variation Into Account” (*ALDEx2* R package version 1.28.0) executed Kruskal-Wallis and generalized linear model tests to identify differentially abundant pathways. Results were visualized via pathway_pca() for principal component analysis, pathway_heatmap() for heatmaps, and pathway_errorbar() for error bar plots. The pathway_annotation() function annotated pathways with detailed descriptions, providing comprehensive overview of functional pathways and their differential abundances relative to studied variables. This functional prediction framework helped understand microbial communities' predicted metabolic capacities and their potential health implications.

II. SHOTGUN BIOINFORMATICS ANALYSIS

3.5.7. MAG workflow for shotgun data

For the shotgun sequences, a Metagenome-Assembled Genome (MAG) workflow (Figure 3.10) was utilized (Bhatt Lab, 2023). This workflow involved the assembly of shotgun reads into longer contiguous sequences, referred to as contigs, and subsequent binning of these contigs into metagenome-assembled genomes, representing individual microbial genomes within the community. The MAG workflow allowed for the analysis of the functional potential and taxonomic composition of the microbial community based on the shotgun sequencing data.

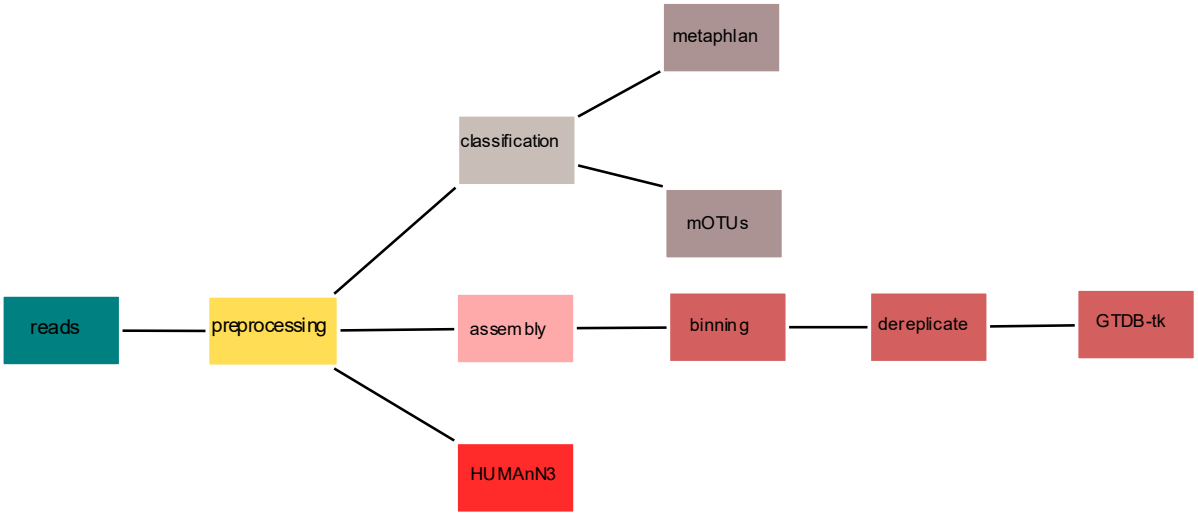


Figure 3.10 Overview of the bioinformatics pipeline for shotgun sequencing reads using the Metagenome-assembled genome (MAG) workflow. Raw sequencing reads were used as input and underwent preprocessing, followed by three separate processes: Classification of taxonomic abundance with metaphlan for species-level profiling and mOTUs for marker gene-based taxonomic profiling; Assembly of contigs followed by binning of metagenome-assembled genomes, dereplicate in order to remove redundant MAGs and GTDB-tk for taxonomic classification; HUMAnN3 was used to analyse functional pathways and gene families.

To provide comparative analyses across different populations, we included two additional cohorts that underwent metagenomic shotgun sequencing and were processed through our MAG workflow. Thirty paired-end metagenomic shotgun sequencing samples from the American Gut Project (AGP) were randomly selected and included for analysis. The samples were sequenced using Illumina HiSeq 2500 technology with 2 x 150 bp paired-end reads at 30X coverage. The AGP, launched in 2012, represents one of the largest crowd-sourced citizen science projects characterizing human microbiome diversity in the United States, collecting samples from over 10,000 participants across the country. The project standardized sample collection using home collection kits and processing at the Knight Lab at UC San Diego. Due to random selection, specific metadata information for our selected samples was limited to American origin.

Fifty paired-end metagenomic shotgun sequencing samples from the AWI-Gen Phase 1 microbiome study were included as an African comparative cohort. The samples were collected from HIV-negative women aged 40 – 72 years from both rural (Bushbuckridge Municipality) and urban (Soweto, Johannesburg) South African communities. The samples were processed using the QIAamp PowerFecal DNA Kit and sequenced on the Illumina HiSeq 4000 platform with 2 x 150 bp paired-end reads at 50X coverage. The samples were part of a larger study investigating the relationship between human genomics, lifestyle factors, and the gut microbiome across different South African ethnicities. Both AGP and AWIGEN1 datasets were processed through identical MAG assembly and binning pipelines as described in the following sections.

i. Preprocessing

In the MAG workflow preprocessing (Figure 3.11), quality control used FastQC tool (Andrews, 2010) to assess sequence quality, detect issues, and obtain quality metrics including base quality, sequence length distribution, and GC content. MultiQC tool (Ewels et al., 2016) generated comprehensive summary reports, compiling multiple analyses for comparative assessment across samples.

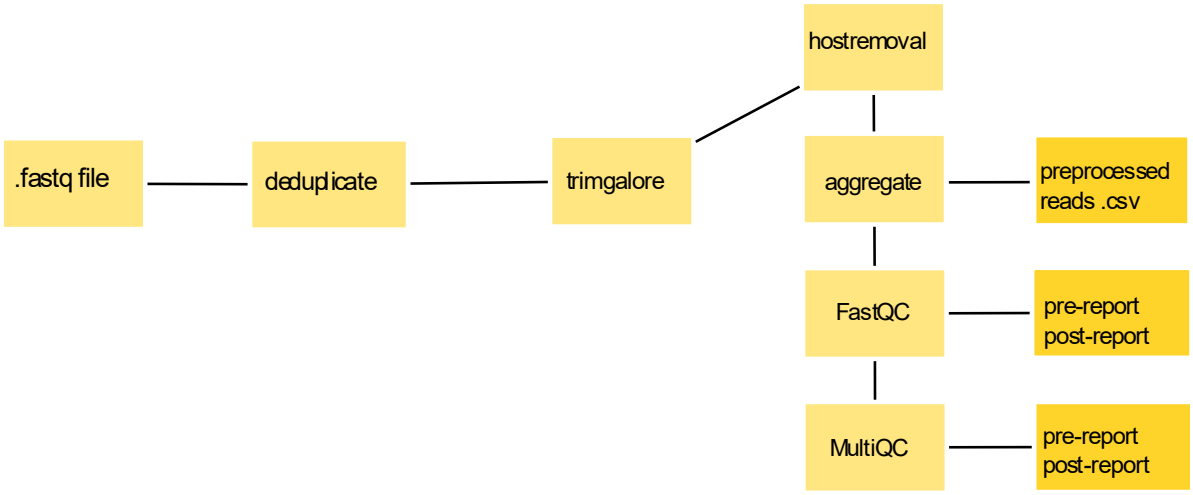


Figure 3.11 MAG preprocessing workflow for shotgun sequencing data. Input .fastq files undergo sequential quality control steps: deduplication to remove duplicate reads, trimgalore for quality and adapter trimming, followed by host sequence removal. The processed reads are then aggregated, producing a preprocessed reads CSV file. Quality control metrics are assessed using both FastQC and MultiQC, generating pre- and post-processing reports for quality assessment.

Deduplication removed duplicate reads to reduce bias and redundancy. "Trimgalore" tool (Martin, 2011) trimmed adaptor sequences and eliminated low-quality sequences, enhancing accuracy by removing artifacts. Low-quality reads were filtered to ensure high-quality data for downstream analysis. Host removal used Burrows-Wheeler Aligner (BWA) tool (H. Li & Durbin, 2009) to eliminate human DNA reads. Post-removal, remaining reads underwent additional FastQC quality check, with results integrated via MultiQC. Reads were aggregated to compile summary statistics (read counts, minimum/maximum values), providing comprehensive quality metrics before and after host DNA removal.

ii. Classification

The taxonomic classification of metagenomic data played a crucial role in understanding the microbial composition within a sample. In this study, two tools, “Metagenomic Phylogenetic Analysis” (*MetaPhlAn*) and *mothur*, were used for taxonomic classification at different levels of resolution. First, MetaPhlAn was utilized to perform species-level profiling of the metagenomic data. MetaPhlAn is a widely-used tool that leverages unique clade-specific marker genes to estimate the abundance of microbial taxa in a sample (Segata et al., 2012). This approach allowed for accurate taxonomic

assignment and quantification of species present in the metagenomic dataset. The resulting output provided valuable insights into the relative abundance and distribution of microbial species within the sample.

To complement the species-level profiling, *mothur* (Schloss et al., 2009) was used for higher-resolution taxonomic classification. It is a versatile tool that offers a range of algorithms and databases for taxonomic assignment and analysis of metagenomic data. It employs a variety of approaches, including Bayesian and phylogenetic methods, to classify the sequences at different taxonomic ranks, such as genus, family, and phylum. The use of *mothur* allowed for a comprehensive taxonomic characterization of the microbial communities present in the metagenomic dataset.

iii. Assembly

The assembly step in the MAG workflow (Figure 3.12) was crucial for reconstructing longer contigs from the short reads, thereby enabling a deeper understanding of the genomic content within the metagenomic dataset. In this study, three tools, namely Megahit, Prodigal, and Quast, were applied for different aspects of the assembly and analysis process.

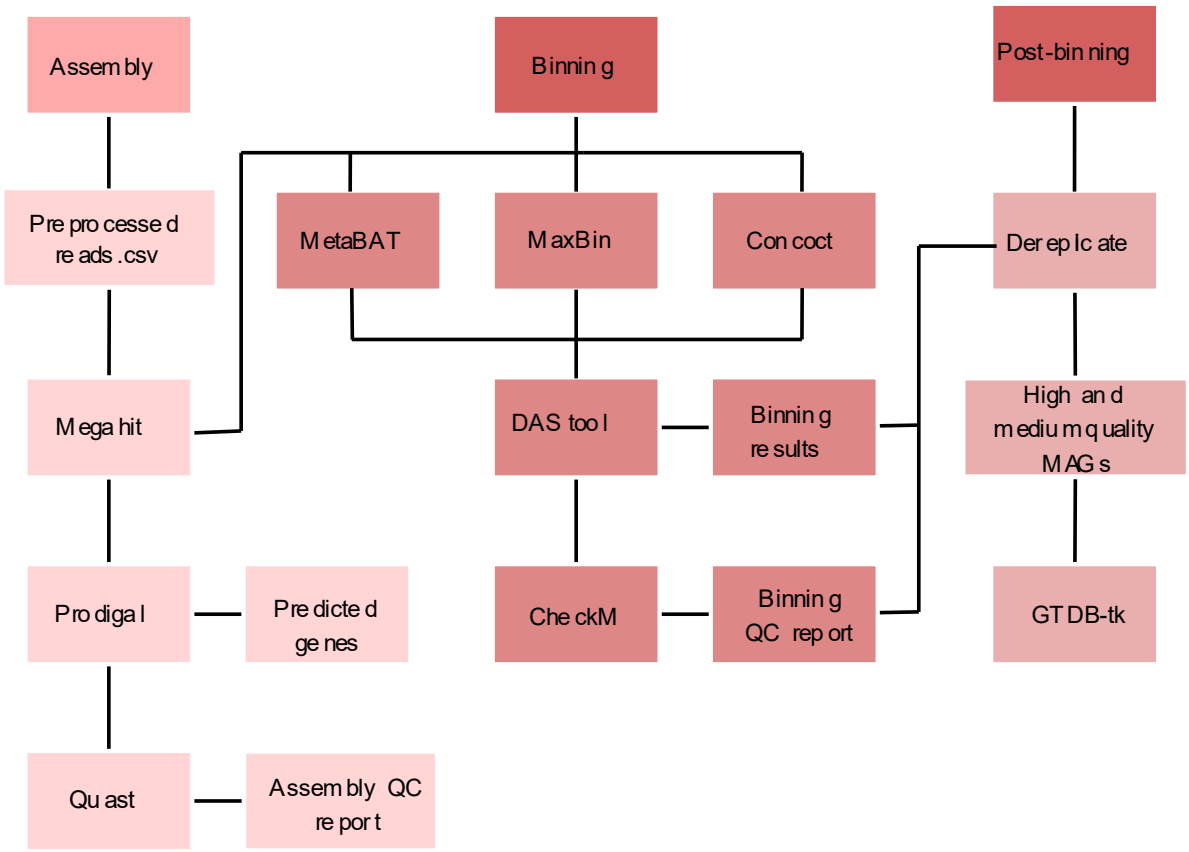


Figure 3.12 MAG workflow for assembly, binning, and post-binning analysis of shotgun metagenomic data. The assembly pipeline begins with preprocessed reads.csv files processed through Megahit for assembly, followed by Prodigal for gene prediction and Quast for assembly quality assessment. The binning process employs three parallel binning tools (MetaBAT, MaxBin, and Concoct), whose outputs are integrated using DAS Tool. CheckM evaluates the quality of bins, generating binning QC reports. In the post-binning phase, the bins undergo dereplication to obtain high and medium quality MAGs, which are then taxonomically classified using GTDB-tk.

The Megahit tool was utilized to perform the assembly of the short-reads into contigs. Megahit is a popular algorithm designed specifically for metagenomic assembly (Hyatt et al., 2010). It utilized a de Bruijn graph-based approach to identify overlapping regions within the reads and construct longer contiguous sequences, ultimately generating contigs representative of the assembled genomes. Megahit's efficient and scalable implementation allowed for the assembly of large and complex metagenomic datasets, providing valuable insights into the genomic content of the microbial communities.

Next, Prodigal was employed to predict protein-coding genes from the assembled contigs. Prodigal is a widely-used gene prediction tool that utilized Hidden Markov Models (HMMs) to identify potential protein-coding sequences within the contigs (Hyatt *et al*., 2010). By identifying and annotating protein-coding genes, Prodigal allowed for functional characterization and further analysis of the assembled genomes, aiding in the understanding of the genetic content and potential functional capabilities of the microbial communities.

To assess the quality of the assembled genomes, the Quast tool was utilized. Quast is a comprehensive quality assessment tool specifically tailored for evaluating genome assemblies (Gurevich et al., 2013). It provided various metrics such as N50 contig length, genome coverage, and mis-assemblies, allowing for a detailed analysis of the assembly quality. Quast's comprehensive evaluation provided valuable insights into the accuracy and completeness of the assembled genomes, aiding in assessing the reliability of the results and facilitating downstream analyses.

iv. Binning

The binning step was performed to group metagenomic contigs into distinct genome bins representing populations of microorganisms with similar genomic features. Three binning programs – MetaBAT, MaxBin, and Concoct – were used for this process.

MetaBAT (D. D. Kang et al., 2015) involves a mixture model-based clustering approach that considers tetranucleotide frequencies, coverage depth, and linkage information. MaxBin (Y.-W. Wu et al., 2016) constructs a k-mer frequency-based algorithm and incorporates marker genes for bin refinement. Concoct (Alneberg et al., 2014) uses a clustering algorithm based on nucleotide composition and coverage across multiple samples, along with marker gene information for bin refinement. To integrate

the binning outputs from multiple tools, the DASTool software was used (Sieber et al., 2018), allowing for the combination of binning results from the preceding tool outputs. Lastly, the quality and completeness of the generated genome bins were assessed using CheckM (Parks et al., 2015), which estimated the completeness, contamination, and strain heterogeneity of the bins, providing an evaluation of the binning process.

v. Post-binning – Genome Taxonomy Database

After the CheckM evaluation, bins were selected for further analysis if they demonstrated a "Completeness" of greater than or equal to 50% and a "Contamination" of less than 10%. This criterion was crucial to ensure the reliability and quality of MAGs for detailed taxonomic investigation (Bowers et al., 2017). For the taxonomic classification of the filtered MAGs, the Genome Taxonomy Database Toolkit (GTDB-Tk) software toolkit v2.1.1 and database version v214 was utilized (Chaumeil et al., 2022). GTDB-TK operates through a three-step classification approach: (1) identifying marker genes using Prodigal and HMMER, specifically targeting 120 bacterial and 122 archaeal marker genes, (2) creating multiple sequence alignments of these markers, and (3) placing genomes into a reference tree using pplacer, which employs maximum-likelihood phylogenetic placement. The classification system uses ANI (Average Nucleotide Identity) and RED (Relative Evolutionary Divergence) values to establish taxonomic ranks, providing standardized bacterial and archaeal taxonomy based on genome phylogeny.

The computational environment was set up using Mamba (v1.4.2) from the conda-forge and bioconda channels (Grüning et al., 2018; QuantStack, 2024). GTDB-TK classified each MAG against the reference database, providing standardized taxonomic classifications aligned with genome-based phylogeny rather than traditional 16S rRNA-based taxonomy. The outputs from this provided a basis for understanding microbial genome identities and relationships.

vi. Observed functional profiling analysis

Functional profiling of shotgun metagenomic data was performed using HUMAnN3 (HMP Unified Metabolic Analysis Network version 3.9). HUMAnN3 and MetaPhlAn 3.0 were installed via Mamba, with database paths configured to local directories on the Wits Core Research Cluster. Required databases – *ChocoPhlAn* (v2019.01.31) and *UniRef90* (v2019.01) – were downloaded and linked using humann_config.

Cleaned reads from MAG workflows were concatenated (paired-end to single files) and processed via a *Nextflow* pipeline for scalable, parallelized execution. Taxonomic and functional profiling combined two aligners: (i) *Bowtie2* (v2.5.3; Langmead & Salzberg, 2012) mapped reads to *ChocoPhlAn* for species identification; and (ii) *DIAMOND* (v2.1.9; Buchfink et al., 2015), a BLAST-compatible tool, aligned unclassified reads to UniRef90 for protein annotation. This hybrid approach

leveraged Bowtie2's precision for known microbes and DIAMOND's breadth for novel functional insights, generating comprehensive taxonomic/functional profiles.

The HUMAnN3 outputs include:

- Gene family abundance, identifying functional potential through conserved gene groups.
- Metabolic pathway abundance and coverage, quantifying biochemical process activity and completeness.

Gene family and pathway abundances were normalized to relative abundance ("RelAb") to enable cross-platform (long-read sequencing) comparisons and mitigate sequencing depth variation. Normalized data were integrated using HUMAnN3 utilities and analysed via *MaAsLin2* 1.16.0 (Mallick et al., 2021) to model associations between microbial functions and host demographics/anthropometrics (Section 3.6.5). This approach elucidated how microbial functional profiles interact with urban South African host traits.

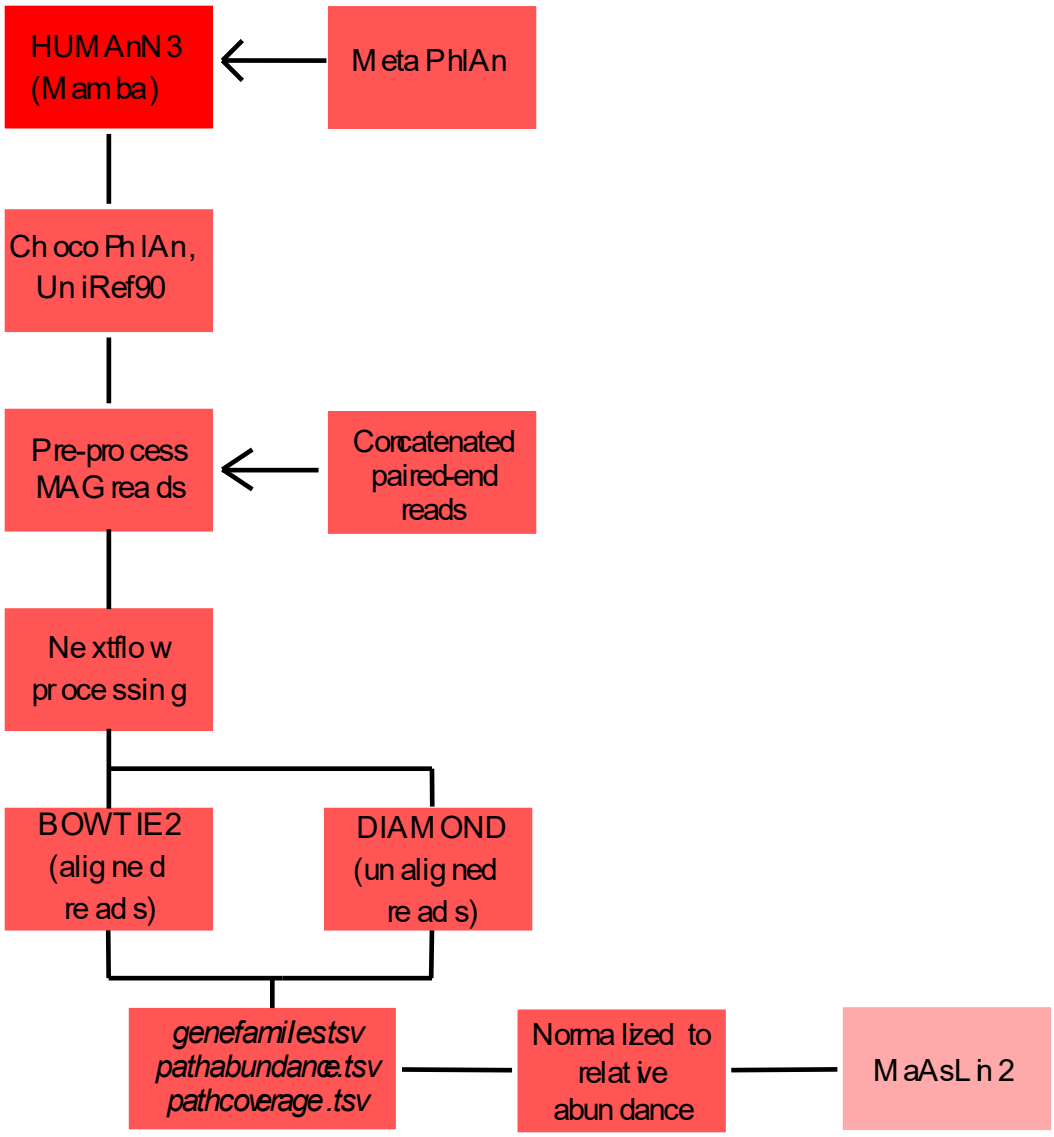


Figure 3.13 Bioinformatics workflow for shotgun metagenomic analysis utilizing HUMAnN3 (with Mamba environment) pipeline. The workflow begins with MetaPhlAn input, followed by ChocoPhlAn and UniRef90 database processing. Raw MAG reads are pre-processed and undergo Nextflow processing before splitting into BOWTIE2 (aligned reads) and DIAMOND (unaligned reads) paths. The workflow generates gene families, path abundance, and path coverage files (.tsv format) which are normalized to relative abundance using MaAsLin2 for downstream statistical analysis.

3.6. Long-read and shotgun metagenomic dataset concordance

3.6.1. Taxonomic concordance assessment

A multi-level approach was employed to evaluate taxonomic concordance between 16S rRNA and shotgun metagenomic datasets (Clooney et al., 2016; Durazzi et al., 2021; Laudadio et al., 2018). Taxonomic classification of shotgun metagenomic data was performed using MetaPhlAn. Initially, taxa containing the suffix '_unclassified' were retained as they indicated classification at higher taxonomic ranks, while entries explicitly labelled as 'Unclassified' were excluded (Schoch et al., 2020). Analysis was restricted to bacterial taxa, with abundance thresholds > 0 considered as present taxa.

Relative abundance comparisons were conducted at genus and species levels, focusing on the top 20 taxa by mean abundance across datasets. Agreement was quantified using Intraclass Correlation Coefficient (ICC) to assess absolute agreement (Koo & Li, 2016) strength of relationships between methods was further assessed using Spearman and Pearson correlations (Mukaka, 2012; Schober et al., 2018). Systematic bias detection was performed using Bland-Altman analysis with LOESS smoothing to visualize method differences (Giavarina, 2015; Singer et al., 2021). Additionally, log2 fold change calculations between methods were used to identify taxa-specific biases, complemented by paired Wilcoxon rank-sum tests to determine the statistical significance of methodological differences.

3.6.2. Distribution and abundance analysis

Abundance distributions were examined using density plots with log10 transformation to visualize detection ranges across methods (Gloor et al., 2017). The analysis focused on characterizing method-specific patterns in the detection of rare taxa (< 1.0×10^{-2}% abundance), distribution continuity, and the impact of rarefaction on abundance profiles. This approach allowed for detailed examination of each method's capability in capturing the full spectrum of microbial diversity, particularly at the extremes of abundance ranges.

3.6.3. Method agreement evaluation

Method agreement was assessed through multiple analytical approaches to provide a comprehensive understanding of methodological concordance. Agreement strength was quantified using R^2 values, interpreted according to established ecological standards (Cohen, 2013). Scatter plot analysis on log-transformed data was employed to evaluate abundance-dependent agreement patterns, revealing specific ranges where methods showed strongest concordance. Systematic bias evaluation through fold-change analysis (Rausch et al., 2019) provided insights into consistent methodological differences across taxonomic groups and abundance ranges.

3.6.4. Functional analysis comparison

i. Differential abundance analysis

Functional profiles were generated from shotgun metagenomic data using HUMAnN3, which maps reads to the UniRef90 protein database and reconstructs metabolic pathways, as previously described (section 3.5.7). The differential abundance analysis was performed using MaAsLin2, implementing a multivariable linear modelling approach. The model was constructed with no additional normalization or transformation of the input abundance data, which was indicated by the "NONE" parameter, since the data was previously normalized during the HUMAnN3 processing step (Figure 3.13).

The final multivariate model included seven fixed effects: *BMI group* classification; *batch* as a technical covariate to account for sequencing variation; *time in traffic* exposure; *cooking* method (electricity usage); *exercise type*; *heating* method (electricity usage); and *commute duration* (long vs short). The model assessed the relationship between these variables and microbial pathway abundances while controlling for potential confounding effects. Statistical significance was determined using default MaAsLin2 parameters, with results adjusted for multiple comparisons using the Benjamini-Hochberg procedure to control the false discovery rate (FDR). Associations were considered significant at an FDR-adjusted p-value < 5.0×10^{-2}.

ii. Comparative functional analysis

To compare functional predictions between methods, 16S-derived pathways were predicted using ggPICRUSt2, while shotgun metagenomic pathways were obtained from HUMAnN3. Pathway abundances were normalized to relative abundance and log10-transformed for comparison. Overlap between methods was assessed using pathway presence / absence analysis, and correlation between abundance predictions was evaluated using Pearson correlation coefficients (S. Sun et al., 2020). Pathways were classified as method-specific or shared based on detection patterns.

iii. Metadata integration analysis

PERMANOVA was performed to examine associations between functional profiles and metadata categories. The four distinct variable groups were analysed (diet, SES, lifestyle and health status), where PERMANOVA analyses were performed independently for both HUMAnN3 and PICRUSt2 results using the *vegan* package in R (Oksanen et al., 2022). The strength of associations was quantified using R^2 values and significance was assessed using FDR-corrected p-values. Visualization of association patterns was generated using ggplot2 (Kassambara, 2023) with separate analyses for each metadata category and sequencing method.

3.7. Core versus temporal microbiome profile: Longitudinal analysis

A longitudinal analysis was conducted using paired stool samples collected from *n* = 22 participants at two timepoints. This design enabled investigation of both core (stable) and temporal (variable) components of the microbiome within the urban professional cohort.

3.7.1. Compositional stability assessment

Microbiome stability was evaluated through multiple analytical approaches. Bray-Curtis dissimilarity metrics were calculated to compare intra-individual (within-person) versus inter-individual (between-people) variation. PERMANOVA was performed and stratified by subject to account for repeated measures. This approach quantified the proportion of variation explained by temporal changes. Then PCoA was applied to Bray-Curtis dissimilarities to visualize compositional patterns.

3.7.2. Diversity and taxonomic analysis

Alpha diversity was assessed using the Shannon diversity index, and temporal changes were evaluated using Wilcoxon rank-sum tests. Taxonomic composition was analysed across taxonomy levels (phylum to species), focusing on taxa present at > 0.1% relative abundance. Statistical comparisons between timepoints were performed using paired tests with FDR correction to control for multiple testing. Additionally, LEfSe was applied to identify discriminative features between timepoints (Segata et al., 2011).

3.7.3. Functional stability analysis

Functional profiles were generated using ggPICRUSt2 to predict KEGG Orthology (KO) pathways. Pathway stability was assessed through variance partitioning into inter- and intra-individual components (Olsson et al., 2022). The Intraclass Correlation Coefficient (ICC), a measure of reliability which quantified the proportion of variance attributable to between-subject differences relative to total variance, was used to classify pathway stability. ICC values ranged from 0 to 1, where values closer to 1 indicated greater stability across timepoints. The relationship between pathway abundance and variance was evaluated, incorporating Shannon diversity metrics to understand the association between community diversity and functional stability (Olsson *et al.*, 2022).

3.7.4. Integration with participant metadata

Subjects in the top quartile of the PCoA Bray-Curtis distances between their timepoints were identified as having the largest compositional shifts. For these subjects, a comprehensive comparison of all survey responses between timepoints, examining changes in health status, lifestyle factors and environmental exposures was conducted. Changes in categorical variables were observed as transitions between response options (e.g., "Unchecked to Checked", "No to Yes"), while changes in continuous variables were observed as direct numerical differences. This systematic comparison allowed the identifying of concurrent changes in subject characteristics that could potentially explain shifts in microbiome composition.

3.7.5. Core and variable component analysis

Core and variable taxa were identified using prevalence thresholds of 80% and 95% (Gacesa *et al.*, 2022), with both categories requiring a minimum mean relative abundance of 0.1%. At each taxonomic level (phylum to species), abundance and prevalence distributions were analysed and visualized to identify taxa meeting the core criteria thresholds.

Temporal dynamics were analysed by calculating log2 fold changes in abundance between timepoints for all taxa. A threshold of ±1 log2 fold change was established to classify taxa as temporally stable or variable. For species-level analysis, taxa were filtered using dual criteria: mean abundance > 10% and log2 fold change > 1. Visualization methods included mean abundance versus fold change plots,

with faceted versions generated to display individual taxa names and temporal patterns (Tamburini et al., 2022).

Functional stability was assessed using KEGG Ortholog (KO) annotations predicted from the taxonomic data (Kim et al., 2023). The analysis framework paralleled the taxonomic approach, applying the same prevalence thresholds (80% and 95%) to identify core functions. Temporal stability of KO functions was evaluated using log2 fold changes. Functions were filtered based on mean abundance > 0.1% and log2 fold change > 4. The identified functions meeting these criteria were compiled for detailed enzyme investigation in supplementary materials (Supplementary Data S4.10).

3.8. Comparative analysis to other cohort populations

This analysis established methodological frameworks for comparing South African (SAN) microbiome profiles with the two reference cohorts described in section 3.5.7: the American Gut Project (AGP) and AWI-Gen Phase 1 (AWI). The analytical approach accommodated the comparison of 16S rRNA gene sequencing data from SAN participants with shotgun metagenomic data from AGP and AWI cohorts.

3.8.1. Data processing and standardization

The initial processing phase focused on achieving compatibility between different sequencing platforms and taxonomic classification systems. SILVA-based taxonomic classifications derived from QIIME2 analysis of the SAN population data were harmonized with MetaPhlAn taxonomic classifications from AGP and AWI cohorts. This process involved standardizing phylum names and taxonomic prefixes across datasets, implementing consistent labelling protocols for unclassified taxa, and verifying taxonomic hierarchy consistency across all datasets (Schoch et al., 2020).

Abundance data processing followed a systematic approach across three taxonomic levels (phylum, genus, and species). The processing pipeline included the removal of taxa with zero counts across all samples, followed by relative abundance calculations. This standardization process enabled cross-platform comparisons while accounting for methodological differences between 16S rRNA gene and shotgun metagenomic sequencing approaches.

3.8.2. Diversity analysis and taxonomic analysis

Alpha diversity was quantified using the Shannon diversity index, while Beta diversity analysis was performed with log-Euclidean distances (Montalban-Arques et al., 2021; Wirbel et al., 2019). Community structure differences were assessed using PCoA at three taxonomic levels (phylum, genus, and species). The significance of between-group differences was evaluated using PERMANOVA with 999 permutations. This analysis was performed both globally across all cohorts and in pairwise comparisons between cohort pairs. The statistical framework included correction for multiple comparisons (Benjamini-Hochberg adjusted p-value < 5.0×10^{-2}) where appropriate. The

beta diversity analysis was structured to examine community differences at increasing taxonomic resolution, enabling an assessment of both broad phylogenetic patterns and fine-scale taxonomic variations. This multi-level approach was identified whether community structure differences were consistent across taxonomic ranks or emerged at specific levels of resolution.

Abundance data were aggregated at species, genus, and phylum levels. Low-abundance taxa were consolidated into a group labelled as "Other," and unclassified sequences were consistently labelled as "Unclassified" across all datasets to maintain uniformity in taxonomic classification. The statistical framework employed a multi-step approach beginning with Kruskal-Wallis tests to identify taxa with significant differences across cohorts. For taxa showing significant differences, pairwise Wilcoxon rank-sum tests were performed. Effect sizes were calculated using Cliff's delta, a measure of how often values in one distribution are larger than values in another distribution. A threshold of ≥ 0.47 was applied to identify biologically meaningful variations (Kargar et al., 2020). To account for multiple comparisons, p-values were adjusted using the Benjamini-Hochberg correction method for FDR.

3.8.3. Functional pathway analysis

Pathway abundances from PICRUSt2 and HUMAnN3 were both normalized to relative abundance, enabling meaningful comparisons of relative differences between methods. The comparative analysis framework included the calculation of log2 ratios between HUMAnN3 and PICRUSt2 predictions, as well as the computation of absolute differences in predicted abundances (Gehrig et al., 2022). A presence/absence analysis was conducted for pathways, with presence defined as abundance exceeding zero. Statistical analyses were performed at both pathway and functional category levels, with categories requiring a minimum of five pathways for inclusion in visualization and statistical testing. Non-parametric statistical approaches were used due to the non-normal distribution of ratios, and results were interpreted in an exploratory context to acknowledge the complexity of functional inference from marker gene data.

Demographic associations with functional profiles were assessed using PERMANOVA analysis, generating R^2 values to quantify the strength of relationships between demographic variables and gut microbiome functional profiles. The analysis included visualization through multiple complementary approaches: strength of association plots comparing R^2 values across demographic variables, method comparison scatter plots examining correlation between HUMAnN3 and PICRUSt2 predictions, significance pattern heatmaps displaying statistical significance using -log10(FDR) transformations, effect size distribution plots showing R^2 value distributions, and statistical summary plots comparing mean R^2 values and significant associations. Statistical significance was assessed using FDR correction, with the significance threshold set at FDR < 5.0×10^{-2}. Density plots visualize the distribution of R^2 values across demographic variables, and heatmaps displayed patterns of statistical significance between prediction methods.

3.9. Exploratory analysis of community clusters

Community stratification analysis in the SAN population investigated clustering formed with PCoA beta diversity from the log-Euclidean distances method. Taxonomic profiles were aggregated at the phylum level – a common practice in microbiome research for exploratory analyses, especially in cohorts of modest size, to enhance statistical power, mitigate sparsity, and reduce multiple-comparison burden resulting from lower-level taxonomic resolution (Busato et al., 2023; Nearing et al., 2022). Abundances were normalized to relative abundances. Differential abundance analysis between clusters utilized Kruskal-Wallis tests, with p-value adjustment using the Benjamini-Hochberg method. Associations between cluster membership and participant characteristics were evaluated using Fisher's exact tests for categorical variables, examining relationships with factors such as *COVID-19 diagnosis, cosmetics use*, dietary habits, and lifestyle choices. For identified associations, subsequent Kruskal-Wallis tests determined which bacterial phyla drove these relationships, with p-values again adjusted using the Benjamini-Hochberg correction for multiple comparisons. This exploratory analysis enabled systematic identification of SAN community patterns and their associations with participant characteristics.

4. Results

4.1. Descriptive analysis

4.1.1. Study cohort characteristics

The study population consisted of n = 220 urban professional adults recruited primarily from the Council for Scientific and Industrial Research (CSIR, Pretoria) and the University of the Witwatersrand (Wits, Johannesburg). Participants' mean age was 38.2 (± 15.5) years, ranging from 18 to 72 years (excluding irregular data entries), with an average BMI of 25.3 (± 5.2) kg/m², indicating a predominantly normal-to-overweight demographic (Table 4.1). A 60.5% majority were of European ancestry (White South Africans). Females represented approximately two-thirds (68%) of the population and most participants had attained high educational levels (71% postgraduate or professional degrees). In addition to being highly educated and predominantly middle-to-high income, (~80%) most participants followed omnivorous diets (93%), rarely excluded dairy (90% did not) or refined sugars (93% did not), and a third reported multivitamin (35%) or probiotic use (44%). Lifestyle exposures included high-fat food consumption (84%), frequent alcohol use (75% overall; 41% red wine), low rates of smoking (11%), and substantial cosmetic use (63%), with most relying on private transport (87%) and spending up to 20% of their day in traffic. Health-related factors with previously studied microbiome links were common: recent antibiotics (53% within the past year), food intolerances (13% lactose; 13% gluten), IBS (22%), mental health conditions (26%), and prior COVID-19 diagnosis (53%). Further demographic details and lifestyle factors, are comprehensively summarized in Table 4.1 and Appendix One (Supplementary Tables S4.1 – S4.2).

Table 4.1 Descriptive characteristics of the urban professional South African study population

Variable	Category	*n*	percent
Age (years)	**Mean ± SD**	**220**	**38.2 ± 15.5**
Age Group	Young adults (18-35 years)	112	50.9%
	Older adults (35+ years)	102	46.4%
	NA	6	2.7%
Gender	Female	151	68.6%
	Male	69	31.4%
BMI (kg/m^2)	**Mean ± SD**	**220**	**25.3 ± 5.2**
BMI Group	Normal weight (<25)	115	52.3%
	Overweight (<30)	60	27.3%
	Obese (30+)	37	16.8%
	Underweight (<18.5)	8	3.6%
Ancestry	European	133	60.5%
	African	64	29.1%
	Asian/Multi-ancestry/Other	23	10.5%
Education	Post Graduate or Professional degree	157	71.4%
	Other	63	28.6%
Household Income	High	110	50%
	Middle	87	39.5%
	Low	23	10.5%
Pets	Yes	139	63.2%
	No	81	36.8%
Work location	Other	166	75.5%
	CSIR	54	24.5%
Refuse	Refuse disposal services	203	92.3%
	Other	17	7.7%
Transport - Private	Yes	192	87.3%
	No	28	12.7%
Transport - Public	No	193	87.7%
	Yes	27	12.3%
Transport - Walk	No	205	93.2%
	Yes	15	6.8%
Time in traffic	Less than 10%	103	46.8%
	10-20%	57	25.9%
	20-40%	39	17.7%
	More than 40%	21	9.5%

Variable	Category	*n*	percent
How long in SA	All my life	175	79.5%
	Other	45	20.5%
Diet classification	Omnivore	205	93.2%
	Other	15	6.8%
High Protein	No	205	93.2%
	Yes	15	6.8%
Exclude Dairy	No	198	90%
	Yes	22	10%
Exclude Refined Sugars	No	204	92.7%
	Yes	16	7.3%
Gluten-free	No	208	94.5%
	Yes	12	5.5%
Multivitamins	No	143	65%
	Yes	77	35%
Herbal supplements	No	152	69.1%
	Yes	68	30.9%
Probiotic	No	123	55.9%
	Yes	97	44.1%
Water source	Tap	148	67.3%
	Non-tap	72	32.7%
Average consumption - high fat diet	Yes	185	84.1%
	No	35	15.9%
Exercise type	Endurance	61	27.7%
	Other	47	21.4%
	Strength	37	16.8%
	Flexibility & Balance	23	10.5%
	Sport	11	5%
	All of the above	31	14.1%
	Not applicable	10	4.5%
Smoke	No	197	89.5%
	Yes	23	10.5%
Cosmetics	Yes	138	62.7%
	No	82	37.3%
Hours of sleep	Less than 7 hours	139	63.2%
	7+ hours	81	36.8%
Consume alcohol	Yes	165	75%
	No	55	25%
Beer/Cider	No	148	67.3%

Variable	Category	*n*	percent
	Yes	72	32.7%
White Wine	No	151	68.6%
	Yes	69	31.4%
Red Wine	No	129	58.6%
	Yes	91	41.4%
Spirits or Hard Liquor	No	147	66.8%
	Yes	73	33.2%
Number of alcoholic drinks consumed	0-2	120	54.5%
	0	53	24.1%
	2-4	36	16.4%
	More than 4	11	5%
Lactose intolerance	No	154	70%
	I don't know	37	16.8%
	Yes	29	13.2%
Gluten intolerance	No	192	87.3%
	Intolerance	28	12.7%
Antibiotics	No	103	46.8%
	Six months or less	82	37.3%
	Year	35	15.9%
Flu vaccination	No	137	62.3%
	Year	48	21.8%
	Six months or less	35	15.9%
Birth control	No	172	78.2%
	Yes	48	21.8%
Appendix removed	No	192	87.3%
	Yes	28	12.7%
Medication for conditions	Yes	123	55.9%
	No	97	44.1%
Born via C-section	No	177	80.5%
	Yes	43	19.5%

Variable	Category	*n*	percent
Infancy diet	Primarily breast milk	109	49.5%
	Primarily infant formula	50	22.7%
	Equal measures of formula and breast milk	34	15.5%
	Not sure	27	12.3%
Allergies	Yes	111	50.5%
	No	109	49.5%
Diagnosed with COVID	Yes	116	52.7%
	No	104	47.3%
IBS	No	171	77.7%
	Yes	49	22.3%
ADHD	No	192	87.3%
	Yes	28	12.7%
Autoimmune disease	No	200	90.9%
	Yes	20	9.1%
Skin condition	No	173	78.6%
	Yes	47	21.4%
Migraines	No	171	77.7%
	Yes	49	22.3%
Lung disease	No	193	87.7%
	Yes	27	12.3%
Fungal infection	No	192	87.3%
	Yes	28	12.7%
Thyroid disease	No	195	88.6%
	Yes	25	11.4%
Mental Health conditions	No	162	73.6%
	Yes	58	26.4%
Depression	No	169	76.8%
	Yes	51	23.2%
GERD	No	141	64.1%
	Yes	79	35.9%

4.1.2. Data collection and sample processing overview

Between September 2021 and September 2023, a total of *n* = 243 biological samples were collected, with *n* = 242 successfully sequenced using 16S rRNA gene long-read sequencing and *n* = 49 shotgun metagenomic short-read sequencing samples. The longitudinal subset involved deeper analysis of *n* = 22 samples (detailed in Appendix One, Supplementary Table S4.1).

4.1.3. Data cleaning and variable selection

Initial survey data (*n* = 123 questions) underwent rigorous cleaning and refinement. Variables with minimal variability or sparse responses were excluded or consolidated, and numeric data were categorized for analytical simplicity. Ultimately, *n* = 101 variables were retained for downstream statistical analyses. Detailed outcomes of data cleaning processes are documented in Appendix One (Supplementary Table S4.2).

4.1.4. Significant demographic associations

Statistical analyses (Fisher's exact tests with Bonferroni corrections) revealed multiple significant associations between demographic variables and socioeconomic status (SES), dietary, lifestyle, exercise, and health-related factors (Appendix One, Tables A1.2 – A1.6). Notably:

- **SES**: Strong associations were identified between demographic variables (age, gender, ancestry) and SES factors; for instance, gender showed robust associations with household income and transport mode (Appendix One, Table A1.2).
- **Diet**: Age and gender prominently influenced dietary preferences and special dietary practices, including high-protein, dairy-free, and gluten-free diets (Appendix One, Table A1.3).
- **Lifestyle**: Lifestyle variables, including alcohol consumption, pet ownership, probiotics usage, and sleep duration, were significantly associated with age, gender, ancestry, and BMI categories (Appendix One, Table A1.4).
- **Exercise**: Strong associations were found between demographic factors and exercise type and location, indicating distinct exercise behaviours across age, gender, and ancestry groups (Appendix One, Table A1.5).
- **Health and clinical status**: Variables such as lactose and gluten intolerance, allergies, mental health, and thyroid conditions significantly correlated with demographic factors (Appendix One, Table A1.6).

These associations highlight demographic factors that were considered as covariates or potential confounders in subsequent microbiome analyses.

4.2. Sequencing data quality control (QC)

4.2.1. 16S rRNA gene sequencing (long-read data)

Detailed QC was conducted using QIIME2-DADA2, resulting in robust sequencing data quality (median sequence length ~1,452 nts). After stringent filtering, primer removal, denoising, and chimera removal, a high-quality dataset was established, suitable for downstream microbial diversity analyses. Rarefaction normalization (to 90 sequences/sample) balanced sample retention and microbial diversity detection, with no significant variation in alpha or beta diversity across rarefaction depths (Appendix One, Figures A1.1 – A1.7; Supplementary Data S4.2 – S4.3).

4.2.2. Shotgun metagenomic sequencing (short-read data)

QC of shotgun metagenomic data employed FastQC and a comprehensive Metagenome-Assembled Genome (MAG) pipeline. Initial dataset size was substantially reduced during deduplication, trimming, and host DNA removal, yielding high-quality, host-free microbial reads. Metagenome assembly produced *n* = 887 high-quality genome bins meeting stringent completeness and contamination thresholds. Genome assembly metrics (N50, L50, contig lengths) demonstrated high-quality assemblies suitable for microbial functional profiling (Appendix One, Figures A1.8 – A1.10; Supplementary Data S4.4).

4.3. Gut microbiome composition in the urban South African population

The characterization of the gut microbiome composition in the urban professional South African population revealed distinct patterns of diversity across this group, using the long-read and shotgun sequencing datasets. Key demographic variables included age group, gender, BMI group, SES, dietary, lifestyle and health factors. These findings provide a foundational understanding that inform the subsequent analyses and discussions regarding the implications of these associations for health and disease within the cohort.

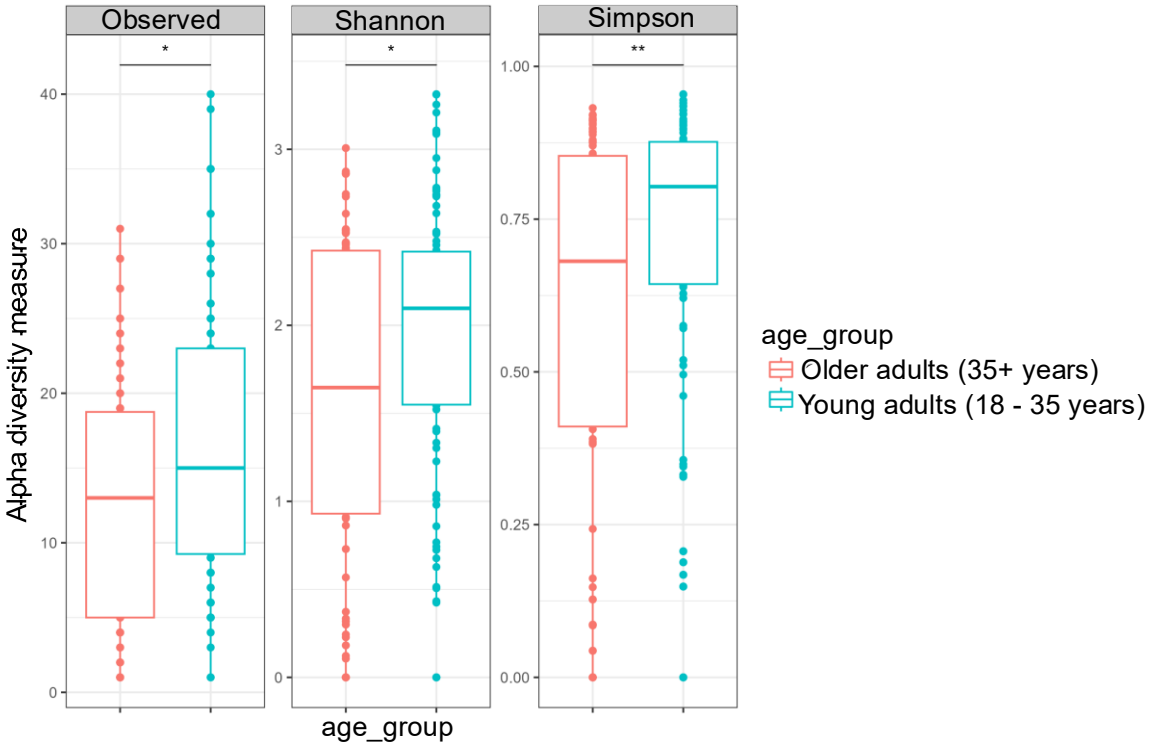
Section 4.3 starting on page 85 focuses on microbial community richness and differential taxonomic abundance analyses using the long-read data. Additional detailed results from the functional prediction analyses, including comparisons with reference groups, are provided in Appendix One; Section 4.4 from page 113 examines community stratification patterns within the cohort and their associations with lifestyle factors; Section 4.5 from page 120 provides an overview of concordance between the long-read and short-read data from a subset of the cohort, further detailed in Appendix One; Section 4.6 from page 121 investigates core versus temporal dynamics of the gut microbiome through longitudinal sample analysis; and Section 4.7 starting on page 150 presents comparative analyses with other cohort populations including Western (AGP) and urban-rural African (AWIGEN1) cohorts.

4.3.1. Microbial community richness estimates and differences using long-read data

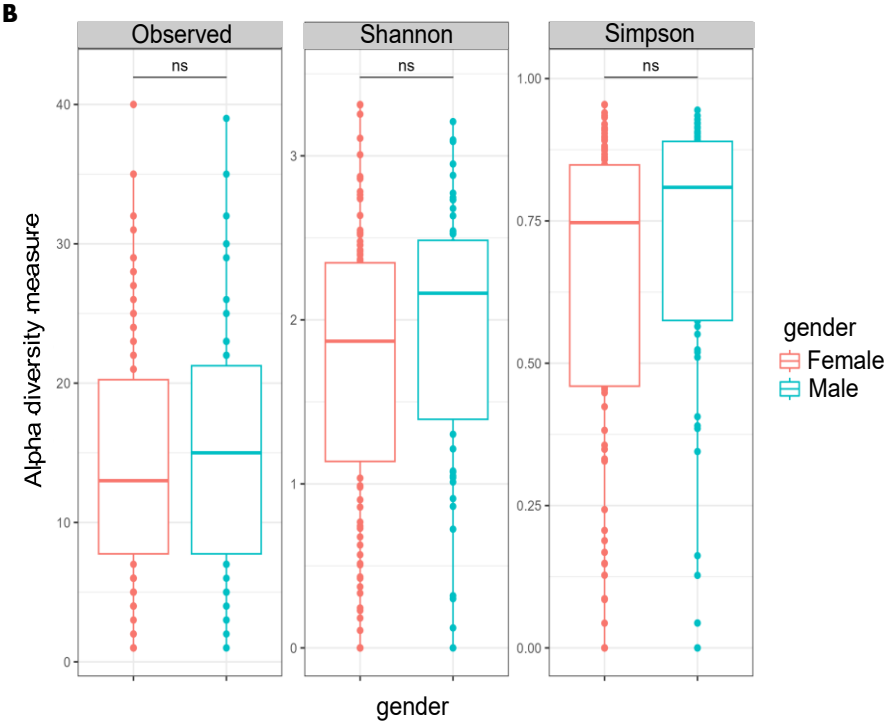
i. Alpha diversity across demographic variables

The distribution of alpha diversity within the gut microbiome across various demographic variables in the urban South African population is depicted in Figure 4.1. Alpha diversity was assessed using three indices – Observed, Shannon, and Simpson – which capture different aspects of microbial richness and evenness. In the age group comparison (Figure 4.1A), significant differences in microbiome diversity between older adults (35+ years) and young adults (18 – 35 years) were identified, with p-values derived from Wilcoxon rank-sum tests that were subsequently adjusted for multiple comparisons using the Benjamini-Hochberg method: Observed ($P = 2.0 \times 10^{-2}$, adj. $P = 2.0 \times 10^{-2}$) and Shannon indices ($P = 2.0 \times 10^{-2}$, adj. $P = 2.0 \times 10^{-2}$) showed concordant shifts, while the Simpson index exhibited stronger differentiation ($P = 9.0 \times 10^{-3}$, adj. $P = 2.7 \times 10^{-2}$), suggesting age disproportionately impacts dominant taxa. However, no significant differences were observed across gender (Figure 4.1B) and BMI (Figure 4.1C) categories. Similarly, the ancestry comparison (Figure 4.1D) revealed no significant differences in alpha diversity between African, Asian/Multiancestry/Other, and European groups. These univariate analysis findings highlighted the complex interplay between age and the gut microbiome, with significant influence on diversity.

A


Observed
Shannon
Simpson
*
*
**
Alpha diversity measure
age_group
age_group
Older adults (35+ years)
Young adults (18 - 35 years)

B


Observed
Shannon
Simpson
ns
ns
ns
Alpha diversity measure
gender
gender
Female
Male

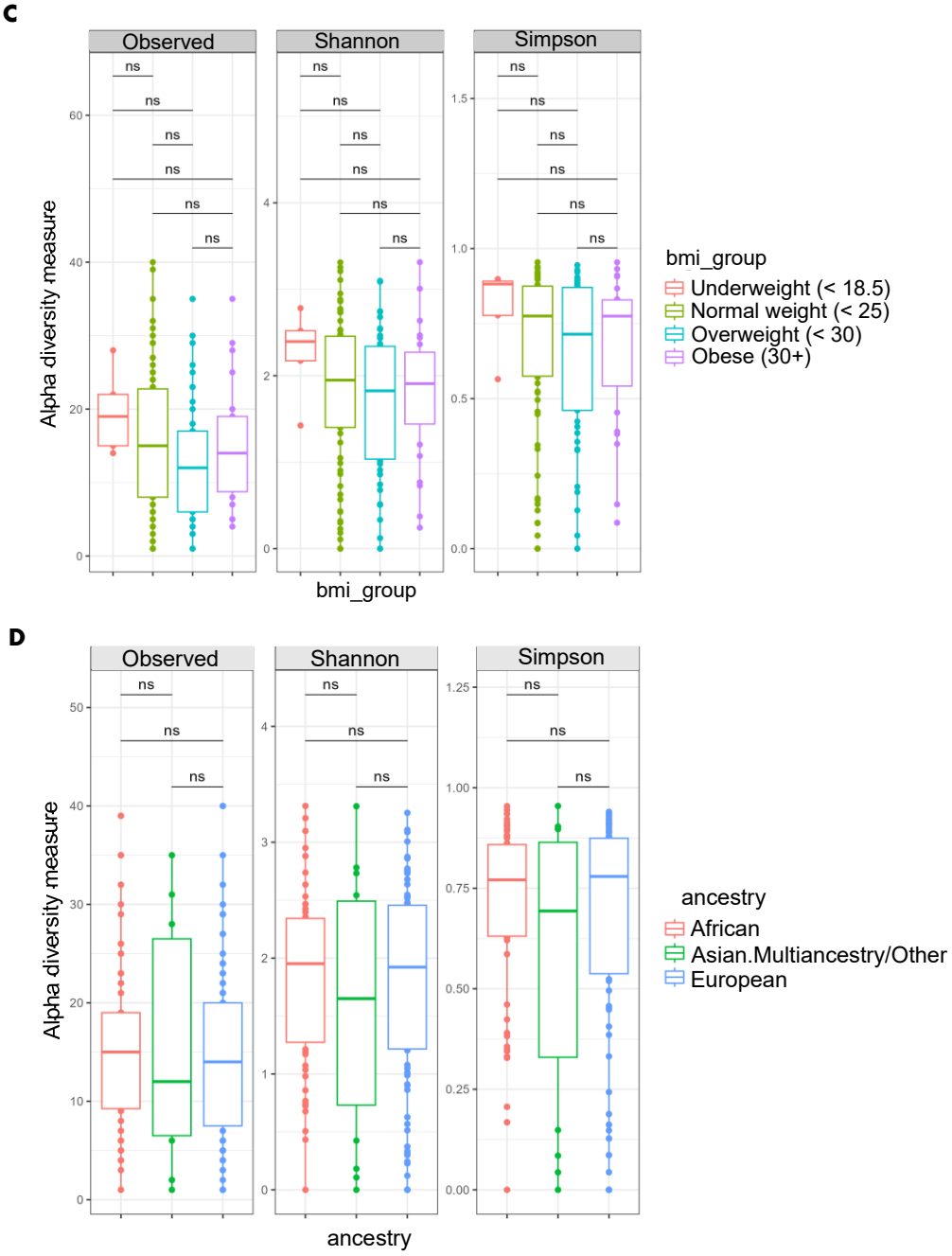


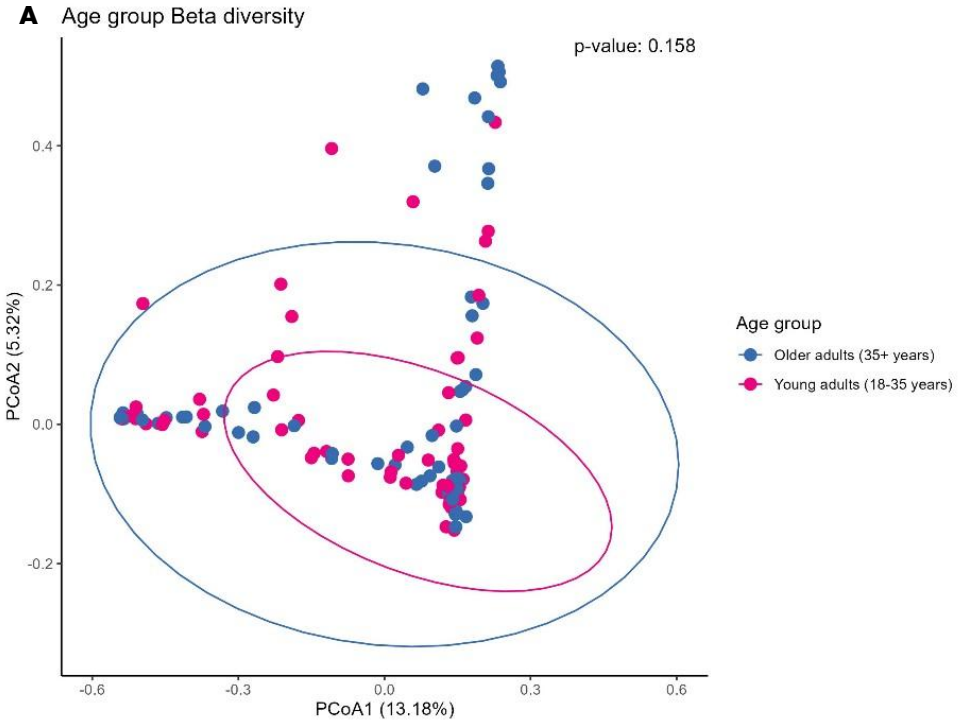
Figure 4.1 Comparison of alpha diversity measures across demographic variables. This figure illustrates the distribution of alpha diversity of A) two demographic age groups: older adults (35+ years) and young adults (18-35 years); B) two gender categories: female and male; C) four BMI categories: underweight (<18.5), normal weight (<25), overweight (<30), and obese (30+); and D) three ancestry categories: African, Asian/Multiancestry/Other, and European. Three different metrics of alpha diversity are presented (left to right): Observed (count of unique taxa), Shannon (taxa abundance and evenness), and Simpson indices (count and abundance of taxa). Each boxplot represents the interquartile range (IQR) of the diversity index, with the median indicated by the horizontal line within the box. Outliers are represented as individual points. The significance of differences between variables is denoted by asterisks above the plots: $p < 5.0 \times 10^{-2}$ = *, $p < 1.0 \times 10^{-2}$ = **, "ns" = non-significant, as measured by the Wilcoxon rank-sum test with a p-value of $< 5.0 \times 10^{-2}$. The Observed and Shannon indices show significant differences between the groups, while the Simpson index shows a more pronounced difference, suggesting variability in community richness and evenness between the age groups only (A). "Asian/Multiancestry/Other": East Asian or Pacific Islander, Indian, Coloured.

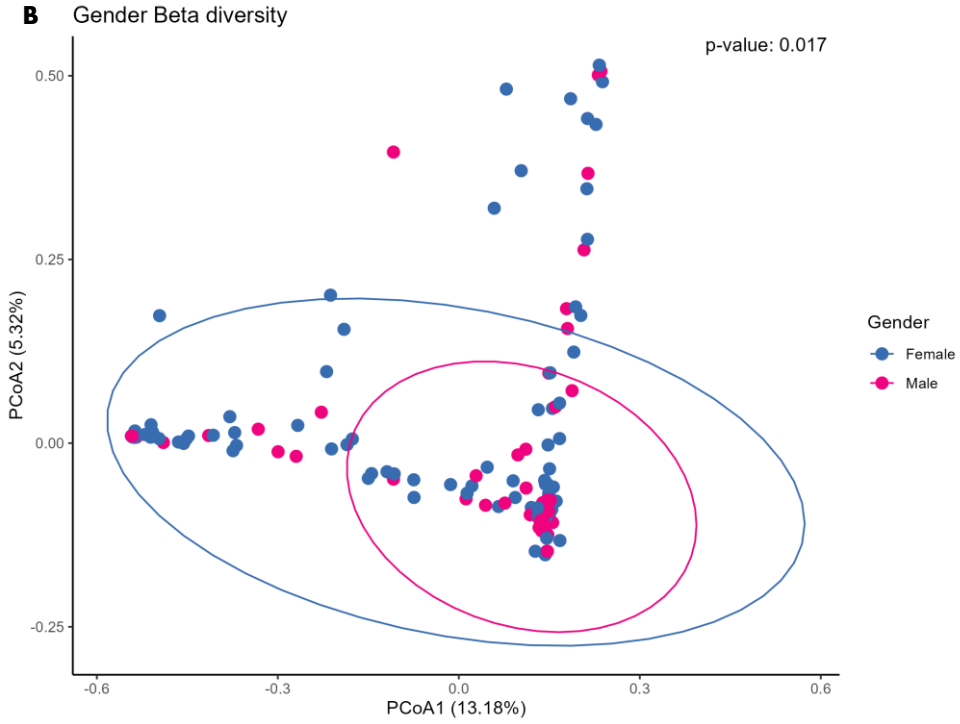
ii. Beta diversity across demographic variables

In Figure 4.2, the Bray-Curtis Dissimilarity-based PCoA analysis revealed nuanced distinctions in microbial community composition across demographic variables within the urban South African population, extending the insights from previous alpha diversity findings in Figure 4.1. The age-related significant differences in alpha diversity, indicative of variations in microbial evenness and richness, are not paralleled in the beta diversity outcomes; Figure 4.2A shows no significant clustering of microbiota between older and younger adults (p-value: 0.2). This lack of distinction suggests that both age groups maintain similar levels of community diversity, though the specific microbial species and their relative abundances may still differ substantially between older and younger adults.

In contrast, gender, which had previously shown no significant alpha diversity differences (Figure 4.1B), now displayed a significant divergence in microbial communities (Figure 4.2B, p-value: 1.7×10^{-2}). The PCoA plot shows female samples occupying a broader area of the ordination space compared to male samples, implying differences in community variability between genders. BMI, while not affecting alpha diversity, hinted at a subtle difference in microbial species composition that was close to reaching statistical significance (Figure 4.2C, p-value: 5.7×10^{-2}), pointing towards a nuanced relationship between body weight and microbial variance.

Ancestry, consistently with Figure 4.1, showed no significant impact on microbial composition (Figure 4.2D, p-value: 0.1), implying a similarity in microbiome diversity across different ancestral backgrounds in the cohort. Collectively, these univariate analysis results emphasized the differential effects demographic factors have on the microbiome's diversity and composition, with some influencing the variety within communities and others shaping the distinctness between them.


A
Age group Beta diversity
p-value: 0.158
PCoA2 (5.32%)
PCoA1 (13.18%)
Age group
Older adults (35+ years)
Young adults (18-35 years)


B
Gender Beta diversity
p-value: 0.017
PCoA2 (5.32%)
PCoA1 (13.18%)
Gender
Female
Male

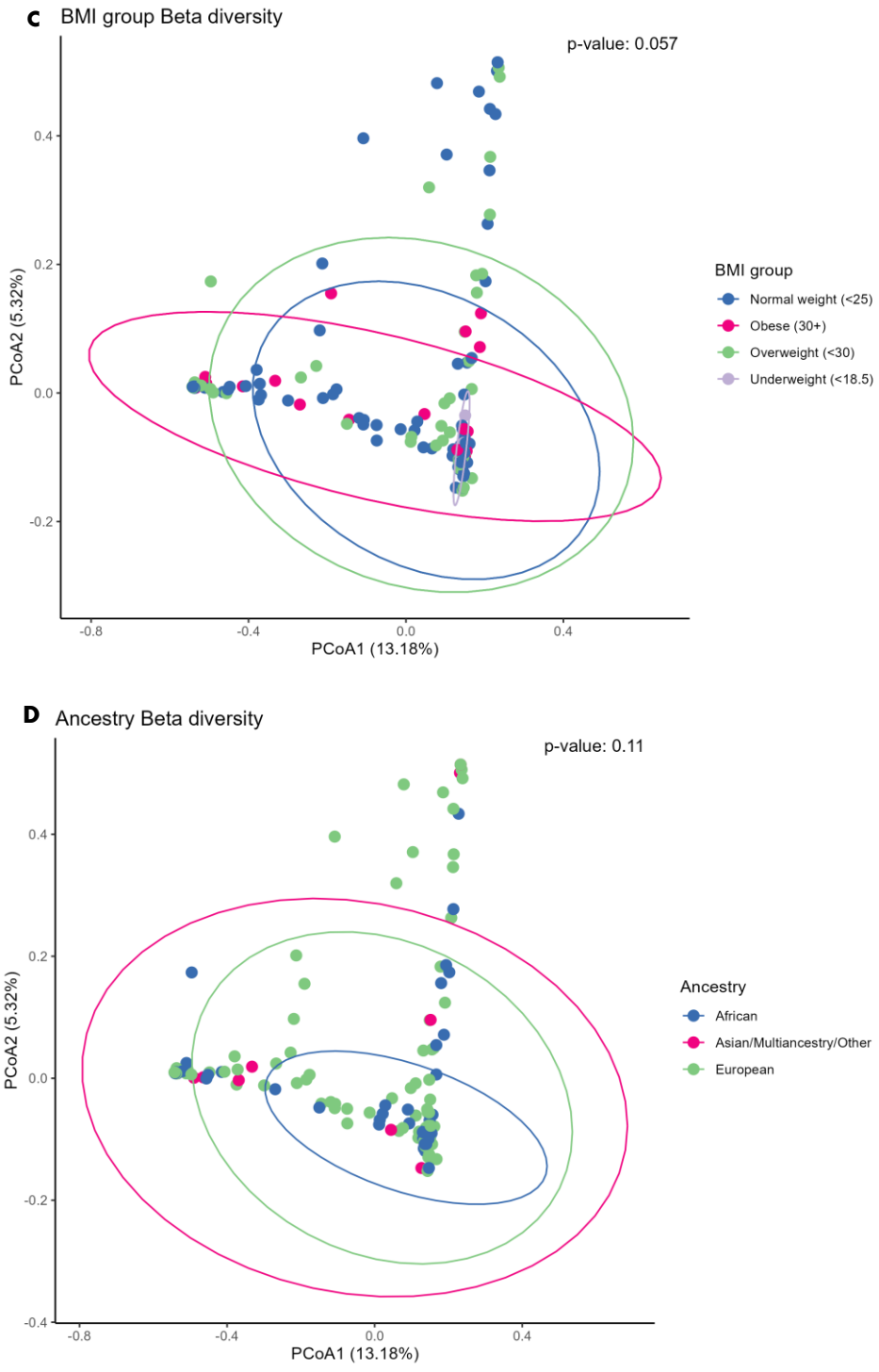


Figure 4.2 Beta diversity principal coordinates analysis (PCoA) plots with Bray-Curtis Dissimilarity measure. This figure showcases the variation in microbial community composition between: A) age groups, B) gender, C) BMI groups, and D) ancestry. The plot illustrates the first two principal coordinates, which explain 13.2% and 5.3% of the variance, respectively. Points represent individual samples, coloured according to the group comparisons in the legend. Ellipses represent the 95% confidence regions for each group. Inset p-values resulted from PERMANOVA analysis between compared groups, where $p < 5.0 \times 10^{-2}$ indicates statistical significance.

iii. Diversity across response variables

For diversity analysis across the SES, dietary, lifestyle and health response categories, variables that met the descriptive statistics criteria of a significant Bonferroni adjusted p-value ($P < 5.0 \times 10^{-2}$) and strong effect size (Cramer V ≥ 0.5), were selected for downstream diversity univariate analysis (Supplementary Table S4.1A).

The alpha diversity figures shown combine a box plot with a violin plot to showcase the Shannon Diversity Index for the compared groups. The biplot provided a detailed view of the distribution and median values of microbial diversity for each subgroup. The violin plot aspect revealed the density of data points and potential multimodal distributions, while the box plot component highlighted the median, quartiles, and outliers. The statistical significance of these comparisons was measured with Wilcoxon-rank sum testing, with the unadjusted p-values denoted on the figures. This visualization was particularly advantageous for emphasizing the differences in microbial diversity within the specific response categories, enabling focused and in-depth subgroup comparative analysis.

For beta diversity, principal coordinates analysis (PCoA) plots were used to visualise the clustering and spread of similar and dissimilar samples between the subgroups respectively, as measured by the Weighted UniFrac Distance. This measure considers the phylogenetic distances between observed ASVs (i.e. taxa). In addition, it is a weighted measure, therefore incorporates both the presence, absence and the relative abundance of ASVs, thus more abundant taxa would contribute more to the patterns observed. Statistical significance of these comparisons was measured with PERMANOVA, with p-values depicted on the PCoA (Supplementary Data S4.7).

Socioeconomic status

The descriptive statistics criteria resulted in four SES variables for diversity analysis, namely: household income, public transport, private transport and walking. The use of public transport was singly significant when measured for Shannon diversity, indicating a greater richness and evenness of diversity in individuals that indicated a positive “checked” response (Figure 4.3).

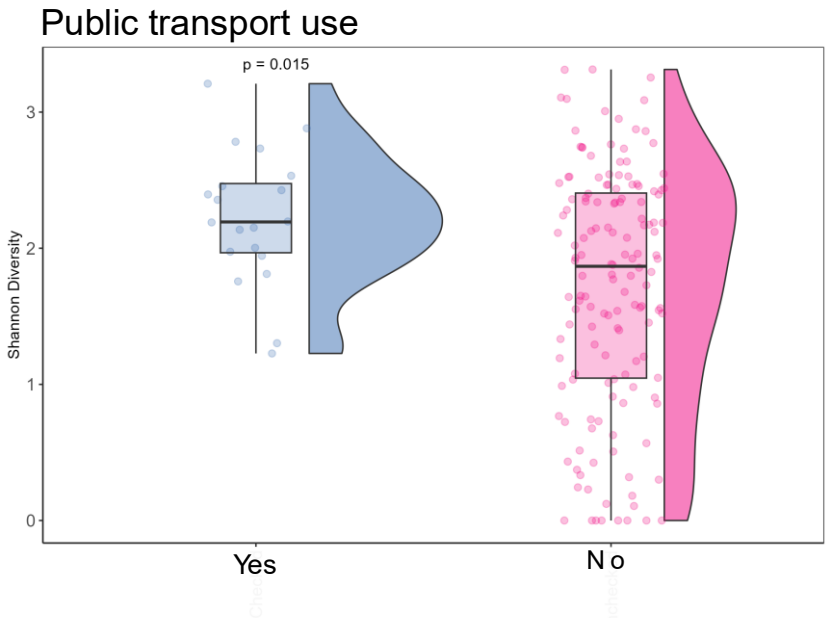


Figure 4.3 Comparison of Shannon diversity indices between individuals who use public transport ("Yes/Checked") and those who do not ("No/Unchecked"). The boxplots represent the interquartile range (IQR) and median Shannon diversity, with the overlaid violin plots indicating the overall data distribution. Individual data points are shown as dots. The statistical significance of the difference in diversity is denoted at the top (Wilcoxon-rank sum test p = 0.015), suggesting a higher Shannon diversity among individuals who use public transport compared to those who do not.

Diet

For diversity analysis across dietary-related responses, six variables met the criteria for inclusion, and "classify diet" was added for testing due to its overarching nature of capturing omnivorous vs other diet responses (Supplementary Data S4.7). The six main variables describing different diets followed in this cohort were: *high protein, exclude dairy, exclude refined sugars, gluten free, specialised diet (none)* and *infant food*.

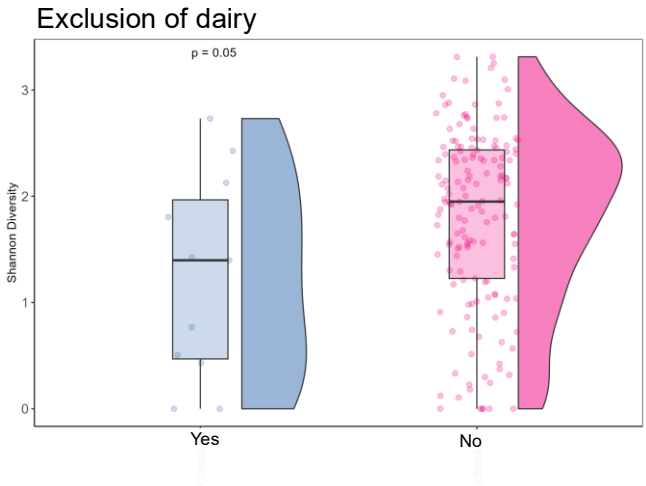


Figure 4.4 Comparison of Shannon diversity indices between individuals who exclude dairy ("Yes/Checked") and those who do not ("No/Unchecked"). The boxplots represent the interquartile range (IQR) and median Shannon diversity, with the overlaid violin plots indicating the overall data distribution. Individual data points are shown as dots. The statistical significance of the difference in diversity is denoted at the top (Wilcoxon-rank sum test p = 0.05), suggesting a higher Shannon diversity among individuals who consume dairy compared to those who do not.

Individuals that included dairy from their diets were uniquely highlighted as having a statistically higher richness and evenness within this subgroup ($P = 5.0 \times 10^{-2}$) as measured by Shannon diversity (Figure 4.4).

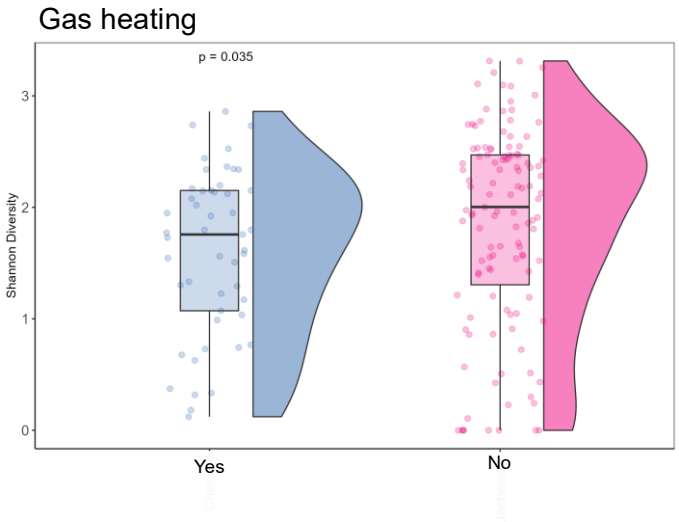


Figure 4.5 Comparison of Shannon diversity indices between individuals who use gas heating ("Yes/Checked") and those who do not ("No/Unchecked"). The boxplots represent the interquartile range (IQR) and median Shannon diversity, with the overlaid violin plots indicating the overall data distribution. Individual data points are shown as dots. The statistical significance of the difference in diversity is denoted at the top (Wilcoxon-rank sum test p = 5.0×10^{-2}), suggesting a higher Shannon diversity among individuals who do not use gas heating compared to those who do.

Lifestyle

In the diversity analysis for lifestyle, *n* = 21 response variables were investigated under this category (Supplementary Data S4.7). Of these, two emerged with a significant p-value for Shannon alpha diversity analysis: individuals who indicated a negative response to the use of *gas heating* ($p = 3.5 \times 10^{-2}$, Figure 4.5) as well as *multivitamin use* ($p = 9.0 \times 10^{-3}$, Figure 4.6), had a statistically higher richness and evenness in microbial diversity. The use of *herbal supplements* (Figure 4.7) was uniquely highlighted in beta diversity lifestyle analysis as having a statistical difference in microbial composition and abundance, with a PERMANOVA p-value of 4.7×10^{-2}.

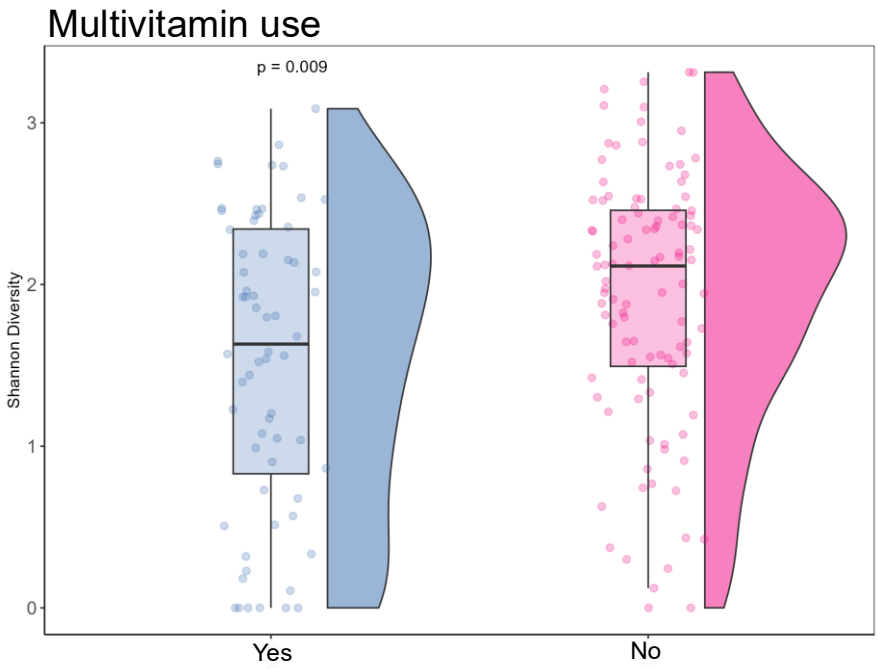


Figure 4.6 Comparison of Shannon diversity indices between individuals who consume multivitamins ("Yes") and those who do not ("No"). The boxplots represent the interquartile range (IQR) and median Shannon diversity, with the overlaid violin plots indicating the overall data distribution. Individual data points are shown as dots. The statistical significance of the difference in diversity is denoted at the top (Wilcoxon-rank sum test $p = 5.0 \times 10^{-2}$), suggesting a higher Shannon diversity among individuals who do not consume multivitamins compared to those who do.

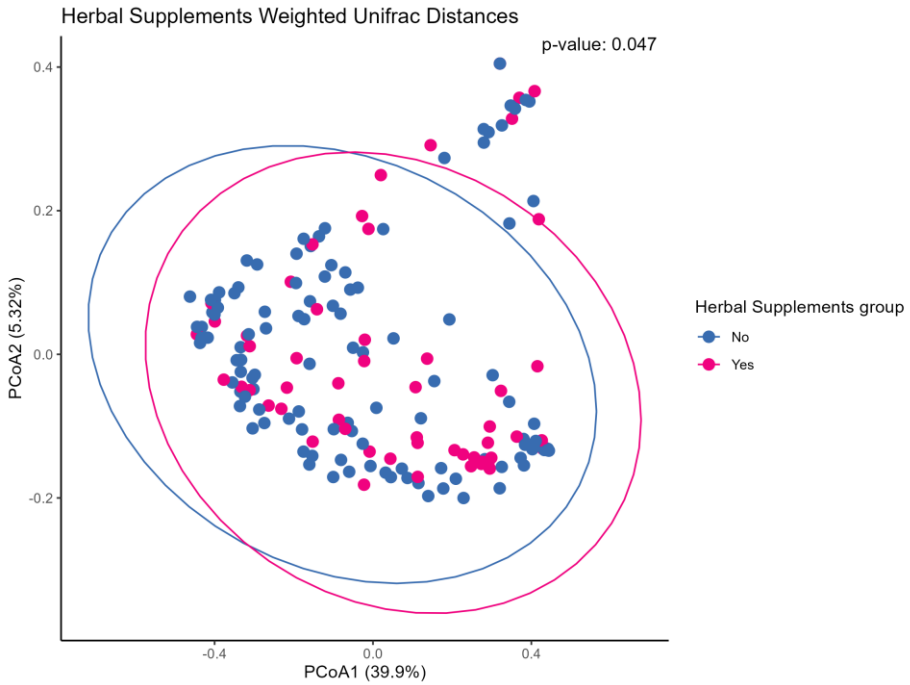


Figure 4.7 Principal Coordinates Analysis (PCoA) plot of Weighted UniFrac Distances comparing microbial community composition between individuals who use herbal supplements ('Yes') and those who do not ('No'). Each point represents an individual's microbial community, with ellipses encapsulating the 95% confidence intervals of the group centroids. The statistical significance of the difference in diversity is denoted at the top (PERMANOVA p-value = 4.7 x 10^{-2}), suggesting a difference in microbial community composition between those who use herbal supplements compared to those who do not.

Health status

In the analysis of health status related variables, *n* = 12 responses met the criteria for selection. Three additional variables related to the use of *antibiotics, irritable bowel syndrome diagnoses* and reporting of any *conditions* were included in this analysis. One variable under the health factors, which was indicating whether individuals had no food allergies, emerged as having a significant difference in beta diversity composition and abundance (Figure 4.8, $P = 4.2 \times 10^{-2}$).

The analyses reveal that specific variables related to socioeconomic status (SES), diet, lifestyle, and health status significantly contribute to the diversity in microbial profiles. These variables, identified through descriptive statistics as having strong and statistically significant associations, serve as proxies for demographic-driven diversity. On a univariate level, the results highlight response subgroups within these demographic categories that reflect differences in their environments, cultures, and lifestyles, even within an urban professional setting. This sets the stage for subsequent investigations into the taxonomic profiles and key bacterial species driving these observed differences.

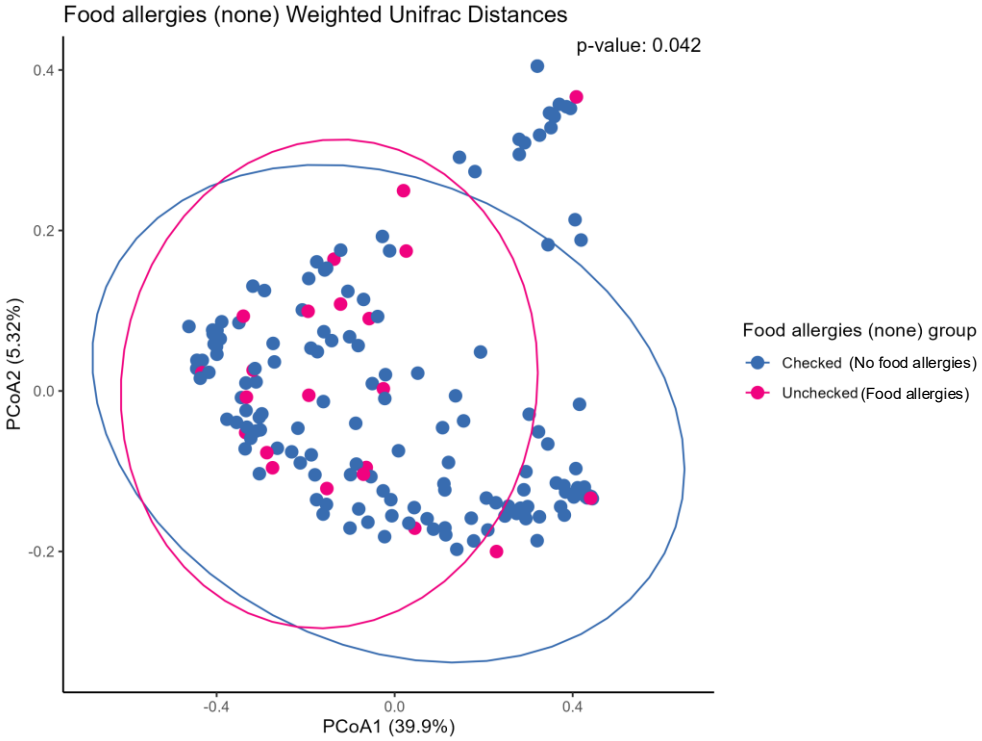


Figure 4.8 Principal Coordinates Analysis (PCoA) plot of Weighted UniFrac Distances comparing microbial community composition between individuals who do not have food allergies ('Checked') and those who do ('Unchecked'). Each point represents an individual's microbial community, with ellipses encapsulating the 95% confidence intervals of the group centroids. The statistical significance of the difference in diversity is denoted at the top (PERMANOVA p-value = 4.7 x 10^{-2}), suggesting a difference in microbial community composition between those who use have food allergies compared to those who do not.

4.3.2. Differential taxonomic abundance analyses with LEfSe

Linear Discriminant Analysis (LDA) Effect Size (LEfSe) is used for identifying biomarkers in microbiome datasets that may distinguish different groups based on their microbial compositions. In this study, LEfSe analysis was utilized to explore the differential abundance of microbial taxa across the demographic and survey-response variables at the species level, or otherwise lowest possible phylogenetic rank. The Kruskal-Wallis (KW) test was used to detect taxa with significantly different abundances across multiple response subgroups. This was followed by the Wilcoxon rank-sum test for pairwise comparisons between these groups when necessary. LEfSe was performed with a significance threshold set at $P < 5.0 \times 10^{-2}$ and a minimum LDA effect size greater than two, allowing for the detection and ranking of microbial taxa based on their effect sizes (Supplementary Data S4.8). This method is particularly useful for highlighting significant taxa that may not be apparent through abundance data alone due to their low relative abundance.

The significance of taxa is visually represented in the following dot plots where the size of each dot correlates with the negative logarithm of the p-value from these statistical tests, primarily from the KW test, or the Wilcoxon test in cases of pairwise comparisons. Although not meeting the threshold criteria for downstream variable selection, the lifestyle and health variables – *antibiotics, irritable bowel*

syndrome (IBS) diagnoses and the indication of any health-related "*conditions*" – were included in the analysis due to their strong links with the gut microbiome (Duan et al., 2024).

i. Differential abundance across demographic variables

LEfSe analysis was conducted to explore the influence of age, gender, BMI group and ancestry on the gut microbiome composition (Supplementary Data S4.8). From this analysis, specific enterotypes (*Bacteroides, Prevotella* and *Ruminococcus*) at the genus level that were associated with the distinct demographic groups were noted. Overall, *n* = 33 species were identified as differentially abundant with statistical significance across these demographic groups, where *n* = 22 of those species were unique.

For age group analysis, *n* = 3 species emerged, namely *Bifidobacterium longum* and *Roseburia faecis* in young adults, and *Barnesiella intestinihominis* in older adults (Figure 4.9). In young adults *n* = 19 genera were noted with *Prevotella* enterotype presence, whereas older adults had *n* = 5 genera with *Ruminococcus* presence, suggesting age-related differences in microbial composition (Supplementary Data S4.8A).

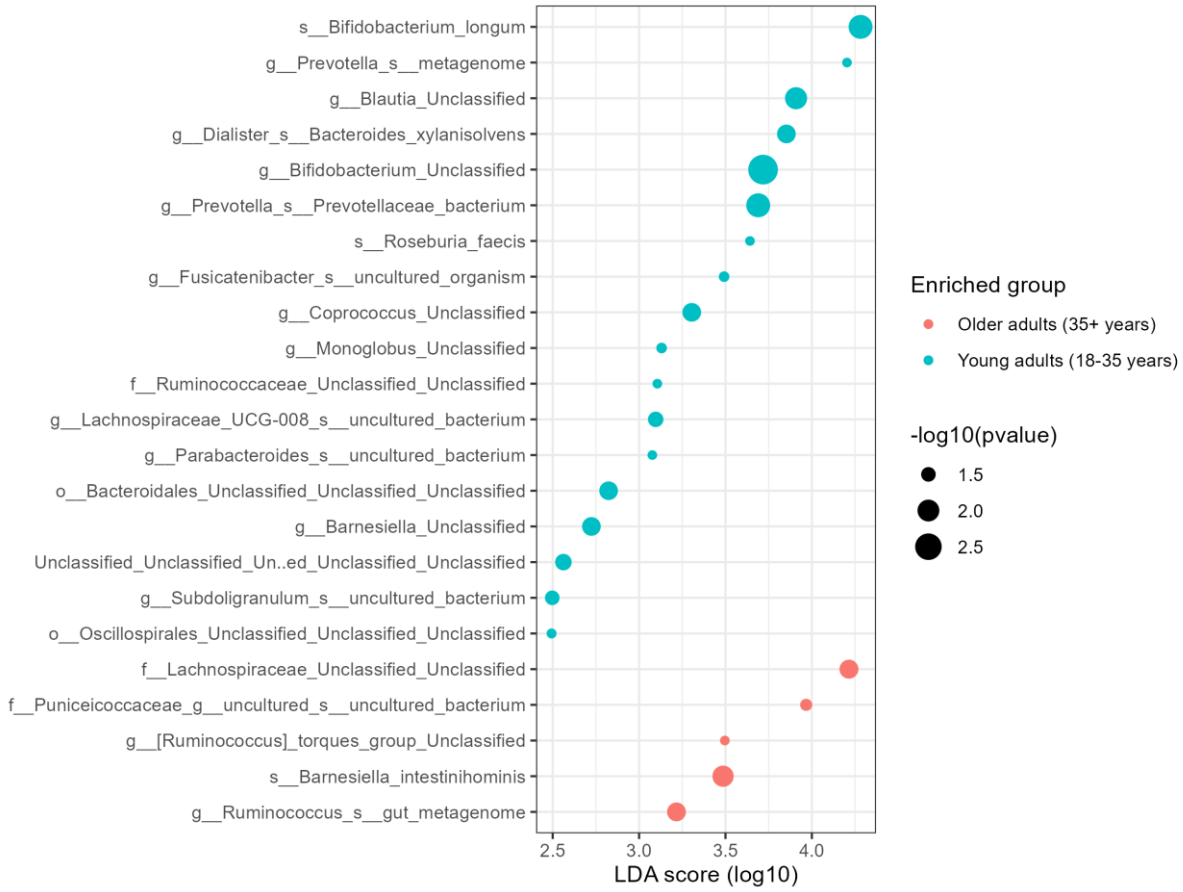


Figure 4.9 Differential abundance of microbial taxa in "age group" within the SAN population. This figure displays a comparison of microbial taxa abundance from 16S long-read sequencing data, between older adults (35+ years, red dots) and young adults (18-35 years, blue dots). The x-axis measures the Linear Discriminant Analysis (LDA) scores in log10, which quantifies the effect size of the abundance differences between groups. The y-axis lists various microbial taxa at the lowest possible phylogenetic level (i.e. species level and above). Dot sizes correlate with the negative logarithm of the p-value (-log10(p-value)), derived from the Kruskal-Wallis rank-sum test used to initially identify differentially abundant taxa across age categories, followed by pairwise Wilcoxon rank-sum tests where applicable. Taxa with larger dot sizes and higher LDA scores represent more statistically significant and biologically meaningful differences in abundance between the age groups.

In the gender analysis, *n* = 2 (*Bacteroides dorei* and *Phascolarctobacterium faecium*) and *n* = 6 (*Prevotella stercorea, Bacteroides cellulosilyticus, Bifidobacterium adolescentis, Massiliprevotella massiliensis, Dialister succinatiphilus, Veillonella atypica*) species were observed as significantly more abundant in females and males, respectively (Figure 4.10). Enterotype variation was observed as well with the female group presenting *Bacteroides*, and males presenting *Prevotella* profiles (Supplementary Data S4.8B).

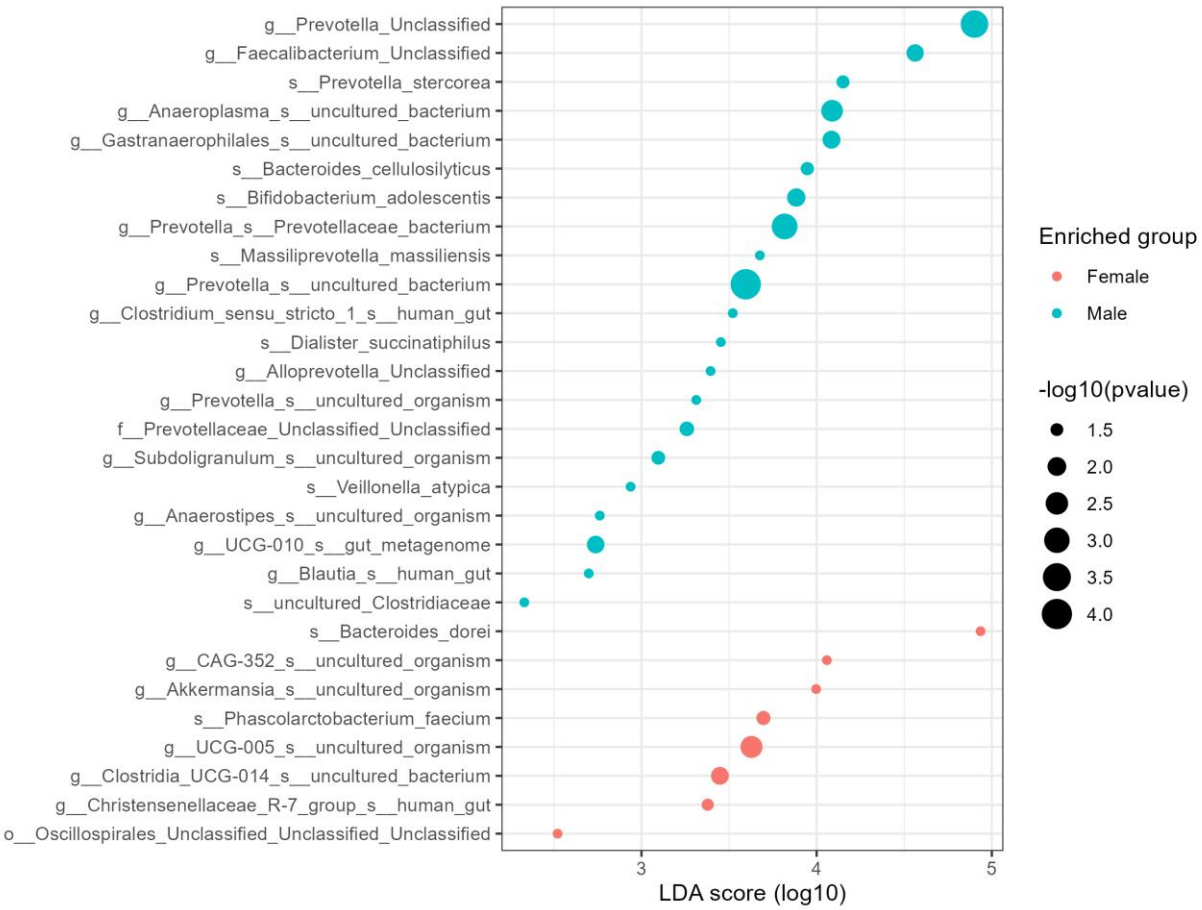


Figure 4.10 Differential abundance of microbial taxa in "gender" groups within the SAN population. This figure displays a comparison of microbial taxa abundance from 16S long-read sequencing data, between females (red dots) and males (blue dots). The x-axis measures the Linear Discriminant Analysis (LDA) scores in log10, which quantifies the effect size of the abundance differences between groups. The y-axis lists various microbial taxa at the lowest possible phylogenetic level (i.e. species level and above). Dot sizes correlate with the negative logarithm of the p-value (-log10(p-value)), derived from the Kruskal-Wallis rank-sum test used to initially identify differentially abundant taxa across gender categories, followed by pairwise Wilcoxon rank-sum tests where applicable. Taxa with larger dot sizes and higher LDA scores represent more statistically significant and biologically meaningful differences in abundance between the gender groups.

For the ancestry comparison, *n* = 12 differentially abundant species were observed across the groups of African (*Prevotella stercorea*, *Bacteroides xylanisolvens*) and Asian/Multiancestry/Other (*Bacteroides plebeius, Fusobacterium gonidiaformans, Bacteroides xylanisolvens, Bifidobacterium catenulatum, Collinsella intestinalis, Blautia glucerasea, Parasutterella excrementihominis, Adlercreutzia equolifaciens, Lactobacillus plantarum, Clostridium glycyrrhizinilyticum*) individuals. The *Prevotella* enterotype and *n* = 13 genera, and the *Bacteroides* enterotype with *n* = 15 genera were correlated with the African and Asian/Multiancestry/Other groups respectively (Supplementary Data S4.8C). Interestingly, only an uncultured species linked to the *Muribaculaceae* genus was identified in the European ancestry group (Figure 4.11).

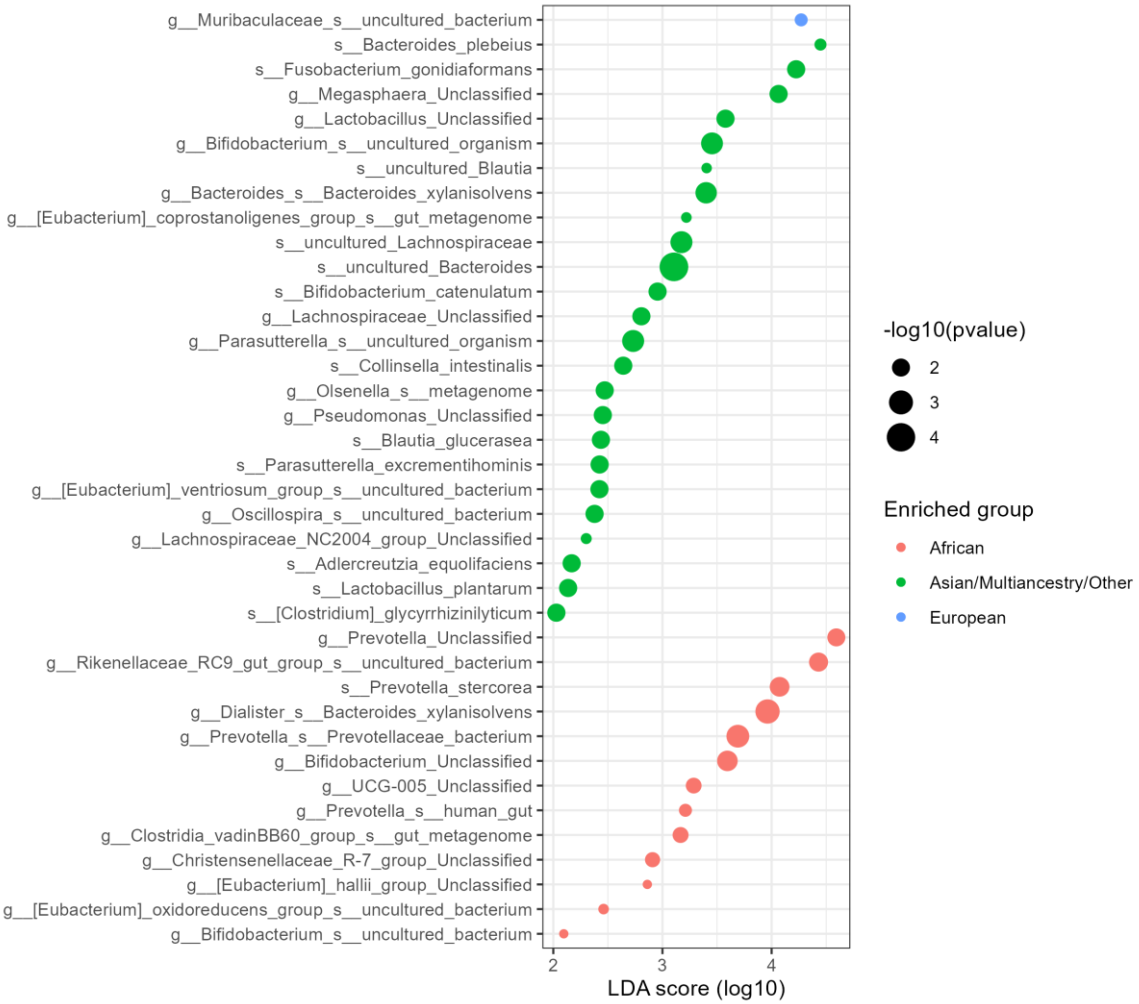


Figure 4.11 Differential abundance of microbial taxa in "ancestry" groups within the SAN population. This figure displays a comparison of microbial taxa abundance from 16S long-read sequencing data, between African (red dots), Asian/Multiancestry/Other (green dots) and European (blue dots) ancestries. The x-axis measures the Linear Discriminant Analysis (LDA) scores in log10, which quantifies the effect size of the abundance differences between groups. The y-axis lists various microbial taxa at the lowest possible phylogenetic level (i.e. species level and above). Dot sizes correlate with the negative logarithm of the p-value (-log10(p-value)), derived from the Kruskal-Wallis rank-sum test used to initially identify differentially abundant taxa across ancestry categories, followed by pairwise Wilcoxon rank-sum tests where applicable. Taxa with larger dot sizes and higher LDA scores represent more statistically significant and biologically meaningful differences in abundance between the ancestry groups.

In the BMI group analysis, combinations of enterotypes were observed across the categories, with *n* = 9 species identified as differentially abundant (Supplementary Data S4.8D). For the underweight group, a *Prevotella/Bacteroides* enterotype presence was noted across the *n* = 7 genera, with *n* = 6 ranked at the species level (*Eubacterium rectale, Bacteroides coprocola, Parasutterella excrementihominis, Faecalicoccus pleomorphus, Alistipes indistinctus and bacterium YE57*) (Figure 4.12).

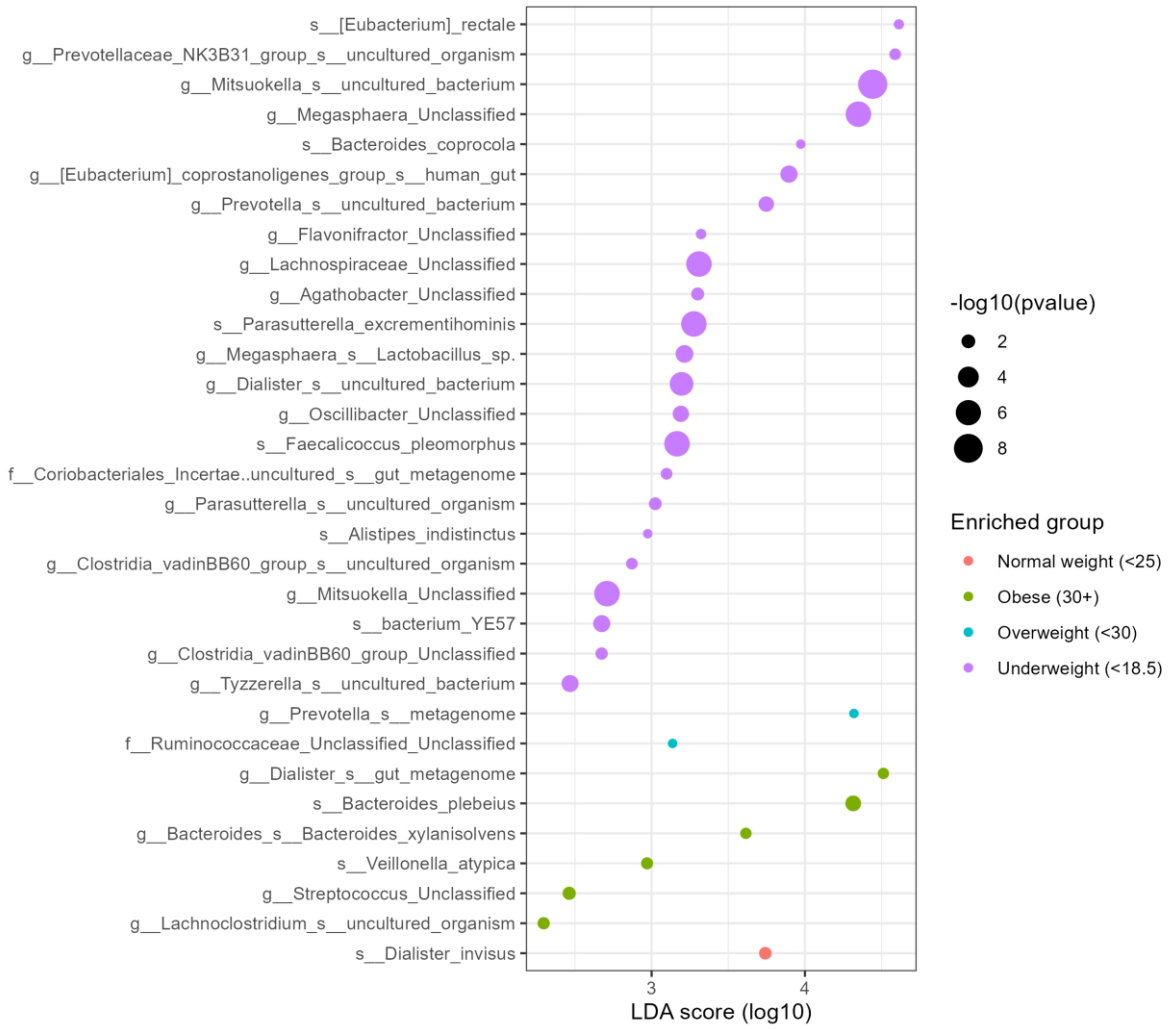


Figure 4.12 Differential abundance of microbial taxa in "BMI group" within the SAN population. This figure displays a comparison of microbial taxa abundance from 16S long-read sequencing data, between Normal weight (<25) (red dots), Obese (30+) (green dots), Overweight (<30) (blue dots) and Underweight (<18.5) (purple dots) individuals. The x-axis measures the Linear Discriminant Analysis (LDA) scores in log10, which quantifies the effect size of the abundance differences between groups. The y-axis lists various microbial taxa at the lowest possible phylogenetic level (i.e. species level and above). Dot sizes correlate with the negative logarithm of the p-value (-log10(p-value)), derived from the Kruskal-Wallis rank-sum test used to initially identify differentially abundant taxa across BMI categories, followed by pairwise Wilcoxon rank-sum tests where applicable. Taxa with larger dot sizes and higher LDA scores represent more statistically significant and biologically meaningful differences in abundance between the BMI groups.

The comprehensive LEfSe analysis of microbial differential abundance across different demographic variables reveals distinct patterns that correspond closely with the statistically significant results observed in the alpha and beta diversity analyses. Age-specific differences in microbial composition are evident and align with the alpha diversity results, which showed age-related variations in microbial richness and evenness. Gender differences further emphasize distinct microbial profiles, which correlate with the beta diversity results indicating significant microbial community differences between genders.

Although the diversity analyses were performed on a rarefied dataset to standardize sequencing depth and the LEfSe analysis utilized the whole cleaned dataset for a broader ecological perspective, the patterns derived offer valuable insights. LEfSe incorporates internal normalization via its Kruskal-Wallis test (accounting for uneven sampling depth) and subsequent LDA effect size estimation, which scales features to minimize depth-related bias (Segata et al., 2011). While residual risks exist (e.g., extreme depth disparities), the concordance of LEfSe results with rarefaction-based diversity metrics strengthened confidence in the biological relevance of identified biomarkers.

The observed correlations in these analyses highlight a complex interplay between demographic factors and microbial dynamics, underscoring their potential implications for human health and the profiling of the urban South African gut microbiome. Nevertheless, given the univariate nature of these analyses, it is crucial to consider that other interacting response variables within these demographic groups might be influencing the observed differences in microbial abundance. The following section portrays the proceeding analyses performed to further disentangle these interactions and to determine if the associations are directly attributable to demographic characteristics or are mediated by other factors within the cohort.

ii. Differential abundance across response variables

In this section, subsequent analysis was performed focusing on the *n* = 42 SES, diet, lifestyle and health variables, where *n* = 293 taxa emerged as differentially and statistically abundant. To further refine the differentially abundant taxa, the resultant species were subset to exclude those present in both the demographic and response variables to pinpoint the microbial associations specific to these variables. These *n* = 226 species were then de-duplicated to account for species unique to each variable (*n* = 43), inferring microbial biomarker identification using a univariate analytical approach (Table 4.2).

SES

The SES analysis, conducted across income and transport modes, highlighted a shared enterotype presence of Bacteroides/Prevotella and Prevotella/Ruminococcus for private and public transport use respectively (Supplementary Data S4.8E). Although *n* = 44 species were identified across all SES subgroups, only *n* = 3 were uniquely present with similar p-values ($p = 7.7 \times 10^{-3}$) in relation to the non-use of private (*Blautia caecimuris*) and use of public transport (*Clostridium scindens, Phascolarctobacterium succinatutens*). Public transport was also shown to have significant differences in alpha Shannon diversity in prior analyses, where its use was correlated with increased abundance and evenness postulated to be linked to the biomarkers shown in Figure 4.13 (Table 4.2).

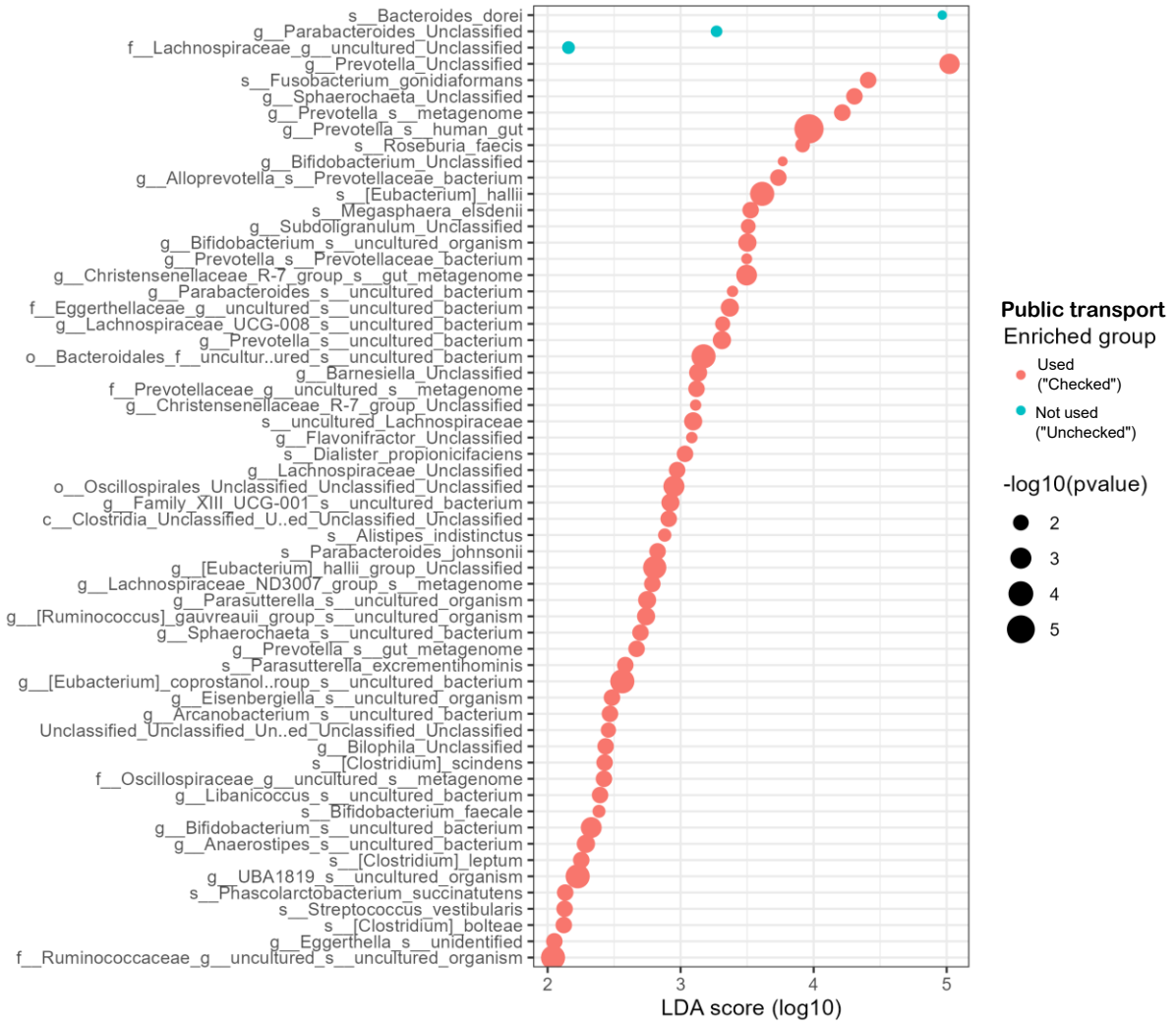


Figure 4.13 Differential abundance of microbial taxa for "public transport" use within the SAN population. This figure displays a comparison of microbial taxa abundance from 16S long-read sequencing data, between individuals who use public transport ("Used/Checked", red dots) and who do not ("Not used/Unchecked", blue dots). The x-axis measures the Linear Discriminant Analysis (LDA) scores in log10, which quantifies the effect size of the abundance differences between groups. The y-axis lists various microbial taxa at the lowest possible phylogenetic level (i.e. species level and above). Dot sizes correlate with the negative logarithm of the p-value (-log10(p-value)), derived from the Kruskal-Wallis rank-sum test used to initially identify differentially abundant taxa across public transport categories, followed by pairwise Wilcoxon rank-sum tests where applicable. Taxa with larger dot sizes and higher LDA scores represent more statistically significant and biologically meaningful differences in abundance between the groups.

Diet

The dietary analysis with LEfSe revealed the Bacteroides enterotype to be present in individuals that did not follow a specialised diet (Supplementary Data S4.8F). A combination of *n* = 41 species emerged across the seven diet variables. The "Classify diet" variable had the most unique species (*n* = 3) with them all related to following a diet other than an omnivorous one: this included vegetarian and vegan diets. In addition, *n* = 2 species correlated with a "high protein" diet, with *Absiella argi* showing a pronounced LDA effect size > 3 and significant p-value ($p = 6.6 \times 10^{-4}$).

Table 4.2 Differentially abundant LEfSe taxa biomarkers uniquely identified in a response variable

Variable	Enriched_group	Species	LDA[1]	p-value
Transport_private	Unchecked[2]	*Blautia_caecimuris*	2.9	7.7×10^{-3}
Transport_public	Checked[3]	*[Clostridium]_scindens*	2.4	7.7×10^{-3}
Transport_public	Checked	*Phascolarctobacterium_succinatutens*	2.1	7.7×10^{-3}
classifydiet	Other	*Lactobacillus_sakei*	3.7	2.8×10^{-4}
classifydiet	Other	*Sutterellaceae_bacterium*	3.4	2.8×10^{-4}
classifydiet	Other	*Lactobacillus_plantarum*	2.3	2.8×10^{-4}
exclude_refined_sugars	Checked	*Bacteroides_plebeius*	4.2	4.7×10^{-2}
gluten_free	Checked	*Bifidobacterium_bifidum*	3.7	1.4×10^{-2}
high_protein	Checked	*[Eubacterium]_rectale*	3.8	4.2×10^{-2}
high_protein	Checked	*Absiella_argi*	3.1	6.6×10^{-4}
infant_food	Not sure	*Parabacteroides_goldsteinii*	2.4	3.3×10^{-3}
pets	No	*Alistipes_inops*	3.9	3.5×10^{-2}
pets	No	*Bacteroides_fragilis*	3.3	3.9×10^{-3}
pets	Yes	*Parabacteroides_merdae*	3.6	3.7×10^{-2}
cook_electricity	Unchecked	*[Bacteroides]_pectinophilus*	3.3	4.4×10^{-2}
cook_electricity	Unchecked	*Collinsella_sp.*	2.6	4.4×10^{-2}
cook_electricity	Unchecked	*Lactobacillus_fermentum*	2.4	4.4×10^{-2}
heat_gas	Checked	*Desulfovibrio_piger*	2.5	3.2×10^{-2}
heat_gas	Checked	*[Clostridium]_aldenense*	2.1	3.2×10^{-2}
multivit	No	*Prevotella_stercorea*	4.1	4.2×10^{-2}
herbalsupplements	Yes	*Bacteroides_eggerthii*	4.1	4.4×10^{-2}
bitefingernails	Yes	*Odoribacter_splanchnicus*	3.0	3.7×10^{-2}
hoursofsleep	7+ hours	*Bacteroides_uniformis*	4.1	1.1×10^{-2}
fabricsoftener	Yes	*Coprococcus_eutactus*	3.3	3.7×10^{-2}
white_wine	Checked	*Blautia_stercoris*	3.0	3.2×10^{-2}
white_wine	Checked	*Turicibacter_sanguinis*	2.8	3.2×10^{-2}
red_wine	Unchecked	*Streptococcus_salivarius*	3.6	3.1×10^{-2}
alcohol_type_none	Checked	*Bacteroides_faecis*	3.2	1.7×10^{-2}
alcoholcount	More than 4	*[Clostridium]_glycyrrhizinilyticum*	2.2	1.6×10^{-3}
alcoholcount	More than 4	*Anaerostipes_rhamnosivorans*	2.1	1.6×10^{-3}
exercisetype	Other (please specify)	*Sutterella_wadsworthensis*	3.4	4.6×10^{-2}
exercisetype	Sport (e.g. squash, soccer)	*Victivallis_vadensis*	2.8	3.2×10^{-2}
exerciselocation	Indoors	*Bacteroides_massiliensis*	4.0	2.4×10^{-2}
exerciselocation	Not applicable	*Parabacteroides_distasonis*	4.3	4.7×10^{-2}
antibiotics	Six months or less	*Bacteroides_cellulosilyticus*	3.9	2.7×10^{-2}
antibiotics	Six months or less	*Bacteroides_clarus*	2.7	3.4×10^{-2}
antibiotics	Six months or less	*Erysipelatoclostridium_ramosum*	2.5	3.4×10^{-2}
antibiotics	Year	*Bacteroides_cutis*	3.9	4.8×10^{-2}
antibiotics	Year	*Adlercreutzia_equolifaciens*	2.0	4.8×10^{-2}
lactoseintol	Yes	*Collinsella_intestinalis*	2.7	3.3×10^{-2}
lactoseintol	Yes	*Blautia_obeum*	2.5	3.3×10^{-2}
glutenintol	Intolerance	*Bacteroides_stercoris*	3.7	2.1×10^{-3}
glutenintol	Intolerance	*Dorea_longicatena*	2.6	6.5×10^{-3}

Variable	Enriched_group	Species	LDA[1]	p-value
bowelqualitychoice	I tend to be constipated (have difficulty passing stool) - Type 1 and 2	*Alistipes_putredinis*	4.2	2.8×10^{-2}
allergiesfood_none	Unchecked	*Bacteroides_coprocola*	3.9	3.6×10^{-2}
allergiesfood_none	Unchecked	*Bifidobacterium_catenulatum*	3.0	1.4×10^{-2}
allergiesnonfood_petdander	Checked	*Phascolarctobacterium_faecium*	4.1	3.0×10^{-2}

[1]LDA: Linear Discriminant Analysis effect size value; [2]Unchecked = unselected response; [3]Checked = selected response.

Lifestyle

The LEfSe analysis for lifestyle responses was performed across *n* = 21 related variables (Supplementary Data S4.8G). All three enterotypes were observed across the variables: where *Bacteroides* presence was linked to herbal supplement use and hours of sleep, as well as in conjunction with *Prevotella* for cooking with electricity, cosmetics use and consumption of beers and/or ciders; *Prevotella* was solely enriched with pets and combined with *Ruminococcus* for alcohol count; and *Ruminococcus* was solely enriched in relation to probiotic and antibiotic use.

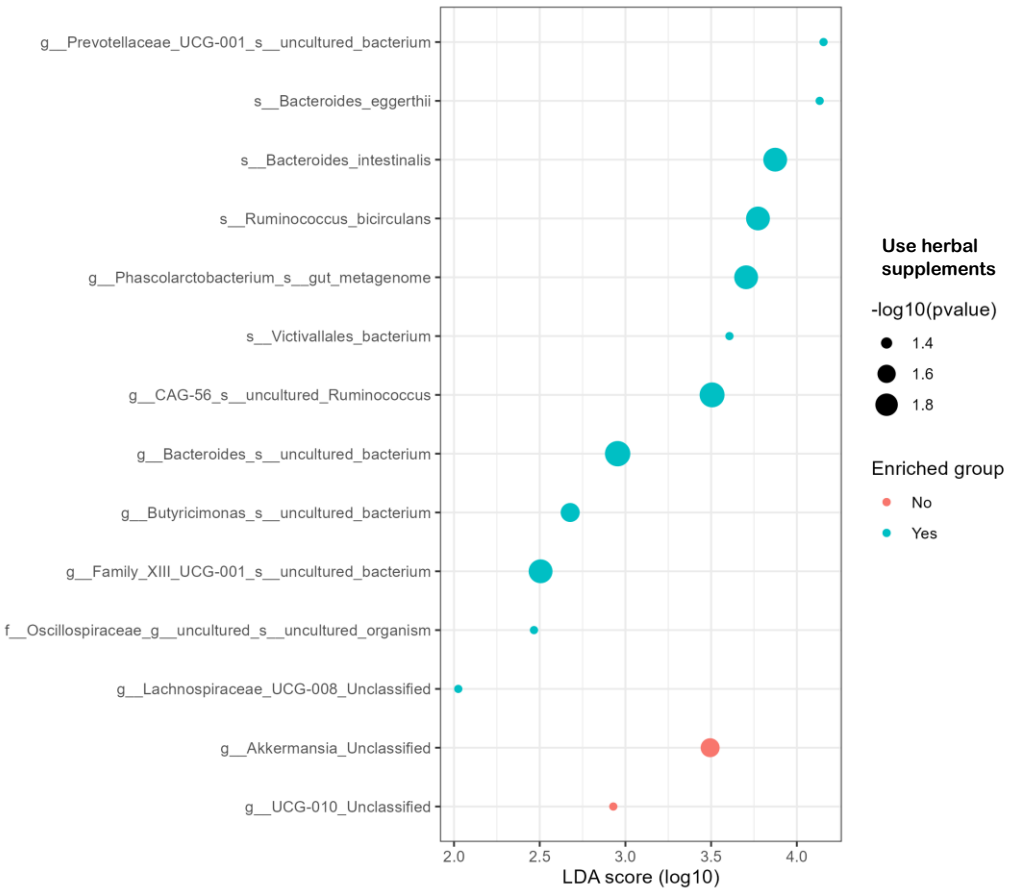


Figure 4.14 Differential abundance of microbial taxa for "herbal supplements" use within the SAN population. This figure displays a comparison of microbial taxa abundance from 16S long-read sequencing data, between individuals who use herbal supplements (Yes, blue dots) and who do not (No, red dots). The x-axis measures the Linear Discriminant Analysis (LDA) scores in log10, which quantifies the effect size of the abundance differences between groups. The y-axis lists various microbial taxa at the lowest possible phylogenetic level (i.e. species level and above). Dot sizes correlate with the negative logarithm of the p-value (-log10(p-value)), derived from the Kruskal-Wallis rank-sum test used to initially identify differentially abundant taxa across herbal supplement use responses, followed by pairwise Wilcoxon rank-sum tests where applicable. Taxa with larger dot sizes and higher LDA scores represent more statistically significant and biologically meaningful differences in abundance between the groups.

The majority of unique species (*n* = 5) were also linked to antibiotic use (Table 4.2), ranging from its use in the past year to as recent as six months from the time of the sample collection. This biomarker discovery was followed by *n* = 3 unique species observed in relation to pets and cooking with electricity; *n* = 2 linked to *gas heating*, consuming *white wine*, *alcohol count*, *exercise type* and *exercise location*; and lastly *n* = 1 count of unique species with significant differential abundance observed over seven other lifestyle related variables (Table 4.2).

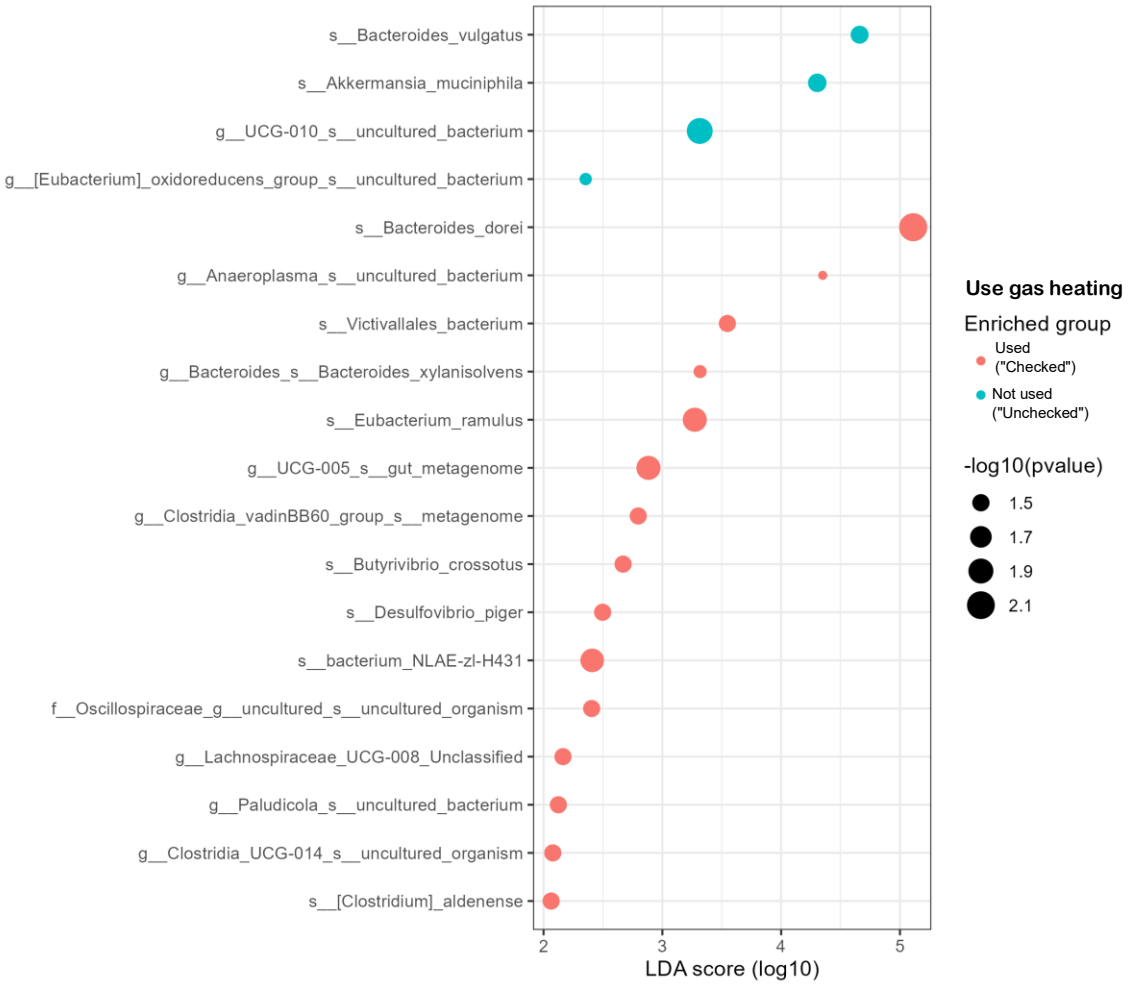


Figure 4.15 Differential abundance of microbial taxa for "gas heating" use within the SAN population. This figure displays a comparison of microbial taxa abundance from 16S long-read sequencing data, between individuals who use gas heating ("Used/Checked", red dots) and who do not ("Not used/Unchecked", blue dots). The x-axis measures the Linear Discriminant Analysis (LDA) scores in log10, which quantifies the effect size of the abundance differences between groups. The y-axis lists various microbial taxa at the lowest possible phylogenetic level (i.e. species level and above). Dot sizes correlate with the negative logarithm of the p-value (-log10(p-value)), derived from the Kruskal-Wallis rank-sum test used to initially identify differentially abundant taxa across gas heating categories, followed by pairwise Wilcoxon rank-sum tests where applicable. Taxa with larger dot sizes and higher LDA scores represent more statistically significant and biologically meaningful differences in abundance between the groups.

One of these variables was herbal supplement use, which was indicated to have a statistical difference in microbial composition measured with beta diversity, postulated to be linked to the observed species (Figure 4.14). The differentially abundant species under gas heating (Figure 4.15) and multivitamin use (Figure 4.16) also revealed a significant association with alpha diversity in the prior diversity analyses (Figure 4.5; 4.6), where those who did not use either showed an increased abundance and evenness of microbial diversity.

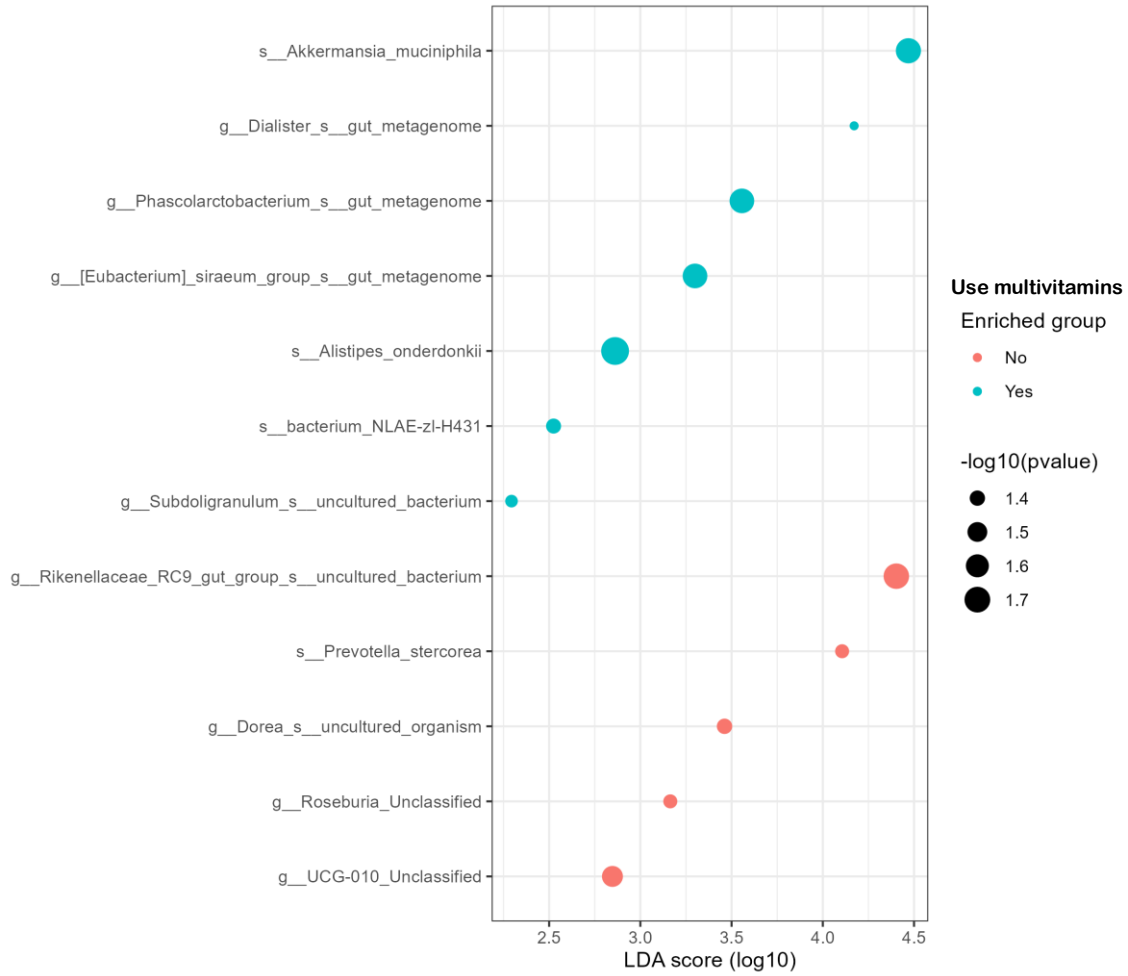


Figure 4.16 Differential abundance of microbial taxa for "multivitamins" use within the SAN population. This figure displays a comparison of microbial taxa abundance from 16S long-read sequencing data, between individuals who use multivitamins (Yes, blue dots) and who do not (No, red dots). The x-axis measures the Linear Discriminant Analysis (LDA) scores in log10, which quantifies the effect size of the abundance differences between groups. The y-axis lists various microbial taxa at the lowest possible phylogenetic level (i.e. species level and above). Dot sizes correlate with the negative logarithm of the p-value (-log10(p-value)), derived from the Kruskal-Wallis rank-sum test used to initially identify differentially abundant taxa across multivitamin use responses, followed by pairwise Wilcoxon rank-sum tests where applicable. Taxa with larger dot sizes and higher LDA scores represent more statistically significant and biologically meaningful differences in abundance between the groups.

Health

The LEfSe analysis was also performed on *n* = 12 health related variables, with the *Ruminococcus* enterotype uniquely being present at genus level when focused on mental health. When delving into the LDA effect size and significance of species identified uniquely across these health variables, *n* = 2 unique biomarkers each emerged in relation to lactose and gluten intolerance. The species *Bacteroides stercoris* had the greatest effect size (LDA = 3.74) and significant p-value (P = 2.1 x 10^{-3}) linked to gluten intolerance, which infers its association with this aetiology (Table 4.2). In addition, having an "unchecked" response in relation to "allergies food none" meant *n* = 2 species (*Bacteroides coprocola and Bifidobacterium catenulatum*) were highlighted to be associated with having some kind of food allergy (Figure 4.17). This variable was also found to have a significant difference in microbiome beta diversity (Figure 4.8), inferring that these species biomarkers may be driving this difference between groups with and without food allergies.

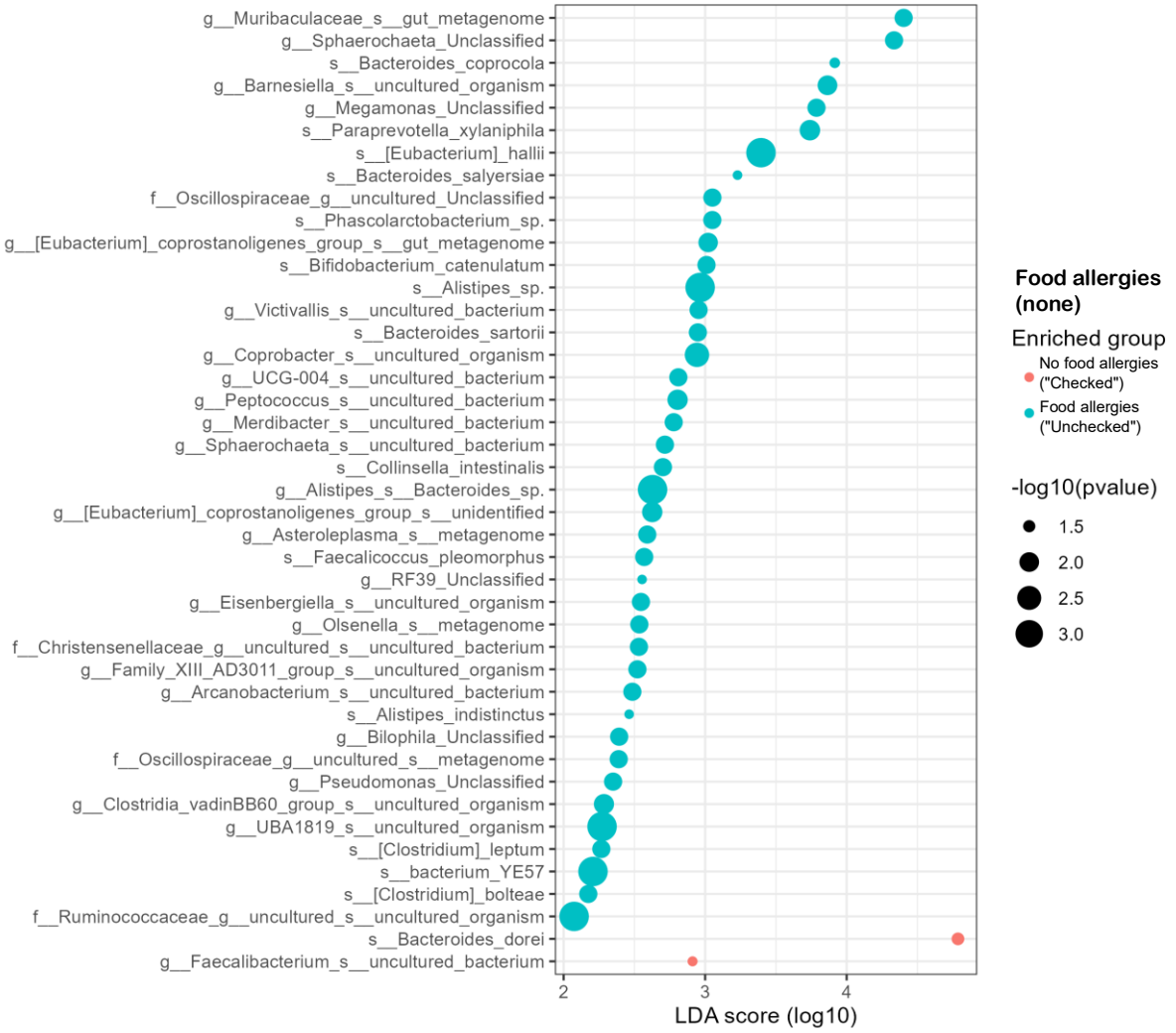


Figure 4.17 Differential abundance of microbial taxa for "food allergies (none)" within the SAN population. This figure displays a comparison of microbial taxa abundance from 16S long-read sequencing data, between individuals who have no food allergies ("Checked", red dots) and those who do have food allergies ("Unchecked", blue dots). The x-axis measures the Linear Discriminant Analysis (LDA) scores in log10, which quantifies the effect size of the abundance differences between groups. The y-axis lists various microbial taxa at the lowest possible phylogenetic level (i.e. species level and above). Dot sizes correlate with the negative logarithm of the p-value (-log10(p-value)), derived from the Kruskal-Wallis rank-sum test used to initially identify differentially abundant taxa across food allergy_none responses, followed by pairwise Wilcoxon rank-sum tests where applicable. Taxa with larger dot sizes and higher LDA scores represent more statistically significant and biologically meaningful differences in abundance between the groups.

The analysis provided a detailed view of how demographic and survey response variables influence the gut microbiome. The identification of specific taxa as biomarkers helps in understanding the biological implications of their abundance across the demographic groups of this cohort, as well as their unique diet, lifestyle, SES and health profiles. These findings strengthen the evidence of prior associations (Chapter 4.3.1) that were observed to be significantly different across the cohort in terms of microbial diversity, which highlights the variation that can be derived in an urban setting. This paves the way for downstream functional analyses to explore whether the metabolic pathways or genes associated with these taxa are influenced by, or contribute to, the observed differences.

4.3.3. Predicted functional profiling analyses with PICRUSt and *ggpicrust2*

The 16S long-read data functional potential was investigated by predictive metagenomic analysis using the PICRUSt workflow, with downstream statistical analysis using *ggpicrust2*. This approach allowed for the estimation of the functional contributions of microbial communities via the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, which are integral to understanding microbial roles in the ecosystem. A total of n = 349 KEGG pathways, with descriptive pathways annotated using the MetaCyc database, were identified across the samples (Supplementary Data S4.9A). "Pathway classes" serve as broad categories that group functionally related biological processes, such as "metabolism" or "cellular processes." Within these classes, individual "pathways" provide detailed molecular maps of specific biological processes. This hierarchical structure allowed navigation from general biological systems down to specific molecular interactions and mechanisms (BV-BRC, 2023). The data for this analysis were derived from enzyme-based predictions, where each pathway's abundance in the dataset was quantified according to the predicted counts of Enzyme Commission (EC) numbers, providing a detailed view of the diverse active metabolic functions present within the microbial communities.

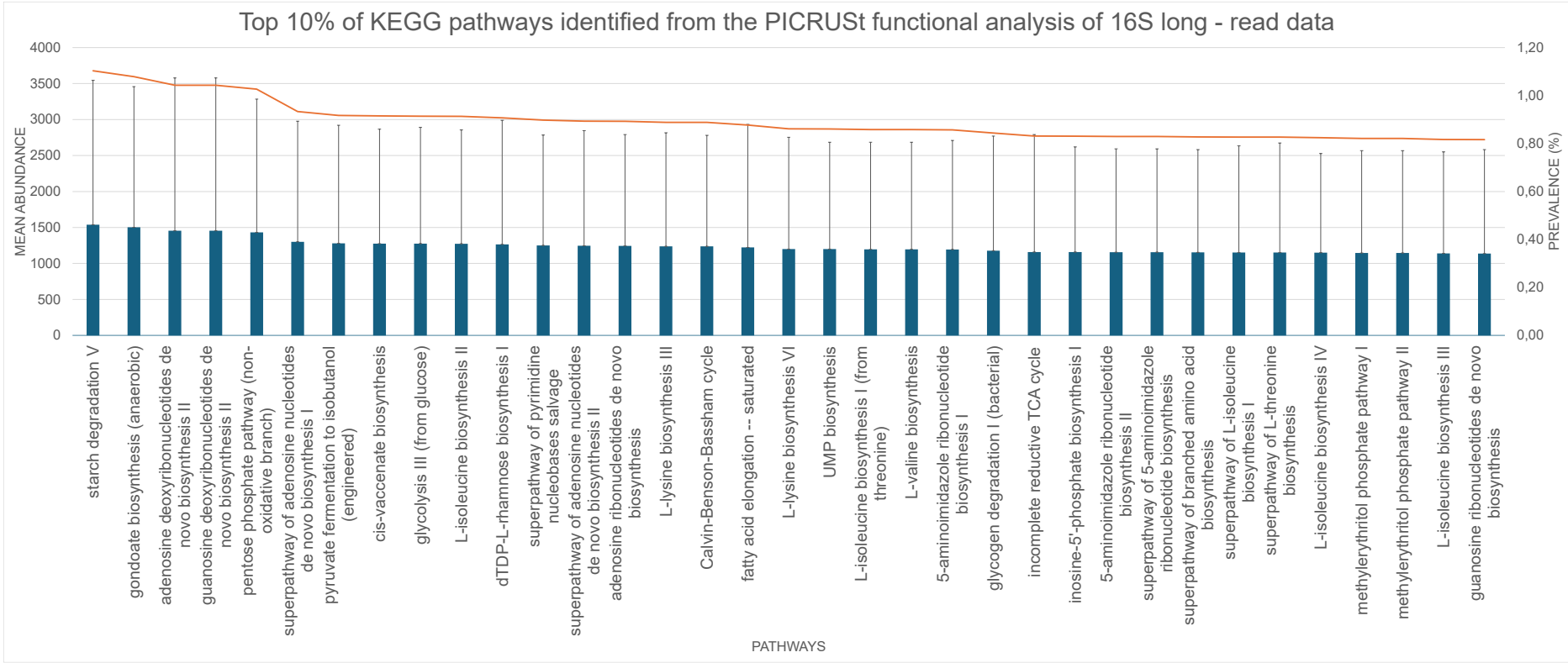


Figure 4.18 Distribution of the top 10% of KEGG (MetaCyc) pathways as identified from PICRUSt functional analysis of 16S long-read data. The graph illustrates the mean abundance (blue bars, left y-axis), and the percent prevalence (orange line, right y-axis) of each pathway, based on coverages per sample of predicted enzyme commission (EC) number abundances, highlighting *n* = 35 predominant metabolic functional KEGG (MetaCyc database annotated) pathways (x-axis) in the SAN population. The error bars represent the standard deviation (stdev.s) of the total abundance of each pathway across the cohort.

The top 10% most prevalent KEGG pathways identified and annotated using the MetaCyc database (n = 35) is illustrated in Figure 4.18. The figure employs a dual-axis format. representing both the mean abundance of each pathway across all samples, and the percentage of the total functional contribution attributed to each pathway. The combined prevalence of these pathways accounted for ~30% of the metabolic functions, with up to ~1.1% of this prevalence being attributable to “starch degradation V”. However, the standard deviations were substantial compared to the means across all identified pathways, indicating significant variation in the functional potential across the cohort.

In addition to starch degradation V, the other top 1% of pathways identified were: “gondoate biosynthesis (anaerobic)”, “adenosine deoxyribonucleotides de novo biosynthesis II”, “guanosine deoxyribonucleotides de novo biosynthesis II”, and “pentose phosphate pathway (non-oxidative branch) I”. Majority of the remaining pathways were also involved in ribonucleotide and amino acid biosynthesis (Figure 4.18). These pathways are integral to various biological processes including lipid metabolism, nucleotide synthesis, and carbohydrate processing, which are key pathways in energy production and cellular growth.

Brief analyses across demographic and response variables identified several intriguing pathway associations within the SAN population, although these were modest and require cautious interpretation due to the predictive nature of the data. Notably, demographic stratifications by age, gender, ancestry, and BMI indicated subtle but significant variations in microbial functional profiles. Similarly, lifestyle and health-related variables such as cosmetics use, dietary choices, and reported health conditions emerged as potential factors linked to differences in microbial functional potential, particularly pathways associated with xenobiotic metabolism, carbohydrate processing, and lipid metabolism.

Given the predictive and exploratory nature of these findings, detailed results, including the comparative analyses with reference cohorts (American athletes and Chinese diarrheal patients) and comprehensive statistical evaluations, are provided in Appendix One. This supplementary information includes extensive visualizations and statistical summaries to facilitate deeper interrogation and validation of the observed relationships.

4.4. Investigation of community stratification in the SAN population

To explore the community stratification in the SAN population, a structured analytical approach focused on taxonomy and metadata profiles was undertaken. Firstly, the defining of distinct community clusters based on compositional similarities. Secondly, a detailed characterization of cluster-specific community structures through taxonomic analyses. Finally, investigations of associations between cluster membership and participant characteristics to understand potential lifestyle and health-related factors underlying this stratification.

4.4.1. Identification and validation of community clusters

Analysis of the SAN population microbiome composition revealed distinct community structures when examining sample distributions in ordination space. To objectively classify these apparent patterns, k-means clustering was applied to the PCoA coordinates resulting in two distinct clusters (Figure 4.19). Cluster 1 contained *n* = 106 samples and was primarily distributed in the upper half of PC2, while Cluster 2 contained *n* = 136 samples and occupied the lower portion of the ordination space (Supplementary Data S4.19). The samples in Cluster 1 showed higher median relative abundance values (median = 2.5 x 10^{-3}) compared to Cluster 2 (median = 1.9 x 10^{-4}). Both clusters maintained similar taxonomic richness with *n* = 14 phyla identified in each cluster after taxonomic grouping, derived from an initial set of 3,485 taxa.

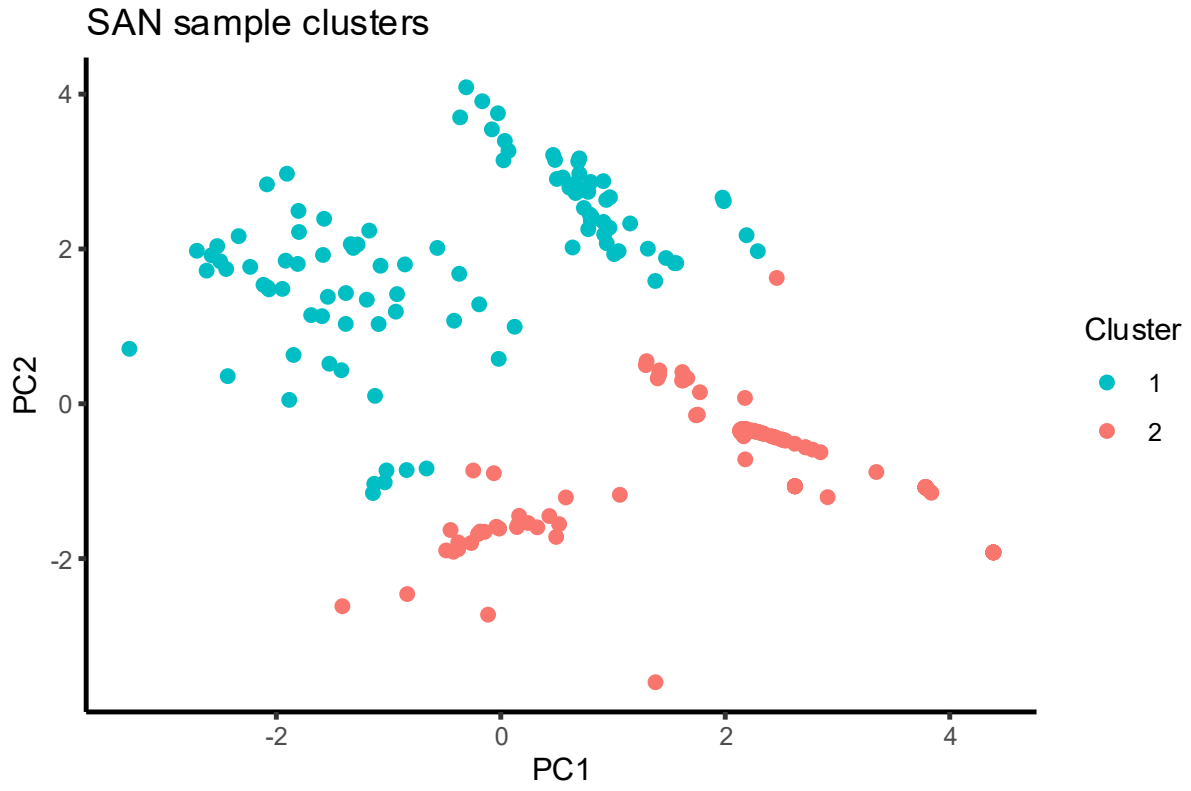


Figure 4.19 Principal Coordinates Analysis (PCoA) plot showing the distribution of SAN population samples clustered using k-means clustering (k = 2). The plot represents the first two principal coordinates (PC1 and PC2) of the Bray-Curtis dissimilarity matrix calculated from phylum-level taxonomic profiles. Each point represents a single sample, with colours indicating cluster membership (Cluster 1 in blue, *n* = 106; Cluster 2 in pink, *n* = 136).

4.4.2. Characterization of cluster-specific community structures

Further analysis of the microbial communities between the two clusters revealed significant differences in phylum-level composition (Figure 4.20). Six out of fourteen phyla showed statistically significant differences between clusters (FDR < 5.0 x 10^{-2}). The most pronounced difference was observed in Actinobacteriota (adjusted p-value = 3.1 x 10^{-41}) which showed markedly higher abundance in Cluster 1 (mean = 13.4%) compared to Cluster 2 (mean = 2.0 x 10^{-2}%). Bacteroidota, the most abundant phylum across all samples, showed significantly higher relative abundance in Cluster 2 (59.3%) compared to Cluster 1 (47.6%; adjusted p-value = 9.0 x 10^{-3}). Additional significant differences were observed in Proteobacteria, Verrucomicrobiota and Desulfobacterota, as well as in the proportion of unclassified phyla (Supplementary Data S4.19C).

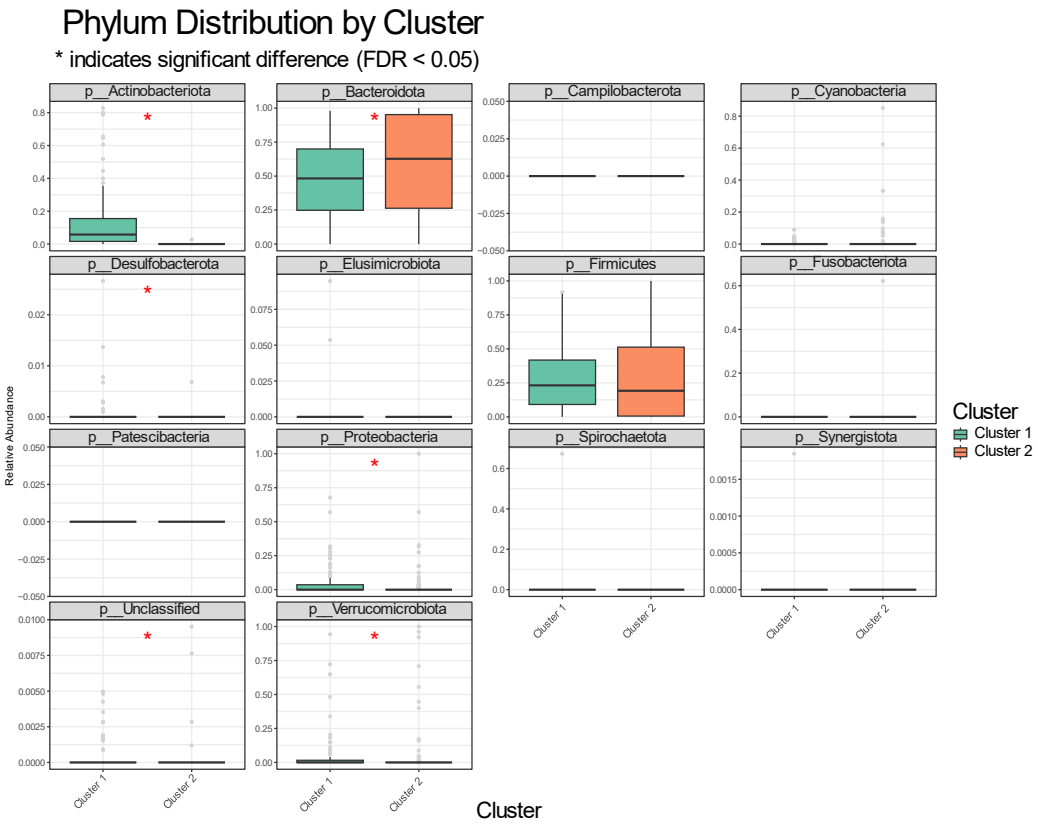


Figure 4.20 Distribution of relative abundances for each bacterial phylum across the two identified clusters in the SAN population. Box plots show the median (horizontal line), interquartile range (box), and range (whiskers) of relative abundances, with outliers represented as grey points. Clusters are colour-coded (Cluster 1 in blue, Cluster 2 in pink). Asterisks (*) indicate phyla showing significant differences between clusters (FDR < 5.0 x 10^{-2}) based on Wilcoxon rank-sum tests with Benjamini-Hochberg correction for multiple comparisons. y-axes are scaled independently for each phylum to optimize visualization of abundance patterns.

4.4.3. Association of clusters with participant characteristics

The analysis of participant characteristics revealed distinct associations between lifestyle factors and microbial community structures in the SAN population (Supplementary Data S4.19D). Fisher's exact tests identified eight metadata variables significantly associated with cluster membership (p-value < 5.0 x 10^{-2}). Notably, *COVID-19 diagnosis* history showed the strongest association (p-value = 4.0 x 10^{-3}) with a higher proportion of previous infections reported in Cluster 2. Participants in Cluster 1 reported higher frequencies of *cosmetics* use (p-value = 2.2 x 10^{-2}) and vivid/frightening *dreams* (p-value = 2.3 x 10^{-2}), while Cluster 2 showed higher proportions of CSIR *post code work* locations (p-value = 2.3 x 10^{-2}) and *dairy exclusion* from their diet (p-value = 2.8 x 10^{-2}). Additional significant associations included *red wine consumption, pool/hot tub use* and *exercise location* preferences (Figure 4.21).

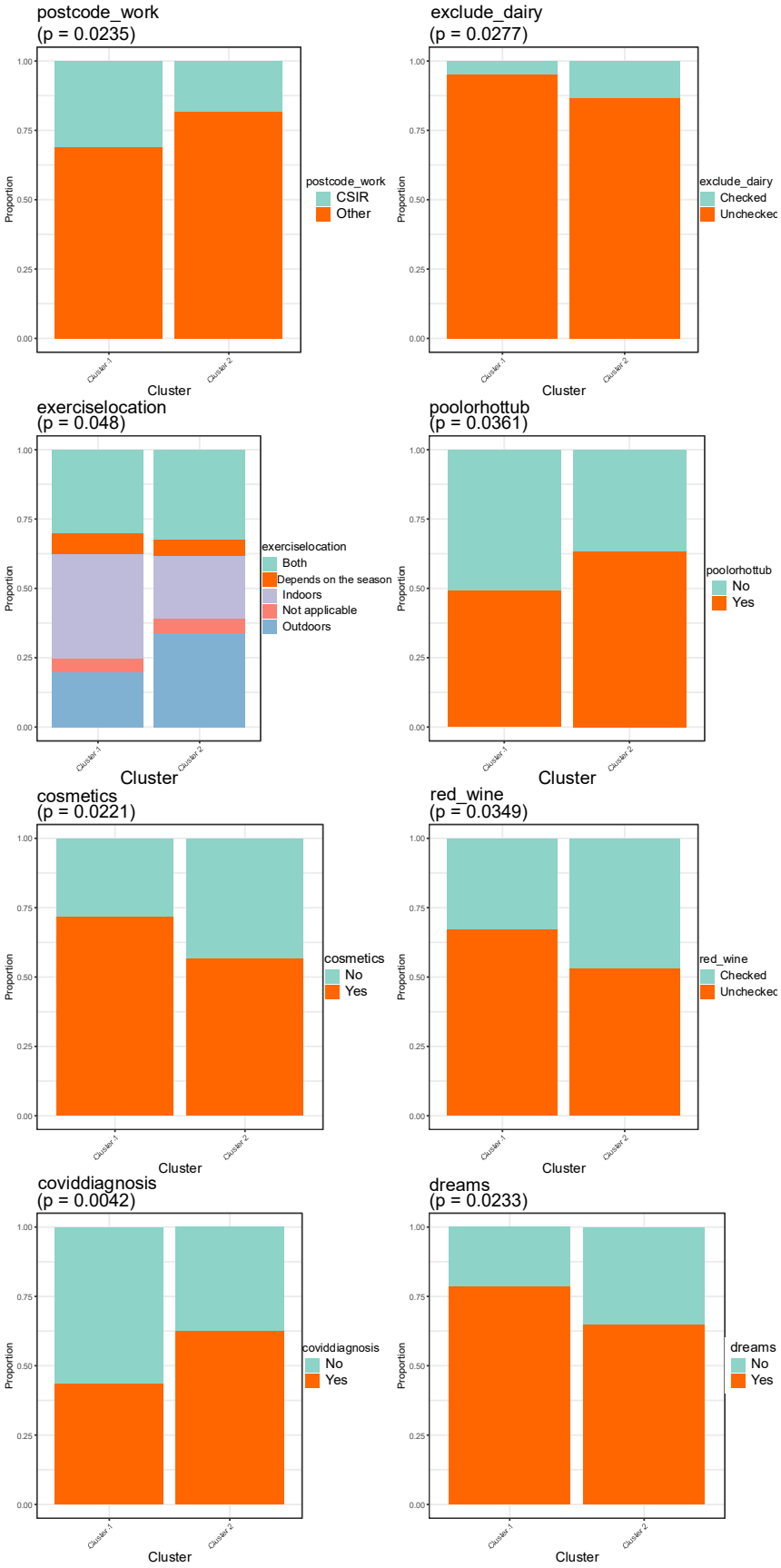

postcode_work
(p = 0.0235)
Proportion
Cluster
postcode_work
CSIR
Other
exclude_dairy
(p = 0.0277)
exclude_dairy
Checked
Unchecked
exerciselocation
(p = 0.048)
exerciselocation
Both
Depends on the season
Indoors
Not applicable
Outdoors
poolorhottub
(p = 0.0361)
poolorhottub
No
Yes
cosmetics
(p = 0.0221)
cosmetics
No
Yes
red_wine
(p = 0.0349)
red_wine
Checked
Unchecked
coviddiagnosis
(p = 0.0042)
coviddiagnosis
No
Yes
dreams
(p = 0.0233)
dreams
No
Yes

Figure 4.21 Distribution of metadata variables significantly associated with microbial community clusters ($p < 5.0 \times 10^{-2}$). Each panel represents the proportion of participant responses (y-axis) between Cluster 1 (*n* = 106) and Cluster 2 (*n* = 136) for different categorical variables. Legend colours indicate response categories. Statistical significance was determined using Fisher's exact tests.

To determine which bacterial phyla were driving these associations, Kruskal-Wallis tests were performed for each significant metadata variable, followed by Benjamini-Hochberg correction for multiple comparisons (Figure 4.22, Supplementary Data S4.19D). This analysis revealed two robust associations: Actinobacteriota showed significantly higher abundance in participants without prior COVID-19 diagnosis (adjusted p-value = 1.2×10^{-2}) and Cyanobacteria demonstrated elevated levels in participants who reported red wine consumption (adjusted p-value = 4.6×10^{-2}).

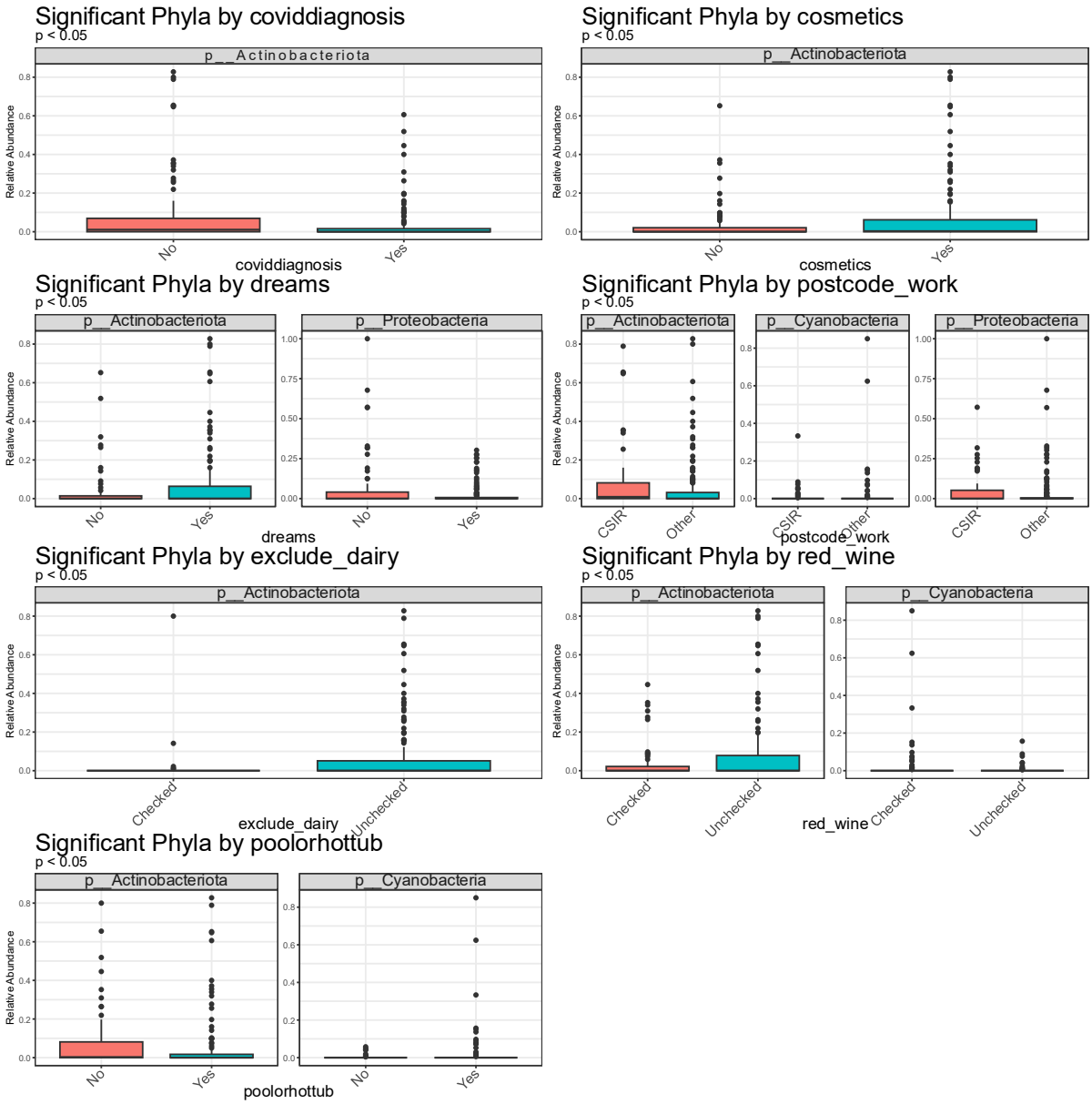


Figure 4.22 Bacterial phyla showing significant associations with lifestyle variables in the SAN population clusters. Relative abundance distributions (y-axis) of bacterial phyla are shown across participant responses (x-axis) for differentially significant lifestyle-related variables across the SAN clusters (Cluster 1, *n* = 106; Cluster 2, *n* = 136). Each point represents an individual participant sample, with boxes indicating median and interquartile range (IQR), and whiskers extending to 1.5 x IQR. Statistical testing was performed using Kruskal-Wallis tests, with Benjamini-Hochberg correction for multiple comparisons. Displayed associations had Kruskal-Wallis $p < 5.0 \times 10^{-2}$, with Actinobacteriota – COVID-19 history and Cyanobacteria – red wine consumption additionally meeting FDR $< 5.0 \times 10^{-2}$ threshold.

Beyond the FDR-significant findings, several other significant phylum-level associations were observed. Actinobacteriota emerged as a key indicator across multiple lifestyle factors. Participants who reported using *cosmetics* showed higher Actinobacteriota abundance compared to non-users (p-value = 7.0×10^{-3}), while those experiencing vivid/frightening *dreams* exhibited elevated levels (p-value = 1.5×10^{-2}). The analysis also revealed that *pool/hot tub* users had lower Actinobacteriota abundance and higher Cyanobacteria levels, particularly evident in several outliers (p-value = 1.5×10^{-2} and p-value = 4.7×10^{-2} respectively). *Work* location analysis showed that CSIR workers had higher relative abundances of both Actinobacteriota and Proteobacteria (p-value = 3.7×10^{-2} and p-value = 1.3×10^{-2}, respectively), while those working in other locations showed higher levels of Cyanobacteria (p-value = 4.3×10^{-2}). *Dairy* consumers ("unchecked" responses) showed higher relative abundances of Actinobacteriota compared to those who excluded dairy from their diet (p-value = 2.8×10^{-2}).

Integration of 16S long-read and short-read shotgun sequencing revealed varying methodological comparisons across the taxonomic resolution levels, with strong functional concordance in profiling the SAN population gut microbiome. Core microbiome analysis identified stable community members across temporal samples, while demonstrating significant inter-individual variation. Natural community stratification exploratory analysis revealed significant associations between specific bacterial phyla and participant characteristics, with Actinobacteriota and Cyanobacteria showing the strongest statistical associations with lifestyle factors. Global population comparisons positioned the SAN population's microbiome profile within established patterns of human gut microbiome diversity. Functional pathway analysis highlighted distinctive metabolic capabilities, particularly in carbohydrate metabolism and biosynthesis pathways, with significant differences observed between participant subgroups. These results collectively provide a comprehensive profile of the urban professional South African gut microbiome across taxonomic, functional, and comparative analyses, which will be discussed in the following chapter.

4.5. Long-read and short-read dataset concordance

A concordance analysis between the 16S rRNA gene long-read sequencing (PacBio) and shotgun metagenomic sequencing (Illumina) datasets was performed on a representative subset (*n* = 29 samples), from the South African (SAN) cohort (Appendix One, Table A1.14).

4.5.1. Taxonomic resolution and classification

At higher taxonomic levels (Kingdom and Phylum), moderate agreement was observed between the 16S and shotgun methods. Shotgun sequencing consistently detected a significantly greater number of unique taxa across all classification levels, particularly evident at genus and species levels. At the species level, shotgun sequencing identified over 125 unique taxa, compared to less than 20 for the 16S method (Appendix One, Figures A1.19 – A1.21). Shotgun sequencing displayed greater sensitivity and resolution, specifically identifying prevalent gut species including *Faecalibacterium prausnitzii* and *Prevotella copri clade A* (Appendix One, Figure A1.26).

4.5.2. Microbial community composition

Both approaches detected similar dominant bacterial phyla (Firmicutes and Bacteroidetes). Notable differences emerged in their relative abundance, with 16S sequencing showing a higher Firmicutes:Bacteroidetes ratio compared to shotgun sequencing (Appendix One, Figure A1.23). Differences became more pronounced at finer taxonomic resolutions (class, order, family), with shotgun sequencing frequently detecting a higher proportion of unclassified taxa, highlighting its deeper resolution and potential to detect novel microbial diversity (Appendix One, Figures A1.24 – A1.25, Supplementary Figures S4.1 – S4.2).

4.5.3. Diversity metrics comparison

Diversity analyses (alpha and beta diversity) revealed significant methodological differences between sequencing approaches across multiple taxonomic ranks. Shotgun sequencing consistently demonstrated higher alpha diversity (Shannon diversity index) than the 16S method at phylum, class, order, and family levels (Kruskal-Wallis $p < 5.0 \times 10^{-2}$; Appendix One, Figures A1.27 – A1.30, Supplementary Figures S4.3 – S4.6). Beta diversity analysis (Bray-Curtis dissimilarity) indicated strong method-specific clustering (PERMANOVA R^2 ranging from 0.3 – 0.5; $p = 1.0 \times 10^{-3}$), emphasizing distinct microbial community structures detected by each sequencing method (Appendix One, Figures A1.27 – A1.30).

4.5.4. Method agreement and bias analysis

Intraclass correlation (ICC) analyses revealed moderate-to-good concordance at higher taxonomic levels (phylum and class; ICC ~0.7), but agreement notably decreased at genus and species levels (ICC < 0.5), indicating increased method-specific biases at finer

resolutions. Bland-Altman and fold-change analyses confirmed systematic abundance-dependent biases between methods, with shotgun sequencing more effectively capturing both low and high abundance taxa (Appendix One, Figures A1.31 –A1.34).

4.5.5. Functional profiling concordance

Predicted functional profiles from 16S data (PICRUSt2) and directly measured functional annotations from shotgun data (HUMAnN3) were compared. Both methods identified a core set of n = 285 common metabolic pathways, while shotgun sequencing uniquely identified an additional n = 506 pathways, substantially broadening functional characterization potential. Moderate correlations were observed between methods ($r = 0.6$; $p < 3.0 \times 10^{-24}$), though substantial method-specific variations exist, especially for pathways at lower abundances (Appendix One, Figures A1.35 – A1.37).

Subsequent analyses examining associations between demographic, socioeconomic, dietary, lifestyle, and health variables and microbial functional profiles showed generally similar trends between methodologies. No associations remained statistically significant after multiple testing corrections (Appendix One, Figures A1.38 – A1.51, Supplementary Data S4.20 – S4.21).

In summary, the comparative analysis highlights methodological differences and specific strengths: shotgun sequencing provides greater taxonomic and functional resolution, while 16S sequencing remains an efficient, cost-effective approach useful for broader ecological and comparative studies. Detailed evaluation of concordance, diversity metrics, and functional profiling methods are documented in Appendix One (Supplementary Results).

4.6. Core versus temporal dynamics of the gut microbiome

Understanding the stability and temporal dynamics of the human gut microbiome is crucial for distinguishing between core community members and transient variations. This section examines the longitudinal changes in gut microbiome composition across two timepoints in n = 22 individuals using the 16S long-read sequencing SAN dataset. We investigate both community-level metrics through diversity analyses and specific taxonomic changes to characterize the balance between stable core microbiome components and temporal fluctuations. The analysis incorporates multiple complementary approaches, including alpha and beta diversity measurements, core microbiome prevalence and differential abundance testing, to provide a comprehensive view of temporal dynamics. By comparing intra-individual changes to inter-individual differences, we aimed to quantify the relative contribution of temporal drift versus host-specific factors in shaping microbiome composition.

4.6.1. Diversity Analysis

i. Taxonomic analysis

To assess microbiome stability within individuals compared to between-individual variation, analysis of paired stool samples collected from *n* = 22 subjects at two timepoints was performed. Bray-Curtis dissimilarity was calculated between all samples, comparing both within-individual temporal changes (intra-individual) and between different individuals (inter-individual) (Figure 4.23). The median Bray-Curtis dissimilarity for intra-individual comparisons (0.8) was significantly lower than inter-individual comparisons (0.9) (Wilcoxon rank-sum test p-value = 1.2×10^{-8}). However, while intra-individual samples were more similar on average (lower median distance), they displayed greater variability in their degree of similarity (SD = 0.2) compared to the more consistently distant inter-individual comparisons (SD = 0.1). This indicates that some individuals maintain highly stable gut microbiomes over time, while others experience substantial compositional shifts, resulting in the wider range of Bray-Curtis values observed in the intra-individual comparison group.

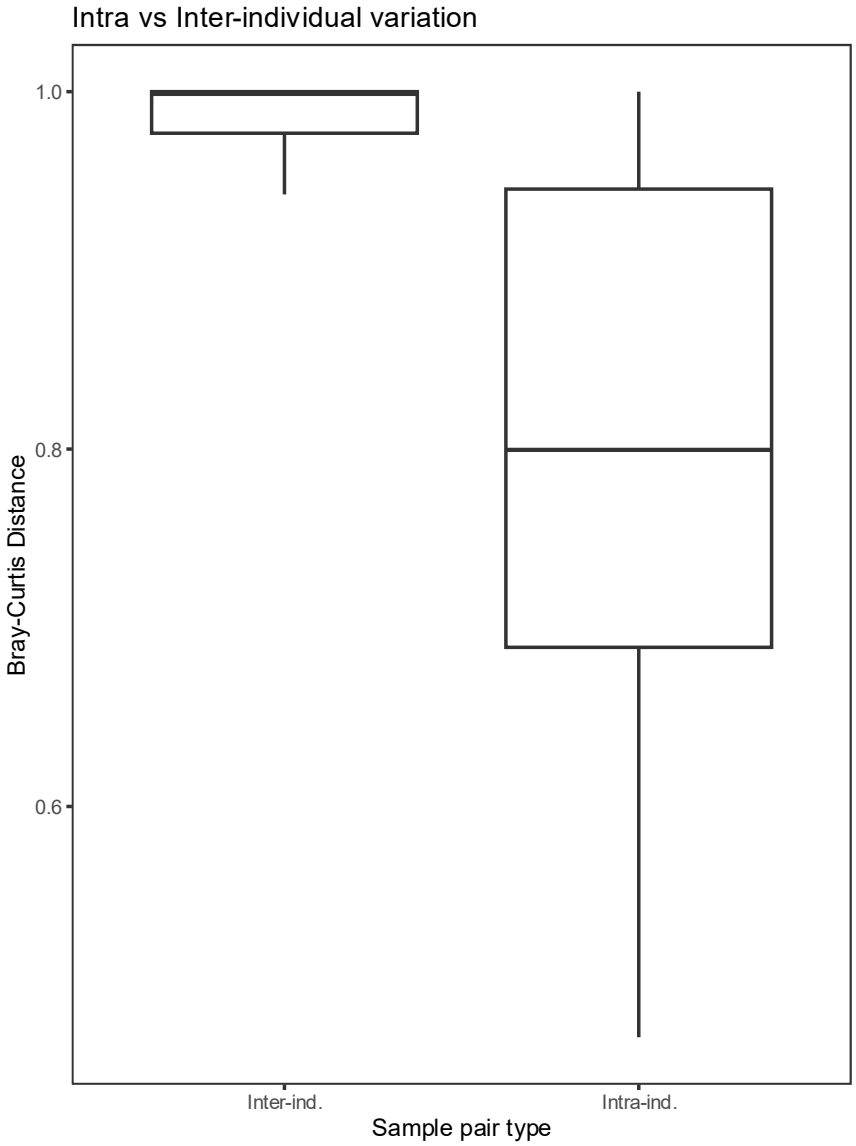


Figure 4.23 Comparison of intra- and inter-individual gut microbiome variation. Boxplots showing Bray-Curtis dissimilarity between stool microbiome samples collected from *n* = 22 subjects. Intra-individual comparisons (n=22) represent temporal changes within the same individual between two timepoints. Inter-individual comparisons (*n* = 231) represent differences between baseline samples from different individuals. Boxes show median and interquartile range; whiskers extend to 1.5 times the interquartile rang.

A permutational multivariate analysis of variance (PERMANOVA) stratified by subject to account for repeated measures revealed that collection timepoint explained only 1.6% of the total variation in microbiome composition ($R^2 = 1.6 \times 10^{-2}$, p-value = 0.5). These results suggest that while individuals' microbiome compositions are generally more similar to their own samples than to others' (as indicated by the lower median dissimilarity), there is substantial and variable temporal change within individuals. The non-significant PERMANOVA result further indicates that these temporal changes do not follow a consistent pattern across

subjects, suggesting that individual-specific factors may drive temporal variability in gut microbiome composition. To further investigate these individual-specific patterns, analysis of temporal changes in alpha diversity at the subject level was performed.

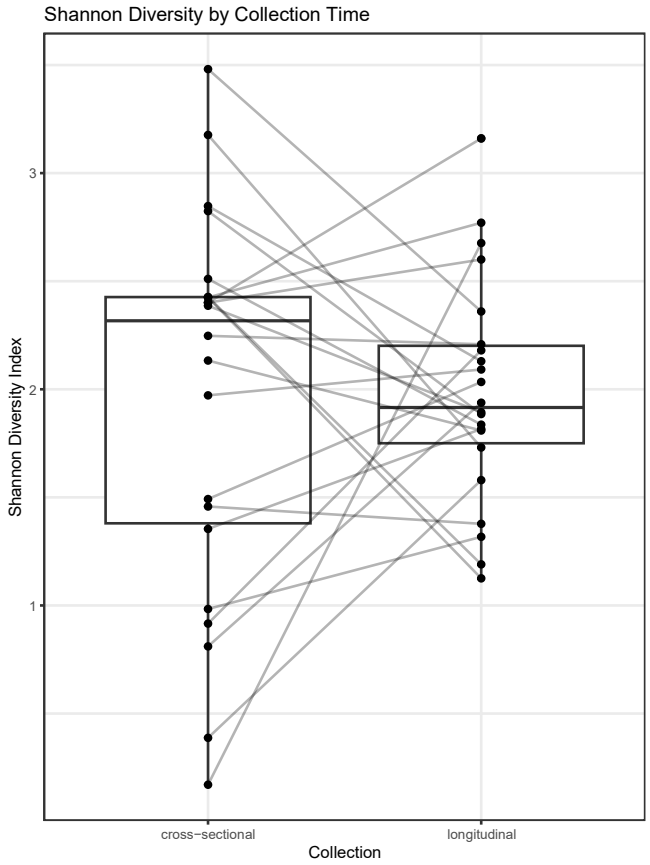


Figure 4.24 Shannon Diversity Index comparison between baseline (cross-sectional) and follow-up (longitudinal) timepoints. Boxplots show the distribution of diversity values at each timepoint, while grey lines connect paired samples from the same individual (*n* = 22), illustrating subject-specific temporal changes.

The overall Shannon diversity indices showed a decrease from baseline (mean = 2.4) to follow-up (mean = 2.0), yet with considerable overlap in distributions and no statistically significant difference between timepoints (Wilcoxon rank-sum p-value = 9.7×10^{-1}; Figure 4.24). Examination of individual-level changes further highlighted substantial heterogeneity in temporal patterns (Figure 4.25). Changes in Shannon diversity varied widely among participants, ranging from a decrease of approximately 1.0 to an increase as large as 2.5, underscoring the highly individualized nature of microbiome dynamics within this cohort. Notably, most participants (68%; 15/22) demonstrated moderate shifts (absolute change < 1.0), whereas roughly a third (32%, 7/22) exhibited pronounced shifts (absolute change > 1.0).

This observed variability was consistent with published longitudinal microbiome studies, where individual microbiome profiles often show personalized and idiosyncratic trajectories over time due to diverse environmental, dietary, lifestyle, and other influencing factors (Björk et al., 2024;

Bourdeau-Julien et al., 2023). Therefore, the lack of significant overall temporal difference was not indicative of uniform stability among individuals but rather reflected averaging out heterogeneous individual-level shifts in opposite directions. This variability was accounted for in subsequent analyses to ensure accurate interpretation of group-level microbiome patterns and associations.

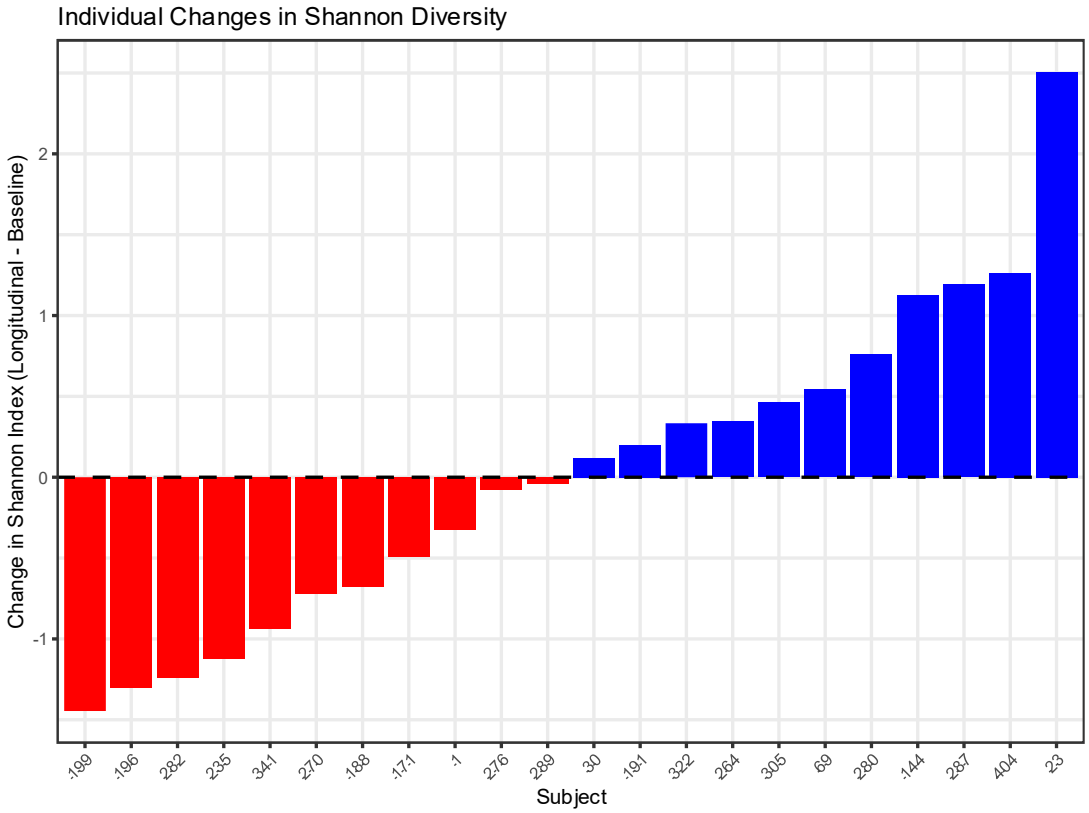


Figure 4.25 Individual changes in Shannon Diversity between timepoints. Bars represent the change in Shannon Diversity Index (longitudinal - baseline) for each subject (*n* = 22). Blue bars indicate increased diversity at follow-up, while red bars indicate decreased diversity. Subjects are ordered by magnitude of change.

To visualize the overall patterns of microbiome composition and validate our previous dissimilarity findings (Figure 4.23), we performed Principal Coordinates Analysis (PCoA) based on Bray-Curtis dissimilarities (Figure 4.26). The first two principal coordinates explained a combined 17.6% of the total variation (PCoA1: 11.1%, PCoA2: 6.5%) and emphasized the high dimensionality of the data. The ordination revealed no clear clustering by collection timepoint, consistent with the PERMANOVA results (Supplementary Data S4.12A).

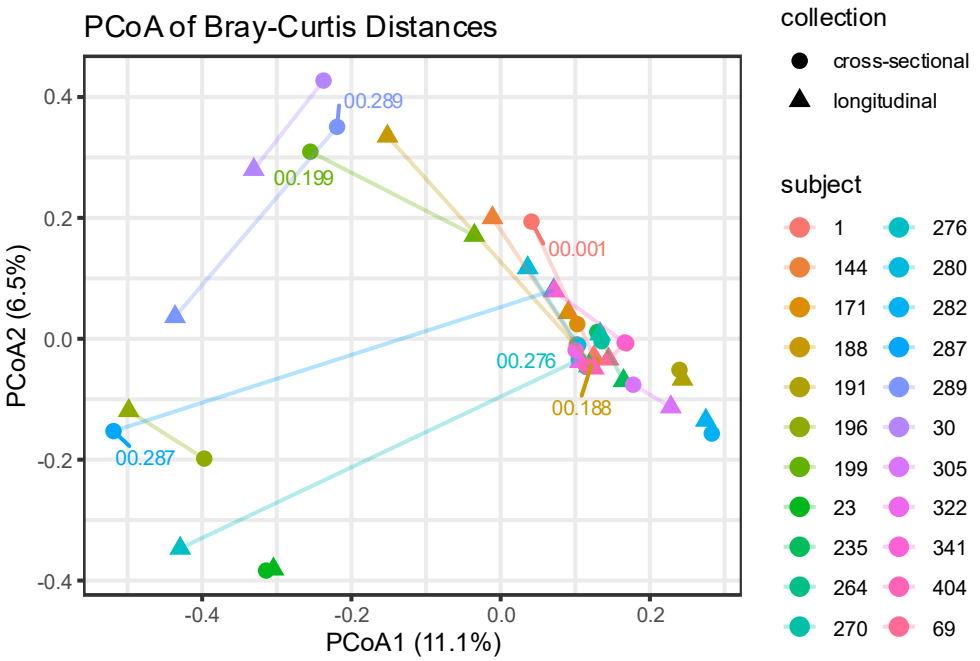


Figure 4.26 Principal Coordinates Analysis (PCoA) of stool microbiome samples based on Bray-Curtis dissimilarities. Points represent individual samples coloured by subject (*n* = 22), with shapes indicating collection timepoint (circles = cross-sectional / baseline, triangles = longitudinal / follow-up). Connected lines show paired samples from the same subject. Subject IDs on cross-sectional samples are shown for individuals with the largest temporal shifts (top 25% of paired distances). The first two principal coordinates explain 11.1% and 6.5% of the total variation, respectively.

Samples from the same subject tended to cluster together, as indicated by the relatively short distances between paired samples (connected by lines) compared to the overall spread of points in the ordination space (Figure 4.26). However, the magnitude and direction of temporal changes varied substantially between subjects with some showing minimal shift in community composition (e.g. subjects 23 and 282 with tightly clustered pairs), while others displayed more pronounced changes. The most substantial temporal shifts (Bray-Curtis distances > 0.9) were observed in subjects *1, 188, 199, 276, 287* and *289*, representing the top quartile of paired sample distances in the ordination space. Further investigation of the six subjects showing the largest compositional shifts revealed concurrent changes in multiple lifestyle and health factors (Table 4.3). These subjects exhibited distinct separation between timepoints in the PCoA ordination, with their temporal changes following subject-specific trajectories.

Table 4.3 Survey responses in subjects with largest Bray-Curtis Dissimilarity temporal shifts

Changes in survey responses for subjects with shifts in Bray-Curtis Distances

Subject ID	Survey variable	Change between timepoints
188	timeintraffic	Changed from Less than 10% to 20-40%
188	heat_gas	Changed from Unchecked to Checked
188	cook_gas	Changed from Unchecked to Checked
188	multivit	Changed from No to Yes
188	vitb	Changed from No to Yes
188	spirits_hard	Changed from Unchecked to Checked
199	antibiotics	Changed from No to Six months or less
199	fluvac	Changed from Year to Six months or less
276	longcommute	Changed from 30-45 minutes to 45+ minutes
276	travelmonth	Changed from NA to Within the year
276	heat_gas	Changed from Unchecked to Checked
276	cook_gas	Changed from Unchecked to Checked
276	bowelqualitychoice	Changed from I tend to be constipated (have difficulty passing stool) - Type 1 and 2 to I tend to have normal formed stool - Type 3 and 4
276	antibiotics	Changed from No to Six months or less
287	weightkgs	Changed from 73 to 63
287	bmi	Changed from 27,14 to 23,42
287	bmi_group	Changed from Overweight (<30) to Normal weight (<25)
287	multivit	Changed from Yes to No
287	vitd	Changed from Yes to No
287	poolorhottub	Changed from Yes to No
287	alcohol	Changed from No to Yes
287	white_wine	Changed from Unchecked to Checked
287	alcohol_type_none	Changed from Checked to Unchecked
287	bowelmovements	Changed from Less than one to One
287	bowelqualitychoice	Changed from I tend to be constipated (have difficulty passing stool) - Type 1 and 2 to I tend to have normal formed stool - Type 3 and 4

Note:
Only subjects in the top 25% of microbiome compositional changes are shown

Notable patterns emerged in the metadata: four subjects (*188, 276, 287* and *199*) reported changes in medication use or medical interventions, particularly antibiotic exposure. Three subjects (*188, 276* and *287*) experienced changes in lifestyle factors affecting daily routines, including commuting patterns and environmental exposures. Changes in gastrointestinal function were reported by two subjects (*276* and *287*), who showed improvements in bowel movement patterns. Subject *287*'s substantial shift coincided with post-partum changes, including significant changes in BMI and dietary habits. To determine whether these subject-specific taxonomic shifts corresponded to functional shifts, we next examined the temporal stability of metabolic pathways across timepoints.

ii. Functional analysis

While taxonomic diversity analyses revealed substantial individual-specific temporal patterns, these compositional changes may not directly reflect functional variations in the gut microbiome. To investigate whether functional capabilities show similar patterns of variation, we examined the diversity and stability of KEGG ortholog (KO) pathways across timepoints (Supplementary Data S4.13). This complementary functional analysis provides insight into whether the observed taxonomic changes translate to alterations in metabolic potential, or if functional redundancy among different taxa maintains stable pathway representation despite compositional shifts.

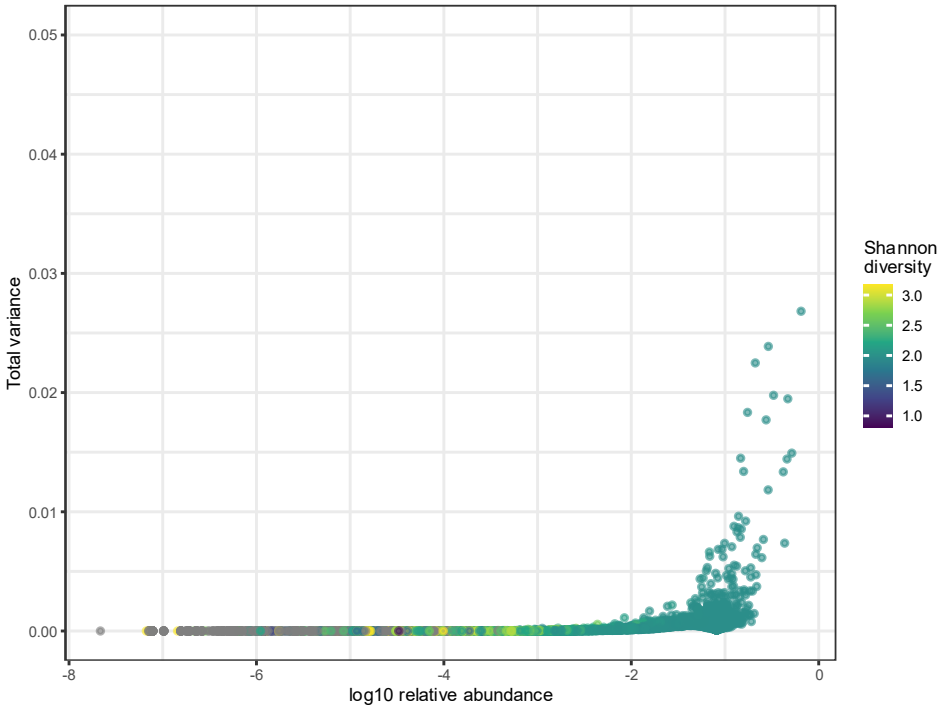


Figure 4.27 Total variance of KEGG ortholog (KO) pathway abundances plotted against their log10-transformed relative abundances. Points are coloured by the weighted Shannon Diversity Index of samples where each pathway is present, showing relationships between pathway abundance, variability, and community diversity.

Functional analysis of gut microbiome KO pathways revealed distinct patterns of abundance and variability. The relationship between pathway abundance and variance (Figure 4.27) demonstrated that pathways with higher relative abundances (log10 relative abundance approaching 0) showed greater total variance, with some highly abundant pathways exhibiting variance values > 2.0×10^{-2}. The majority of pathways clustered at lower abundance levels (log10 relative abundance < -2) with correspondingly low variance (< 1.0×10^{-2}). Shannon diversity indices were relatively consistent across most pathways (~1.5 – 2), with only a few pathways showing notably different diversity values.

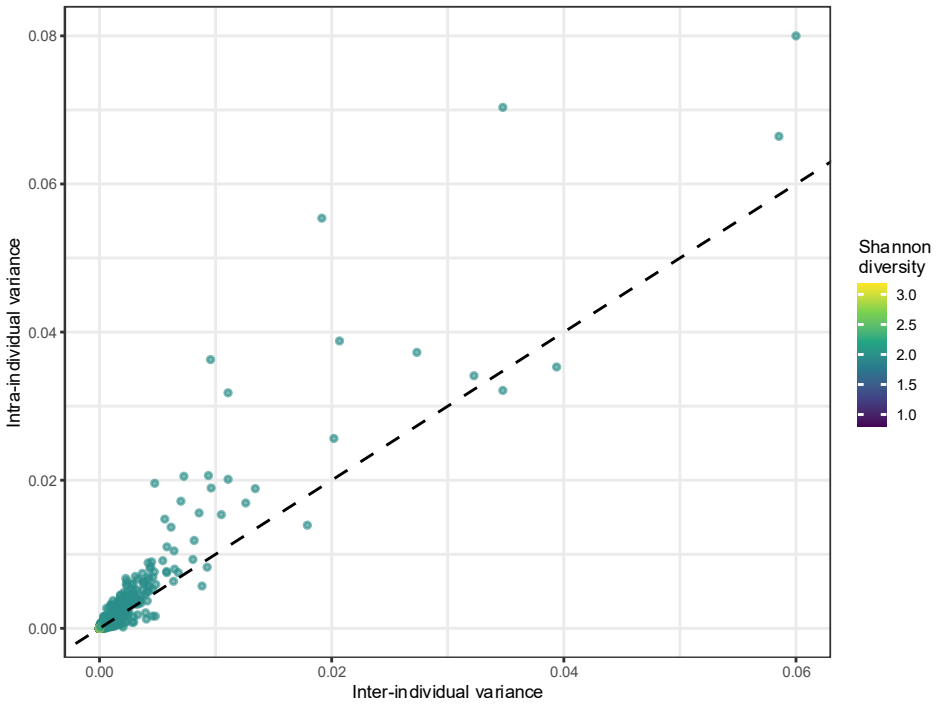


Figure 4.28 Intra- versus inter-individual variance components of KO pathway abundances. Each point represents a KO pathway, with its position determined by its inter-individual variance (x-axis) and intra-individual variance (y-axis). The intraclass correlation coefficient (ICC), defined as the ratio of inter-individual variance to total variance (ICC = inter-individual variance / [inter-individual variance + intra-individual variance]), quantifies the proportion of total variance attributable to between-individual differences. The dashed line indicates where intra-individual equals inter-individual variance (ICC = 0.5), with points below the line having ICC > 0.5 (predominant inter-individual variation) and points above having ICC < 0.5 (predominant intra-individual variation). Points are coloured by weighted Shannon Diversity Index of samples containing each pathway.

Analysis of variance components revealed most KO pathways had low variability (both inter- and intra-individual variance < 2.0 x 10^{-2}). A subset of pathways showed higher variance, with some exhibiting predominantly intra-individual variation and others showing greater inter-individual variation (Figure 4.28; Table 4.4). These high variance pathways had maintained Shannon diversity values similar to low variance pathways. The variance distribution indicates that while most functional pathways remain stable, specific pathways show distinct patterns of temporal or inter-individual variation.

Having established these overall patterns of compositional and functional diversity changes, we next investigated which specific taxa contributed to the observed temporal dynamics. While diversity analyses revealed broad patterns of stability and variation in both taxonomic structure and functional potential, examining individual taxonomic shifts is crucial for understanding which community members drive these temporal changes and may contribute to functional variations.

4.6.2. Taxonomic composition analysis

To characterize the temporal dynamics of gut microbiome composition, analysis of the distribution and stability of bacterial taxa across taxonomic levels was performed, beginning with phylum-level patterns. This analysis focused on comparing bacterial communities between cross-sectional and longitudinal timepoints, considering both collection time group trends and subject-specific variations (Supplementary Data S4.12).

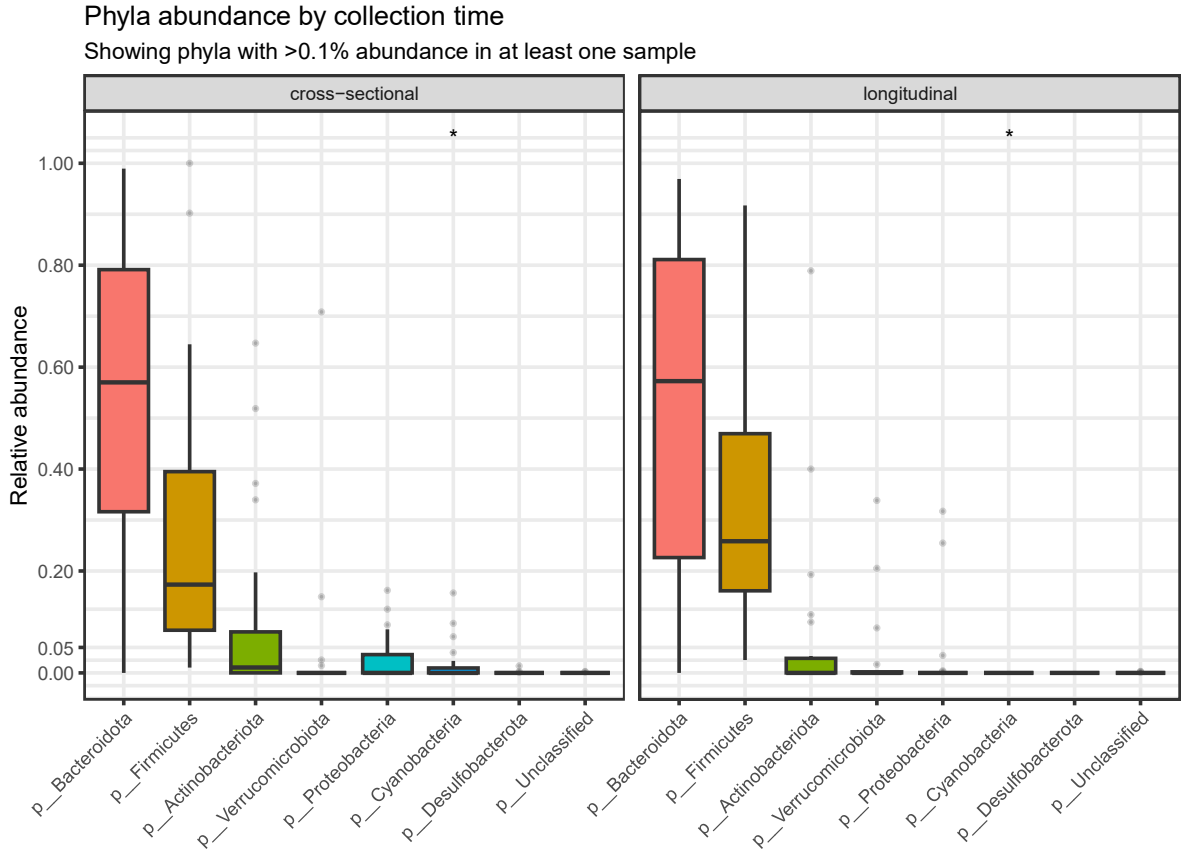


Figure 4.29 Distribution of bacterial phyla relative abundance across collection times. Box plots show the relative abundance of bacterial phyla present at > 0.1% in at least one sample, comparing cross-sectional and longitudinal collection times. Asterisks indicate significant differences between timepoints (Wilcoxon test with FDR correction: * $p < 5.0 \times 10^{-2}$).

Eight bacterial phyla were present at > 0.1% relative abundance in at least one sample. Bacteroidota and Firmicutes dominated both cross-sectional and longitudinal samples (median abundances ~0.6 and ~0.2 respectively), while Actinobacteriota, Verrucomicrobiota, and Proteobacteria showed lower abundances (median < 5.0×10^{-2}). The remaining phyla had minimal representation (Figure 4.29). Abundance patterns were largely consistent between timepoints, though Cyanobacteria showed significant temporal variation (FDR-adjusted p < 5.0×10^{-2}).

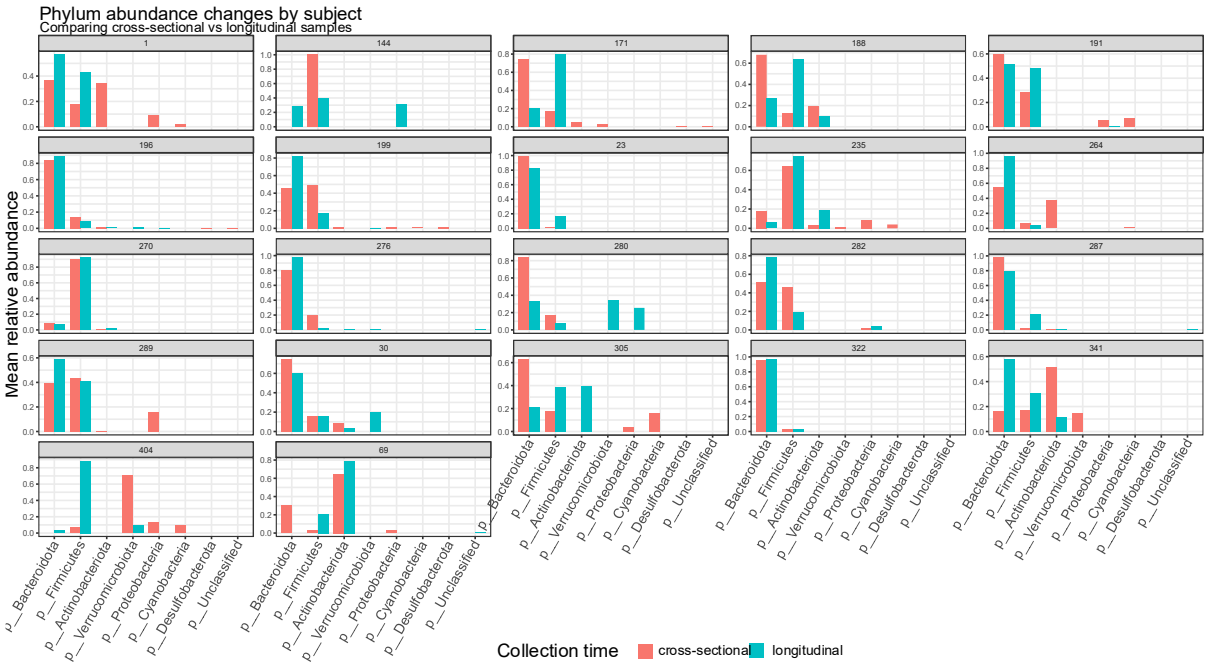


Figure 4.30 Subject-specific mean relative abundance of bacterial phyla across collection times. Faceted bar plots display the mean relative abundance of phyla (> 0.1%) for each subject (*n* = 22), comparing cross-sectional (red) and longitudinal (blue) timepoints.

Subject-level analysis revealed considerable inter-individual variation in phyla abundance and temporal dynamics (Figure 4.30). While some subjects maintained relatively stable phyla proportions between timepoints, others showed marked changes, particularly in the dominant Bacteroidota and Firmicutes phyla. This variability was evident across all 22 subjects with no consistent pattern of increase or decrease for any particular phylum.

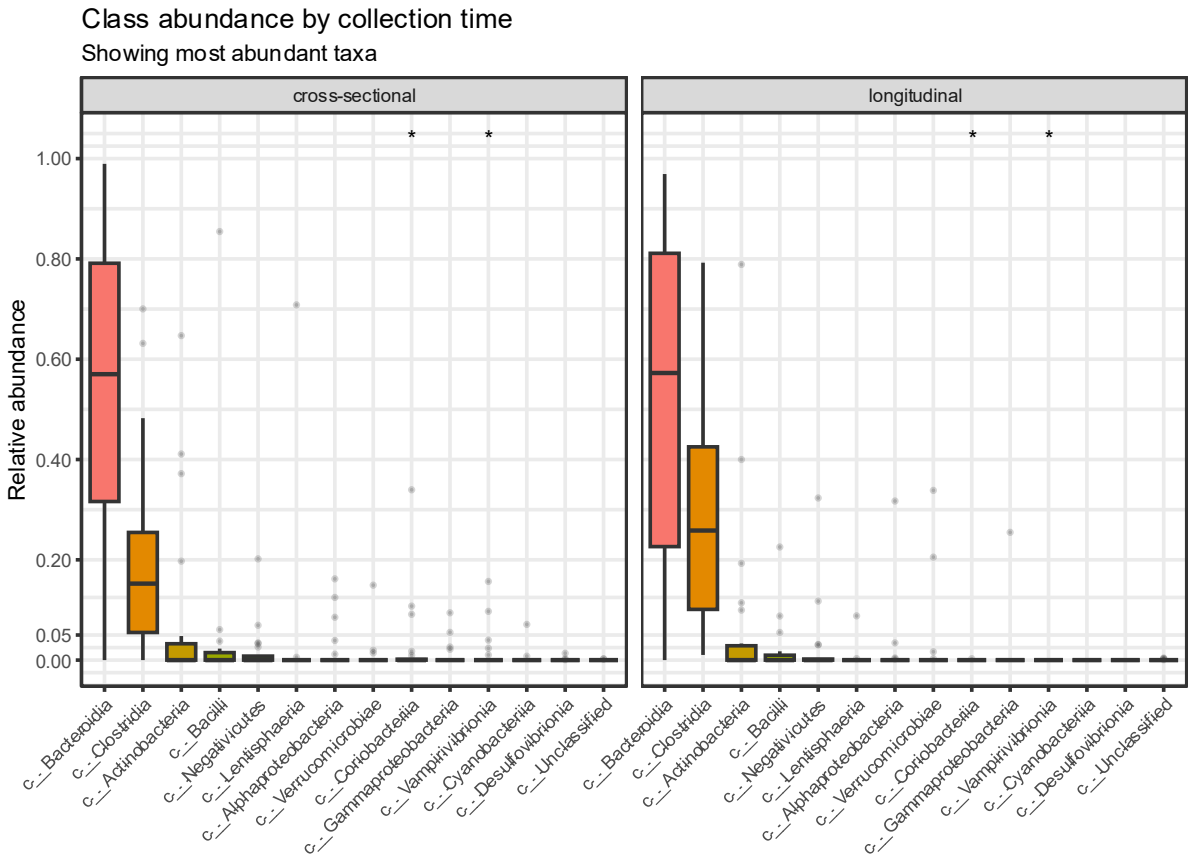


Figure 4.31 Distribution of the 14 most abundant bacterial classes across collection timepoints. Boxplots showing relative abundances in cross-sectional and longitudinal samples. Asterisks indicate significance levels from Wilcoxon tests (*p < 5.0 x 10^{-2}, unadjusted p-values).

At the class level the gut microbiome was predominantly composed of Bacteroidia and Clostridia. Bacteroidia showed the highest relative abundance (~60% median abundance), followed by Clostridia (~25% median abundance). Several other classes were consistently present at lower abundances, including Actinobacteria, Bacilli and Negativicutes, each typically representing 1 – 5% of the community (Figure 4.31). Statistical comparison between timepoints revealed significant differences (unadjusted p < 5.0 x 10^{-2}) in the relative abundances of two classes: Vampirivibrionia and Coriobacteriia. However, these differences did not remain significant after FDR correction for multiple comparisons.

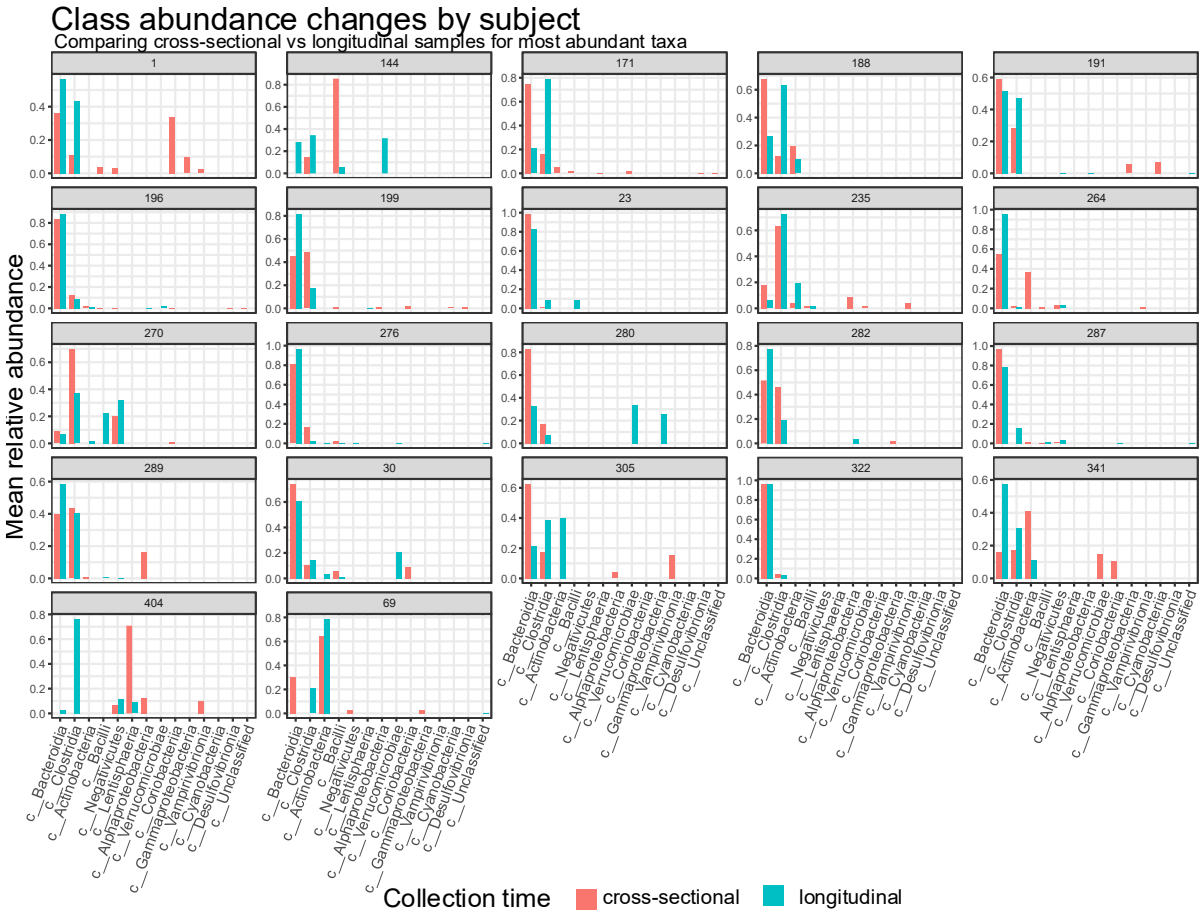


Figure 4.32 Class-level bacterial abundance changes by subject. Faceted bar plots display the mean relative abundance of the most abundant classes for each subject (*n* = 22), comparing cross-sectional (red) and longitudinal (blue) timepoints. Numbers at the top of each facet plot indicate the sample identification number.

Subject-level analysis demonstrated substantial inter-individual variability in class-level compositions. While the dominance of Bacteroidia and Clostridia was consistent across most subjects, their relative proportions and temporal stability varied considerably between individuals (Figure 4.32).

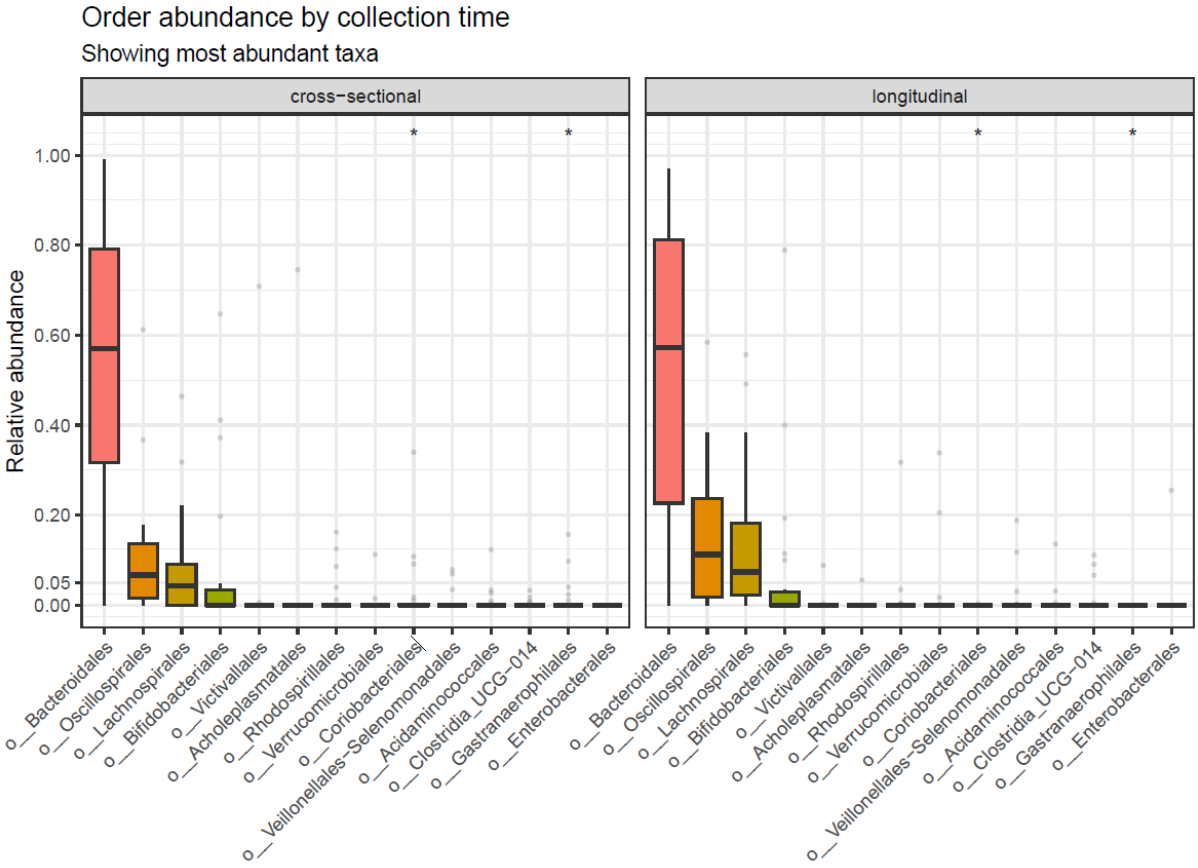


Figure 4.33 Distribution of the 14 most abundant bacterial orders across collection timepoints. Boxplots showing relative abundances in cross-sectional and longitudinal samples. Asterisks indicate significance levels from Wilcoxon tests (*p<0.05, unadjusted p-values).

At the order level, Bacteroidales remained the dominant group (60% median relative abundance), followed by Oscillospirales and Lachnospirales (approximately 5 – 20%). The next most prevalent orders included Bifidobacteriales and Victivallales, which were consistently present but at lower abundances (1-10%) (Figure 4.33). Statistical comparison between timepoints identified significant differences (unadjusted $p < 5.0 \times 10^{-2}$) in the relative abundances of Gastranaerophilales and Coriobacteriales. Examining the subject-level plots (Figure 4.34), these differences appear to be driven by specific individuals in cross-sectional samples.

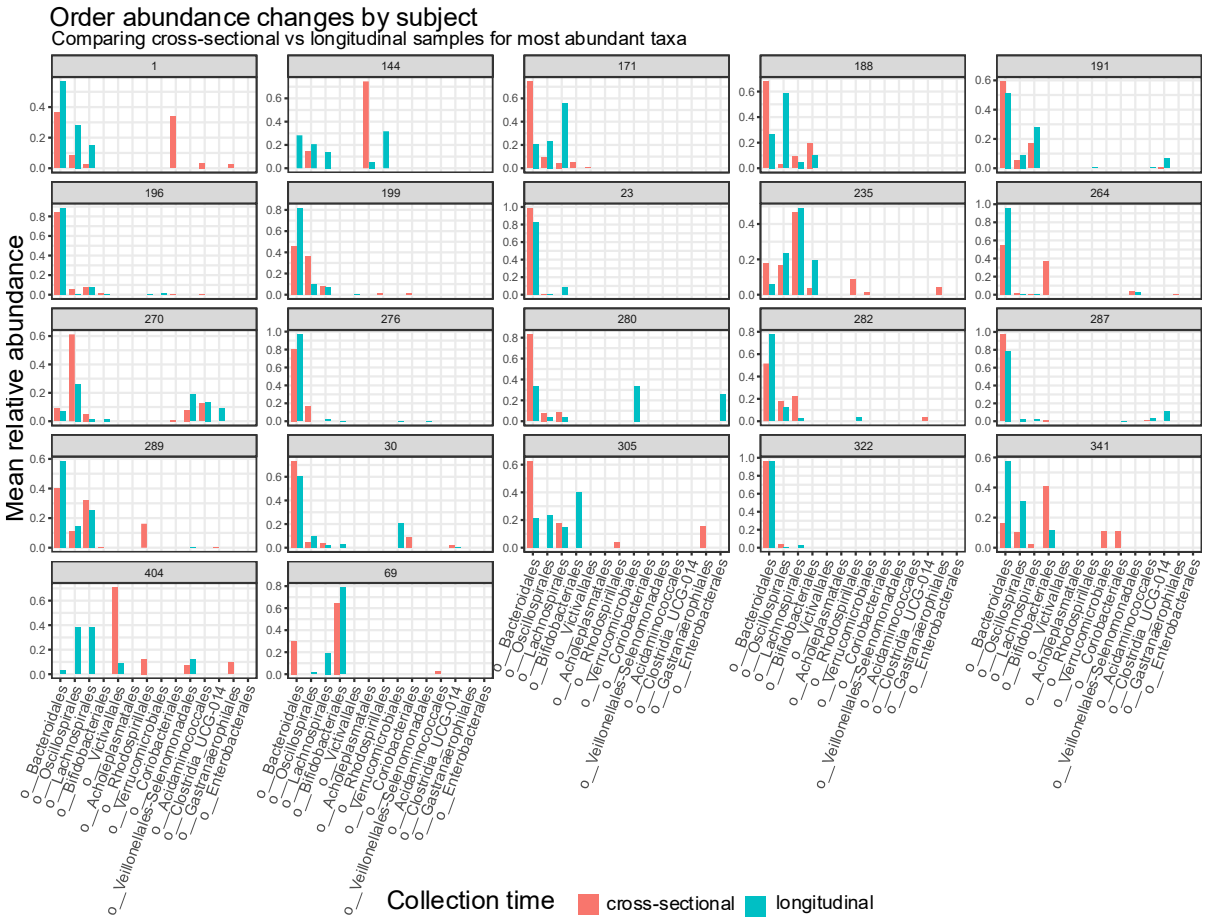


Figure 4.34 Order-level bacterial abundance changes by subject. Faceted bar plots display the mean relative abundance of the most abundant orders for each subject (n=22), comparing cross-sectional (red) and longitudinal (blue) timepoints.

For Gastranaerophilales, subjects *1, 235, 305* and *404* showed notably higher abundances; and for Coriobacteriales, subjects *1, 30* and *341* showed a similar pattern (Figure 4.34). However, as with the class level, these differences did not remain significant after FDR correction for multiple comparisons. Furthermore, the subject-specific analysis continued to demonstrate substantial inter-individual variability in community composition. While Bacteroidales consistently dominated across most subjects, its relative abundance and temporal stability varied considerably between individuals. ranging from highly stable (e.g. subjects *322, 276*) to showing marked temporal changes (e.g. subjects *305, 341*).

At the family level (Supplementary Figures S4.15 – 16), Bacteroidaceae emerged as the predominant family (up to 55% relative abundance) followed by Lachnospiraceae and Ruminococcaceae up to ~25%). This distribution suggests a Bacteroides-dominated enterotype pattern typical of urbanized populations. While individual subjects showed varying patterns of family-level composition and temporal dynamics, no significant differences were detected between cross-sectional and longitudinal timepoints. The high inter-individual

variability observed at higher taxonomic levels was maintained, with subjects displaying distinct patterns of family-level community composition and stability over time.

At the genus level, we identified four taxa showing significant differences between timepoints (unadjusted $p < 5.0 \times 10^{-2}$): *Blautia, Gastranaerophilales, Sutterella* and *Butyricimonas* (Figure 4.35). Linear discriminant analysis effect size (LEfSe) analysis supported these findings with *Blautia* showing the strongest effect size (LDA score >4.4) and enrichment in longitudinal samples (Figure 4.36).

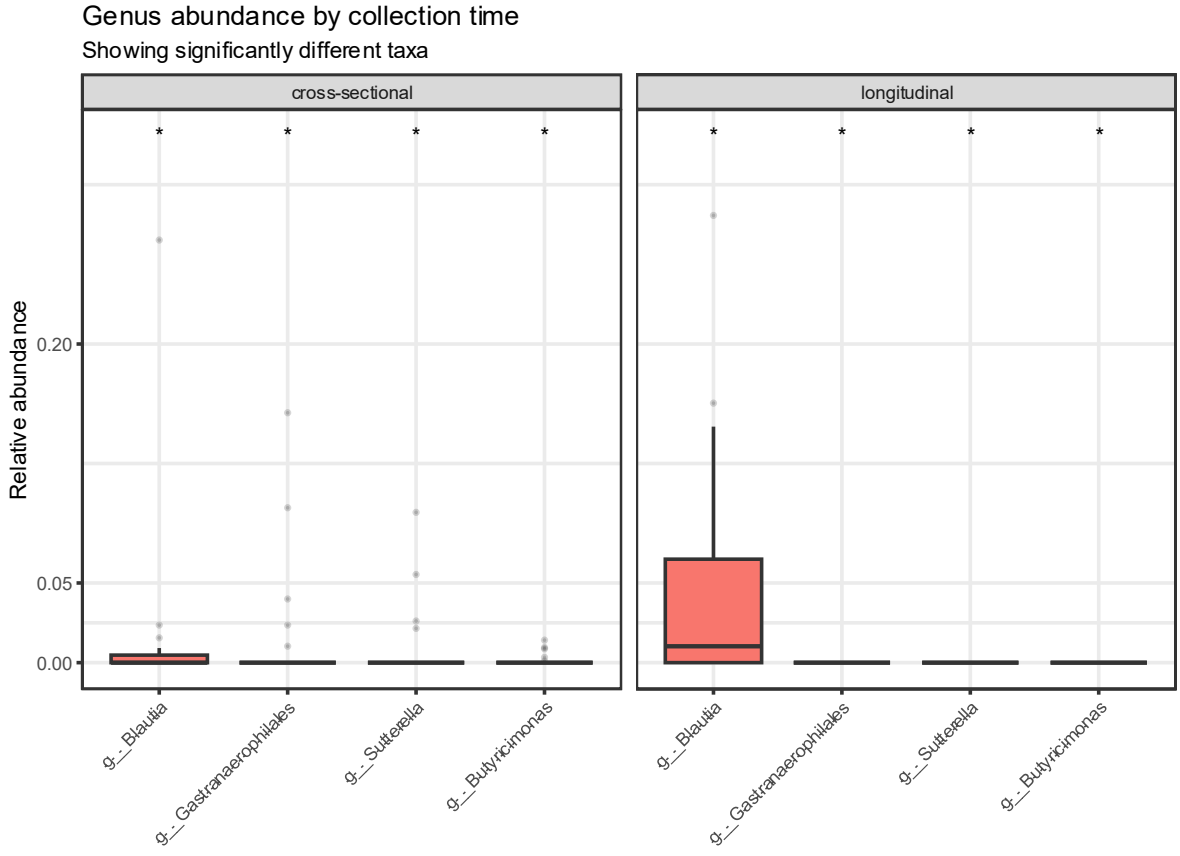


Figure 4.35 Genus-level bacterial differences between timepoints. Boxplots showing relative abundances of significantly different genera between cross-sectional and longitudinal samples (Wilcoxon test, p<0.05, unadjusted). Asterisks indicate significance levels.

Notably, *Gastranaerophilales*, which falls under the phylum Cyanobacteria (the only phylum showing significant changes after FDR correction), demonstrated marked differential abundance between timepoints (LDA score ~4.1). The remaining genera showed moderate effect sizes (LDA score < 3.9) (Figure 4.36).

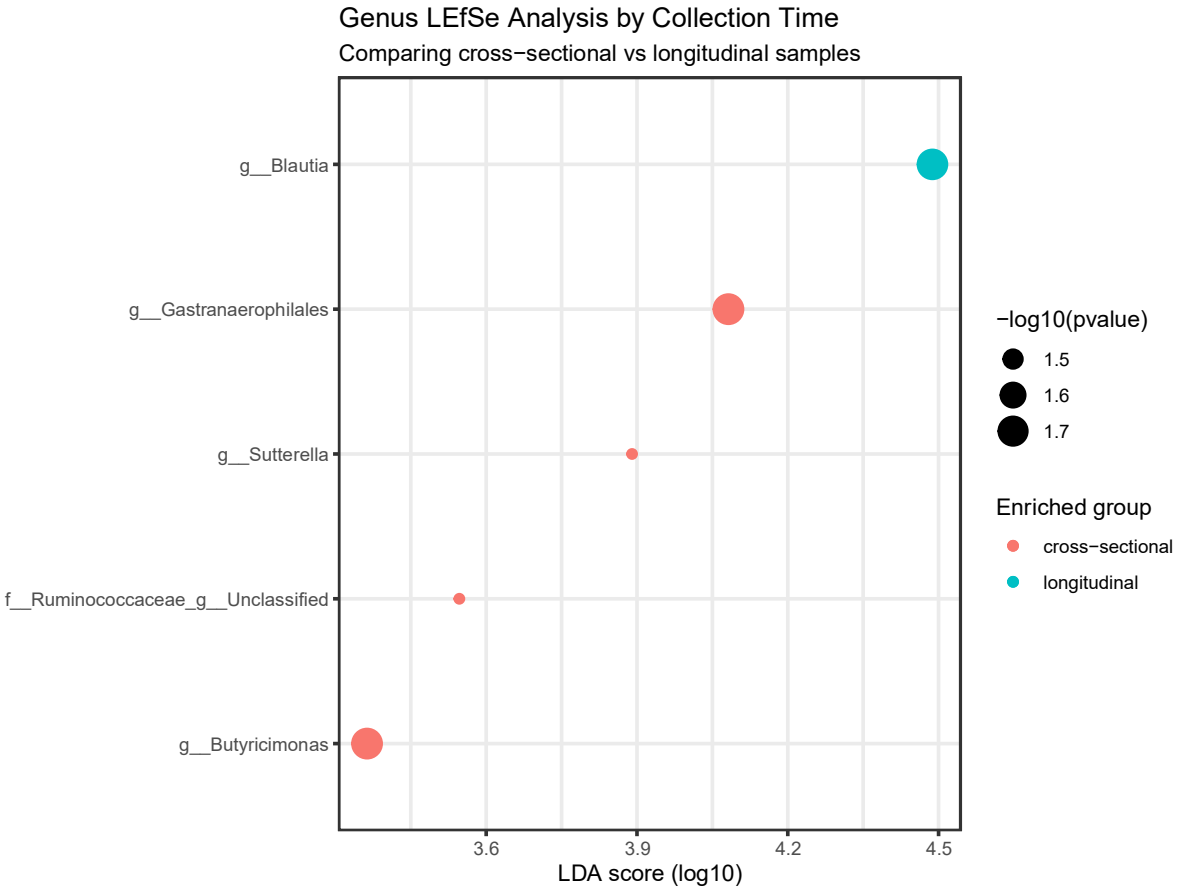


Figure 4.36 LEfSe analysis results displaying differentially abundant genera between timepoints. Point size indicates statistical significance (-log10(p-value)), and position shows effect size (LDA score). Colours indicate the enriched timepoint for each genus.

Subject-level analysis revealed that these genus-level differences were driven by distinct patterns in specific individuals (Figure 4.37). For example, subjects *1, 144, 196, 199* and *289* showed markedly higher *Blautia* abundance in longitudinal samples, while subject *270* exhibited the opposite pattern. *Butyricimonas* changes were most prominent in subjects *270* and *30*. This high inter-individual variability in temporal dynamics suggests subject-specific variability at the genus level.

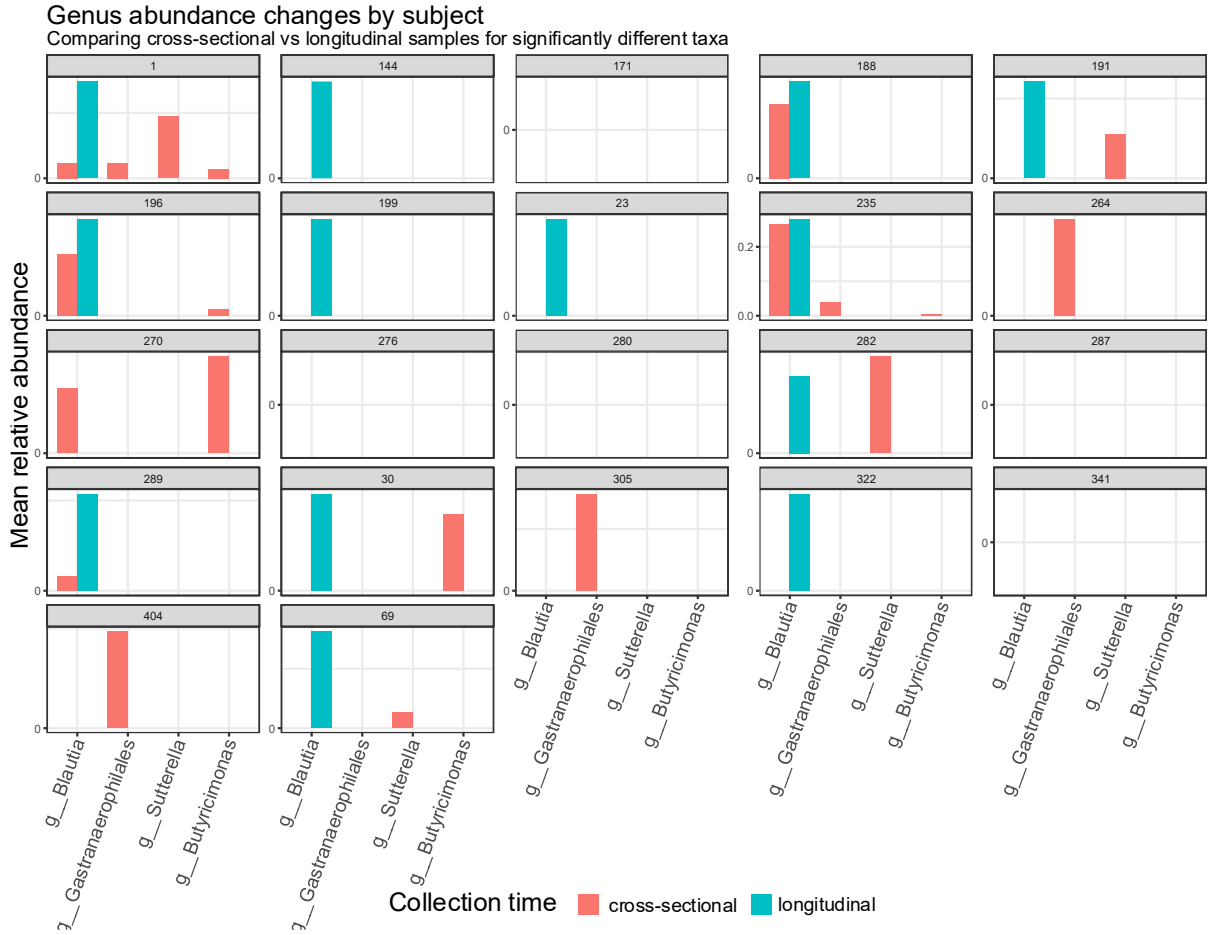


Figure 4.37 Genus abundance changes by subject. Faceted bar plots display the mean relative abundance of significantly different genera for each subject (n=22), comparing cross-sectional (red) and longitudinal (blue) timepoints.

At the species level, only *Bacteroides thetaiotaomicron* showed significant differences between timepoints (unadjusted $p < 5.0 \times 10^{-2}$). LEfSe analysis revealed additional species-level patterns, with unclassified *Blautia* species showing the strongest effect size (LDA score > 4.4) and enrichment in longitudinal samples. *B. thetaiotaomicron* demonstrated moderate effect size (LDA score ~3.6) with enrichment in cross-sectional samples (Figure 4.38).

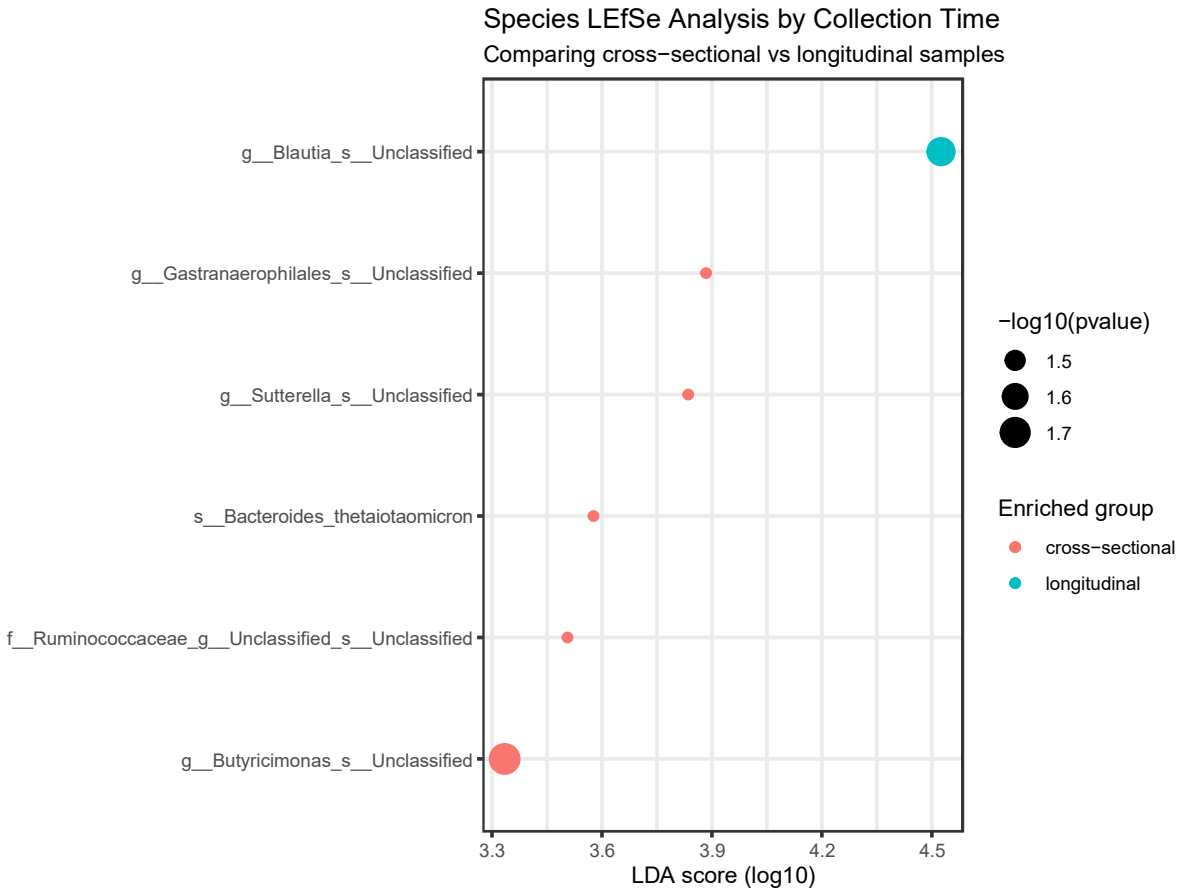


Figure 4.38 LEfSe analysis results displaying differentially abundant species between timepoints. Point size indicates statistical significance (-log10(p-value)), and position shows effect size (LDA score). Colours indicate the enriched timepoint for each species.

Subject-level analysis showed that *B. thetaiotaomicron* presence and abundance was primarily driven by a subset of individuals, with subjects *188, 199, 276* and *289* highlighting this in cross-sectional samples (Figure 4.39). This pattern of subject-specific variation at the species level further emphasizes the trend of inter-individual variability in temporal dynamics.

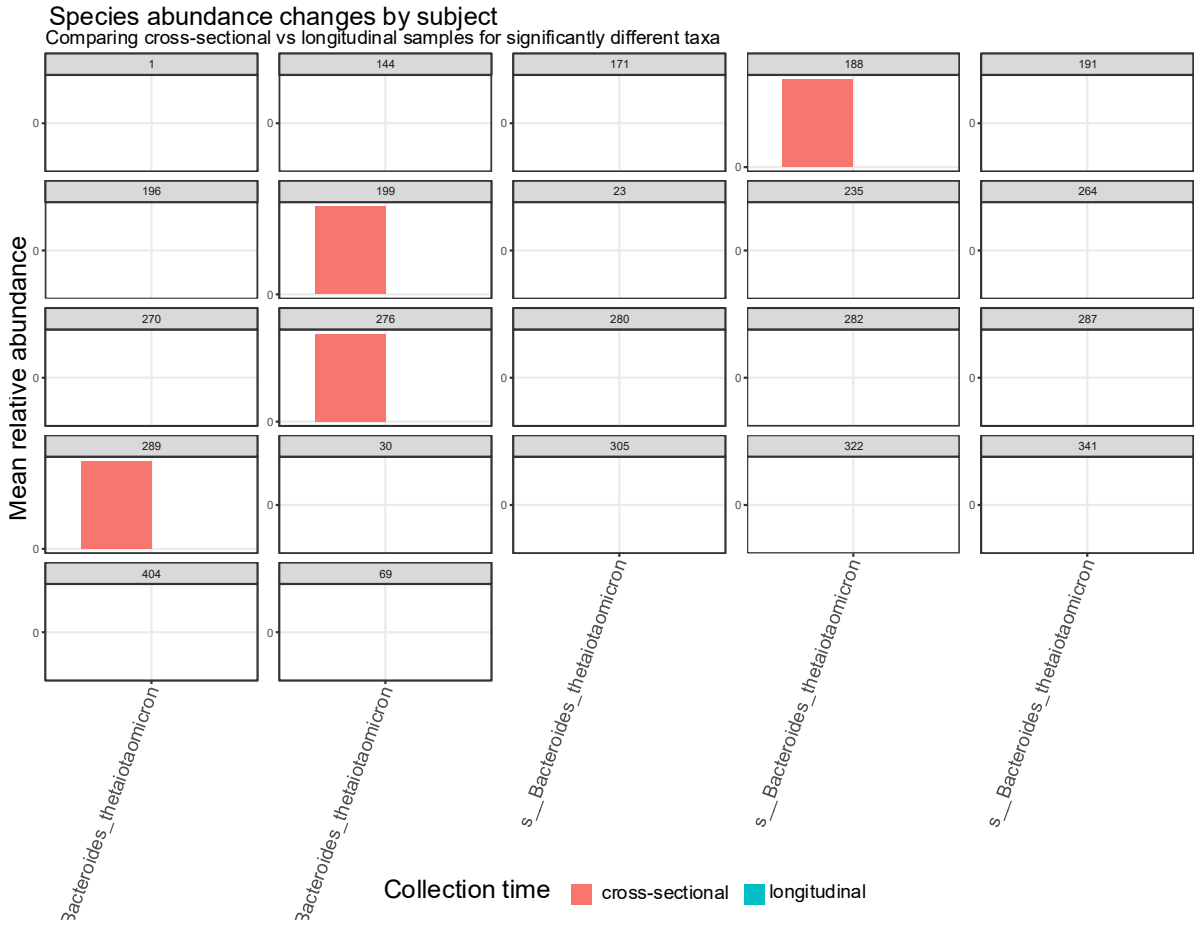


Figure 4.39 Species abundance changes by subject. Faceted bar plots display the mean relative abundance of significantly different species for each subject (n=22), comparing cross-sectional (red) and longitudinal (blue) timepoints.

While these taxonomic analyses revealed significant temporal patterns and highlighted substantial inter-individual variability, they do not fully address the stability of the core gut microbiome across timepoints. To better understand which bacterial taxa constitute the stable core microbiome in our cohort and how this core community might differ between cross-sectional and longitudinal samples, we next performed a detailed core microbiome analysis.

4.6.3. Core microbiome analysis

i. Taxonomic analysis

To identify the consistent microbial signatures in the longitudinal subset of the SAN population, a core microbiome analysis framework incorporating both prevalence and abundance criteria across all taxonomic levels was implemented. Variable and core taxa were defined using two prevalence thresholds: a moderate threshold of 80% and a stringent threshold of 95% presence across all samples (Falony et al., 2016), while maintaining a minimum mean relative abundance of 0.1% to exclude rare taxa. This dual-threshold approach allows us to distinguish between widely distributed taxa (i.e. variable taxa present in most individuals) and highly

conserved taxa (i.e. core taxa present in nearly all individuals). We further characterized the temporal stability of identified taxa by quantifying their abundance fluctuations between timepoints, with a log2 fold change threshold of ±1 distinguishing stable from variable community members.

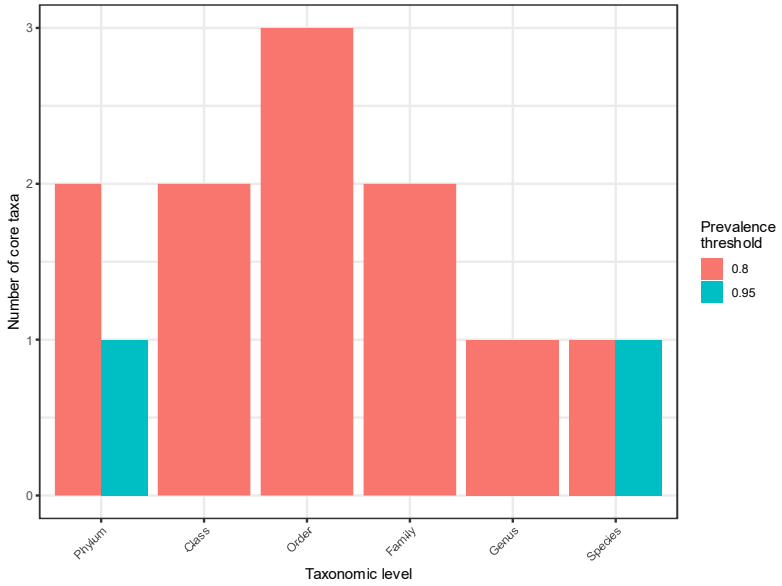


Figure 4.40 Distribution of core taxa across taxonomic levels identified using two prevalence thresholds. Bars show the number of core taxa at each taxonomic rank that meet the prevalence criteria of 80% (pink) and 95% (blue) while maintaining a minimum mean relative abundance of 0.1%.

The analysis revealed distinct patterns in the distribution and abundance of core taxa across taxonomic levels (Figure 4.40). At the moderate prevalence threshold (80%), core taxa were identified across all taxonomic ranks with the highest number observed at the order level (3 taxa), followed by phylum and class levels (2 taxa each) and family level (2 taxa). Fewer core taxa were identified at the genus and species levels (1 taxon each). When applying the more stringent prevalence threshold of 95%, the number of core taxa decreased substantially, with only the phylum and species levels retaining core taxa (1 each).

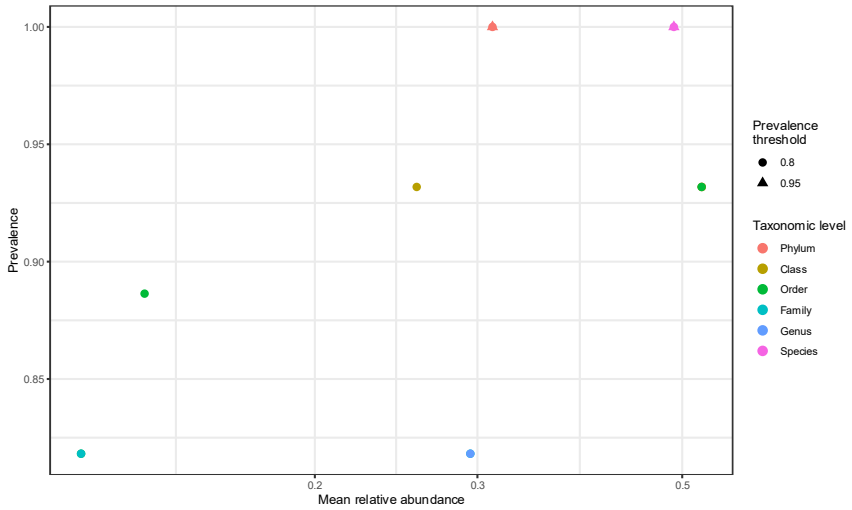


Figure 4.41 Mean relative abundance versus prevalence of core taxa across taxonomic levels. Points represent individual taxa coloured by taxonomic level, with circles indicating the 80% prevalence threshold and triangles indicating the 95% prevalence threshold.

The abundance-prevalence relationship (Figure 4.41) revealed that the most abundant core taxa were identified at the species and phylum levels, with mean relative abundances reaching approximately 0.5. This distribution pattern suggests that while core community membership is more conserved at broader taxonomic levels, the most abundant individual taxa can be found across the taxonomic hierarchy. Notably, taxa meeting the 95% prevalence threshold showed relatively high mean relative abundances, indicating that the most consistently present community members also tend to be among the most abundant.

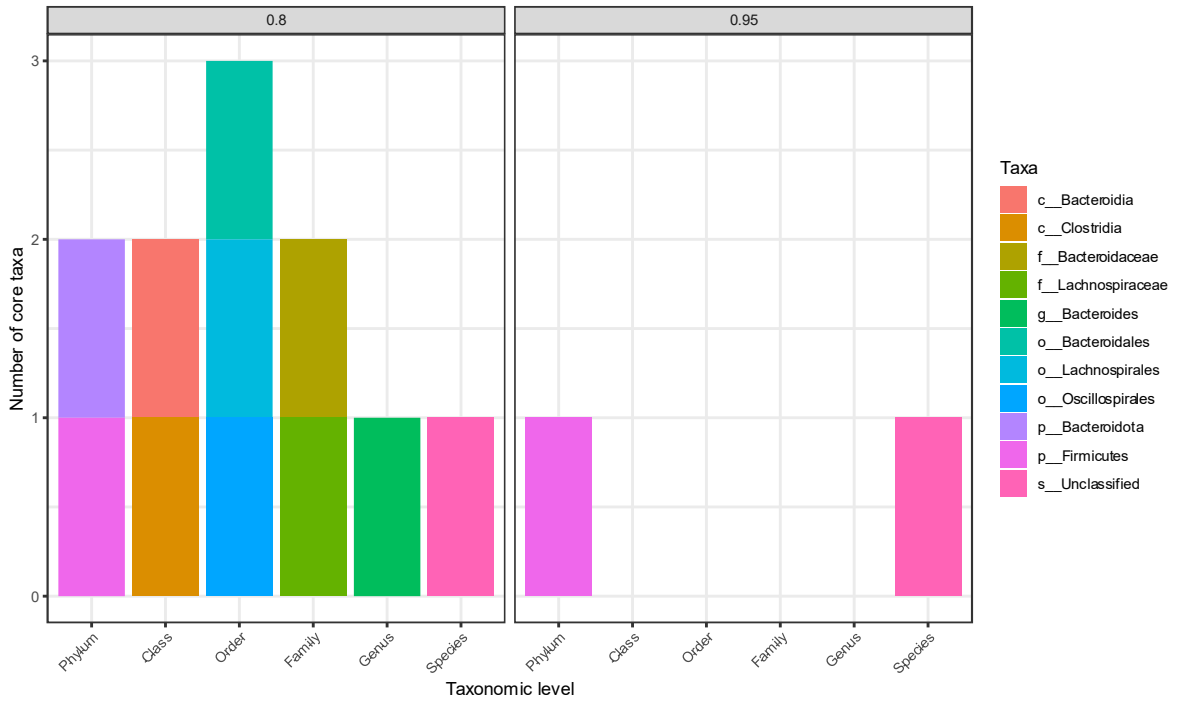


Figure 4.42 Taxonomic composition of core microbiome members at different prevalence thresholds. Left panel shows taxa meeting the 80% prevalence threshold, while right panel shows taxa meeting the more stringent 95% threshold. Colours indicate different taxonomic groups, with prefixes corresponding to taxonomic levels.

A detailed examination of the taxonomic composition across prevalence thresholds (Figure 4.42) further clarified the core microbiome structure. At the 80% threshold, two phyla emerged as core taxa: Bacteroidota and Firmicutes. Within Bacteroidota, core members included class Bacteroidia, order Bacteroidales, family Bacteroidaceae, and genus Bacteroides. Within Firmicutes, the class Clostridia, orders Lachnospirales and Oscillospirales, and family Lachnospiraceae were identified as core taxa. At the species level, a category labelled “unclassified species” was identified, representing an aggregate of sequences that could not be assigned clear species-level taxonomy rather than a distinct microbial species.

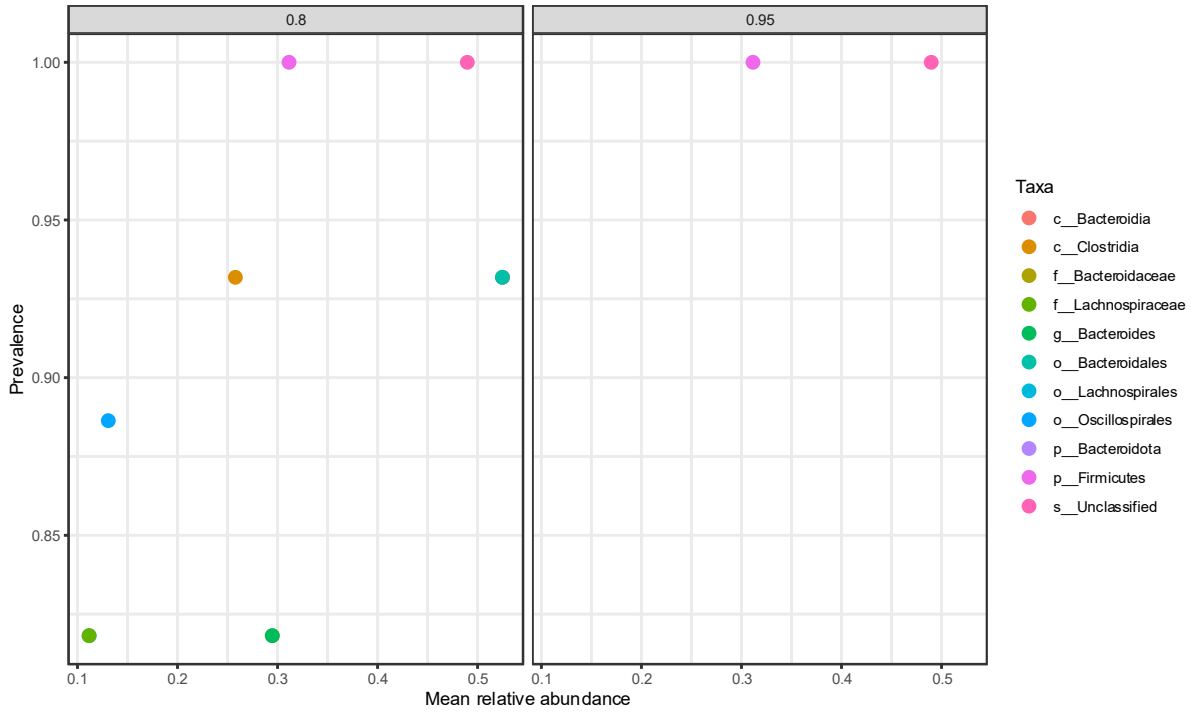


Figure 4.43 Relationship between prevalence and mean relative abundance of core taxa at different taxonomic levels. Data points are coloured by taxonomic classification, with separate panels showing taxa meeting 80% and 95% prevalence thresholds. The x-axis represents mean relative abundance, and the y-axis shows the prevalence of each taxon.

The abundance-prevalence relationships revealed interesting patterns across taxonomic levels (Figure 4.43). At both prevalence thresholds, the Firmicutes phylum and the aggregated "unclassified species" demonstrated the highest prevalence (100%), with the latter group also showing the highest mean abundance (~0.5). This high prevalence and abundance of unclassified sequences reflects the cumulative contribution of multiple taxa without species-level identification rather than a single abundant taxon. The Bacteroides genus and Lachnospiraceae family showed the lowest prevalence (~82%) among visible core taxa, with the latter maintaining relatively lower abundance (0.1). The Clostridia class showed intermediate prevalence (93%), with mean relative abundance approximately 0.25. Two additional core taxa (Bacteroidia class and Lachnospirales order) were not visible in the plot due to mean relative abundances below 0.1.

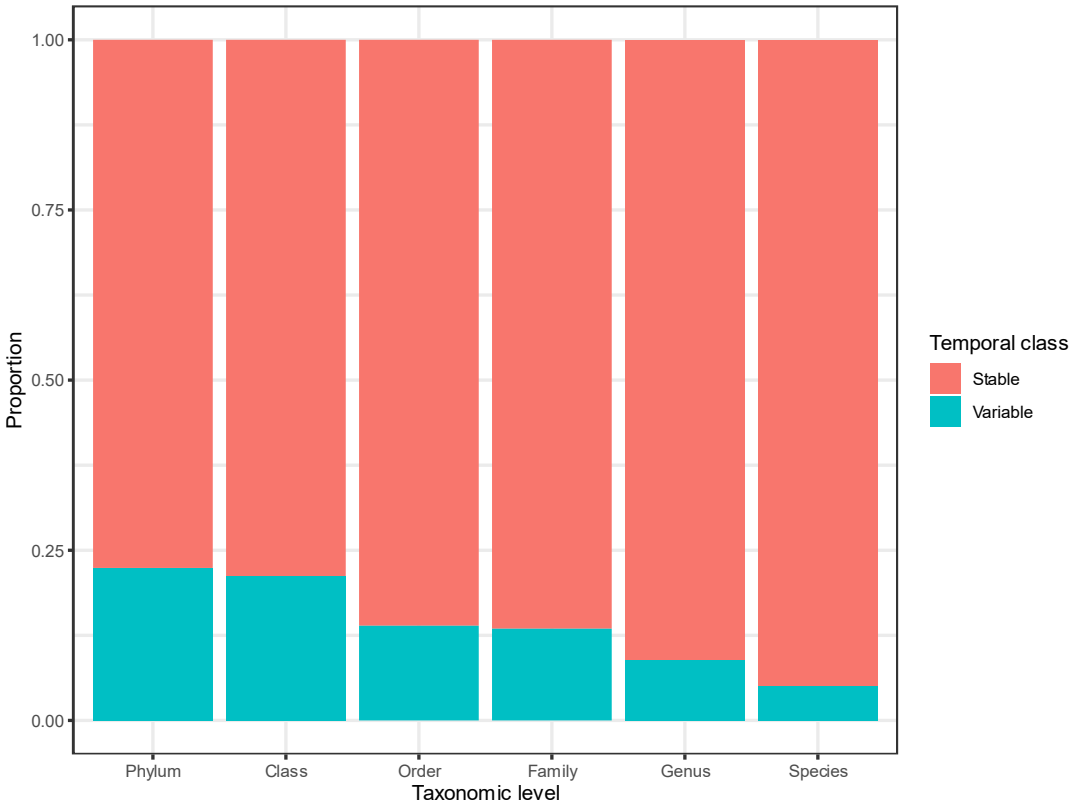


Figure 4.44 Proportion of stable and variable taxa across taxonomic levels. The stacked bar plot shows the relative proportion of stable (95% prevalence, pink) versus variable (80% prevalence, blue) taxa at each taxonomic rank from phylum to species level.

Analysis of taxonomic stability specifically within identified core taxa across hierarchical levels (Figure 4.44) revealed a clear trend. At higher taxonomic ranks (phylum and class), approximately 20% of core taxa were classified as variable, whereas this proportion progressively decreased at finer taxonomic ranks (order and family: ~13%; genus: ~8%; species: ~5%). Conversely, the proportion of stable core taxa increased from roughly 80% at higher ranks to around 95% at the species level. This hierarchical stability pattern aligns with ecological expectations, as broader taxonomic levels may aggregate multiple taxa with different ecological behaviours, resulting in greater apparent variability. At finer resolutions (species level), taxa represent narrower ecological niches, thus exhibiting greater temporal stability.

The temporal analysis of gut microbiota across n = 22 subjects revealed distinct patterns of microbial stability between baseline and follow-up samples (Supplementary Data S4.12l). Analysis across taxonomic ranks showed increasing variability from higher to lower taxonomic levels, with species-level classifications demonstrating the most dynamic temporal patterns (Figure 4.92). At the species level, several abundant gut microbes showed notably different temporal behaviours across individuals. *Bacteroides dorei* exhibited the most striking variation with log2 fold changes ranging from -10 to +10 across different subjects. Other species

including *Bifidobacterium pseudocatenulatum, B. adolescentis* and *Akkermansia muciniphila* also showed substantial subject-specific temporal variation.

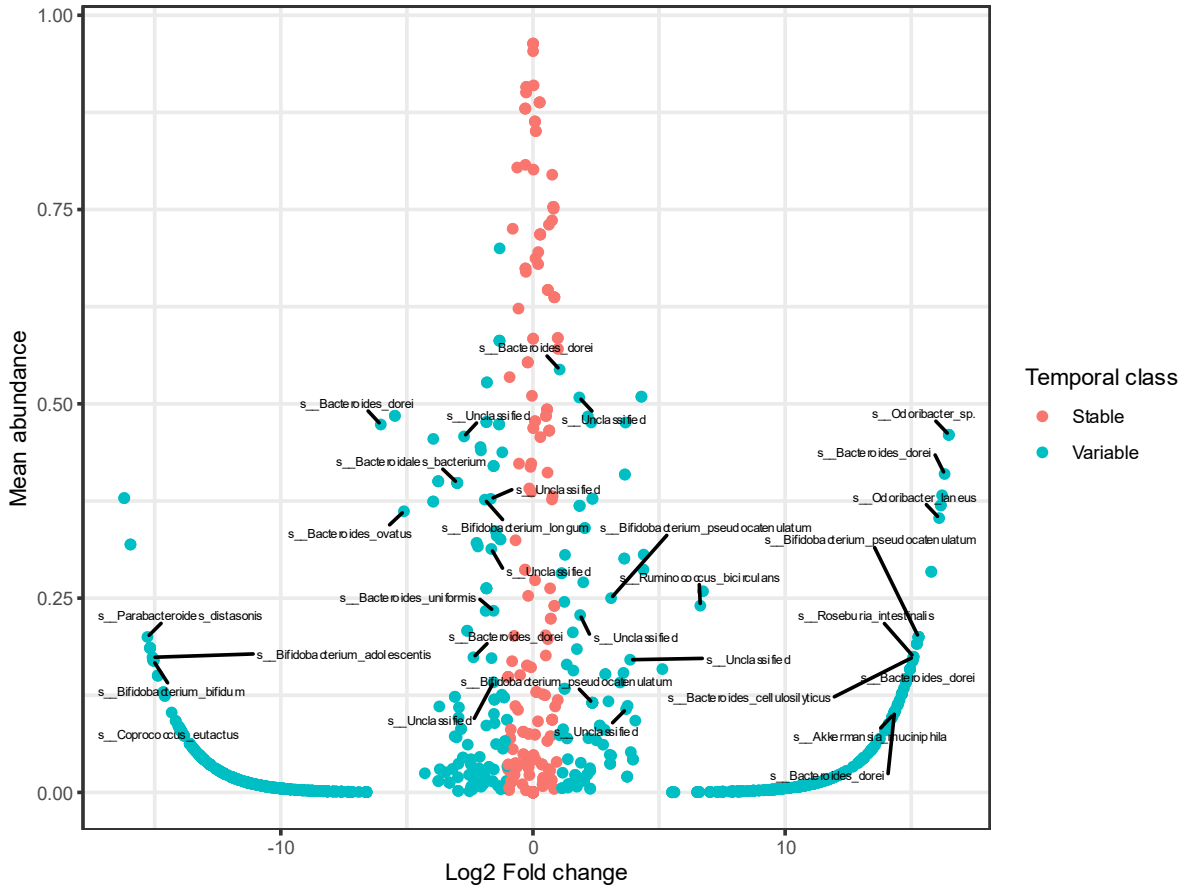


Figure 4.45 Volcano plot depicting individual-level temporal dynamics of microbial taxa between baseline and follow-up samples (*n* = 22 subjects). Each point represents a single taxon in a single subject, with mean abundance (y-axis) versus log2 fold change (x-axis). Points are coloured by temporal stability classification (pink: stable, log2 fold change ≤ 1; blue: variable log2 fold change > 1). Labels indicate species with mean abundance >0.1 and |log2FC| > 1. Multiple occurrences of the same species (e.g. *Bacteroides dorei*) reflect different temporal behaviours across individuals, demonstrating subject-specific dynamics. The plot reveals both consistent patterns (clustering of stable taxa near centre) and extreme variation (taxa showing >10-fold changes in some individuals).

The relationship between abundance and temporal stability showed that while most variable taxa had mean abundances below 0.25, some highly abundant species (mean abundance > 0.5) also displayed significant temporal fluctuations (Figure 4.45). Multiple unclassified taxa across different abundance levels showed significant temporal variation, with fold changes comparable to those of well-characterized species. The same species often showed different temporal patterns across individuals, with some subjects showing strong positive fold changes while others showed negative fold changes for the same organism. These contrasting patterns were particularly evident among the *Bacteroides* and *Bifidobacterium* genera, where individual species demonstrated both stability and variability depending on the subject.

ii. Functional analysis

Building upon the taxonomic temporal patterns observed, investigations on whether similar dynamics were present at the functional level by examining KEGG Ortholog (KO) annotations were performed. This analysis aimed to determine if the observed species-level variations translated to functional changes, testing the hypothesis of potential functional redundancy in the gut microbiome. These parallelled approaches to the taxonomic analysis, first identifying core functions at different prevalence thresholds (80% and 95%), then examining their temporal stability. This functional assessment provides insight into whether the considerable temporal variation observed in taxonomic composition is reflected in, or buffered at, the functional level.

To refine the functional analysis approach, the distribution of KO abundances and fold changes across all 128,546 function-subject pairs were examined first. The data revealed a highly skewed abundance distribution with a median of 1.3×10^{-6} and mean of 1.7×10^{-4}, suggesting most functions occur at very low abundances. The maximum abundance observed was 8.3×10^{-3}, providing a reference point for defining "high-abundance" functions. Fold changes ranged from -6.7 to 7.2 with a median near zero and mean of $8.0 \times 10^{-3,}$ indicating that most functions remain relatively stable over time (Supplementary Data S4.13C).

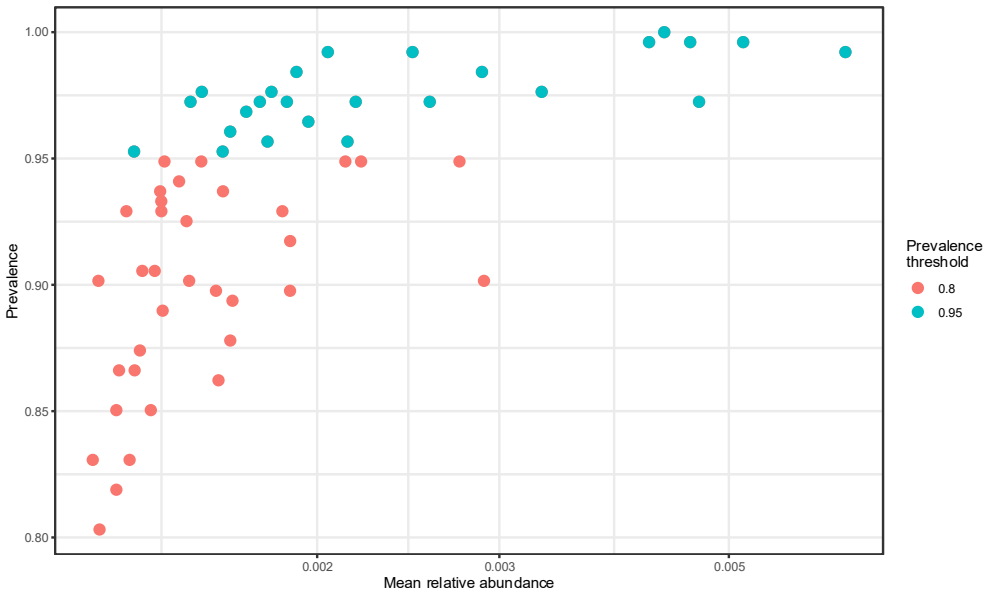


Figure 4.46 Distribution of core KO functions by prevalence and abundance. Each point represents a unique KO function, with mean relative abundance (x-axis) versus prevalence (y-axis) across subjects. Points are coloured by prevalence threshold (pink: 80%; blue: 95%). Functions present in higher proportions of subjects (95%) generally show higher mean abundances.

Stringent thresholds set to > 0.1% mean abundance (1.0×10^{-3}) and log2 fold change (log2FC) > 4 set focus on the most dramatically varying functions. These refined criteria identified key functions showing substantial temporal dynamics, providing a focused set of KOs for detailed investigation. Analysis of functional stability across each individual revealed distinct patterns between core and variable KO functions (Figure 4.46). At the 95% prevalence threshold, *n* = 105,044 (~82%) KO functions were identified as core (stable) components, while *n* = 23,502 (~18%) functions met the 80% (variable) threshold (Supplementary Data S4.13C).

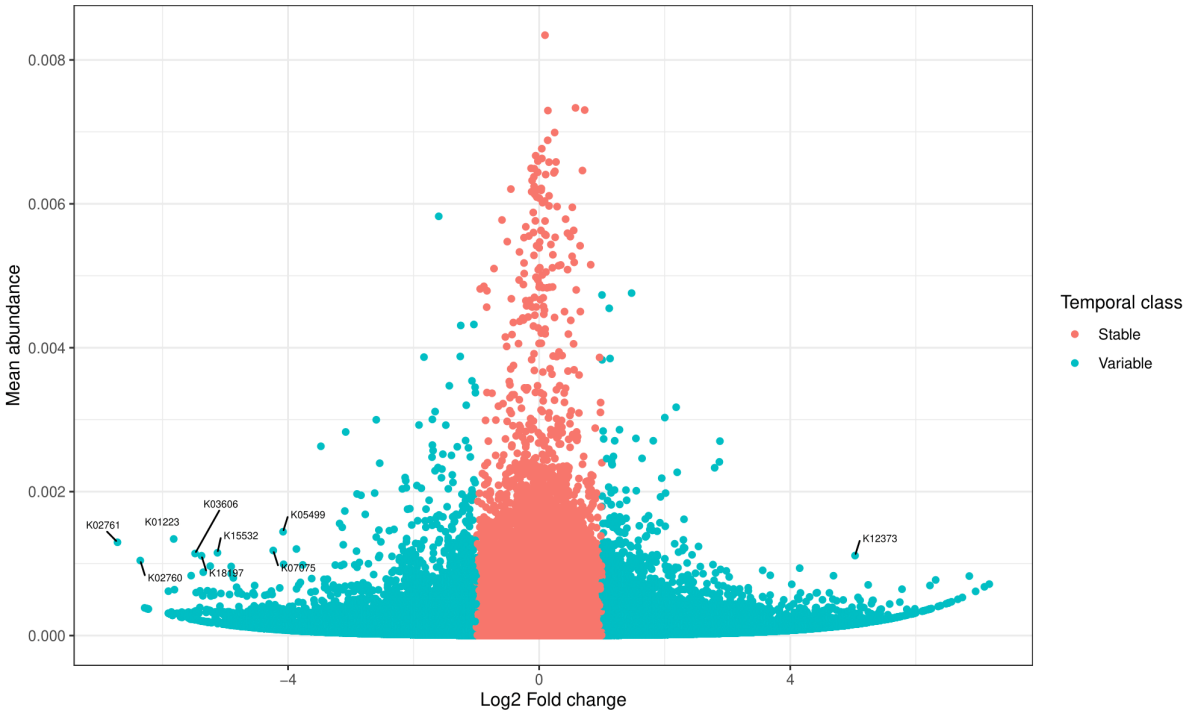


Figure 4.47 Volcano plot of KO functional temporal dynamics. Mean abundance versus Log2 fold change (log2FC) between baseline and follow-up samples for all KO functions. Points are coloured by temporal stability classification (pink: stable log2FC ≤ 4; blue: variable log2FC > 4). Labelled points indicate highly abundant (> 0.1%) and highly variable functions (log2FC > 4).

Core functions at both thresholds showed a positive correlation between prevalence and mean abundance (Figure 4.46), with higher-prevalence functions typically maintaining higher abundances. To identify the most relevant stable functions from the stable KO functions meeting, increasingly stringent filtering criteria to include only KO functions with mean abundance ≥ 0.5% and minimal temporal variation (log2FC between ± 1.0×10^{-2}), resulted in *n* = 3 highly stable pathways (*K02004; K01990; K06147*). This extremely narrow log2FC range indicates remarkable temporal consistency, with abundance variations of less than 1% (~0.7%) between timepoints, suggesting these functions represent fundamental metabolic processes that were tightly regulated and essentially invariant over time in the gut microbiome.

All three core functions belong to the ATP-binding cassette (ABC) transporter family. *K02004* ("putative ABC transport system permease protein") exhibited temporal stability (log2FC = 1.7×10^{-3}) with a relative abundance of 0.6%. *K01990* ("ABC-2 type transport system ATP-binding protein") demonstrated exceptional temporal consistency (log2FC = 7.9×10^{-3}) with a mean abundance of 0.5%. *K06147* ("ATP-binding cassette, subfamily B, bacterial") was the most stable core function (log2FC = 3.1×10^{-4}), maintaining a consistent mean abundance of 0.5%. The predominance of *ABC transporters* among the most stable core functions reflects the fundamental importance of nutrient acquisition and molecular transport systems in the gut microbiome.

Temporal stability analysis identified nine KO functions exhibiting both high abundance (> 0.1%) and substantial temporal variation (log2FC > 4) (Figure 4.47). These functions spanned diverse metabolic processes, highlighted in Table 4.4.

Table 4.4 High-abundance KO functions exhibiting substantial temporal variation

KO ID	Function description	Enzyme classification	Primary metabolic role
K01223	6-phospho-beta-glucosidase (bglA)	EC:3.2.1.86	Glycolysis / Gluconeogenesis; Starch and sucrose metabolism
K02760	Cellobiose PTS system EIIB component (celA. chbB)	EC:2.7.1.196 2.7.1.205	Starch and sucrose metabolism; Phosphotransferase system (PTS)
K02761	Cellobiose PTS system EIIC component (celB. chbC)	-	Starch and sucrose metabolism; Phosphotransferase system (PTS)
K12373	Hexosaminidase (HEXA_B)	EC:3.2.1.52	Other glycan degradation; Various types of N-glycan biosynthesis; Amino sugar and nucleotide sugar metabolism; Glycosaminoglycan degradation; Sphingolipid metabolism; Glycosphingolipid biosynthesis - globo and isoglobo series; Glycosphingolipid biosynthesis - ganglio series; Metabolic pathways; Lysosome
K03606	Putative colanic acid biosynthesis UDP-glucose lipid carrier transferase (wcaJ)	-	Exopolysaccharide biosynthesis; Biofilm formation - Vibrio cholerae
K15532	Unsaturated rhamnogalacturonyl hydrolase (yteR. yesR)	EC:3.2.1.172	Unclassified: metabolism; Enzymes with EC numbers; Hydrolases - Glycosylases
K18197	Rhamnogalacturonan endolyase (yesW)	EC:4.2.2.23	Unclassified: metabolism; Enzymes with EC numbers; Lyases - Carbon-oxygen lyases
K07075	Uncharacterized protein	-	Poorly characterized; Function unknown
K05499	LacI family transcriptional regulator. repressor for deo operon (cytR)	-	Transcription factors; Prokaryotic type; Helix-turn-helix; LacI family

Functional characterization of the identified KOs (Table 4.4) revealed three primary categories: transport/metabolism (44%), cell surface modification (22%), and regulatory functions (11%). Key components included "rhamnogalacturonyl hydrolase" and "hexosaminidase" operating within transport networks, "cytR" regulator controlling nucleoside

metabolism, and “wcaJ” involved in biofilm formation. The remaining 11% comprised uncharacterized proteins, suggesting potential novel metabolic functions.

This longitudinal analysis revealed contrasting stability patterns: dynamic taxonomic composition with only two universal core taxa at 95% prevalence, yet functionally stable profiles with 82% of KO functions maintained. The nine highly variable, abundant functions primarily involved carbohydrate metabolism and transport systems, suggesting functional redundancy despite taxonomic volatility. The snapshot of these baseline dynamics in an urban professional South African population provide context for comparative analyses with urban-rural African and Western populations, with the potential to elucidate urbanization's impact on microbiome composition.

4.7. Comparative analysis with other cohort populations

To contextualize the findings within the broader human gut microbiome landscape, the urban professional South African population (SAN) was compared with two distinct populations: the American Gut Project (AGP) and the AWIGEN1 (AWI) Microbiome study. These cohorts represent industrialized Western and urban-rural African populations respectively, offering complementary reference points for potentially revealing how urbanization influences gut microbiome composition and function. The comparative analysis utilized complementary approaches, beginning with recovery and analysis of metagenome-assembled genomes (MAGs) across all cohorts, followed by integration of data from different sequencing platforms – 16S rRNA long-read sequencing for the SAN population and shotgun metagenomic sequencing for AGP and AWI cohorts. This tiered approach allowed for direct genome-level comparisons while maximizing the statistical power of the larger 16S dataset. The MAG analysis revealed distinct taxonomic signatures across cohorts, with the SAN population demonstrating a notably higher proportion of Actinomycetota (~15%) compared to AGP and AWI cohorts (<5%), while maintaining comparable genome quality metrics across populations. Species-level analysis showed the strongest cohort-specific associations (Cramer's V = 0.619), suggesting finer taxonomic resolution reveals more population-specific patterns than broader classifications.

While methodological differences between platforms may influence some observations, the analytical approach was guided by several practical considerations. First, the substantially larger sample size of the SAN 16S long-read dataset (n = 221) compared to the shotgun metagenomics subset (n = 49) provided greater statistical power to detect biologically relevant patterns. Second, the application of PacBio full-length 16S sequencing represented a primary focus of the study design, allowing more comprehensive taxonomic resolution than traditional short-read 16S approaches.

Cross-platform comparisons examining community diversity, taxonomic composition, genome novelty and functional capabilities were performed (Supplementary Data S4.14-18). The comparative findings were carefully interpreted in light of these methodological differences, particularly where they involved low-abundance taxa or functional predictions, which previous concordance analysis identified as areas where shotgun metagenomics demonstrates higher sensitivity. Despite these limitations, this systematic comparison provided valuable insights into how the SAN population gut microbiome relates to both Western and other South African cohorts, while maintaining appropriate caution regarding technology-driven differences.

4.7.1. MAG quality and novelty analysis

i. Quality analysis

Analysis of *n* = 887 medium to high-quality bacterial genomes across the populations revealed distinctive genomic features and potentially novel lineages specific to the SAN population. The following sections examine the quality distribution of these MAGs across cohorts, their taxonomic classifications, and highlight key genomic elements unique to each population. From the shotgun metagenomic data, *n* = 881 MAGs meeting quality thresholds across the three cohorts were recovered. The AWI cohort yielded the highest number of MAGs (*n* = 535), followed by SAN (*n* = 258) and AGP (*n* = 88). Each cohort showed similar proportions of high-quality (≥ 90% completeness, ≤ 5% contamination) and medium-quality (≥ 50% completeness, ≤ 10% contamination) MAGs (Figure 4.48).

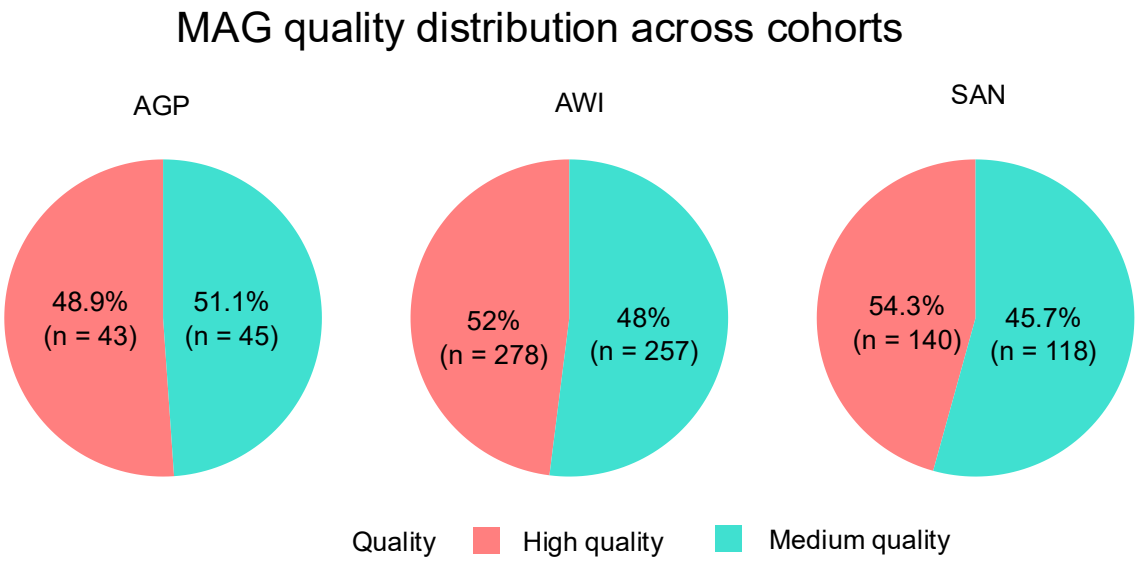


Figure 4.48 Distribution of metagenome-assembled genome (MAG) quality across three human gut microbiome cohorts (AGP, AWI, and SAN). MAGs were classified as high-quality (≥ 90% completeness, ≤ 5% contamination) or medium-quality (≥ 50% completeness, ≤ 10% contamination) based on CheckM quality metrics. Values shown represent the percentage and absolute count (*n*) of MAGs in each quality category per cohort.

The proportion of high-quality MAGs remained relatively consistent across all three cohorts, ranging from 45.7% to 51.1%, despite substantial differences in cohort sizes. AWI contributed the largest number of MAGs (*n* = 535), followed by SAN (*n* = 258) and AGP (*n* = 88). The comparable quality distributions across cohorts, confirmed by statistical testing (Fisher's exact test p-value > 5.0 x 10^{-2}), suggest that methodological biases were minimal and that downstream comparative analyses could be continued with confidence (Table 4.5).

Table 4.5 Statistical comparison of MAG quality distributions between cohorts.

Statistical test	p-value
Overall tests	
Quality distribution across cohorts (Fisher's)	0.7
Proportion of high vs medium quality MAGs across cohorts	0.7
Pairwise cohort comparisons	
AGP vs AWI	0.7
AGP vs SAN	0.5
AWI vs SAN	0.6

ii. Taxonomic analysis

Taxonomic analysis of the MAGs revealed *n* = 12 distinct phyla across the cohorts (Figure 4.49). The most abundant phyla were Bacillota_A, Bacteroidota, Actinomycetota and Pseudomonadota, collectively representing over 80% of MAGs in each cohort. The SAN population showed a notably higher proportion of Actinomycetota (~15%) compared to AGP and AWI cohorts (< 5%).

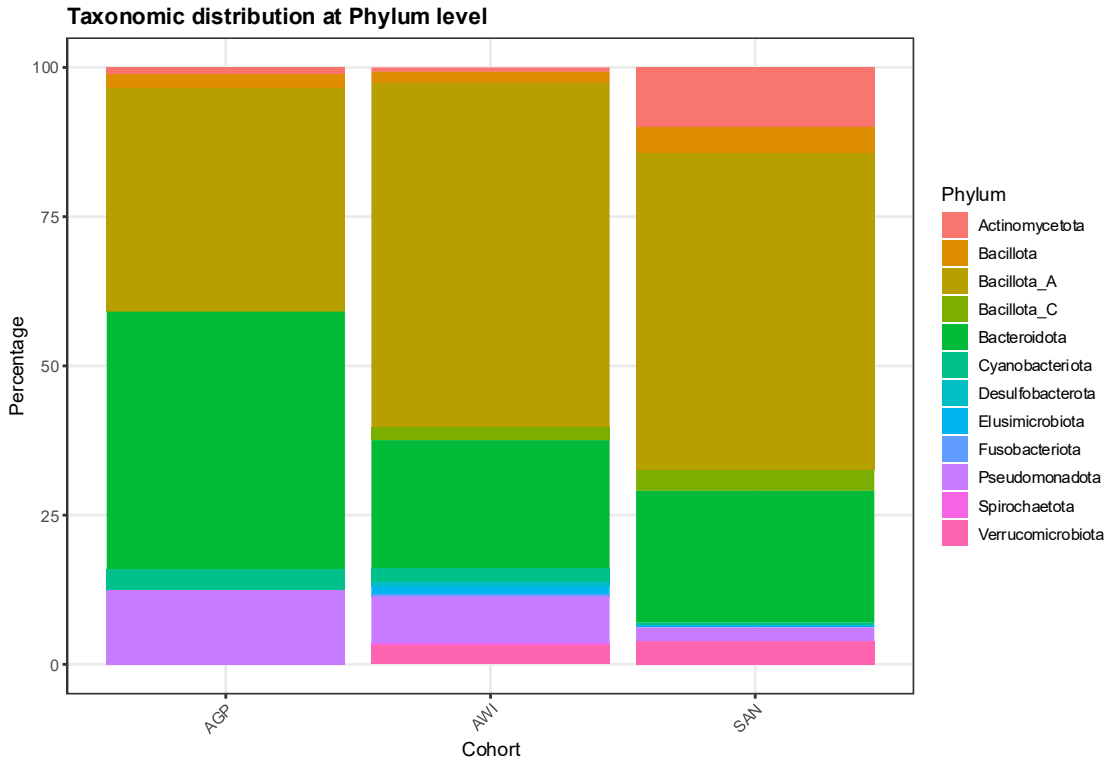


Figure 4.49 Taxonomic composition at phylum level for metagenome-assembled genomes (MAGs) recovered from three human gut microbiome cohorts – American Gut Project (AGP), AWIGEN1 (AWI) and the urban professional South African population (SAN). Classification was performed using Genome Taxonomy Database-Toolkit (GTDB-Tk) taxonomy assignments based on placement in the GTD reference tree.

At the genus level, *n* = 176 unique genera with *n* = 31 genera present at > 1% relative abundance in at least one cohort (Figure 4.50) were identified. Several uncharacterized genera particularly within the Bacillota_A phylum were prevalent across all cohorts. The most abundant genera included *Bacteroides, Faecalibacterium, Alistipes*, and members of *Ruminococcus*, collectively dominating the microbiome profiles. Notable differences were observed between cohorts, with the SAN population showing higher proportions of *Bifidobacterium* compared to AGP and AWI cohorts.

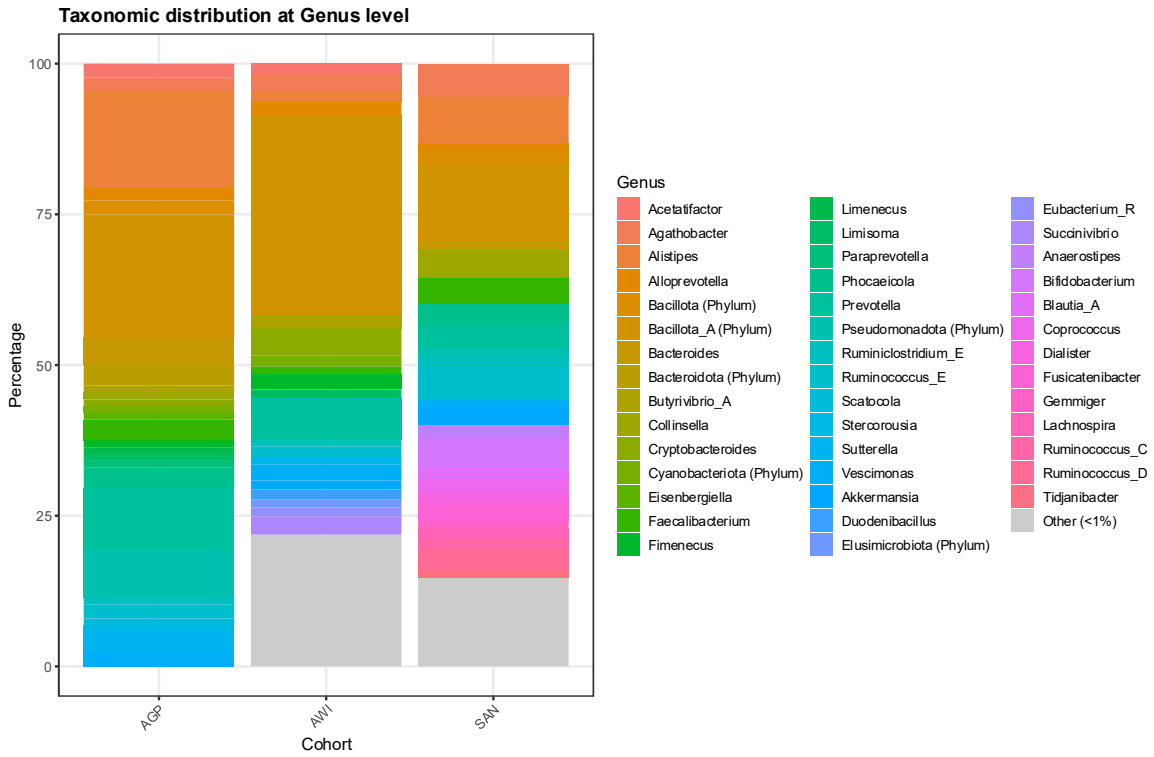


Figure 4.50 Taxonomic composition at genus level for metagenome-assembled genomes (MAGs) recovered from three human gut microbiome cohorts (AGP, AWI, and SAN). Only genera representing > 1% relative abundance in at least one cohort are shown individually, while those < 1% are grouped as "Other" (grey). Uncharacterized genera are designated by their higher taxonomic classification (indicated with "Phylum"). All classifications were performed using GTDB-Tk taxonomy assignments based on placement in the GTD reference tree.

Taxonomic analysis across different classification levels revealed distinct microbial community structures among the three cohorts, as visualized in the stacked bar plots (Figure 4.51) and supported by statistical analysis (Fisher's exact test, $p < 1 \times 10^{-4}$). The increasing Cramer's V values from phylum (0.290) to species (0.619) level indicated stronger cohort-specific associations at finer taxonomic resolutions (Table 4.6).

While the stacked bar visualizations provide a comprehensive overview of taxonomic distributions, the accompanying ranked table format offered precise quantification of the top 5 taxa at each level, enabling clearer interpretation of the complex microbial communities. Looking at the hierarchical relationships through the taxonomic levels, several key patterns emerged: The phylum-level dominance of Bacteroidota in AGP (43.2%) cascaded through class Bacteroidia and order Bacteroidales, maintaining similar proportions. Similarly, the Bacillota_A predominance in AWI and SAN (57.8% and 53.1%) flowed through to Clostridia at class level, followed by Oscillospirales becoming the dominant order (30.7% and 27.1% respectively).

The class-to-order transition revealed interesting shifts, particularly in the AWI cohort where Clostridia (57.8%) diversified into multiple orders, with Oscillospirales (30.7%) and Lachnospirales (10.5%) being most prominent. The family level introduced greater complexity, with Bacteroidaceae, Bacillota_A-affiliated families, and Lachnospiraceae showing varying distributions across cohorts. A notable feature in the SAN population was the consistent presence of Actinomycetota-related taxa across levels, from phylum (10.1%) through class Actinomycetia (4.7%) to genus *Bifidobacterium* (4.7%), suggesting a stable and cohort-specific microbial signature.

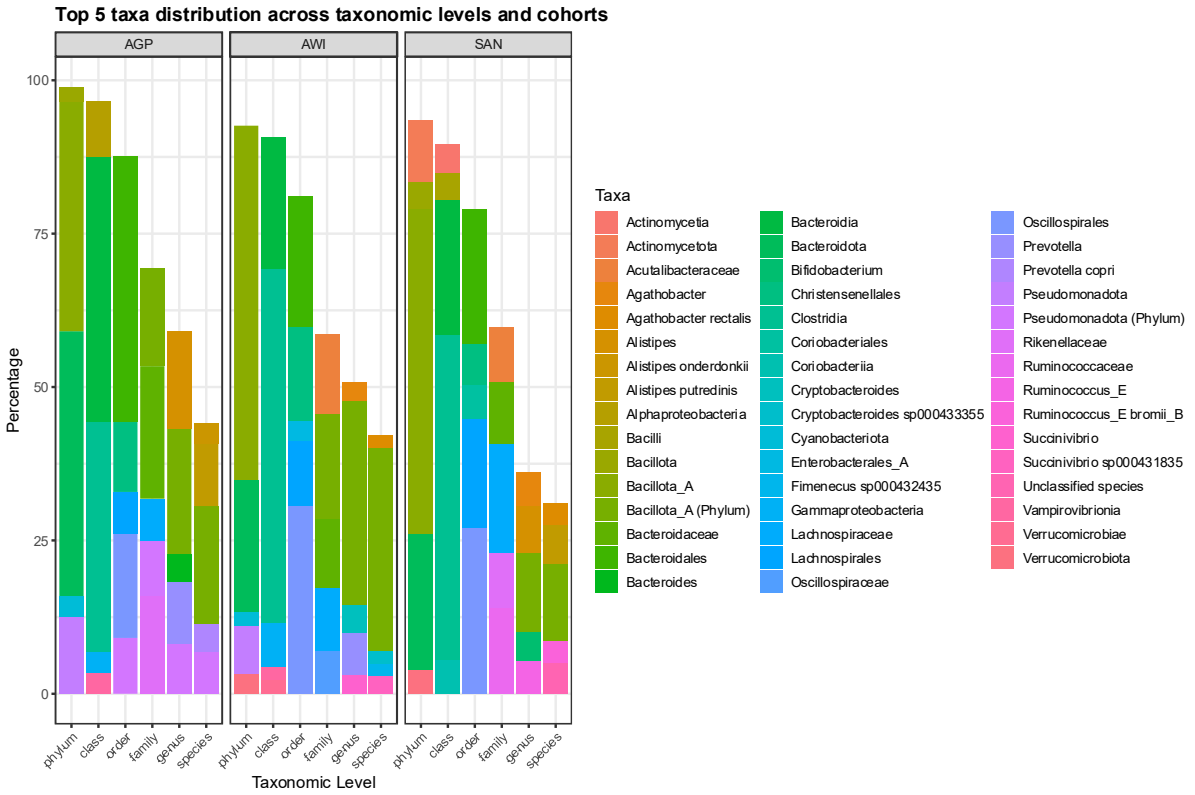


Figure 4.51 Distribution of the five most abundant taxa at each taxonomic level (phylum to species) across three human gut microbiome cohorts (AGP, AWI, and SAN). The relative abundance of each taxon is shown as a percentage of the total MAGs within each cohort. Uncharacterized genera with alphanumeric identifiers are grouped under their respective phyla. Classification was performed using GTDB-Tk taxonomy assignments based on placement in the Genome Taxonomy Database reference tree.

In addition, the table reveals a substantial proportion of “unclassified” species particularly in the SAN population (ranked 3rd at 5%), suggesting potential novel taxa, while specific strains like *Succinivibrio sp000431835* (3%) and *Cryptobacteroides sp000433355* (2.1%) were identified as key members in the AWI cohort. After examining the taxonomic distribution patterns across cohorts, we next investigated the novelty of the recovered MAGs by assessing their similarity to known reference genomes, providing insight into potentially undescribed taxa within these communities.

Table 4.6 Top 5 taxa distribution across taxonomic ranks in three human gut microbiome cohorts.

Top 5 taxa distribution across cohorts

Rank	AGP (n, %)	AWI (n, %)	SAN (n, %)
Phylum (p-value = <1e-04, Cramer's V (effect size) = 0.290)			
1	Bacteroidota (38, 43.2%)	Bacillota_A (309, 57.8%)	Bacillota_A (137, 53.1%)
2	Bacillota_A (33, 37.5%)	Bacteroidota (115, 21.5%)	Bacteroidota (57, 22.1%)
3	Pseudomonadota (11, 12.5%)	Pseudomonadota (42, 7.9%)	Actinomycetota (26, 10.1%)
4	Cyanobacteriota (3, 3.4%)	Verrucomicrobiota (17, 3.2%)	Bacillota (11, 4.3%)
5	Bacillota (2, 2.3%)	Cyanobacteriota (12, 2.2%)	Verrucomicrobiota (10, 3.9%)
Class (p-value = <1e-04, Cramer's V (effect size) = 0.364)			
1	Bacteroidia (38, 43.2%)	Clostridia (309, 57.8%)	Clostridia (137, 53.1%)
2	Clostridia (33, 37.5%)	Bacteroidia (115, 21.5%)	Bacteroidia (57, 22.1%)
3	Alphaproteobacteria (8, 9.1%)	Gammaproteobacteria (38, 7.1%)	Coriobacteriia (14, 5.4%)
4	Gammaproteobacteria (3, 3.4%)	Vampirovibrionia (12, 2.2%)	Actinomycetia (12, 4.7%)
5	Vampirovibrionia (3, 3.4%)	Verrucomicrobiae (12, 2.2%)	Bacilli (11, 4.3%)
Order (p-value = <1e-04, Cramer's V (effect size) = 0.327)			
1	Bacteroidales (38, 43.2%)	Oscillospirales (164, 30.7%)	Oscillospirales (70, 27.1%)
2	Oscillospirales (15, 17%)	Bacteroidales (115, 21.5%)	Bacteroidales (57, 22.1%)
3	Christensenellales (10, 11.4%)	Christensenellales (82, 15.3%)	Lachnospirales (46, 17.8%)
4	Pseudomonadota (Phylum) (8, 9.1%)	Lachnospirales (56, 10.5%)	Christensenellales (17, 6.6%)
5	Lachnospirales (6, 6.8%)	Enterobacterales_A (17, 3.2%)	Coriobacteriales (14, 5.4%)
Family (p-value = <1e-04, Cramer's V (effect size) = 0.540)			
1	Bacteroidaceae (19, 21.6%)	Bacillota_A (Phylum) (92, 17.2%)	Lachnospiraceae (46, 17.8%)
2	Bacillota_A (Phylum) (14, 15.9%)	Acutalibacteraceae (69, 12.9%)	Ruminococcaceae (36, 14%)
3	Rikenellaceae (14, 15.9%)	Bacteroidaceae (60, 11.2%)	Bacteroidaceae (26, 10.1%)
4	Pseudomonadota (Phylum) (8, 9.1%)	Lachnospiraceae (55, 10.3%)	Acutalibacteraceae (23, 8.9%)
5	Lachnospiraceae (6, 6.8%)	Oscillospiraceae (37, 6.9%)	Rikenellaceae (23, 8.9%)
Genus (p-value = <1e-04, Cramer's V (effect size) = 0.596)			
1	Bacillota_A (Phylum) (18, 20.5%)	Bacillota_A (Phylum) (178, 33.3%)	Bacillota_A (Phylum) (33, 12.8%)
2	Alistipes (14, 15.9%)	Prevotella (37, 6.9%)	Alistipes (20, 7.8%)
3	Prevotella (9, 10.2%)	Cryptobacteroides (24, 4.5%)	Agathobacter (14, 5.4%)
4	Pseudomonadota (Phylum) (7, 8%)	Agathobacter (16, 3%)	Ruminococcus_E (14, 5.4%)
5	Bacteroides (4, 4.5%)	Succinivibrio (16, 3%)	Bifidobacterium (12, 4.7%)
Species (p-value = <1e-04, Cramer's V (effect size) = 0.619)			
1	Bacillota_A (Phylum) (17, 19.3%)	Bacillota_A (Phylum) (177, 33.1%)	Bacillota_A (Phylum) (33, 12.8%)
2	Alistipes putredinis (9, 10.2%)	Succinivibrio sp000431835 (16, 3%)	Alistipes putredinis (16, 6.2%)
3	Pseudomonadota (Phylum) (6, 6.8%)	Agathobacter rectalis (11, 2.1%)	Unclassified species (13, 5%)
4	Prevotella copri (4, 4.5%)	Cryptobacteroides sp000433355 (11, 2.1%)	Agathobacter rectalis (9, 3.5%)
5	Alistipes onderdonkii (3, 3.4%)	Fimenecus sp000432435 (10, 1.9%)	Ruminococcus_E bromii_B (9, 3.5%)

iii. Novelty analysis

To assess potential taxonomic novelty, comparison of the high-quality MAGs against the GTDB reference database was performed (Supplementary S4.17). Fast Alignment-free Sequence-based Average Nucleotide Identity (FastANI) analysis of bacterial MAGs revealed consistently high sequence identity with known reference genomes, with all comparisons exceeding the 95% ANI species delineation threshold (Figure 4.52A). The alignment fraction (AF) versus sequence identity plots demonstrated robust genomic matches (AF > 0.6) across all three cohorts (Figure 4.52B), with all cohort AGP MAGs showing alignment fractions between 0.6 and 0.9. The distribution of FastANI values peaked between 97 – 98% sequence identity, indicating strong concordance with previously characterized bacterial species.

Similarly, archaeal MAGs demonstrated high sequence similarity to known reference genomes, with FastANI values ranging from 97 – 98.5% (Figure 4.53A). All archaeal MAGs exhibited alignment fractions above 0.6 across the three cohorts (Figure 4.53B). Notably, the archaeal representation was considerably lower than bacterial MAGs, with only five archaeal genomes meeting the high quality thresholds.

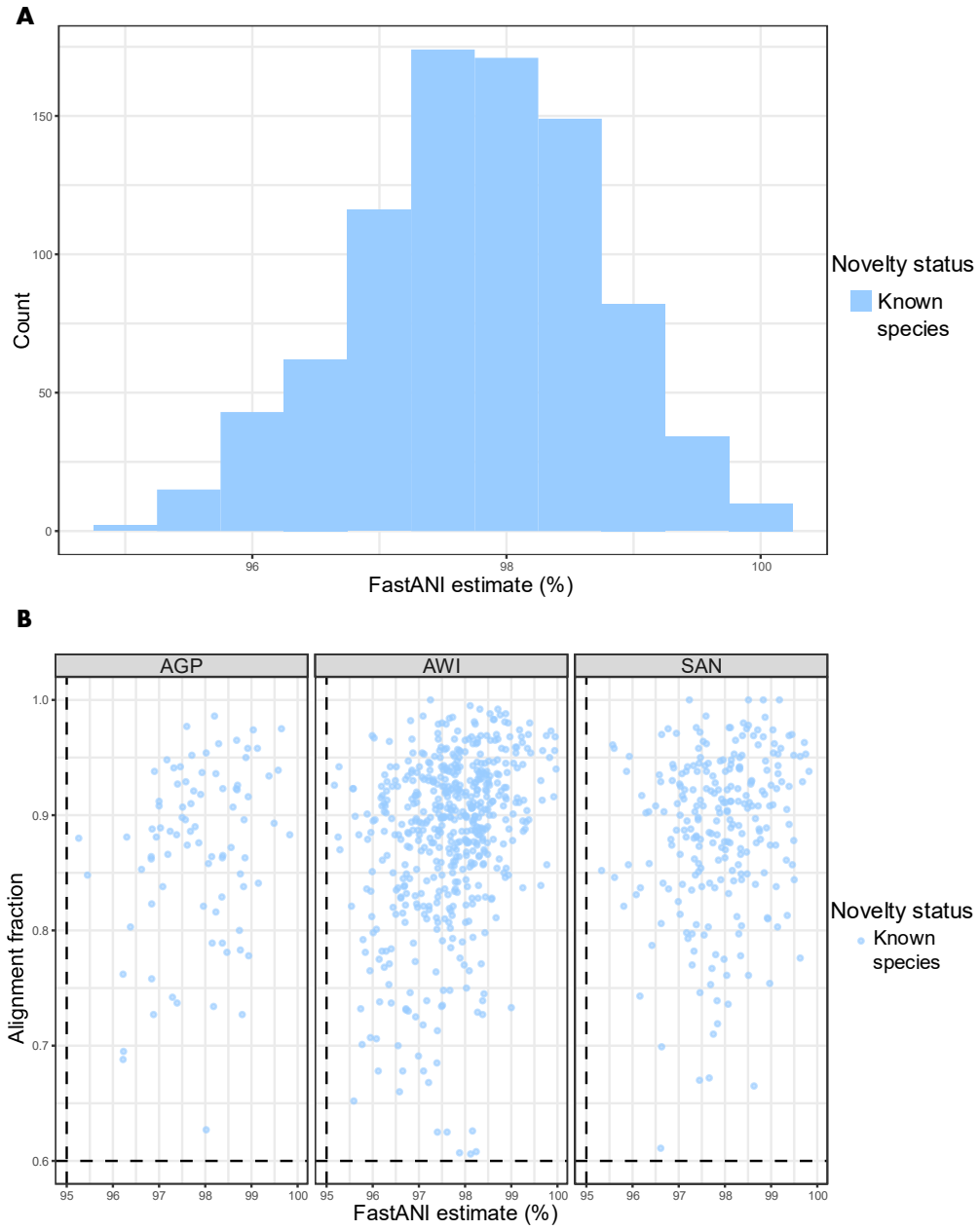


Figure 4.52 Novelty assessment of bacterial MAGs against GTDB reference genomes. (A) Distribution of FastANI sequence identity values for high-quality bacterial MAGs (> 90% completion, < 5% contamination). All comparisons exceeded the species delineation threshold (95% ANI, vertical dashed line) with minimum alignment fraction of 0.6. (B) FastANI alignment fraction versus sequence identity (FastANI estimate) plots across AGP, AWI, and SAN populations. Each point represents a pairwise comparison between a query MAG and its closest reference genome, faceted by each cohort. Horizontal and vertical dashed lines indicate thresholds for reliable comparison (AF = 0.6) and species delineation (ANI = 95%), respectively.

Comparison of GTDB-Tk classifications with FastANI taxonomy assignments showed complete concordance at the species level for both bacterial and archaeal MAGs, further supporting the robust taxonomic placement of the recovered genomes within known species. Having established the taxonomic compositions and confirmed the identity of the MAGs

against known reference genomes across all three cohorts, whether the functional potential captured by shotgun metagenomics differed from that inferred from 16S long-read data was investigated. This comparison is particularly relevant given the distinct taxonomic patterns observed, including the higher proportion of Bacillota_A in both AWI and SAN compared to AGP and allows us to assess whether methodological differences between sequencing approaches impact functional predictions.

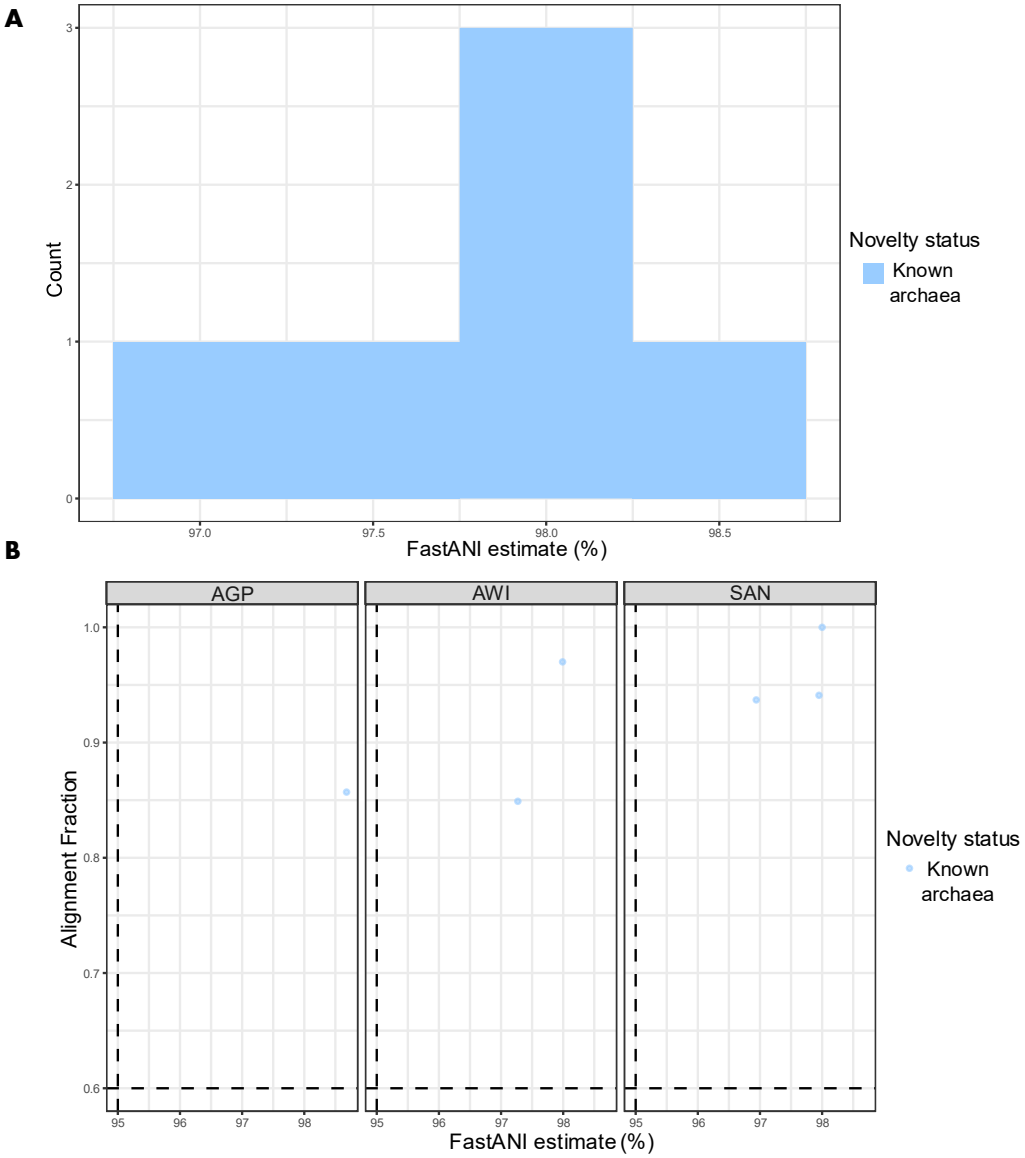


Figure 4.53 Novelty assessment of archaeal MAGs against GTDB reference genomes. (A) Distribution of FastANI sequence identity values for high-quality archaeal MAGs (> 90% completion, < 5% contamination). All comparisons exceeded the species delineation threshold (95% ANI, vertical dashed line) with minimum alignment fraction of 0.6. (B) FastANI alignment fraction versus sequence identity (FastANI estimate) plots across AGP, AWI, and SAN populations. Each point represents a pairwise comparison between a query MAG and its closest reference genome. Horizontal and vertical dashed lines indicate thresholds for reliable comparison (AF = 0.6) and species delineation (ANI = 95%), respectively.

4.7.2. Diversity patterns across cohorts

i. Alpha diversity

Analysis of alpha diversity using the Shannon diversity index revealed significant differences in microbial community structure across the three cohorts at the phylum level (Figure 4.54, Supplementary Data S4.14A – F). The SAN population, despite having the fewest unique phyla (12), exhibited significantly lower sample-level diversity (median Shannon index ≈ 0.6) compared to both the AGP cohort (17 unique phyla; median ≈ 1.1, p-value = 1.1×10^{-12}) and the AWIGEN1 cohort (20 unique phyla; median ≈ 1.2, p-value = 4.8×10^{-19}). While both reference cohorts showed higher diversity and greater unique phylum counts, there was no significant difference in diversity between AGP and AWIGEN1 cohorts (p-value > 5.0×10^{-2}). The similarity between shotgun-sequenced cohorts, despite their distinct geographical origins, suggests that platform differences in detection sensitivity may contribute to the observed diversity patterns.

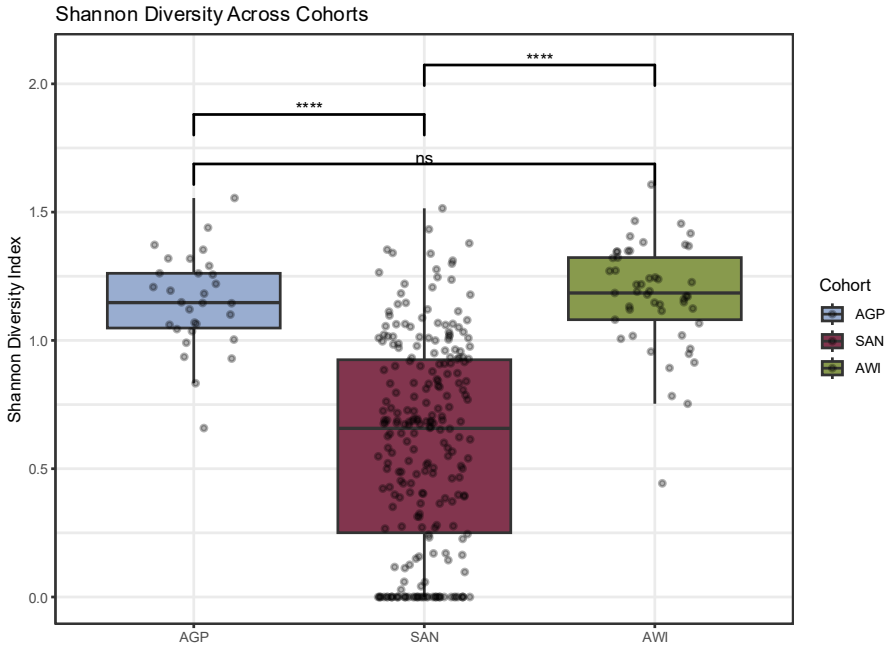


Figure 4.54 Shannon diversity comparison across three human gut microbiome cohorts at the phylum level. Box plots show the distribution of sample-level Shannon diversity indices for the American Gut Project (AGP, blue, 17 unique phyla), urban professional South African (SAN, red, 12 unique phyla), and AWIGEN1 (AWI, green, 20 unique phyla) cohorts. Each point represents an individual sample's diversity measurement. The boxes represent the interquartile range (IQR) with the median indicated by the horizontal line; whiskers extend to 1.5 x IQR. **: Wilcox pairwise test p < 1.0×10^{-4}; ns: not significant (p > 5.0×10^{-2}).**

At the genus level, diversity patterns showed different trends (Figure 4.55, Supplementary Data S4.14G – L). The SAN population (165 unique genera) exhibited intermediate sample-

level Shannon diversity (median ≈ 1.0) with no significant differences compared to either the AGP cohort (150 unique genera; median ≈ 1.2, p-value = 0.9) or the AWIGEN1 cohort (173 unique genera; median ≈ 0.9, p-value = 0.4). A marginally significant difference was observed between AGP and AWIGEN1 cohorts (p-value = 1.2 x 10^{-2}) with AGP showing slightly higher genus-level diversity, despite having fewer unique genera. The similarity in distribution patterns and narrower variance observed in both shotgun-sequenced cohorts (AGP and AWI) compared to the wider spread in the 16S-sequenced SAN population further suggests that sequencing platform differences may influence the diversity measurements beyond biological variation alone.

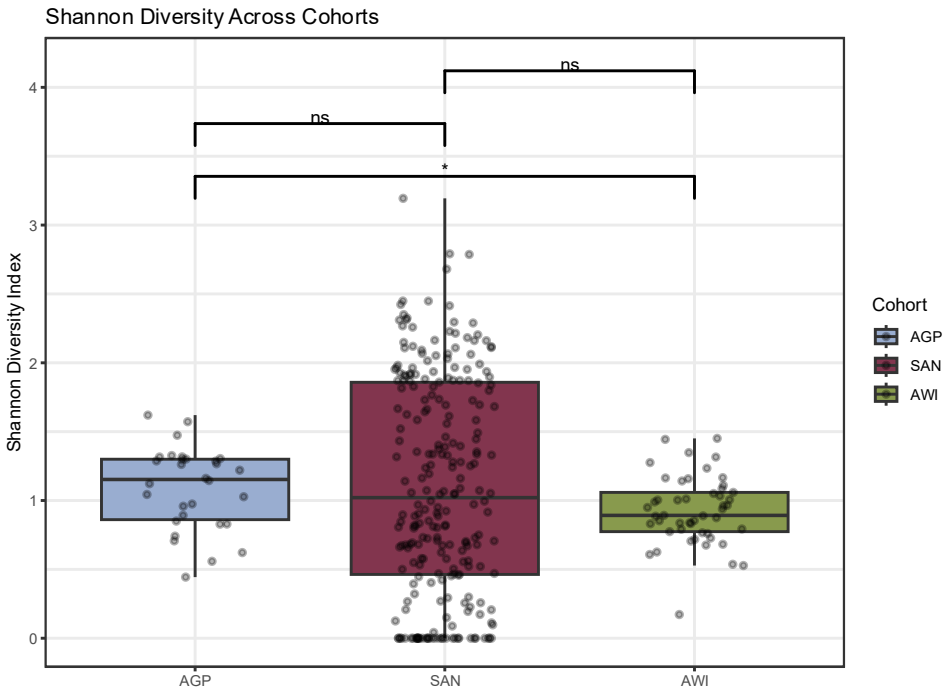


Figure 4.55 Shannon diversity comparison across three human gut microbiome cohorts at the genus level. Box plots show the distribution of sample-level Shannon diversity indices for the American Gut Project (AGP, blue, 150 unique genera), urban professional South African (SAN, red, 165 unique genera), and AWIGEN1 (AWI, green, 173 unique genera) cohorts. Each point represents an individual sample's diversity measurement. The boxes represent the interquartile range (IQR) with the median indicated by the horizontal line; whiskers extend to 1.5 x IQR. *: Wilcox pairwise test p < 5.0 x 10^{-2}; ns: not significant (p > 5.0 x 10^{-2}).

At the species level, diversity analysis revealed distinct patterns (Figure 4.56, Supplementary Data S4.14M – R). The SAN population, had the lowest number of unique species (163) and showed an insignificant but higher sample-level Shannon diversity (median ≈ 0.8), compared to AWIGEN1 (401 unique species; median ≈ 0.7, p-value = 0.3). Similarly, no significant differences were observed between SAN and AGP cohorts (297 unique species: median ≈ 1.0. p-value = 0.4). However, highly significant differences in Shannon diversity were observed between AWIGEN1 and AGP cohorts (p > 3.6 x 10^{-3}), suggesting that the richness and

evenness of unique species in these cohorts are most pronounced when comparing Western and rural-urban African populations.

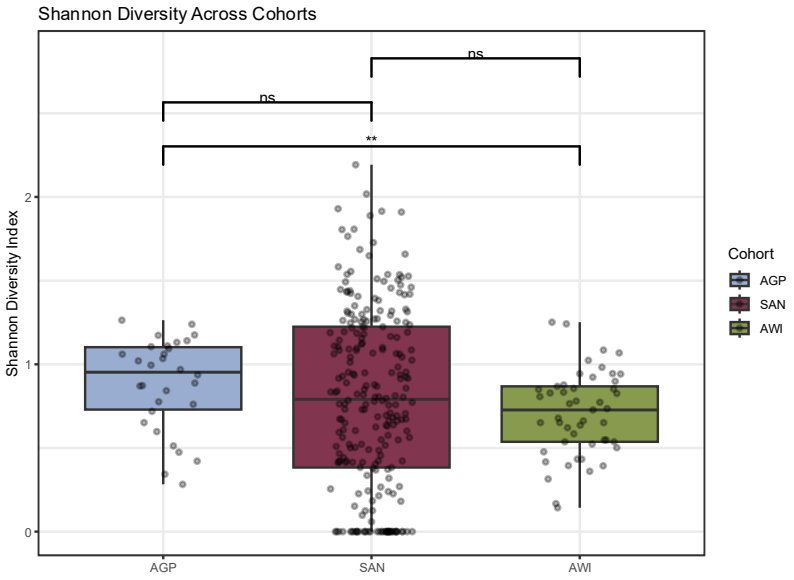


Figure 4.56 Shannon diversity comparison across three human gut microbiome cohorts at the species level. Box plots show the distribution of sample-level Shannon diversity indices for the American Gut Project (AGP, blue, 297 unique species), urban professional South African (SAN, red, 163 unique species), and AWIGEN1 (AWI, green, 401 unique species) cohorts. Each point represents an individual sample's diversity measurement. The boxes represent the interquartile range (IQR) with the median indicated by the horizontal line; whiskers extend to 1.5 x IQR. **: Wilcox pairwise test $p < 1.0 \times 10^{-2}$; ns: not significant ($p > 5.0 \times 10^{-2}$).

ii. Beta diversity

Beta diversity analysis using log-Euclidean distances revealed distinct clustering patterns among the three cohorts (PERMANOVA p-value = 1.0×10^{-3}) (Supplementary Data S4.15, Figure 4.57). Principal Coordinate Analysis (PCoA) explained 53.9% of the total variance in the first two axes (PCo1: 38.3%, PCo2: 15.6%). The overlapping clustering of AGP and AWIGEN1 cohorts, distinct from the SAN population, may be partially attributed to their shared Illumina short-read shotgun sequencing methodology, compared to the PacBio 16S long-read sequencing used for the SAN population.

Pairwise PERMANOVA tests (Supplementary Figure S4.18) showed significant differences between the SAN population and both AGP and AWIGEN1 cohorts (p-value = 1.0×10^{-3}), while AGP and AWIGEN1 cohorts showed no significant difference in community structure (p-value = 0.2). While these results indicate that the urban professional South African gut microbiome composition is distinctly different from both Western (AGP) and urban-rural African (AWIGEN1) populations at the phylum level, the previously identified bi-clustering pattern in the SAN population (Section 4.1) provides additional context. The stratification of

the SAN population into two distinct community clusters, differentiated primarily by Actinobacteriota and Bacteroidota abundances and associated with specific lifestyle factors such as COVID-19 history and cosmetics use, suggests that within-population variation in this urban professional cohort may be driven by modern lifestyle factors rather than geographical distinctions alone.

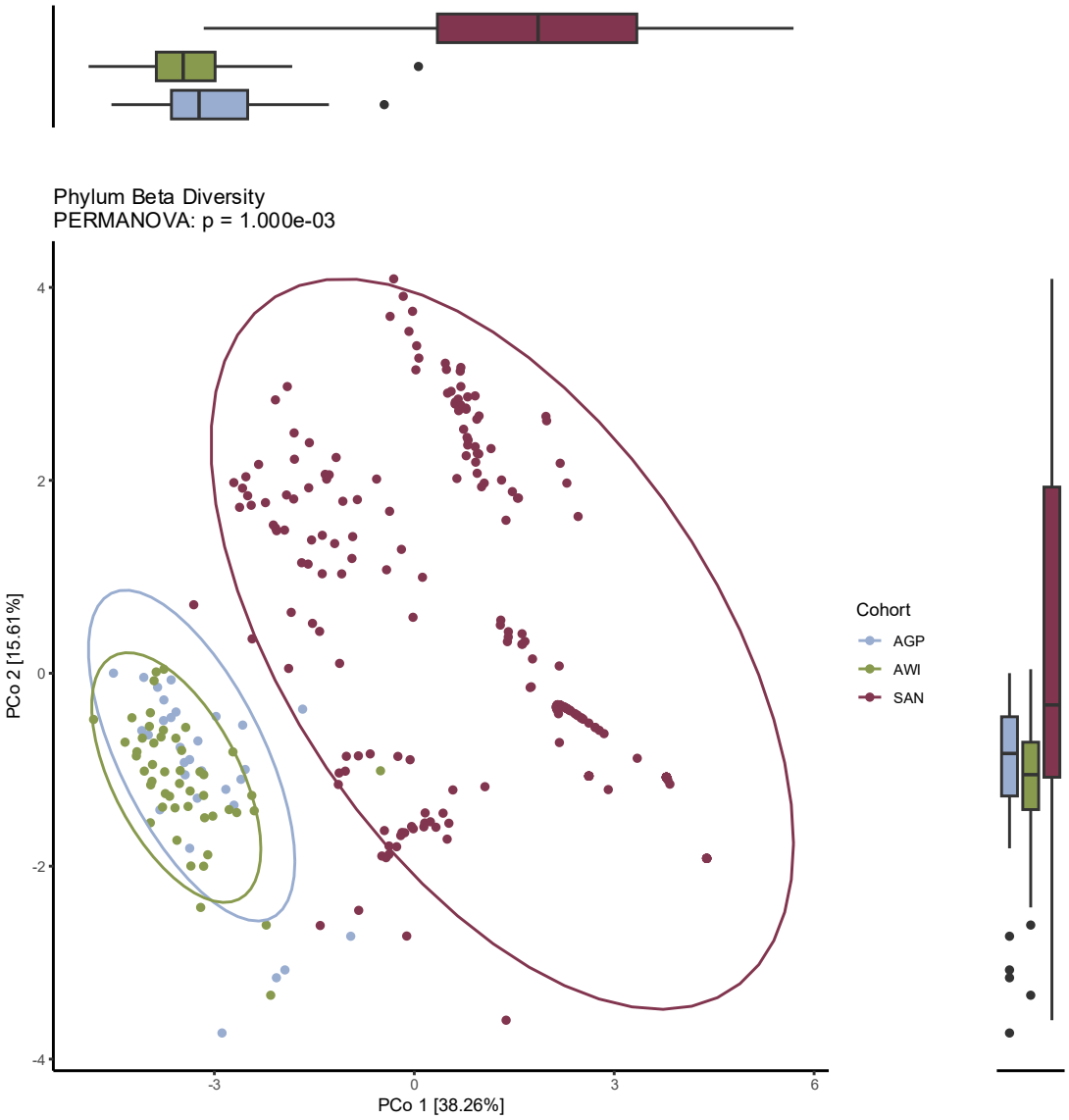


Figure 4.57 Beta diversity analysis using log-Euclidean distances visualized through Principal Coordinate Analysis (PCoA) across three human gut microbiome cohorts at the phylum level. The plot shows sample clustering patterns for the American Gut Project (AGP, blue), urban professional South African (SAN, red), and AWIGEN1 (AWI, green) cohorts. The first two principal coordinates explain 53.9% of the total variance. PERMANOVA test indicates significant differences between cohorts ($p = 1.0 \times 10^{-3}$).

Beta diversity analysis at the genus level using log-Euclidean distances demonstrated significant separation among all three cohorts (PERMANOVA p-value = 1.0×10^{-3}). The PCoA

in Figure 4.58 explained 36% of the total variance in the first two axes (PCo1: 24.8%, PCo2: 11.1%), showing lower total variance explained compared to the phylum level analysis (53.9%). Unlike the phylum level where AGP and AWIGEN1 showed similar community structures, pairwise PERMANOVA tests at the genus level revealed significant differences between all cohort pairs (all p = 1.0 x 10^{-3}, Supplementary Figure S4.19). This suggests that while broader taxonomic patterns may be similar between Western and urban-rural South African populations using shotgun sequencing data, finer-scale taxonomic resolution reveals distinct community structures among all three populations.

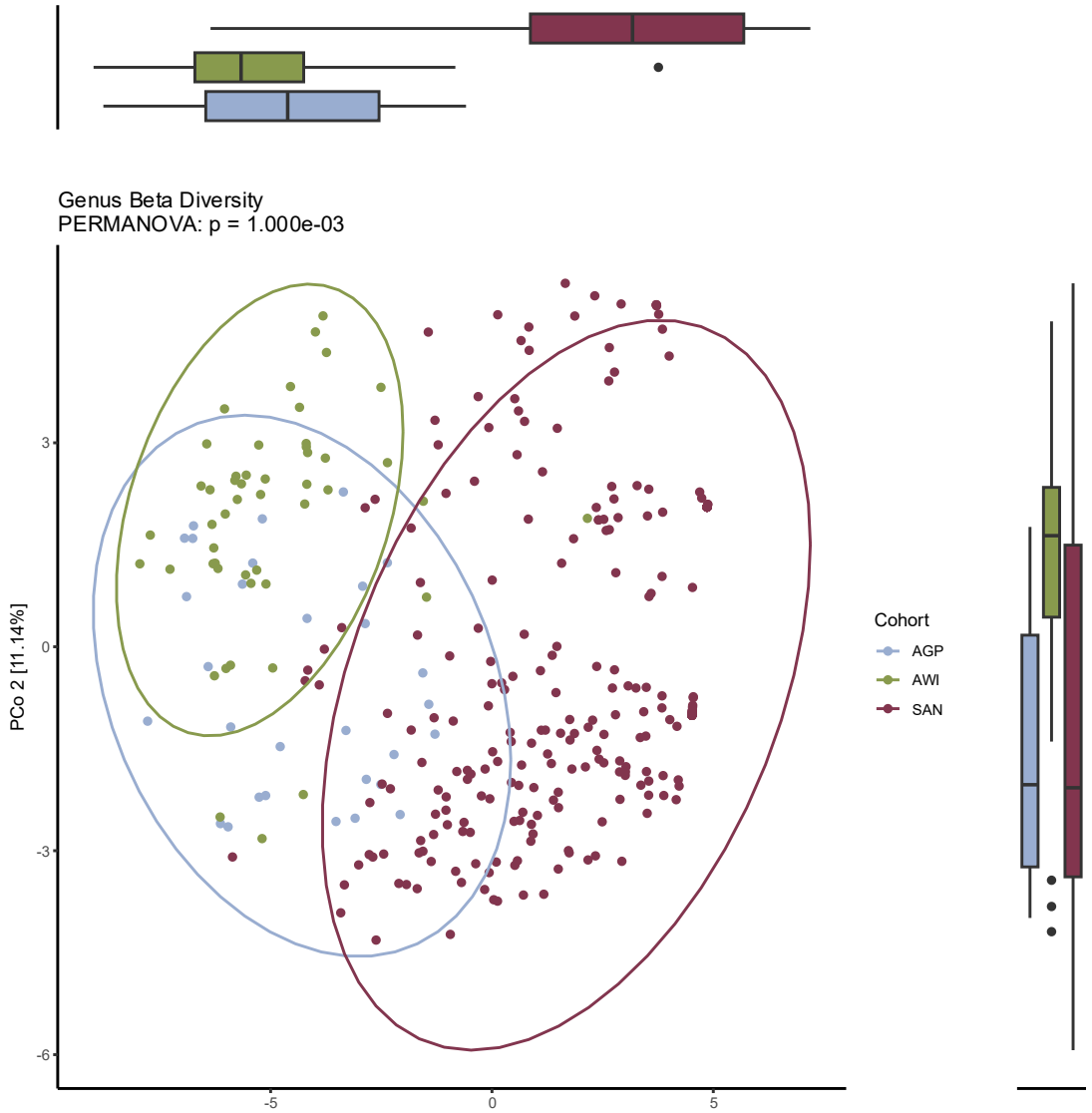


Figure 4.58 Beta diversity analysis using log-Euclidean distances visualized through Principal Coordinate Analysis (PCoA) across three human gut microbiome cohorts at the genus level. The plot shows sample clustering patterns for the American Gut Project (AGP, blue), urban professional South African (SAN, red), and AWIGEN1 (AWI, green) cohorts. The first two principal coordinates explain 36% of the total variance. PERMANOVA test indicates significant differences between all cohorts (p = 1.0 x 10^{-3}).

Beta diversity analysis at the species level showed significant separation among the three cohorts (PERMANOVA p-value = 1.0×10^{-3}). The PCoA explained 27.9% of total variance in the first two axes (PCo1: 20.3%, PCo2: 7.6%) representing a further decrease in explained variance compared to genus (35.9%) and phylum (53.9%) levels (Figure 4.59). Pairwise PERMANOVA tests (Supplementary Figure S4.20, Supplementary Data S4.15F) revealed significant differences between all cohort pairs (p-value = 1.0×10^{-3}). The decreasing proportion of variance explained at finer taxonomic resolution suggests increasing complexity in community structure differences among AGP, SAN and AWIGEN1 populations at the species level.

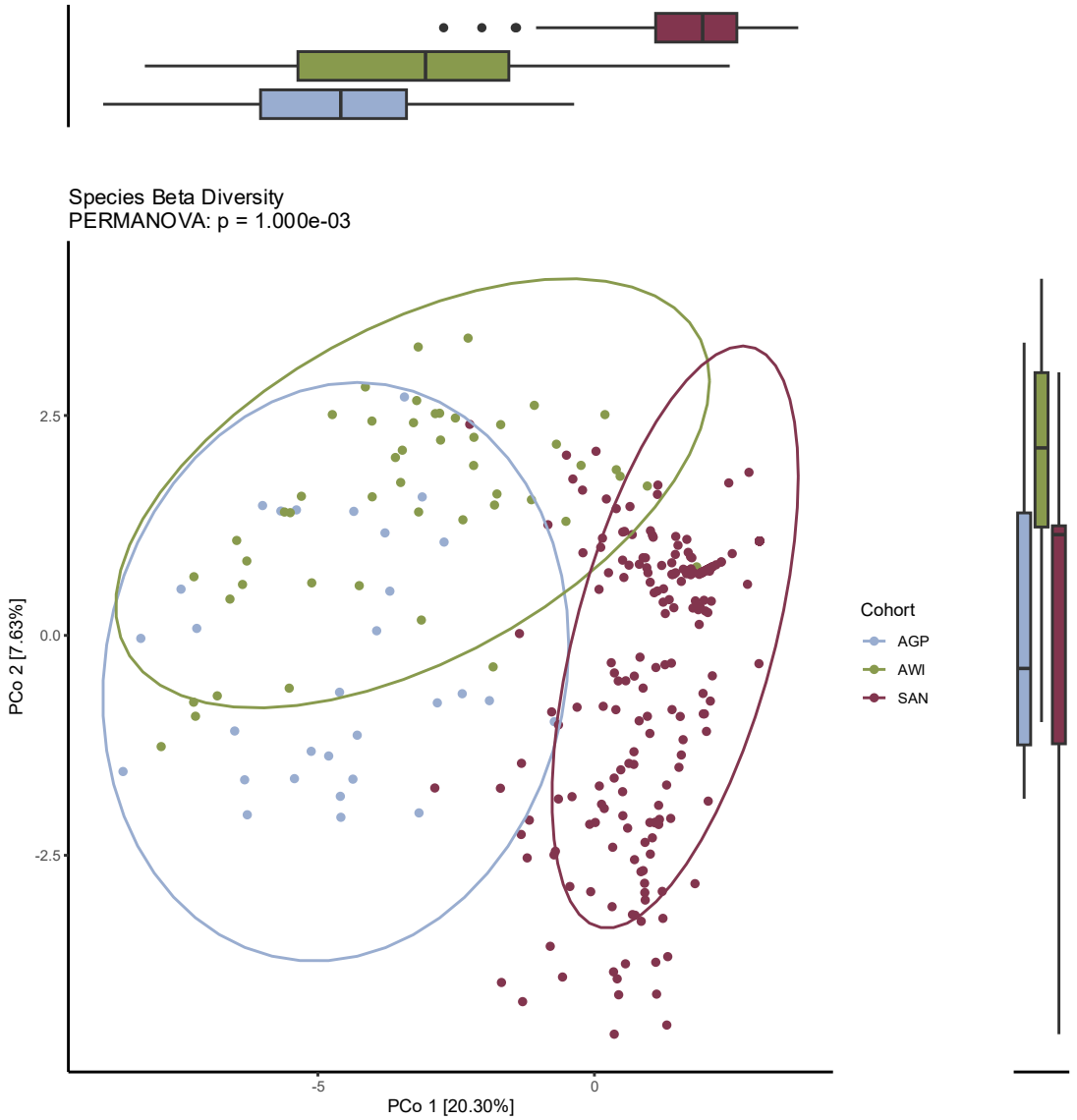


Figure 4.59 Beta diversity analysis using log-Euclidean distances visualized through Principal Coordinate Analysis (PCoA) across three human gut microbiome cohorts at the species level. The plot shows sample clustering patterns for AGP (blue), SAN (red), and AWIGEN1 (green) cohorts. The first two principal coordinates explain 27.9% of the total variance (PCo1: 20.3%, PCo2: 7.6%). PERMANOVA test indicates significant differences between all cohorts (p = 1.0 x 10^{-3}).

To identify the specific taxonomic drivers of these community-level differences, a comprehensive analysis of differential abundance patterns across taxonomic ranks was performed, with particular focus on taxa showing both statistical significance and large effect sizes in comparisons involving the SAN population.

4.7.3. Taxonomic and differential abundance

Building on these observed differences in community structure, a systematic approach combining taxonomic classification at phylum, genus and species levels with differential

abundance testing was conducted. To ensure robust identification of potential biologically meaningful differences, both Kruskal-Wallis and pairwise Wilcoxon tests with Benjamini-Hochberg adjusted p-values were tested, focusing particularly on comparisons involving the SAN population (Supplementary Data S4.16). Effect sizes were calculated using Cliff's delta statistic (δ) with a threshold of 0.47 to identify taxa showing substantial population-specific differences (Kargar et al., 2020).

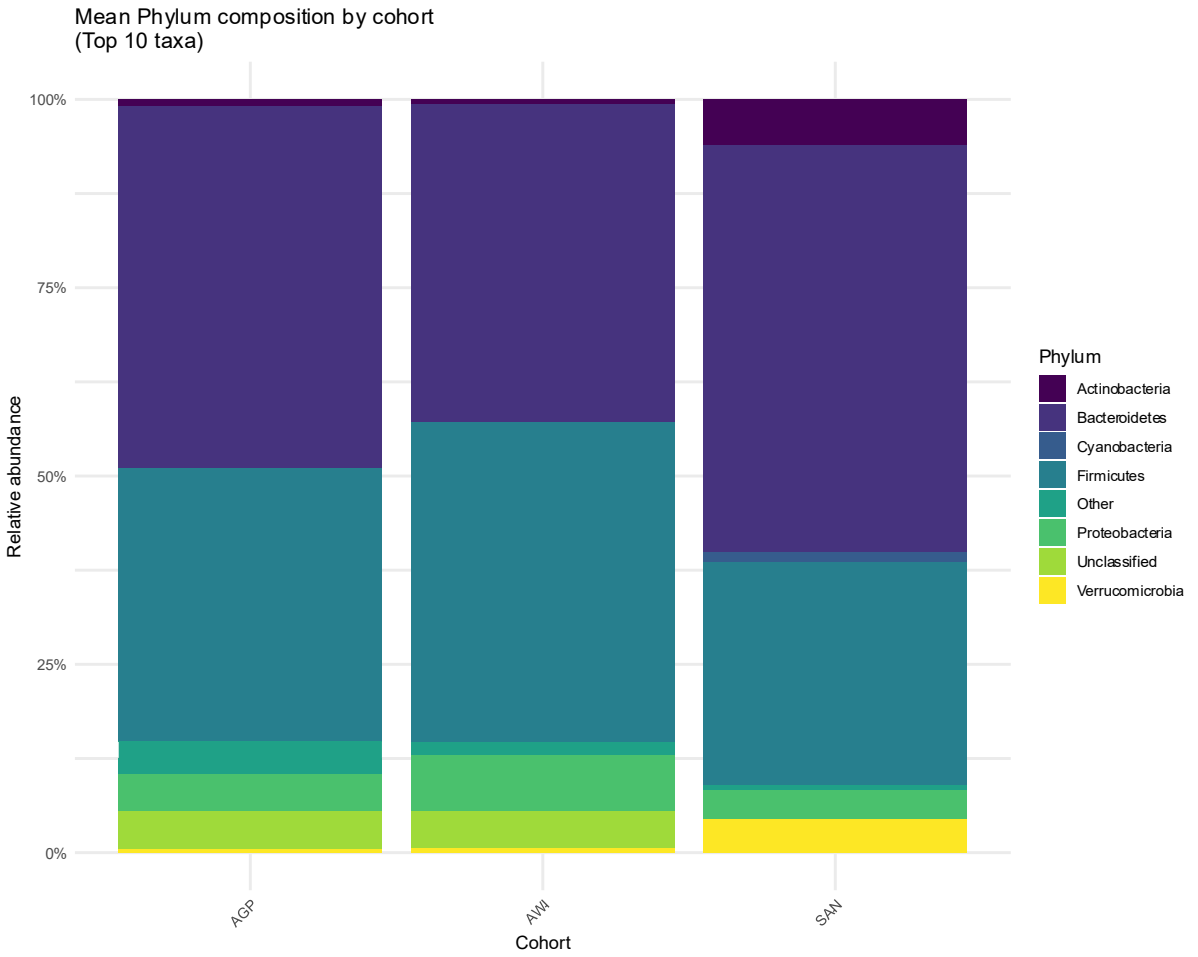


Figure 4.60 Mean relative abundance of bacterial phyla across three cohorts. The stacked bar plot shows the proportional composition of phyla with > 1% mean relative abundance across all samples, based on MetaPhlAn taxonomic profiling of shotgun metagenomic data (AGP and AWI cohorts) and SILVA-based taxonomic classification of 16S long-read sequencing data (SAN population). Less abundant phyla were grouped as "Other", while "Unclassified" represents sequences that could not be confidently assigned to known bacterial phyla.

At the phylum level all three cohorts were dominated by Firmicutes and Bacteroidetes, together comprising the majority of the bacterial composition (Figure 4.60). However, the proportional distribution of these major phyla varied substantially across populations. Comparison of the urban SAN population against AGP and AWIGEN1 populations revealed several significant differences in phylum-level abundance (Figure 4.61, Supplementary Data S4.16B).

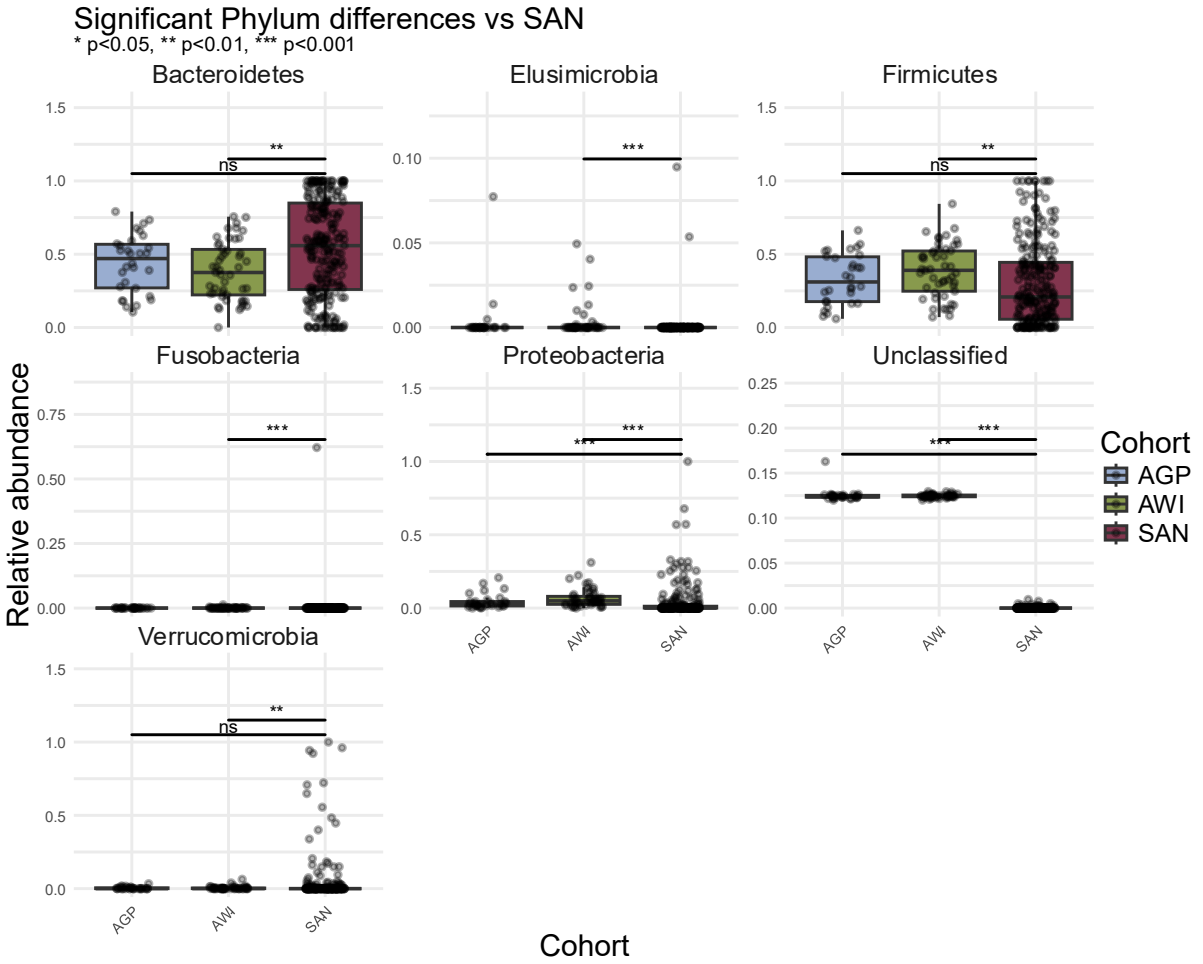


Figure 4.61 Differential abundance analysis comparing phyla between SAN and other cohorts (AGP, AWIGEN1). Only phyla with significant differences (Wilcoxon rank-sum test, Benjamini-Hochberg adjusted $p < 5.0 \times 10^{-2}$) and effect size (Cliff's delta ≥ 0.1) are shown. Statistical significance indicated by asterisks (*: $p < 5.0 \times 10^{-2}$, **: $p < 1.0 \times 10^{-2}$, ***: $p < 1.0 \times 10^{-3}$, ns: not significant).

Firmicutes showed significantly higher abundance in AWIGEN1 cohorts compared to SAN (adjusted p-value = 2.9×10^{-3}), while Bacteroidetes displayed markedly higher levels in the SAN population compared to AWIGEN1 (adjusted p-value = 3.7×10^{-3}). Notably, the Firmicutes and Bacteroidetes difference between SAN and AGP was not significant, suggesting potential dietary or lifestyle factors that might align these populations for these particular phyla. Among the less abundant phyla, several distinctive patterns emerged. Proteobacteria showed significantly higher levels in both AGP and AWIGEN1 cohorts compared to SAN (adjusted p-value = 4.69×10^{-9}, adjusted p-value = 1.43×10^{-14}, respectively. Verrucomicrobia. Fusobacteria and Elusimicrobia demonstrated significant differences between SAN and the AWI cohort (adjusted p-value = 2.1×10^{-3}; 3.3×10^{-11}; 3.7×10^{-11}, respectively). The proportion of unclassified bacteria varied significantly between populations with higher levels observed in both AGP and AWIGEN1 cohorts compared to SAN (adjusted p-value = 1.6×10^{-39}; 2.4×10^{-48}).

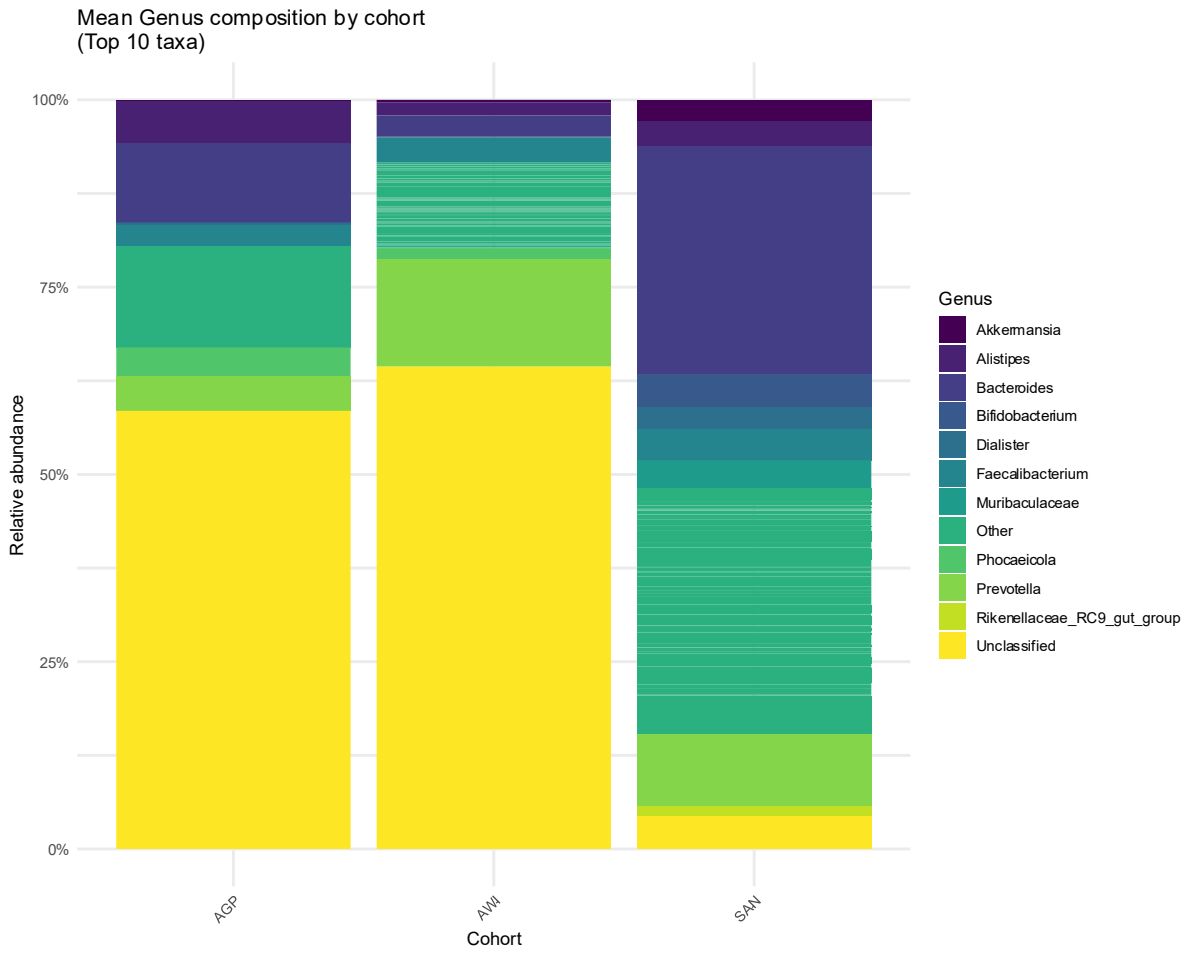


Figure 4.62 Mean relative abundance of bacterial genera across three cohorts. The stacked bar plot shows the proportional composition of genera with > 1% mean relative abundance across all samples, based on MetaPhlAn taxonomic profiling of shotgun metagenomic data (AGP and AWI cohorts) and SILVA-based taxonomic classification of 16S rRNA long-read sequencing data (SAN population). Less abundant genera were grouped as "Other", while "Unclassified" represents sequences that could not be confidently assigned to known bacterial genera.

At the genus level, marked differences in bacterial composition were observed across the three cohorts (Figure 4.62). The AGP and AWIGEN1 cohorts showed broadly similar profiles dominated by unclassified bacteria (approximately 60-65% relative abundance). The SAN population displayed a distinctly different pattern with Bacteroides encompassing ~25% abundance and a more even distribution of the remaining genera, with a notably lower proportion of unclassified sequences.

Differential abundance analysis revealed numerous significant differences between cohorts (Figure 4.63). The most pronounced differences were observed in unclassified bacteria, which showed significantly lower abundance in SAN compared to both AGP ($\delta = -1$, adjusted p-value = 4.2×10^{-19}) and AWIGEN1 ($\delta = -1$, adjusted p-value = 7.6×10^{-28}). *Oscillibacter* demonstrated strong differential abundance, being significantly lower in SAN versus AWIGEN1 ($\delta = -0.9$, adjusted p-value = 6.6×10^{-39}).

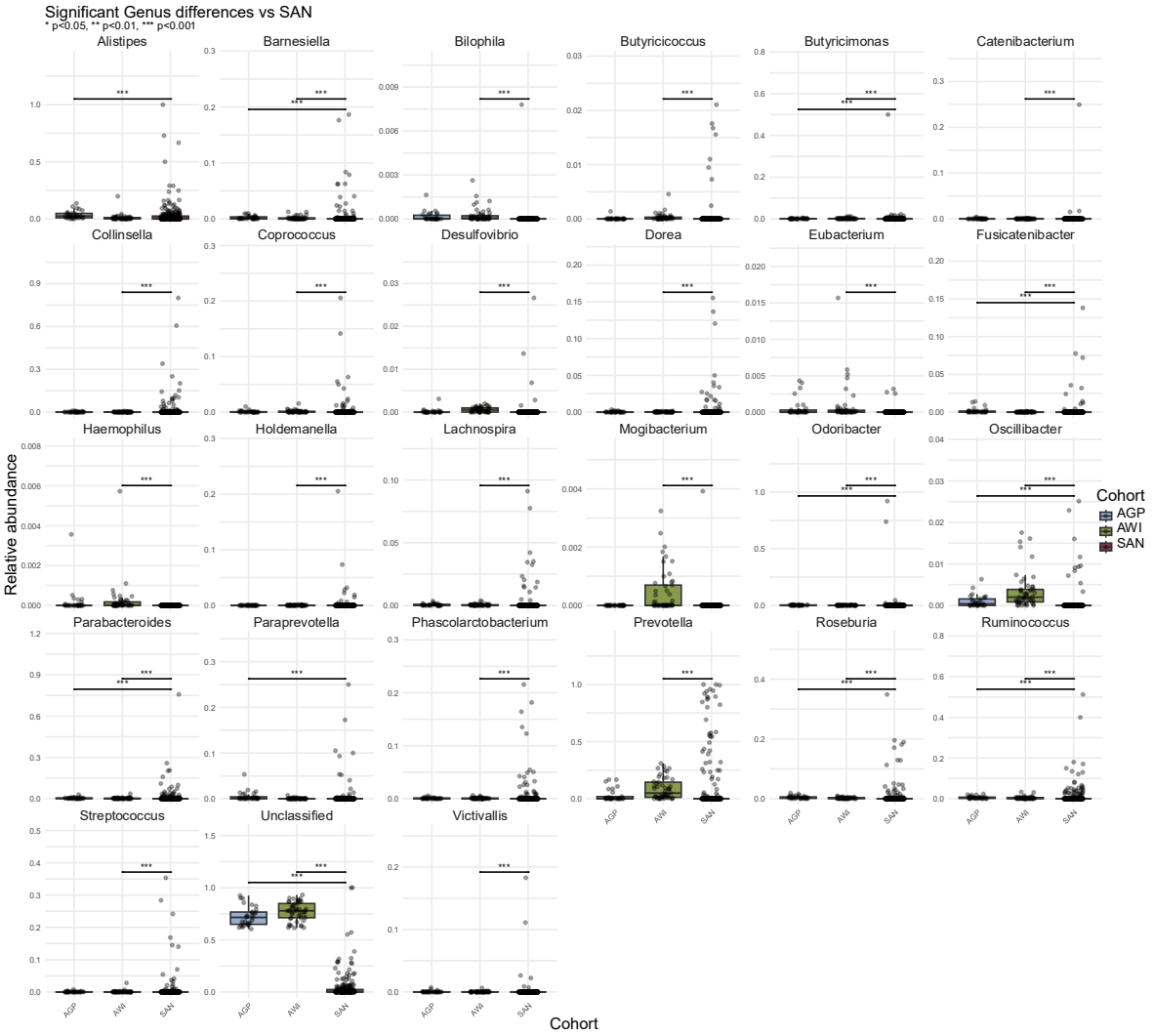


Figure 4.63 Differential abundance analysis comparing genera between SAN and other cohorts (AGP, AWIGEN1). Only genera with significant differences (Wilcoxon rank-sum test, Benjamini-Hochberg adjusted $p < 5.0 \times 10^{-2}$) and effect size (Cliff's delta ≥ 0.1) are shown. Statistical significance indicated by asterisks (*: $p < 5.0 \times 10^{-2}$, **: $p < 1.0 \times 10^{-2}$, ***: $p < 1.0 \times 10^{-3}$, ns: not significant).

Similarly, *Butyricimonas* showed markedly lower levels in SAN compared to AWIGEN1 (δ = -0.8, adjusted p-value = 2.5×10^{-34}). Among other major genera, *Coprococcus* (δ = -0.7, adjusted p-value = 3.9×10^{-26}), *Desulfovibrio* (δ = -0.7, adjusted p-value = 1.7×10^{-36}), and Roseburia (δ = -0.7, adjusted p-value = 2.5×10^{-21}) all showed significantly lower abundance in the SAN population compared to AWIGEN1 (Supplementary Data S4.16D).

At the species level, interesting compositional patterns emerged across the three cohorts (Figure 4.64). The AGP and AWI cohorts showed similar profiles dominated by “unclassified species” (> 70% relative abundance), in addition to *Prevotella copri clade A* and “Other” species comprising < 1% total mean relative abundance. The SAN population also had a high

abundance of “unclassified species” (~30%), in addition to a ~20% abundance of *Bacteroides dorei* and unique presence of > 1% abundance of *Akkermansia muciniphila.*

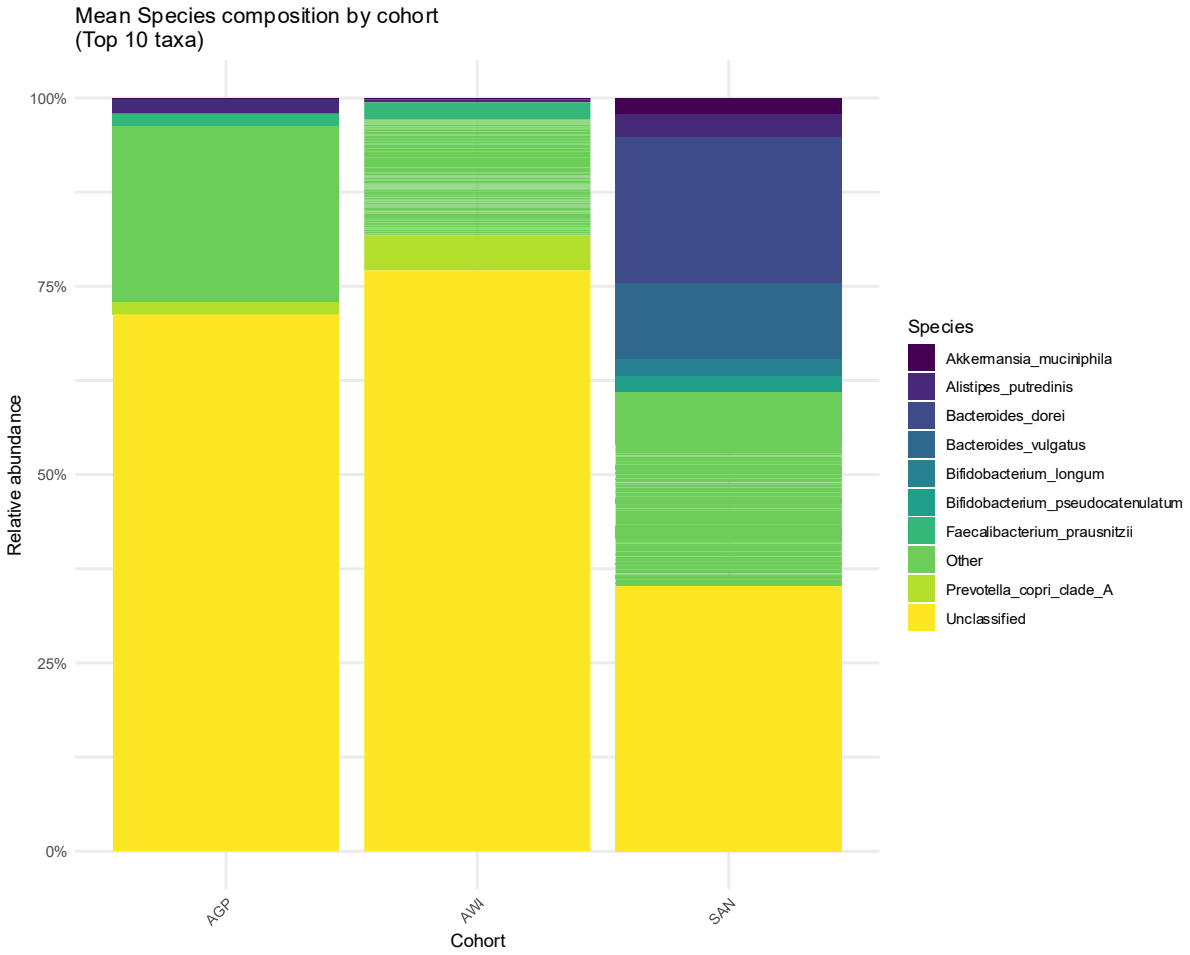


Figure 4.64 Mean relative abundance of bacterial species across three cohorts. The stacked bar plot shows the proportional composition of species with > 1% mean relative abundance across all samples, based on MetaPhlAn taxonomic profiling of shotgun metagenomic data (AGP and AWI cohorts) and SILVA-based taxonomic classification of 16S rRNA long-read sequencing data (SAN population). Less abundant species were grouped as "Other", while "Unclassified" represents sequences that could not be confidently assigned to known bacterial species.

Species-level differential abundance analysis revealed numerous significant differences between cohorts (Figure 4.65). In line with the above, “unclassified species” showed significantly lower abundance in SAN versus both AGP (δ = -0.6, adjusted p-value = 2.6 x 10^{-7}) and AWIGEN1 (δ = -0.7, adjusted p-value = 7.2 x 10^{-13}). Among classified species, *Parabacteroides distasonis* showed strong differences between SAN and both AWIGEN1 (δ = -0.9, adjusted p-value = 3.4 x 10^{-48}) and AGP (δ = -0.8, adjusted p-value = 1.3 x 10^{-37}). *Bacteroides uniformis* also showed substantial differences between SAN versus AGP (δ = -0.9, adjusted p-value = 1.4 x 10^{-34}) and AWIGEN1 (δ = -0.8, adjusted p-value = 4.1 x 10^{-33}). Similarly, *Odoribacter splanchnicus* demonstrated significant differences between SAN and AWIGEN1 (δ = -0.8, adjusted p-value = 3.3 x 10^{-39}) and AGP (δ = -0.7, adjusted p-value = 2.6

x 10^{-31}). Among other species, *Dorea longicatena* (δ = -0.8, adjusted p-value = 1.5 x 10^{-41}) and *Alistipes shahii* (δ = -0.7, adjusted p-value = 8.2 x 10^{-37}) showed significant differences between SAN and AWIGEN1.

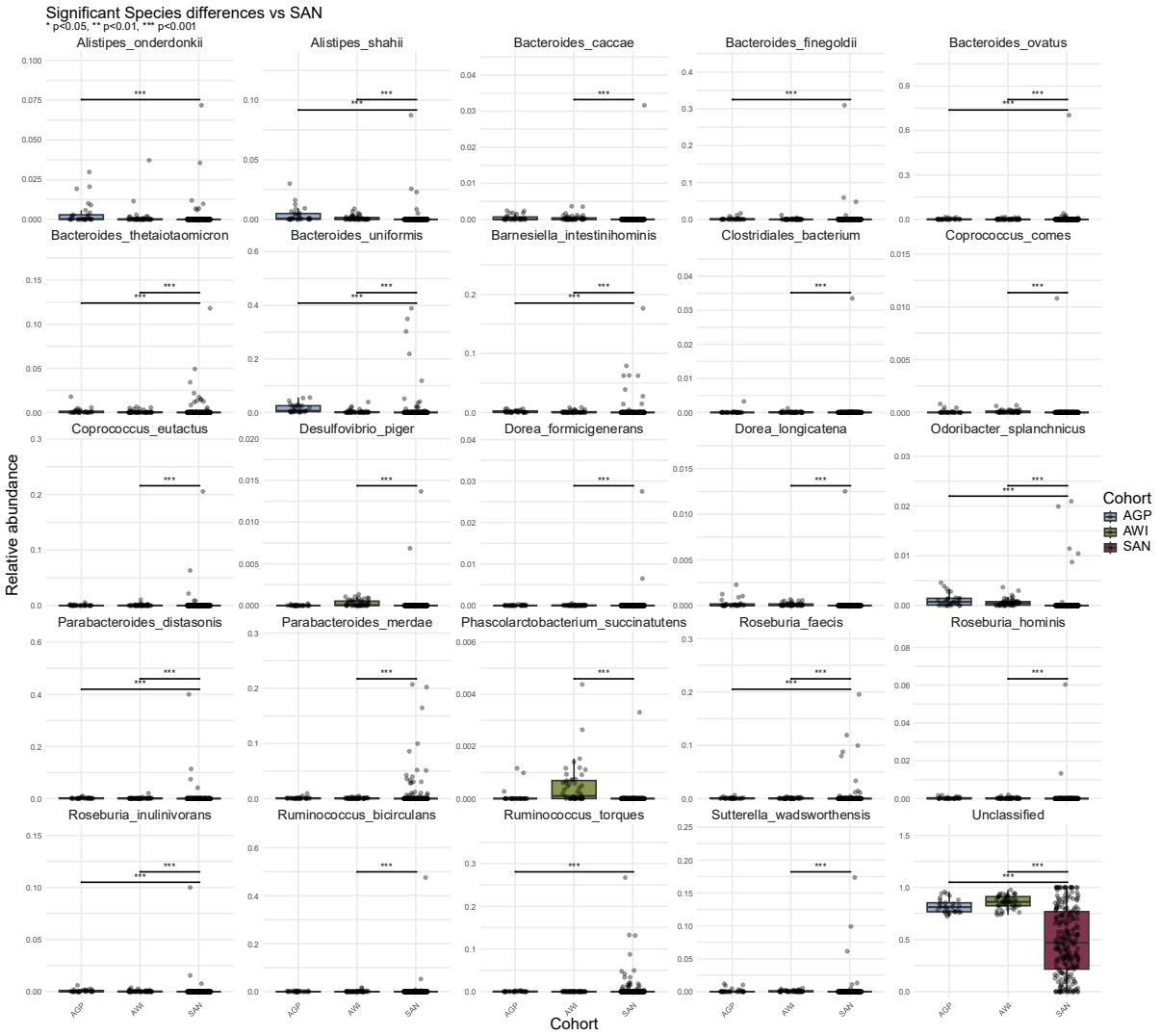


Figure 4.65 Differential abundance analysis comparing species between SAN and other cohorts (AGP, AWIGEN1). Only species with significant differences (Wilcoxon rank-sum test, Benjamini-Hochberg adjusted p < 5.0 x 10^{-2}) and effect size (Cliff's delta ≥ 0.1) are shown. Statistical significance indicated by asterisks (*: p < 5.0 x 10^{-2}, **: p < 1.0 x 10^{-2}, ***: p < 1.0 x 10^{-3}, ns: not significant).

Following the detailed examination of species-level taxonomic composition across the three cohorts, a deeper genomic investigation using metagenome-assembled genomes (MAGs) was performed. While the previous sections utilized 16S long-read sequencing data for the SAN population, a complementary shotgun metagenomic dataset from the same population to enable MAG recovery and comparative analysis with the AGP and AWI cohorts was performed.

4.7.4. Functional diversity across cohorts

To evaluate the consistency of functional predictions between shotgun metagenomics and 16S-based approaches, we compared the abundance profiles of *n* = 292 shared metabolic pathways across all three cohorts (Supplementary Data S4.18). Comparison of log10-transformed pathway abundances revealed significantly strong positive correlations between PICRUSt2 predictions from the SAN population and HUMAnN3 results from both AGP (Spearman's ρ = 0.8, p-value = 9.1×10^{-75}) and AWI (Spearman's ρ = 0.8, p-value = 8.7×10^{-74}) cohorts. The scatter plots in Figure 4.66 demonstrate consistent abundance estimates across most pathways, particularly for highly abundant pathways (log10 abundance > -4) with greater variability observed in low-abundance pathways.

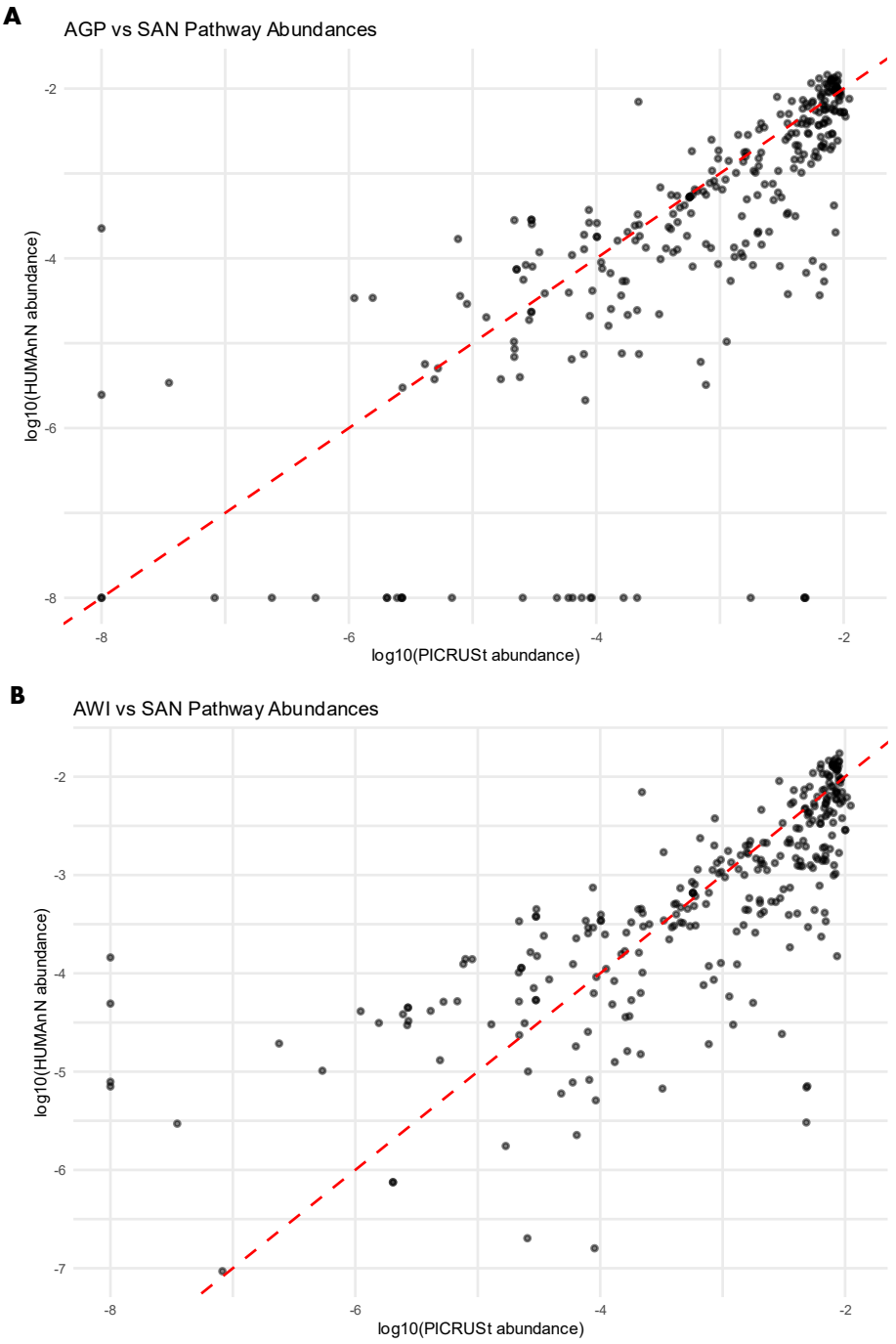


Figure 4.66 Correlation of predicted metabolic pathway abundances between shotgun metagenomics and 16S approaches. (A) Scatter plot comparing pathway abundances between AGP (HUMAnN3) and SAN (PICRUSt2) cohorts. (B) Scatter plot comparing pathway abundances between AWI (HUMAnN3) and SAN (PICRUSt2) cohorts. Each point represents a distinct metabolic pathway (*n* = 292). Red dashed lines indicate theoretical perfect correlation (y=x). Both axes show log10-transformed relative abundances of pathways, with HUMAnN3 observations on y-axis and PICRUSt2 predictions on x-axis.

To quantify the magnitude of differences between methods, we examined the distribution of log2 ratios between HUMAnN3 and PICRUSt2 predictions. The density plots reveal predominantly symmetric distributions centred near zero for both AGP/SAN and AWI/SAN

comparisons (Figure 4.67A-B). The majority of pathways show less than 2-fold difference (log2 ratio between -1 and 1) between methods though some outlier pathways exhibit larger discrepancies, particularly in low-abundance pathways. The AWI/SAN comparison showed a slightly broader distribution of ratios compared to AGP/SAN, suggesting marginally higher variability in predictions between these cohorts.

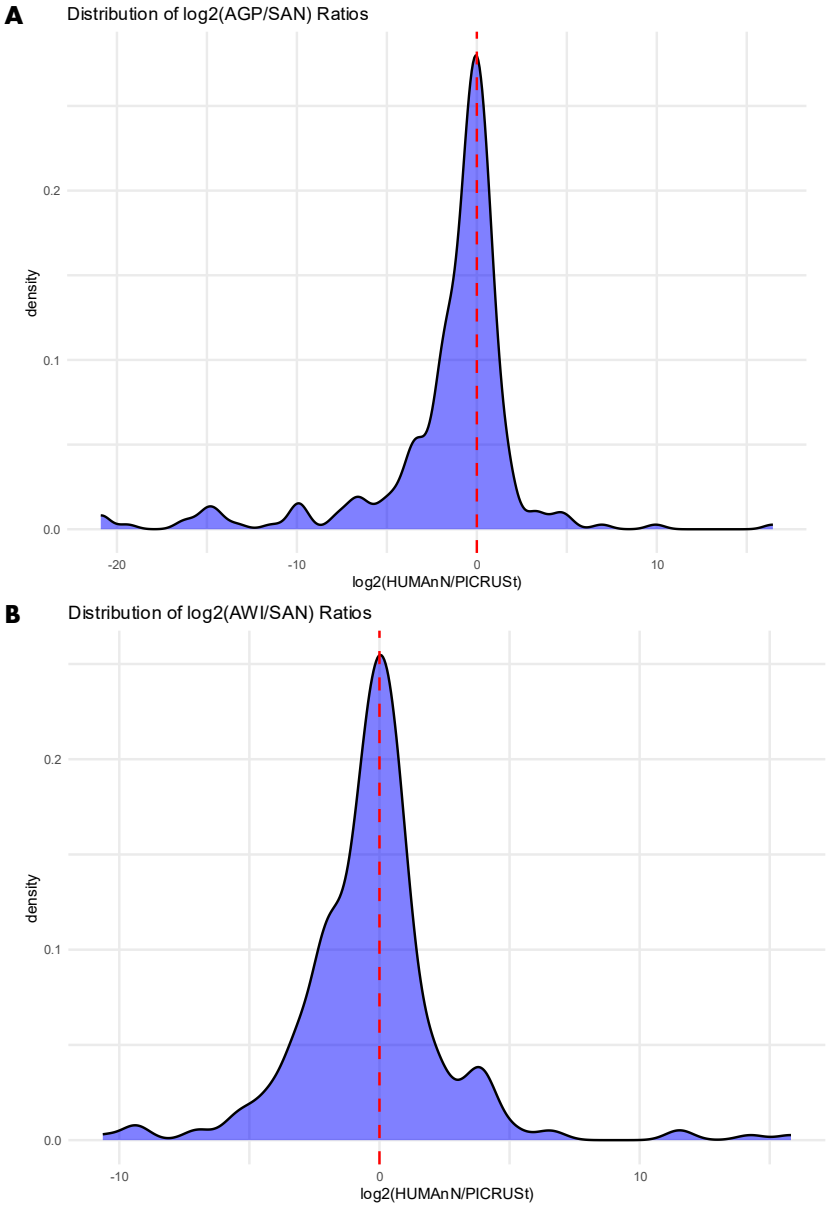


Figure 4.67 Distribution of abundance differences between prediction methods. (A) Density plot showing the distribution of log2 ratios between AGP (HUMAnN3) and SAN (PICRUSt2) pathway predictions. (B) Density plot showing the distribution of log2 ratios between AWI (HUMAnN3) and SAN (PICRUSt2) pathway predictions. Red dashed vertical lines indicate equal abundance (log2 ratio = 0). Positive values indicate higher HUMAnN3 observations, while negative values indicate higher PICRUSt2 predictions. Density plots represent all *n* = 292 shared pathways between methods.

To investigate potential systematic biases between methods, the most divergent pathway predictions between HUMAnN3 and PICRUSt2 were examined. The top 20 most different pathways revealed distinct patterns in both AGP and AWI comparisons (Figure 4.115 – 116). “CMP-pseudaminate biosynthesis” showed consistently higher predictions in HUMAnN3 across both datasets (log2 ratio > 10).

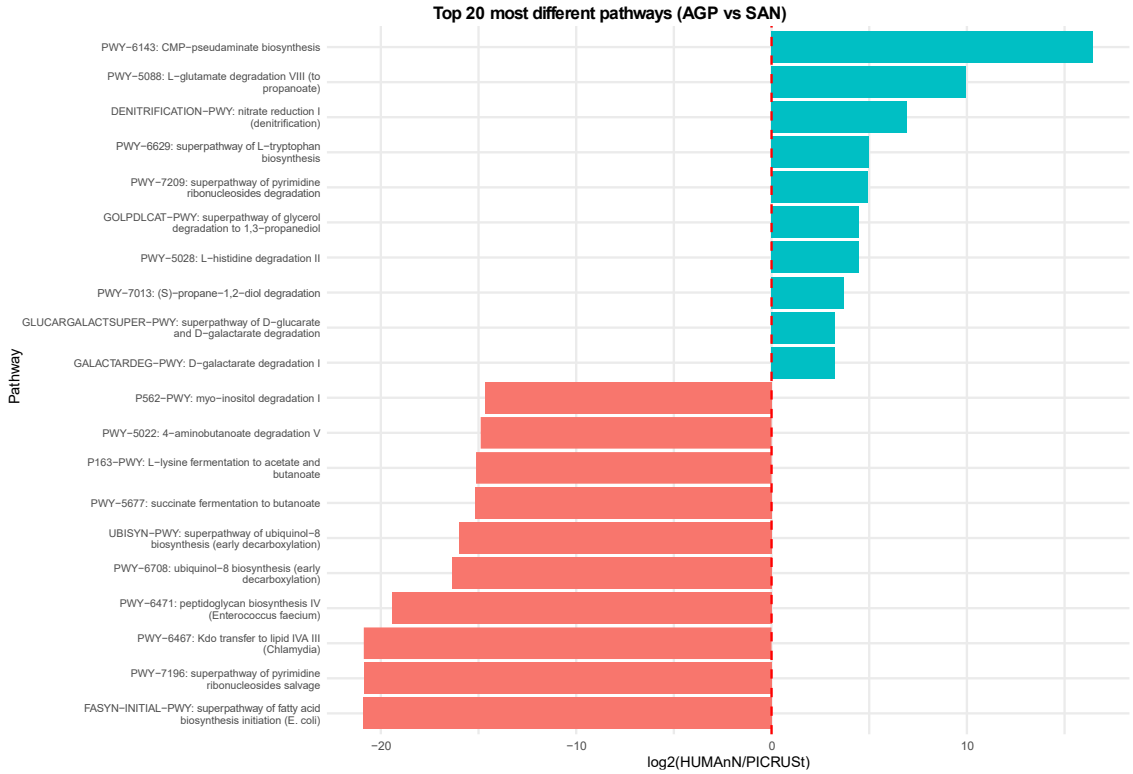


Figure 4.68 Top 20 pathways showing largest abundance differences between methods. For AGP (HUMAnN3) and SAN (PICRUSt2) cohorts. Log2 ratio positive values (blue) indicate higher HUMAnN3 observations, negative values (pink) indicate higher PICRUSt2 predictions. Pathways are ordered by magnitude of difference.

Conversely, pathways like “fatty acid biosynthesis initiation” and “Kdo transfer” showed substantially lower HUMAnN3 predictions (log2 ratio < -20). Notably, several pathways appeared among the most divergent in both comparisons, including “L-glutamate degradation VIII” and “pyrimidine ribonucleoside degradation” pathways.

When examining pathway presence/absence patterns across functional categories, most categories showed high concordance between methods with > 75% of pathways detected by both approaches (Figure 4.70A – B, Supplementary Data S4.18L). However, notable differences emerged in specific categories.

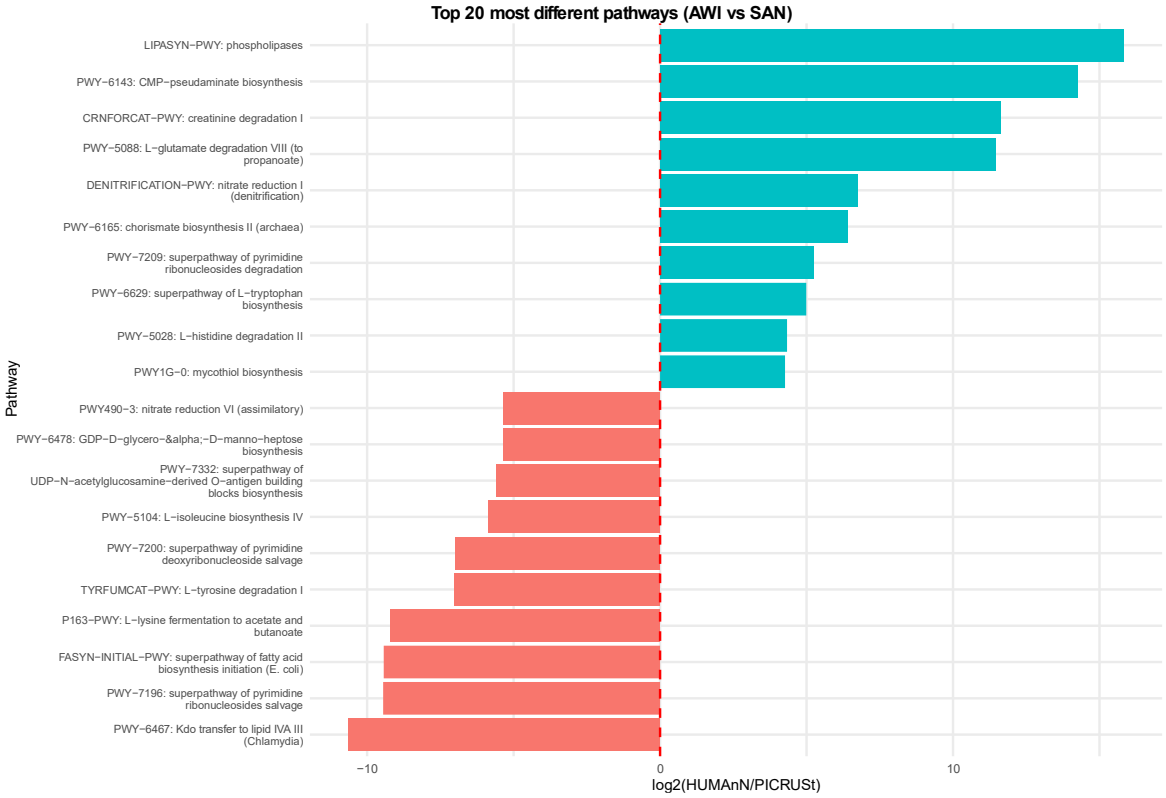


Figure 4.69 Top 20 pathways showing largest abundance differences between methods. For AWI (HUMAnN3) and SAN (PICRUSt2) cohorts. Log2 ratio positive values (blue) indicate higher HUMAnN3 observations, negative values (pink) indicate higher PICRUSt2 predictions. Pathways are ordered by magnitude of difference.

"Lipid Metabolism" showed the highest proportion of PICRUSt-only predictions in the AGP comparison (~25% of pathways) while showing more HUMAnN-only predictions in the AWI comparison (~20% of pathways). "Cell Wall & Membrane" pathways also showed substantial method-specific predictions (Figure 4.70A). In contrast, "Energy Metabolism", "Secondary Metabolism" and "Vitamin & Cofactor" pathways showed complete overlap between methods.

To quantify systematic biases in abundance predictions, the analysis of the distribution of log2 ratios between methods across functional categories was performed (Figure 4.71A – B). Both comparisons showed significant differences between categories (Kruskal-Wallis test; AWI vs SAN: p-value = 6.3×10^{-3}, AGP vs SAN: p-value = 2.2×10^{-2}) (Supplementary Data S4.18K). "Amino Acid Metabolism" showed the most consistent predictions between methods (median log2 ratio near 0), while "Carbohydrate Metabolism" displayed larger systematic differences (median log2 ratio < -1.5 in AGP).

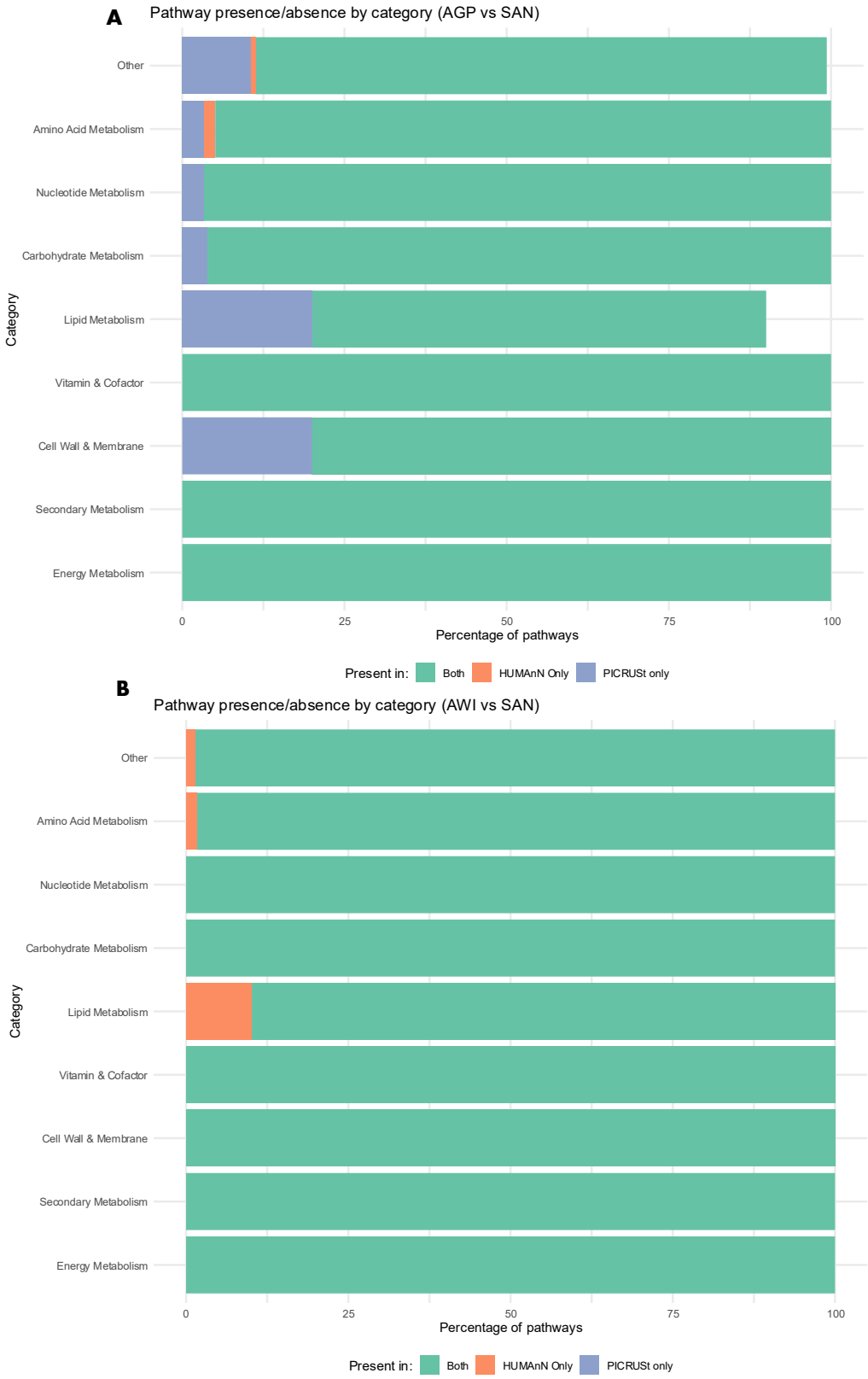


Figure 4.70 Distribution of pathway detection across functional categories. (A) AGP vs SAN and (B) AWI vs SAN comparisons showing percentage of pathways detected by both methods (green), HUMAnN3 only (orange), or PICRUSt2 only (blue).

Pathways categorised as “Other” (*n* = 151) due to not falling within these groups exhibited the highest prediction variability, as indicated by large interquartile ranges and numerous outliers, highlighting that the functional analysis demonstrated systematic differences between functional profiling methods.

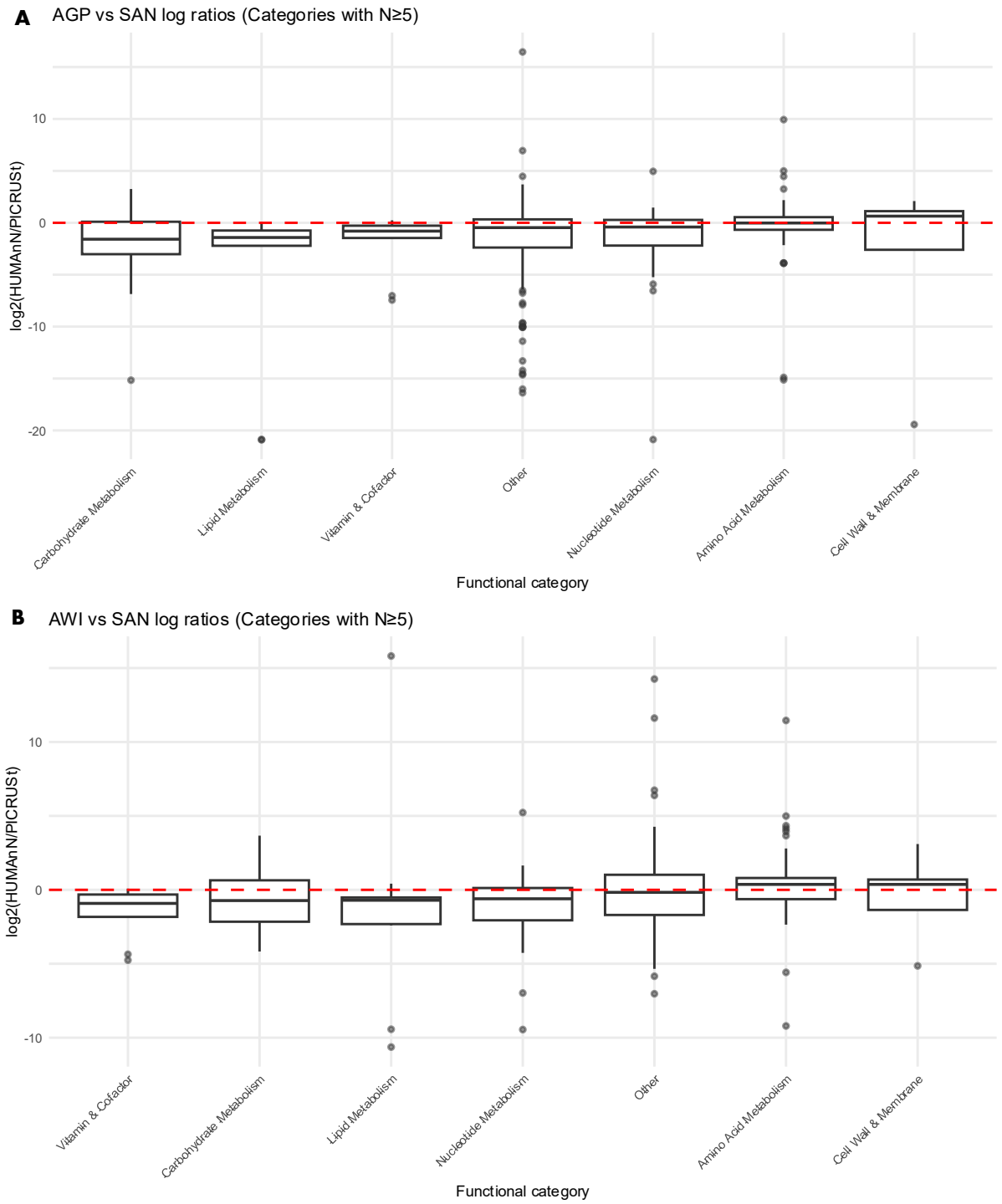


Figure 4.71 Distribution of abundance differences across functional categories. Box plots showing log2 ratios (HUMAnN/PICRUSt) for (A) AGP vs SAN and (B) AWI vs SAN comparisons. Only categories with ≥ 5 pathways are shown. Red dashed line indicates equal abundance (log2 ratio = 0). Outliers shown as individual points.

In summary, this comprehensive analysis of the urban professional South African population reveals a complex interplay between demographic variables, lifestyle factors, and gut microbiome composition (Figure 4.72). The long-read 16S rRNA sequencing data demonstrated significant associations between age and alpha diversity, with older adults exhibiting greater microbial richness and evenness, while gender significantly influenced community composition as measured by beta diversity. Differential abundance analysis identified distinct microbial biomarkers across demographic groups, including species specifically associated with age, gender, ancestry, and BMI classifications. Response variables related to socioeconomic status, diet, lifestyle, and health factors further revealed significant associations with specific microbial taxa and diversity patterns, with variables such as public transport use, dairy consumption, gas heating, and multivitamin intake emerging as particularly influential factors. Predicted functional profiling indicated considerable variation in metabolic potential across the cohort, with pathways related to carbohydrate metabolism, nucleotide synthesis, and amino acid biosynthesis being most prevalent. The top pathways identified, including starch degradation and various biosynthetic processes, reflect key metabolic functions essential for energy production and cellular growth. While these findings provide valuable insights into the urban South African gut microbiome, the subsequent analysis of concordance between long-read and short-read sequencing approaches offers additional methodological considerations and deeper functional investigations to contextualise these observations. Collectively, these results establish a foundational understanding of the urban South African gut microbiome landscape and its associations with key demographic and lifestyle factors.

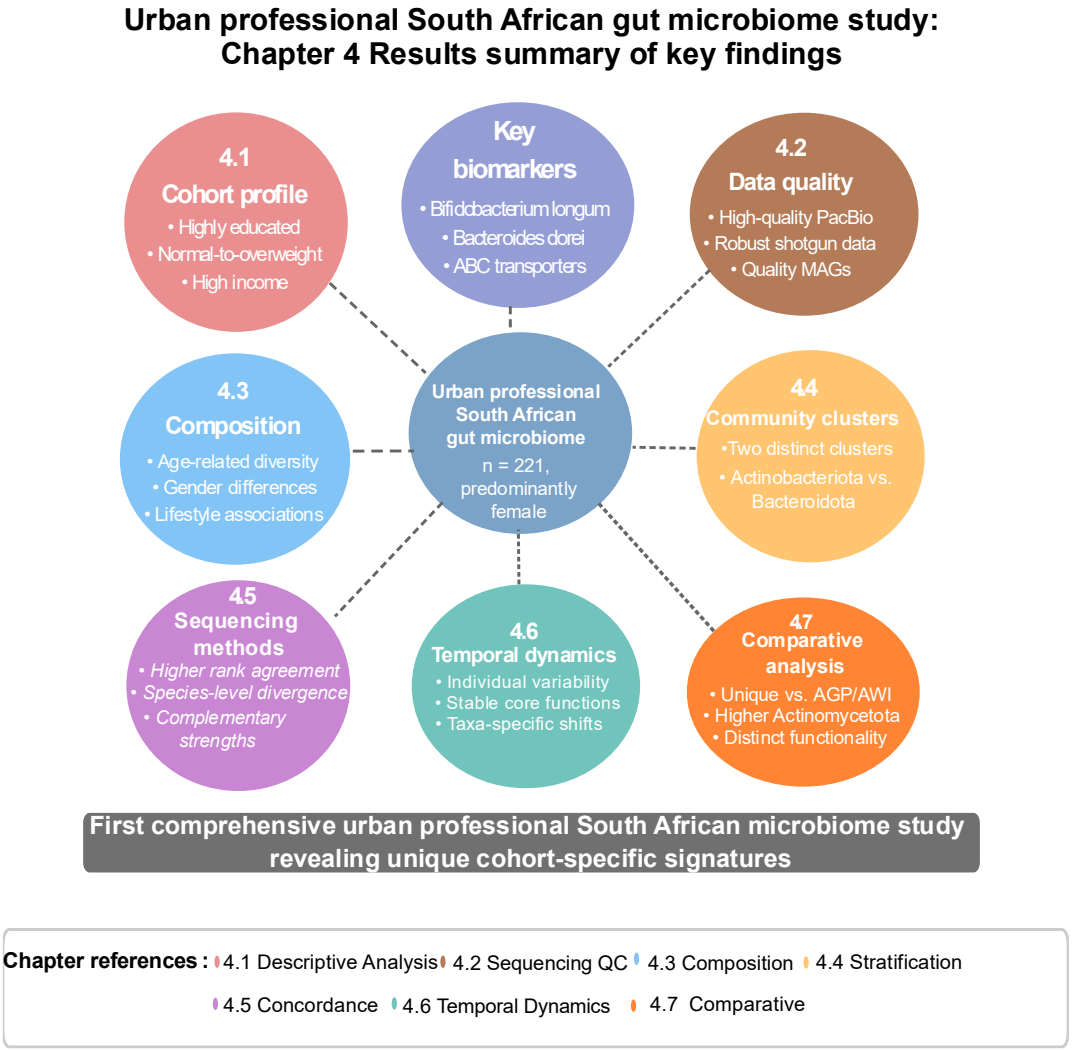


Figure 4.72 Summary of key findings from the urban professional South African gut microbiome study. The diagram illustrates the major results across all chapter sections (4.1-4.7), highlighting cohort characteristics, microbiome composition, community stratification, sequencing method concordance, temporal dynamics, and comparative analyses with other populations. This represents the first comprehensive characterization of the gut microbiome in an urban professional South African population, revealing distinct microbial signatures and demographic associations unique to this population.

5. Discussion

5.1. Overview of findings

This study presents a snapshot characterization of the gut microbiome in *n* = 221 urban professional South Africans (SAN) recruited from the CSIR Pretoria campus and University of the Witwatersrand (Johannesburg). The analysis of *n* = 101 variables, refined from *n* = 123 survey questions, demonstrated associations between demographic factors (*age group, gender, BMI group*, and *ancestry*) and various socioeconomic, dietary, lifestyle, and health-related factors. Influences on microbial community structure were revealed, with differential

abundance analysis identifying specific taxonomic signatures associated with both demographic factors and survey responses. Furthermore, clustering analysis revealed natural community stratification within this SAN population, associated with specific lifestyle factors such as COVID-19 diagnosis history and dietary habits.

The methodological comparison between 16S long-read and shotgun metagenomic sequencing highlighted important technical considerations for microbiome studies. While the approaches showed moderate concordance in functional profiling, systematic differences were observed in taxonomic resolution and community composition assessment, particularly at finer taxonomic levels. These findings highlighted the importance of considering methodological biases when contextualising the results against other cohorts and interpreting microbiome data from different sequencing platforms. The comparative analyses with shotgun metagenome-sequenced Western (AGP) and urban-rural South African (AWIGEN1) populations revealed distinct characteristics of the urban South African gut microbiome. Notable differences included lower phylum-level diversity, but higher Bacteroidetes abundance compared to other cohorts. Metagenome-assembled genome analysis identified cohort-specific taxonomic signatures, particularly within Actinomycetota-related (Actinobacteria) taxa.

Temporal analysis of the SAN population microbiome revealed substantial inter-individual variability in taxonomic composition, with only a limited core microbiome maintained across individuals over time. In contrast, functional profiles showed remarkable stability with 82% identified as core and stable functions, suggesting functional redundancy and resilience within the gut microbiome despite dynamic taxonomic changes.

These findings provided a foundation for understanding the unique features of the SAN gut microbiome and its relationship with various environmental and host factors, shaped by rapid urbanization, dietary shifts, and local health challenges. For example, unlike Western cohorts, SAN exhibits elevated *Bacteroidetes* and transient taxonomy with stable function – a resilience strategy potentially critical for South Africa's public health priorities. This specificity highlights the need to move beyond universal microbiome models toward locally tailored interventions and suggests potential targets for future investigations into the role of the microbiome in health and disease in this understudied population.

5.2. Methodological considerations and limitations

This study of the urban South African gut microbiome requires careful consideration of methodological limitations when interpreting results. Heterogeneous sequencing platforms, bioinformatic pipelines, and cohort designs introduce complexities examined below through technical, study design, and analytical perspectives.

Technical considerations

The study employed 16S long-read sequencing for the SAN population versus shotgun metagenomics for AGP and AWIGEN1 cohorts, introducing technical biases. While 16S offers cost-effective profiling with improved species-level classification, its resolution remains limited compared to shotgun metagenomics' strain-level functional insights (Johnson et al., 2019). This complicates cross-cohort comparisons, as 16S data infer function indirectly via tools like PICRUSt2, while shotgun data derive it directly from characterized protein families (Treiber et al., 2020). Differences between SILVA (16S) versus MetaPhlAn and GTDB-Tk (shotgun) classifications may skew taxonomic conclusions due to varying reference genome coverage (Balvočiūtė & Huson, 2017; Durazzi et al., 2021).

The 16S approach detected only ~7 species versus shotgun's 140, despite high correlation (ICC = 0.745) for overlapping taxa, masking 16S's inability to resolve >90% of diversity, particularly low-abundance taxa critical for niche adaptation (Asnicar et al., 2017). QIIME2's OTU/ASV clustering and the MAG pipeline's genome reconstruction differ fundamentally: OTU methods may over-split related taxa, while MAG methods risk collapsing strain-level variation (Yakobi & Nwodo, 2025). *Prevotella* dominance identified via 16S might mask functional sub-strains detectable only through shotgun sequencing.

Study design

The SAN population's size (n = 221) and urban homogeneity limited statistical power, particularly for sparse demographic subgroups. Collapsing variables into broader groups reduced granularity, increasing false discovery risks despite Benjamini-Hochberg correction (Lin & Peddada, 2020). For example, gas heating's association with *Desulfovibrio piger* may conflate air quality, socioeconomic status, and unmeasured variables (Vujkovic-Cvijin et al., 2020). Small microbiome samples inflate both Type I and II errors when testing multiple covariates with modest effect sizes (Ferdous et al., 2022).

The cross-sectional design precludes causal inference, as temporal dynamics were captured in only 22 longitudinal samples. While associations like transport mode's impact on diversity were identified, determining whether microbial shifts precede or result from lifestyle changes requires longitudinal frameworks (Vangay et al., 2018). The focus on urban professionals limits generalizability to broader South African populations with different microbiome-influencing exposures (Gacesa et al., 2022).

Analytical challenges

Multiple testing remains concerning despite FDR correction. Some significant associations (e.g., cosmetics and apoptosis pathways) may represent false positives, particularly for weakly associated variables like refuse disposal ($R^2 < 1.0 \times 10^{-2}$) where effect sizes approach

technical noise (Saseendran et al., 2019). Conversely, stringent correction may obscure biologically meaningful signals in smaller subgroups (Nearing et al., 2022).

Unmeasured confounders (e.g., air pollution, occupational chemicals) were incompletely addressed. Gas heating's association with microbial diversity might conflate air quality, temperature, and socioeconomic status (Fouladi et al., 2020; Kwak et al., 2024). Similarly, cosmetics usage and Actinobacteriota links may reflect chemical exposure rather than direct microbial selection, unresolved by multivariate adjustments (Aldirawi & Morales, 2023; Chiu et al., 2020; Rio et al., 2024).

The 16S sequencing provided comparable taxonomy at higher levels but resolution declined substantially at finer levels, classifying only 132 ASV-associated bacteria versus shotgun's 608 species, with ~6% more unclassified taxa. This limitation reflects incomplete representation of African microbiomes in reference databases (Maghini et al., 2025), emphasizing the need for shotgun metagenomics and improved databases for African microbial diversity.

Sparse metadata and cohort heterogeneity complicated cross-study interpretations and AGP/AWI's limited descriptors restricted confounding adjustment. Methodological differences between 16S and shotgun sequencing hindered direct comparisons (Truong et al., 2015) and imputation methods were limited by absent foundational metadata (Costea et al., 2017). Significant PICRUSt pathways may reflect SAN divergence from Chinese cohort dysbiosis or athlete lifestyle extremes rather than purely SAN features, highlighting sensitivity to cohort-specific subpopulations without sufficient metadata (Lampeter et al., 2023).

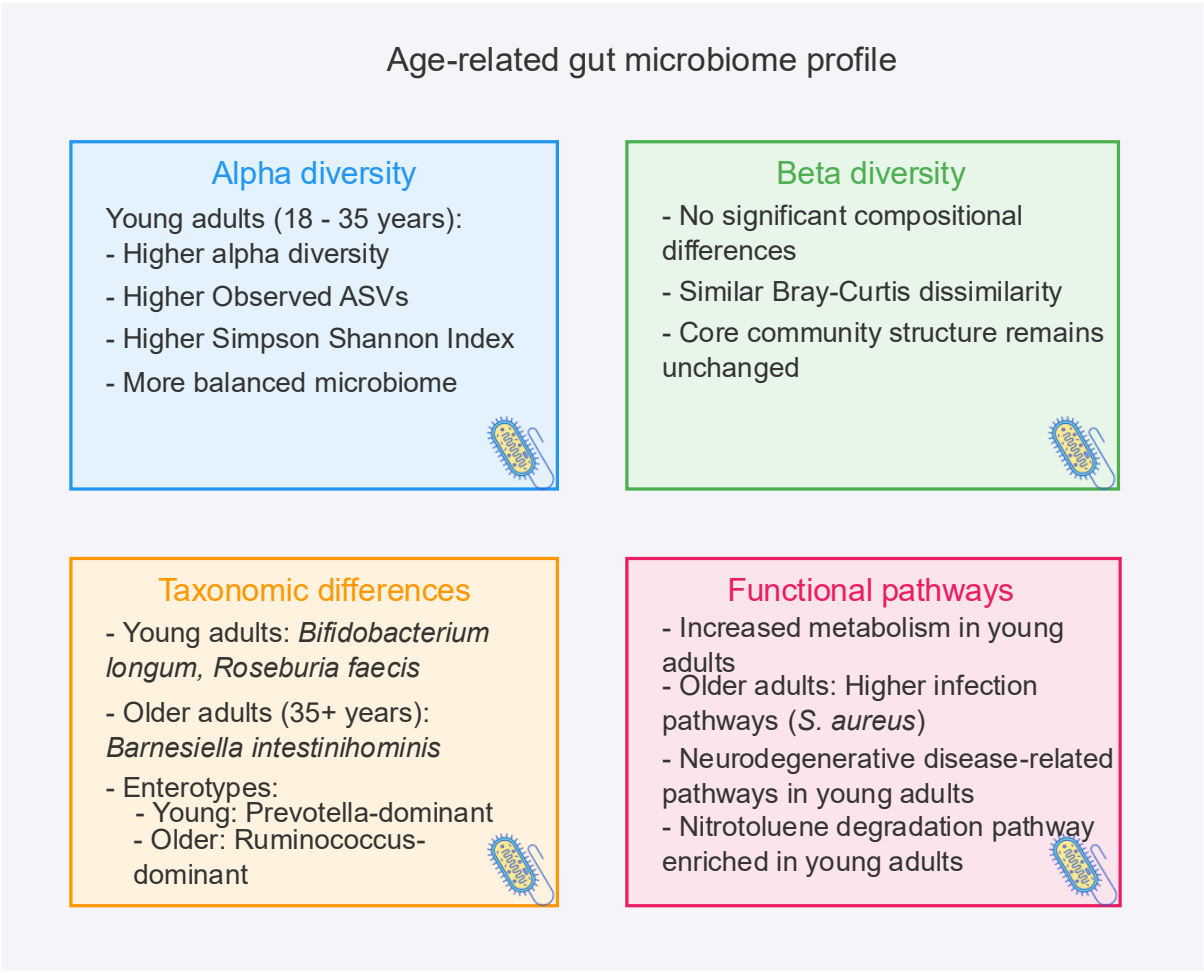
Despite these limitations, this work contributes meaningful insights into urbanization and gut microbiome composition in the African context, particularly within underrepresented urban professional South Africans. Future research would benefit from standardized protocols, expanded shotgun metagenomics, longitudinal sampling, and causal inference modelling.

5.3. Demographic drivers of microbiome composition

The characterization of the gut microbiome in the SAN population revealed distinct patterns of microbial diversity and composition that reflected the complex interplay with host demographics. This investigation uncovered several key findings that warranted detailed interpretation.

i. Age-related patterns

Figure 5.1 Age-related gut microbiome findings showing key differences in microbial diversity, taxonomic composition, and functional pathways between age groups.



Diversity

The *age group* analysis revealed significant associations between age and microbial alpha diversity. Older adults exhibited lower microbial richness and evenness compared to younger adults across Observed OTUs (ASVs), Simpson, and Shannon diversity indices (Figure 4.1). Young adults demonstrated higher numbers of unique bacterial species, better species evenness, and more complex, balanced microbial ecosystems. This finding aligns with published research showing alpha diversity peaks in younger adults and plateaus around age 40 (de la Cuesta-Zuluaga *et al.*, 2019).

In contrast, when examining community composition using Bray-Curtis dissimilarity, no statistically significant differences were observed between young and older adults (PERMANOVA p-value > 5.0×10^{-2}). The absence of distinct clustering in beta diversity ordination confirmed that age groups maintained similar taxonomic profiles at the community level despite differences in diversity metrics (Yan, Qin, Yan, et al., 2022). These findings indicate young adults maintain a more diverse gut microbiome than older adults, even as fundamental community structure remains comparable.

Differential taxonomy

Differential abundance analysis using LEfSe on the unrarefied dataset (Figure 4.9) identified three bacterial species significantly different between age groups: *Bifidobacterium longum* and *Roseburia faecis* enriched in young adults, and *Barnesiella intestinihominis* in older adults. The enrichment of *B. longum* in young adults corresponds with literature showing higher abundance in younger populations, though this species persists across all ages (0 – 104 years), reflecting its lifelong role in gut ecology (Arboleya et al., 2016; Kato et al., 2017; Odamaki et al., 2016). *B. longum* produces beneficial bioactive molecules like short-chain fatty acids and associates with higher gut microbiome diversity (Ladeira et al., 2023; Mills et al., 2023). *R. faecis*, a butyrate-producing anaerobe enriched in young adults, offers anti-inflammatory properties and potential immunity benefits (Tamanai-Shacoori et al., 2017). Elderly individuals generally exhibit decreased butyrate production capacity compared to younger individuals (Hippe et al., 2011).

Young adults exhibited *Prevotella*-dominant enterotypes while older adults showed Ruminococcus-dominant patterns. Age-related enterotype associations vary considerably across studies, with some reporting decreased *Prevotella* in elderly populations (Li et al., 2022; Pérez Martínez et al., 2014), while others found no consistent age-specific patterns (Jeffery et al., 2016). The coexistence of species-level differences alongside similar community-wide structures demonstrates how microbiome variations manifest at different analytical scales, highlighting the importance of multiple analytical approaches to fully understand microbial community dynamics (Kers & Saccenti, 2022).

Functional pathways

Functional analysis revealed significant age-related differences in gut microbiome pathway abundance. Three key pathway classes – “infectious disease (bacterial)," "neurodegenerative disease," and "xenobiotic biodegradation and metabolism" – were differentially abundant between age groups. The most significant pathway, "Nitrotoluene degradation" (adjusted p-value = 1.0×10^{-2}), associated with the *lacA* gene (galactoside O-acetyltransferase), was enriched in young adults. Studies in Fulani and Jarawa populations linked nitrotoluene degradation pathways to dietary and lifestyle variations (Afolayan et al., 2019), while research on microplastic exposure highlighted xenobiotic degradation pathways indicating exposure to toxic compounds and altered energy metabolism (Blot et al., 2023; X. Zhang et al., 2022).

In paediatric populations, nitrotoluene degradation pathways appear enriched in children with autism spectrum disorder (ASD), alongside metabolic pathways related to endocrine-disrupting chemicals (Bhusri et al., 2025). The *lacA* gene, linked to carbohydrate metabolism and adaptation to environmental stressors, suggests connections between chemical

exposure, gut microbiota dysbiosis, and neurodevelopmental disorders. Compounds like p-nitrotoluene, associated with hyperactivity in animal models, may have behavioural implications for individuals with ASD (Bhusri et al., 2025).

Young adults showed enriched pathways linked to metabolism and neurodegenerative diseases, while older adults exhibited stronger associations with *Staphylococcus aureus* infection pathways. Younger individuals often harbour microbiomes optimized for metabolic flexibility, reflecting urban dietary patterns high in processed foods (Ravikrishnan et al., 2024; Wang et al., 2024). Studies show younger adults exhibit higher microbial diversity and energy metabolism pathways supporting gut health (Khaledi et al., 2024; Ravikrishnan et al., 2024). The enrichment of neurodegenerative-related pathways in young adults mirrors evidence that early-life microbiome alterations may prime long-term neurological risks. Gut microbiota dysbiosis in younger populations can produce metabolites like lipopolysaccharides and short-chain fatty acids influencing neuroinflammation through the gut-brain axis (Fu et al., 2024; Liu et al., 2024). Specific microbial genera like *Bacteroides*, enriched in younger adults, associate with pathways linked to Alzheimer's and Parkinson's diseases (Liu et al., 2024; Ravikrishnan et al., 2024).

The *S. aureus* infection pathways in older adults likely reflect age-related immunosenescence and urban comorbidities increasing susceptibility to opportunistic pathogens. Older adults often exhibit decreased microbial diversity, increased pathobionts, and reduced beneficial taxa crucial for gut barrier integrity (Chen & Boyle, 2024; Ravikrishnan et al., 2024). These changes worsen with frailty, chronic inflammation, and medication use (Khaledi et al., 2024; Pu et al., 2024).

South Africa's unique dietary transitions and disease burden warrant deeper exploration to disentangle environmental, lifestyle, and aging interactions (Ravikrishnan et al., 2024; Wang et al., 2024). Functional pathways PCA revealed considerable overlap between age groups, indicating shared core microbial functions across demographics. This suggests factors beyond age – dietary habits, medication use, environmental exposures – may significantly influence microbial community functions (Ghosh et al., 2022; Shanahan et al., 2021). The lack of distinct separation between age groups highlights the complexity of microbiome-host interactions and the need to consider these factors beyond age when interpreting results (Wilmanski et al., 2021).

These findings contribute to understanding the gut microbiome's role in age-related health outcomes and provide a foundation for exploring how additional demographic factors, such as gender, shape microbial community composition and function.

ii. Gender-related patterns

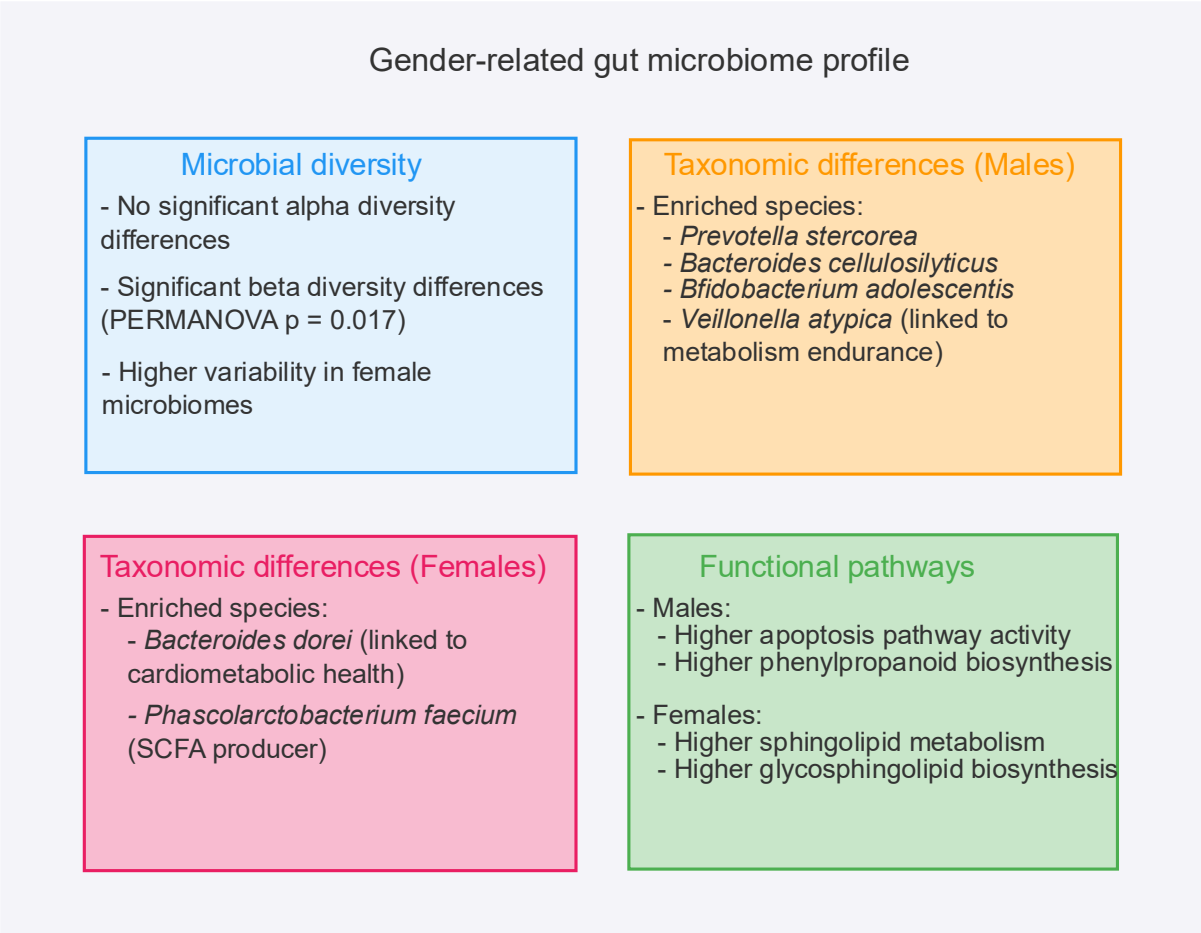


Figure 5.2 Gender-related findings revealing differences in gut microbiome composition and functionality between males and females.

Diversity

Gender-related gut microbiome analysis revealed no significant differences in alpha diversity indices (Shannon, Observed OTUs, Simpson), indicating comparable species richness and evenness within individual samples. However, beta diversity analysis showed significant differences in microbial community structure between genders (PERMANOVA $p = 1.7 \times 10^{-2}$). This finding potentially suggests either systematic compositional differences between males and females, or greater within-group variability in one gender. PERMANOVA alone cannot distinguish between these possibilities, but the lack of distinct clustering in ordination plots and prior evidence suggest sex-related differences may stem from increased variability among females rather than consistent divergence from males (Kim et al., 2020). This aligns with hypotheses that hormonal fluctuations or lifestyle diversity in females could drive broader inter-individual differences in microbial communities (Santos-Marcos et al., 2023).

Differential taxonomy

Differential abundance analysis identified six species enriched in males: *Prevotella stercorea, Bacteroides cellulosilyticus, Bifidobacterium adolescentis, Massiliprevotella massiliensis,*

Dialister succinatiphilus, and *Veillonella atypica. Prevotella stercorea*, with unique carbohydrate-metabolizing enzymes (Yeoh et al., 2022), aligns with *Prevotella* enterotype dominance in males (Del Castillo-Izquierdo et al., 2022). *Massiliprevotella massiliensis* has been linked to systemic T cell activation in immunocompromised males (Armstrong et al., 2018; Ndongo et al., 2017), though with limited cases in our cohort, this association remains speculative.

Bacteroides cellulosilyticus enrichment may reflect both androgen influences on *Bacteroides* abundance (Ma et al., 2021) and dietary factors, as males were overrepresented in non-omnivorous diets. This aligns with findings that older men on plant-based diets exhibit elevated *B. cellulosilyticus* levels (Li et al., 2021), suggesting interplay between hormonal regulation and dietary habits in shaping microbial communities (Feraco et al., 2024).

Bifidobacterium adolescentis, enriched in males, produces gamma-aminobutyric acid (GABA) (Duranti et al., 2020) and may modulate the gut-brain axis. Despite males reporting higher rates of mental health issues, *B. adolescentis* enrichment suggests complex interactions between gender-associated factors and microbial colonization (Argentini et al., 2024; Ladeira et al., 2023).

Dialister succinatiphilus, a succinate-metabolizing bacterium, has been negatively correlated with glycaemic response and linked to inflammatory bowel disease, potentially prolonging succinate accumulation (Anthamatten et al., 2024; Portincasa et al., 2024). *Veillonella atypica* shows contrasting roles – associated with obesity-related metabolic risks in men but beneficial in athletes by converting lactate to propionate for enhanced endurance (Haro et al., 2016; Lundberg et al., 2021; Scheiman et al., 2019).

In females, *Bacteroides dorei* and *Phascolarctobacterium faecium* were enriched, with *Bacteroides*-dominant enterotypes observed. *B. dorei* is implicated in cardiometabolic health (Yoshida et al., 2018), while showing stable abundance across BMI ranges in women compared to men (Haro et al., 2016). *Bacteroides* species are correlated to female-specific conditions, though *B. dorei's* direct role needs further investigation (Siddiqui et al., 2022; Xiao et al., 2024).

Phascolarctobacterium faecium produces beneficial short-chain fatty acids from succinate (Rachek et al., 2021; Wu et al., 2017). While enriched in females with metabolic syndrome (Santos-Marcos et al., 2019), its efficient succinate metabolism contrasts with male-associated *D. succinatiphilus* (Anthamatten et al., 2024; Ogata et al., 2019). This suggests sex-specific succinate-handling strategies: females may favour rapid conversion to anti-inflammatory SCFAs, while males' reliance on slower *Dialister* species may increase inflammatory risks.

Functional pathways

Functional pathway analysis revealed significant gender-specific differences, with the relative abundance of most pathways dominating in males. The most significant pathway was "Apoptosis" (adjusted $p = 8.0 \times 10^{-4}$), critical for tissue homeostasis and immune regulation (Tang et al., 2021). Microbial modulation of apoptosis influences epithelial permeability, with pathogens exerting stronger pro-apoptotic effects than probiotics or commensals (Altonsy et al., 2010; Buret & Bhargava, 2014). Descriptive statistics showed males were more likely to use probiotics, potentially influencing gut apoptotic pathways. While probiotics can modulate apoptosis with therapeutic benefits, excessive apoptosis may contribute to epithelial dysfunction (Daniluk et al., 2012; Wang et al., 2021).

Males also showed increased abundance in "Phenylpropanoid biosynthesis" (adjusted $p = 4.4 \times 10^{-3}$), involving dietary antioxidants with structural and signalling functions (Navarre et al., 2022). This pathway highlights bidirectional interactions between polyphenols and gut microbiota, where microbes metabolize polyphenols into bioactive compounds affecting host health (Espín et al., 2017). Phenylpropanoid-related pathways have been linked to prostate cancer risk (Fujita et al., 2023), suggesting potential roles in male-specific processes.

Females exhibited higher abundance in "Glycosphingolipid biosynthesis - ganglio series" (adjusted $p = 4.1 \times 10^{-2}$) and "Sphingolipid metabolism" (adjusted $p = 1.7 \times 10^{-2}$). These pathways are central to lipid metabolism and linked to conditions like polycystic ovary syndrome, where gut microbiota imbalances correlate with dysregulated metabolism (Zhu et al., 2021). Sphingolipids, produced by bacteria like *Bacteroides*, influence host metabolic pathways including ceramide levels and insulin sensitivity (Johnson et al., 2020). In females, bacterial-derived sphingolipids may modulate hepatic lipid synthesis and glucose homeostasis, as demonstrated in mouse models where *Bacteroides thetaiotaomicron* reduced hepatic sphingolipid production (Johnson et al., 2019).

These gender-related differences highlight sex as a critical variable in microbiome research. While males and females share core microbial activities, distinct functional enrichments describe how sex-specific factors – hormonal profiles, dietary habits, immune responses – shape the gut ecosystem, emphasizing the need for sex-stratified approaches to develop tailored therapeutic strategies (Valeri & Endres, 2021).

iii. BMI group-related patterns

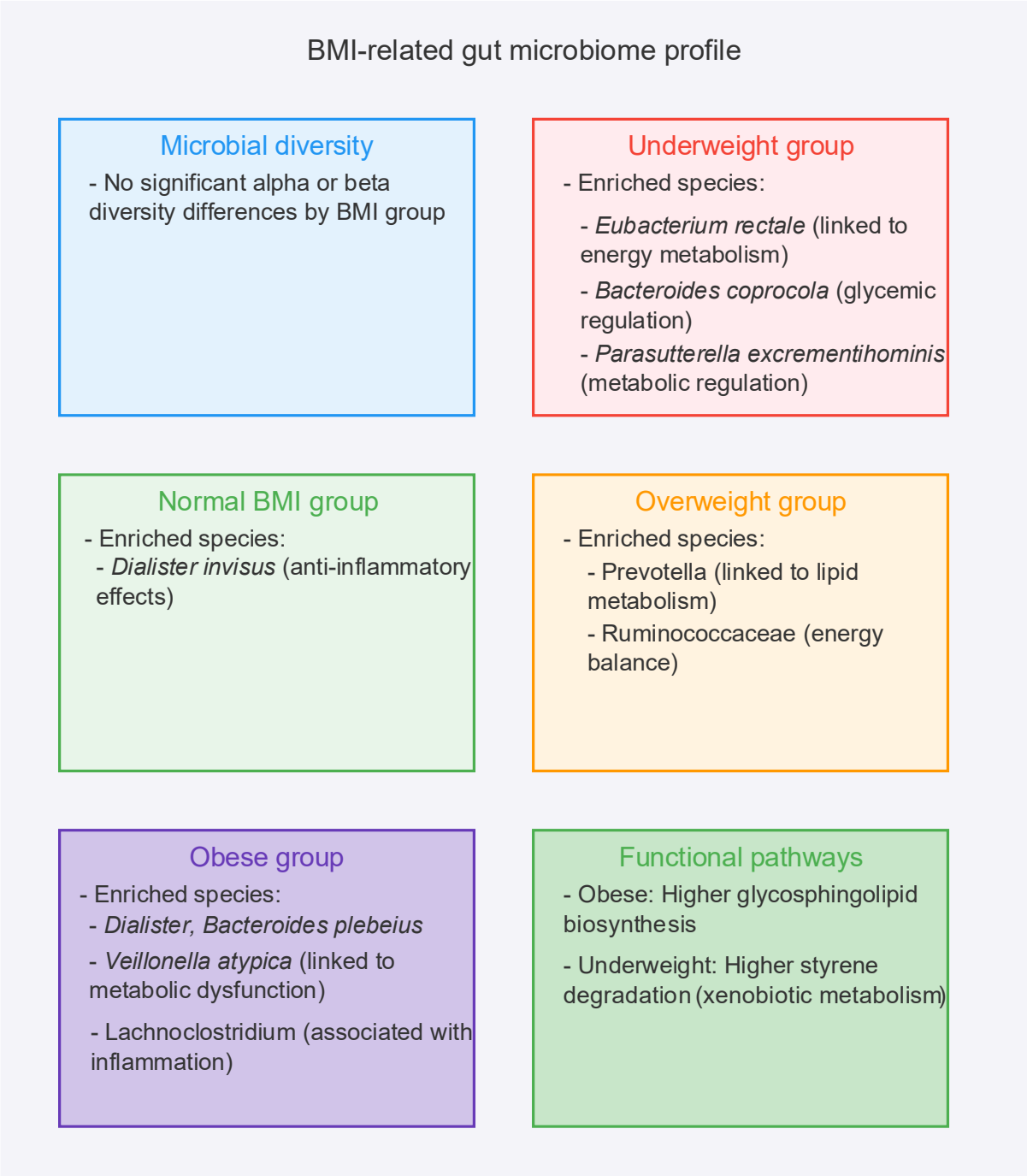


Figure 5.3 BMI-associated variations in gut microbiome characteristics showing relationships between body mass index groups and microbial communities.

Diversity and differential taxonomy

BMI-related gut microbiome analysis revealed distinct patterns while showing no significant differences in diversity indices. The underweight group exhibited a *Prevotella/Bacteroides* enterotype with six differentially abundant species: *Eubacterium rectale, Bacteroides coprocola, Parasutterella excrementihominis, Faecalicoccus pleomorphus, Alistipes indistinctus*, and *bacterium YE57*. Normal weight subjects showed one differentially abundant

species*, Dialister invisus*. The overweight category demonstrated enrichment of *Prevotella* metagenome and unclassified Ruminococcaceae, while the obese group showed increased abundance of *Dialister* gut metagenome, *Bacteroides plebeius, Bacteroides xylanisolvens, Veillonella atypica*, unclassified *Streptococcus*, and uncultured Lachnoclostridium.

E. rectale enrichment in underweight individuals suggests roles in energy regulation, negatively correlating with ghrelin – a hunger-stimulating hormone – potentially suppressing appetite signalling (Komodromou et al., 2024). Its association with body fat distribution parameters indicates influence on lipid metabolism. *B. coprocola* inversely associates with glycaemic responses to artificial sweeteners (Feng et al., 2024) and links to SNPs in type 2 diabetes, particularly in glycosyl hydrolase genes (Chen et al., 2017), potentially contributing to reduced energy extraction in underweight individuals.

P. excrementihominis presents contrasting associations – linked to fatty acid biosynthesis and weight gain in obesity (Henneke et al., 2022) yet elevated in underweight individuals, suggesting context-dependent metabolic roles. Its elevation in Crohn's disease indicates potential adaptation to altered gut environments (Yan et al., 2023). *F. pleomorphus* has appears in undernutrition-associated microbiomes with reduced abundance in breast cancer patients with delayed diagnosis (Kamil et al., 2021; Nguyen et al., 2024). *A. indistinctus* shows reduced abundance in liver cirrhosis and healthier metabolic profiles in obesity studies (Guohong-Liu et al., 2019; Phuong-Nguyen et al., 2024), while *bacterium YE57* co-occurs with *Alistipes* and Lachnospiraceae, suggesting roles in metabolic adaptations (Lee et al., 2023).

D. invisus enrichment in normal BMI subjects suggests mediation between prudent diet patterns and anti-inflammatory biomarkers like adiponectin and IL-6 (Shi et al., 2022). Its sensitivity to antibiotics like ciprofloxacin (Patangia et al., 2022) highlights environmental responsiveness, potentially reflecting balanced gut ecosystems supporting metabolic health. *Prevotella* and Ruminococcaceae enrichment in overweight individuals, particularly in Black populations, links to altered lipid metabolism and inflammation. *Prevotella* positively associates with triglycerides and hs-CRP (Hu et al., 2015), with stronger BMI associations in Black populations (Stanislawski et al., 2019).

In obese individuals, *Dialister* enrichment appears consistent across studies, including increased abundance in rheumatoid arthritis patients and pre-pregnancy obese women (Mena-Vázquez et al., 2023; Sugino et al., 2019). *B. plebeius*, despite associations with normal weight and lower visceral fat (Chen, et al., 2022), may reflect strain-specific adaptations in carbohydrate metabolism (Hehemann et al., 2010). *B. xylanisolvens* shows enrichment in obese individuals across diverse populations (Aljazairy et al., 2022; Liu et al., 2022). *V. atypica* and *Streptococcus* demonstrate significant abundance in obese individuals

with gender-specific associations (Haro et al., 2016; Pinart et al., 2021), while Lachnoclostridium shows variable obesity links (Amadou et al., 2016; Salazar-Jaramillo et al., 2024).

Functional pathways

Functional pathway analysis identified two significant pathways differentially abundant across BMI groups: "Glycosphingolipid biosynthesis – globo and isoglobo series" and "Styrene degradation" (adjusted $p = 4.0 \times 10^{-2}$). The glycosphingolipid pathway, more abundant in obese subjects, synthesizes complex lipids critical for cell membrane structure and signalling (Cumin et al., 2021). This pathway positively associates with elevated cholesterol and supplement use (Manor et al., 2020), suggesting links to metabolic dysregulation. Mouse studies show glycosphingolipid biosynthesis enrichment in high-fat diet groups, correlating with increased lipid metabolism and inflammatory markers (Portela et al., 2023). Upregulation in obese individuals may reflect altered lipid handling contributing to obesity's pro-inflammatory and metabolic disturbances.

The "styrene degradation" pathway enrichment in underweight individuals suggests roles in xenobiotic metabolism and gut adaptation. Though typically associated with high BMI and gut dysbiosis (Yin et al., 2024; Zhu et al., 2024), its presence in underweight subjects may indicate distinctive microbial strategies for handling toxins or metabolites. This pathway negatively associates with bile acids like deoxycholic acid and 7-ketolithocholic acid (Zhang et al., 2024), which link to obesity and insulin resistance. In underweight individuals, this upregulation could represent protection against metabolic disturbances or adaptation to low energy availability.

Method comparison revealed BMI group showed strongest associations with pathway abundances across both HUMAnN3 ($R^2 = 9.4 \times 10^{-2}$) and PICRUSt2 ($R^2 = 8.9 \times 10^{-2}$), finding robustness and BMI's importance as a determinant of microbial functional capacity (Douglas et al., 2020). These differences in pathway abundances suggest metabolic health significantly influences gut microbiome functional potential, affecting nutrient metabolism, detoxification, and inflammation processes.

These findings demonstrate the dynamic interplay between metabolic status and gut microbiome. While overall microbial diversity remains stable across BMI groups, specific taxa and functional pathways exhibit significant variations reflecting adaptations to differing energy availability and metabolic demands, providing a foundation for further exploration of microbiome-metabolic health interactions.

iv. Ancestry-related patterns

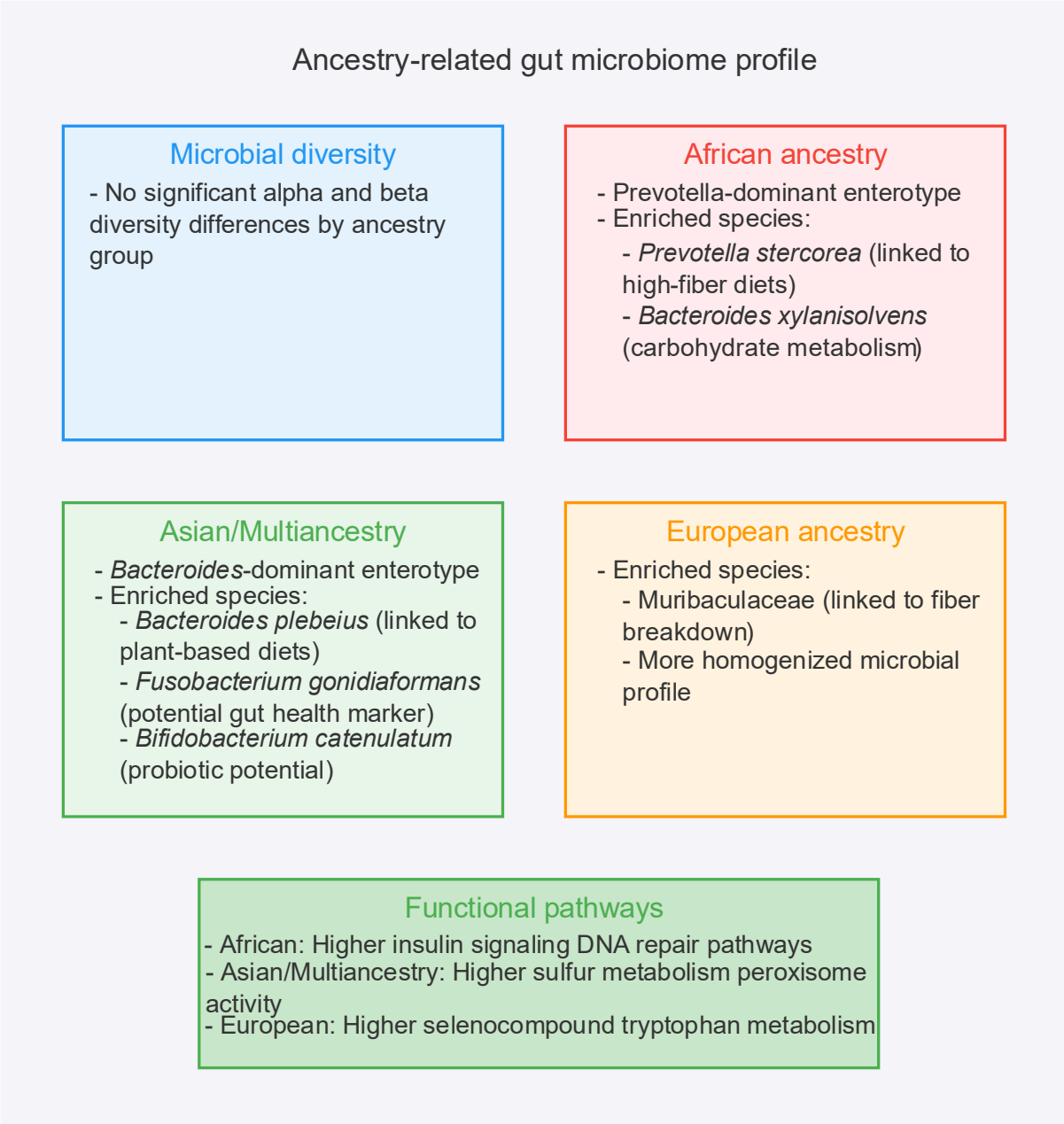


Figure 5.4 Ancestry-related patterns in gut microbiome composition demonstrating differences across various ancestral groups.

Diversity and differential taxonomy

Differential abundance analysis revealed the African group exhibited two differentially abundant species: *Prevotella stercorea* and *Bacteroides xylanisolvens*, with a *Prevotella* enterotype dominated by 13 genera. *P. stercorea* enrichment aligns with populations consuming high-fibre, plant-based diets characteristic of non-industrialized lifestyles (de Goffau et al., 2022; Lozupone, Stombaugh, et al., 2013). In adults, *P. stercorea* links to carbohydrate-rich diets with unique metabolic capabilities including carbohydrate esterases but lacking some hemicellulose-degrading enzymes (Yeoh et al., 2022).

B. xylanisolvens enrichment further highlights dietary and lifestyle influences on microbial composition. Though associated with obesity in some populations (Aljazairy et al., 2022; Liu et al., 2022), its presence may reflect strain-specific adaptations in carbohydrate metabolism. The persistence of *Prevotella*-dominant enterotypes in this adult African cohort despite urbanization suggests resilience of traditional dietary patterns, though increasing *Bacteroides* dominance in urbanized populations raises concerns about potential loss of microbial diversity linked to metabolic health (van der Vossen et al., 2023).

The Asian/Multiancestry/Other group exhibited several differentially abundant species: *Bacteroides plebeius, Fusobacterium gonidiaformans, Bifidobacterium catenulatum, Clostridium glycyrrhizinilyticum, Collinsella intestinalis*, and *Lactobacillus plantarum*, with a *Bacteroides* enterotype dominated by 15 genera. *B. plebeius* enrichment aligns with Asian ancestry and dietary habits (Peters et al., 2020), potentially showing dual health roles – beneficial in chronic kidney disease (Pei et al., 2022) but potentially detrimental in acute pancreatitis (Qu et al., 2024). *F. gonidiaformans* has been implicated in postpartum bacteraemia (Vidaurrazaga et al., 2020), while *B. catenulatum* metabolizes plant-derived glycans reflecting dietary adaptations (X. Liu et al., 2021b). *C. intestinalis* increases with processed food components (Wolf et al., 2019), with links to higher insulin levels and lower fibre intake (Gomez-Arango et al., 2018). *L. plantarum* demonstrates probiotic properties with better colonization in *Prevotella*-dominated microbiomes (Echegaray et al., 2023; Isenring et al., 2022). These patterns highlight ancestry as a proxy for lifestyle and dietary variations even within urban communities (Dwiyanto et al., 2021).

The European group exhibited only one differentially abundant species: uncultured Muribaculaceae, which degrades complex carbohydrates and produces SCFAs from dietary fibres (Zhu et al., 2024). This limited diversity contrasts with more varied profiles in other ancestry groups, likely reflecting the homogenizing effects of urbanization and Westernized lifestyles (Yatsunenko et al., 2012).

Functional pathways

Functional pathway analysis identified three significant pathways: "Selenocompound metabolism," "Tryptophan metabolism," and "Nitrotoluene degradation" (adjusted $p = 1.7 \times 10^{-2}$), associated with enzymes gentisate 1,2-dioxygenase, cysJ (sulfite reductase), and lacA (galactoside O-acetyltransferase), respectively.

Selenium metabolism involves gut microbiome conversion of this essential trace element into bioactive selenocompounds like selenoproteins, critical for reducing oxidative stress and modulating immunity (Kieliszek & Błażejak, 2016). Ancestry-specific variations may influence selenium metabolism as dietary intake varies across populations with different traditional diets

(Fairweather-Tait et al., 2011). Tryptophan metabolism produces metabolites like indole, indole-3-lactic acid, and indole-3-propionic acid that regulate immune responses, gut barrier integrity, and neuro-behavioural functions (Agus et al., 2018). Ancestry-related microbiota differences, such as *Prevotella* enrichment in African populations or *Bacteroides* in European populations, can lead to variations in tryptophan metabolite production, with *Prevotella*-dominant microbiomes linked to higher anti-inflammatory metabolites (Roager & Licht, 2018). Nitrotoluene degradation detoxifies environmental pollutants, varying significantly across populations based on exposures (Claus, 2014). Ancestry-specific adaptations, such as *Clostridium* enrichment, may enhance xenobiotic metabolism, potentially protecting against environmental toxins (Koppel et al., 2017). These ancestry-specific profiles persist despite urbanization (Filippo et al., 2010; Yatsunenko et al., 2012), influencing functional capacity across metabolic pathways.

The African ancestry group showed higher abundance in "Insulin signalling," "DNA replication," and "Homologous recombination" pathways. Insulin signalling, influenced by gut microbiome metabolites like SCFAs, modulates host metabolism (Canfora et al., 2015). Retention of *Prevotella*-dominant features despite urbanization may sustain SCFA production, potentially mitigating metabolic dysregulation in urban settings (van der Vossen et al., 2023; Vinogradova et al., 2024). DNA replication and homologous recombination pathways might reflect adaptations to urban stressors, with metabolites like butyrate potentially enhancing oxidative stress resilience (Krautkramer et al., 2021).

The Asian/Multiancestry/Other group showed enrichment in "Alanine, aspartate, and glutamate metabolism" and "Peroxisome" pathways, potentially reflecting plant-based proteins and fermented foods increasingly common in urbanizing populations (LeBlanc et al., 2024). Glutamate metabolism, critical for nitrogen balance (Wu et al., 2011), might align with dietary protein diversity or stem from pollution-induced oxidative stress and peroxisome pathways may buffer against processed diets or pollution (van der Vossen et al., 2023; Wanders et al., 2016). Sulphur metabolism links to both dietary intake and environmental pollutant detoxification. Sulphur-metabolizing bacteria influence colorectal cancer risk, particularly with high-sulphur diets (David et al., 2014). In urban South Africa, this pathway's enrichment may reflect traditional dietary practices and urban pollutant exposure (Koppel et al., 2017). Protein digestion pathway prominence aligns with dietary patterns prioritizing animal or legume proteins, though excessive protein fermentation can produce harmful metabolites (Wu et al., 2011; Krautkramer et al., 2021).

These findings emphasize the importance of contextualizing microbiome research within ancestry, diet, and urbanization interactions for advancing precision medicine. Future

research should explore strain-level analysis and track how urbanization affects ancestral microbial traits in diverse populations (David et al., 2014; van der Vossen et al., 2023).

5.4. Natural community stratification suggests underlying biological differences

The SAN population's division into Actinobacteriota-dominant (Cluster 1) and Bacteroidota-dominant (Cluster 2) clusters highlights the interplay between urban environmental exposures, dietary habits, and microbial ecology. Cluster 1's Actinobacteriota enrichment was observed in individuals from the CSIR campus, a site characterized by industrial research facilities. Actinobacteriota are known for their metabolic versatility in xenobiotic degradation (Carvalho et al., 2022; Kontomina et al., 2022), which may confer resilience in environments with bioaerosols or airborne particulates from industrial activity (Sun et al., 2023; Zampolli et al., 2024). This aligns with findings that urban microbiomes adapt to anthropogenic stressors through functional redundancy (Zampolli et al., 2024). While the CSIR cohort includes a greater proportion of older adults and occupational diversity, causal links between specific environmental factors and Actinobacteriota dominance require further investigation and validation.

Conversely, the dominance of Bacteroidota in Cluster 2 potentially reflects a dietary pattern associated with Westernized diets, which are typically high in processed foods (Clemente-Suárez et al., 2023). This dietary characteristic may contribute to reduced mucosal protection and increased inflammation, raising concerns for health management in urban settings. Additionally, the presence of Proteobacteria and Desulfobacterota in this cluster highlights the impact of low-fibre, high-fat diets that are prevalent in urbanized South African populations, leading to dysbiotic states that could exacerbate gastrointestinal disorders (Ramaboli et al., 2024; Z. Zhao et al., 2024).

The variation in microbial composition between different urban locations characterised in the Clusters, such as CSIR in Pretoria and academic settings in Johannesburg, further illustrates the potential significant role of microenvironmental factors. These findings may suggest that local environmental conditions, beyond general urbanization, critically influence gut microbiota profiles and, consequently, health outcomes (Rio et al., 2024). Additionally, Cyanobacteria have been identified as one of the phyla present during grape must fermentation in wine production, though their role in human gut microbiota following red wine consumption requires further investigation to establish clear relationships (Piao et al., 2015). The critical role of these microbial taxa in response to urban environmental pressures, dietary habits, and lifestyle choices accentuates the need for exploring targeted interventions, aimed at reducing urbanization-related exposures to support the resilience of beneficial microbes.

5.5. Implications of diet, lifestyle, SES and health factors

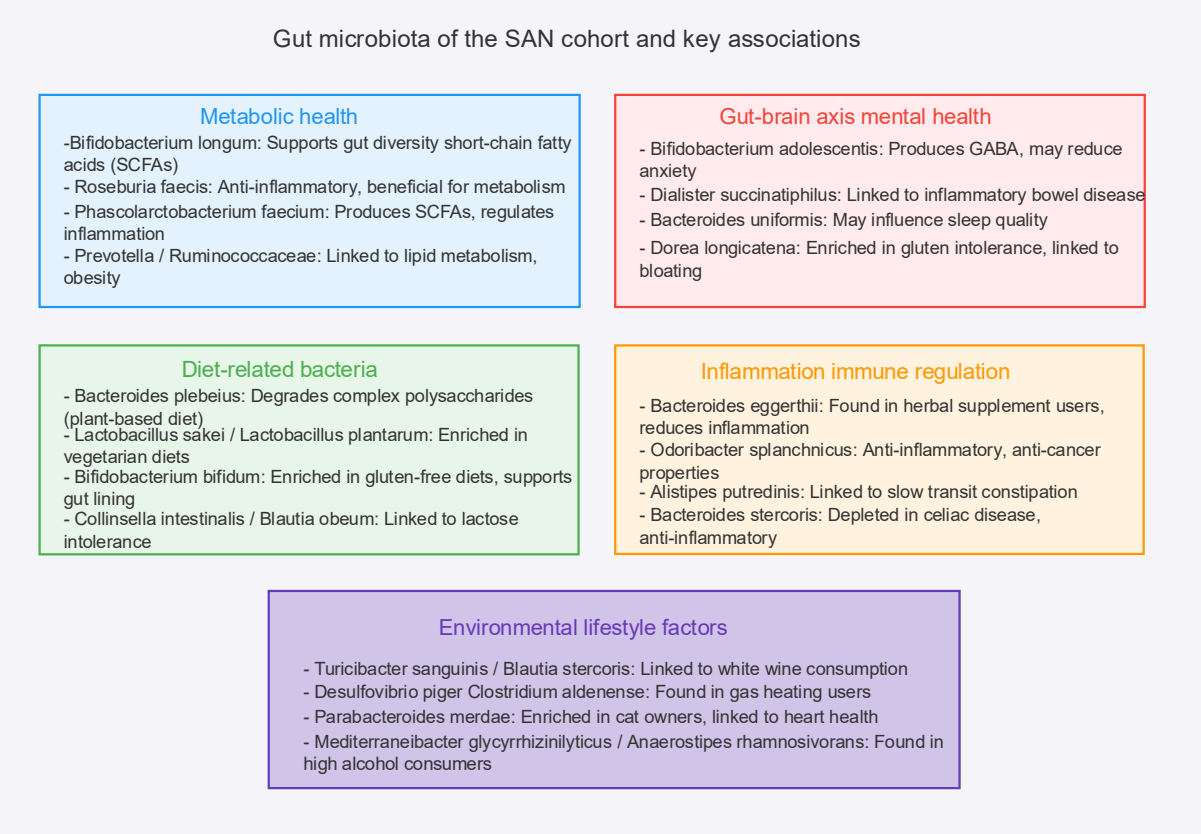


Figure 5.5 Key gut microbiome findings and associations of taxa with diet, lifestyle, socioeconomic status and health variables.

i. SES variables

In our urban professional South African population, SES variables differentially associated with gut microbial taxa. *Dialister invisus*, typically linked to lower SES in high-income countries (Kwak et al., 2024), was paradoxically enriched in higher-income individuals, suggesting context-specific dietary habits may drive this reversal. *Fusobacterium gonidiaformans*, previously associated with low SES in Chinese populations (Yeoh et al., 2020), aligned with lower-income groups, potentially reflecting delayed infection management. *Anaerostipes caccae* was depleted in lower-SES individuals, consistent with global patterns where processed diets dominate (Li et al., 2024).

Household income emerged as the strongest predictor of microbial functional variation in PICRUSt2 (R^2 = 9.5 × 10^{-2}, p-value = 2.8 × 10^{-2}), though this association didn't persist after FDR correction (FDR = 0.2). The stronger association in PICRUSt2 versus HUMAnN3 suggests methodological differences in capturing income-related shifts (Douglas et al., 2020; Kwak et al., 2024). The "middle income males" subgroup drove the relationship between

income and gender, potentially reflecting reduced dietary diversity in lower income groups (Bowyer et al., 2019).

Public transport users exhibited significantly higher Shannon diversity (p-value = 1.5×10^{-2}) with enrichment of *Clostridium scindens* and *Phascolarctobacterium succinatutens* (p-value = 7.7×10^{-3}). In low-income households, transport costs may redirect funds from fresh to processed foods, favouring *C. scindens* – a bile acid transformer linked to high-fat metabolism (Knipe & Krygsman, 2024; Ridlon et al., 2015). *P. succinatutens*, a succinate-utilizing commensal, relies on microbially derived succinate from transit-associated taxa (Bucher-Johannessen et al., 2024; Watanabe, 2011). This paradox – higher diversity alongside depletion of beneficial taxa – highlights how public transport exposure may introduce environmental microbes without offsetting structural inequities in metabolic health (Nagao-Kitamoto et al., 2020).

Private transport users showed enrichment of *Blautia caecimuris* ($p = 7.7 \times 10^{-3}$) and *Bacteroides/Prevotella* enterotypes, potentially reflecting reduced environmental microbial exposure (Gorvitovskaia et al., 2016; Lagkouvardos et al., 2016; Weiss et al., 2023).

Functionally, public transport use associated strongly with the "Phosphotransferase system" (PTS) pathway (BH adjusted p-value = 2.1×10^{-3}), a carbohydrate uptake system critical for microbial competition. This pathway is overexpressed in obese individuals, reflecting its role in carbohydrate breakdown and pathogen virulence (Palmas et al., 2021). A murine model demonstrated that Western diets enrich PTS activity (Turnbaugh et al., 2008). SES analysis revealed public transport users were less likely to be normal weight, suggesting potential links between transport mode, PTS pathways, and metabolic health.

ii. Diet variables

High-protein diets showed enrichment of *Eubacterium rectale* ($p = 4.2 \times 10^{-2}$) and *Absiella argi* ($p = 6.6 \times 10^{-4}$, LDA > 3). *E. rectale*, typically associated with carbohydrate metabolism, may demonstrate metabolic flexibility under protein-rich conditions (Adrian et al., 2023), though this contrasts with rodent studies where high-protein diets reduced *Roseburia/Eubacterium* abundance, correlating with diminished butyrate and increased inflammation (Mu et al., 2017; Russell et al., 2011). *Absiella argi* showed strong protein intake associations though its functional role remains unclear (Paek et al., 2017). These findings align with evidence that protein-centric diets favour taxa capable of metabolizing amino acids, often at the expense of saccharolytic species (David et al., 2014).

Dairy consumers demonstrated higher Shannon diversity ($p = 5.0 \times 10^{-2}$) and elevated Actinobacteria abundance ($p = 2.8 \times 10^{-2}$), consistent with studies linking dairy intake to *Bifidobacterium* enrichment via lactose and galacto-oligosaccharides (Aslam et al., 2021; Lau et al., 2018; Oduaran et al., 2020). In urban professionals, processed dairy may favour broader Actinobacteria proliferation beyond *Bifidobacterium* (Lee et al., 2016). Reduced diversity in non-consumers aligns with diminished prebiotic substrates, though cardiometabolic implications require further investigation (Shuai et al., 2021).

Individuals excluding refined sugars showed enrichment of *Bacteroides plebeius* ($p = 4.7 \times 10^{-2}$), a gut symbiont specialized in degrading complex polysaccharides. *B. plebeius* thrives in low-sugar, high-fibre environments due to its genomic abundance of enzymes targeting indigestible plant polysaccharides (Yun et al., 2021). The paradoxical association with *S. aureus* infection pathways (adj. $p = 3.8 \times 10^{-2}$) could stem from *B. plebeius* overgrowth alongside reduced diversity, potentially weakening colonization resistance against opportunistic pathogens (Liu et al., 2024).

Gluten-free diets associated with *Bifidobacterium bifidum* enrichment ($P = 1.4 \times 10^{-2}$), potentially reflecting reduced gliadin-triggered inflammation (de Almeida et al., 2020) or increased prebiotic-rich alternatives (Xiao et al., 2024). Enrichment of "histidine metabolism" and "aromatic amino acid biosynthesis" pathways (adj. $p = 1.7 \times 10^{-2}$) suggests microbial adaptation to reduced nitrogen availability, as gluten-free diets often prompt gut microbes to upregulate amino acid synthesis pathways (Bonder et al., 2016; Hansen et al., 2018).

Non-omnivorous diets associated with *Lactobacillus sakei* and *Lactobacillus plantarum* (p-value $= 2.8 \times 10^{-4}$). *L. sakei*, though traditionally linked to meat fermentation, may reflect consumption of fermented vegetables where it inhibits pathogens and modulates metabolic health (Lim et al., 2020; Won et al., 2020). *L. plantarum* encodes enzymes that metabolize indigestible fibres prevalent in plant-based diets (David et al., 2014; Garcia-Gonzalez et al., 2022) and mitigates oxidative stress (Gupta et al., 2024). Sphingolipid metabolism enrichment (adj. $p = 7.1 \times 10^{-3}$) aligns with plant-based diets' reliance on microbial-derived sphingolipids, supporting mucosal integrity (Lee et al., 2021; Rohrhofer et al., 2021). The ALS pathway association (adj. p-value $= 7.1 \times 10^{-3}$) reflects altered sphingolipid-driven neuroinflammatory crosstalk, as ALS patients have been shown to exhibit reduced butyrate producers and dysregulated lipid metabolism (Berding et al., 2021; Boddy et al., 2021; McGrattan et al., 2022).

iii. Lifestyle variables

Lifestyle factors significantly shaped gut microbiome composition. Gas heating users showed enrichment of *Desulfovibrio piger* and *Clostridium aldenense*. *D. piger*, a sulphate-reducing

bacterium, likely thrives from trace sulphur compounds released by LPG combustion (Kalantar-Zadeh et al., 2019; Kushkevych et al., 2019). Though associated with harmful hydrogen sulphide production (Pimentel et al., 2013), its co-occurrence with commensals suggests context-dependent roles. Similarly, multivitamin users exhibited reduced alpha diversity (p-value = 9.0×10^{-3}), paralleling findings where high vitamin D intake suppressed microbial richness (Mandal et al., 2016), suggesting selective pressures from both environmental inputs and dietary supplements.

Herbal supplement use associated with significant beta diversity shifts (PERMANOVA p-value = 4.7×10^{-2}), enriching *Bacteroides eggerthii* and a *Bacteroides*-dominant enterotype. Polysaccharide and polyphenol-rich herbal compounds selectively promote *Bacteroides* species (Karim et al., 2023; Lyu et al., 2017). *B. eggerthii* abundance correlates with reduced inflammation and improved gut barrier integrity (Shinde & Deokar, 2024), potentially explaining therapeutic efficacy in gastrointestinal disorders (Lv et al., 2017). Cat ownership associated with *Parabacteroides merdae* enrichment, possibly through cross-species microbial exchange (Jha et al., 2020). While pet ownership doesn't significantly alter overall gut diversity, specific taxa including *Parabacteroides* may be differentially abundant in owners (Kates et al., 2020). *P. merdae*, linked to cardiovascular protection via BCAA catabolism (Qiao et al., 2022), could colonize human hosts from shared environments.

Nail biting linked to differential abundance of *Odoribacter splanchnicus*, an anti-inflammatory commensal (Hiippala et al., 2020; Oh et al., 2021). This behaviour may introduce oral pathogens competing with *O. splanchnicus* (Lima et al., 2021; Vyas, 2017). *Bacteroides uniformis* may influence sleep quality through SCFA production and gut-brain axis interactions (Seong et al., 2024), though context-dependent effects exist, as higher gut diversity correlates with better sleep but elevated SCFAs in older adults associate with poorer sleep quality (Holingue et al., 2021; Smith et al., 2019). White wine consumption linked to *Turicibacter sanguinis* and *Blautia stercoris*. *T. sanguinis* shows dose-dependent responses – enriched in heavy drinkers (Wang et al., 2023) but reduced with low-dose consumption (Lee et al., 2020). *Blautia* species increase with moderate beer and wine intake, correlating with beneficial metabolic effects (González-Zancada et al., 2020; Ozato et al., 2019). Higher alcohol consumption (≥4 servings weekly) increased *Mediterraneibacter glycyrrhizinilyticus*, which produces hepatoprotective compounds (Chigurupati et al., 2018; Sakuma et al., 2013), and *Anaerostipes rhamnosivorans*, which may enhance protective butyrate production (Bui et al., 2021).

Exercise enriched *Sutterella wadsworthensis* and *Victivallis vadensis*. *S. wadsworthensis* is more abundant in resistance athletes, likely due to bile resistance (Jang et al., 2019;

Mukhopadhya et al., 2011). While linked to IgA degradation in IBD (Kaakoush, 2020), it may play commensal roles under exercise-induced anti-inflammatory conditions (Varghese et al., 2024). Aerobic exercise increases *Bacteroides* abundance, potentially enriching *B. massiliensis* (Morita et al., 2019), while sedentary lifestyles may favour *P. distasonis*, which thrives in inflammatory conditions linked to inactivity (Cullen et al., 2023; Ezeji et al., 2021).

Post-antibiotic gut microbiota showed *Bacteroides* species persistence, likely due to *tetQ* resistance genes (Veloo et al., 2019; Wexler, 2007) and metabolic adaptations (Nilson et al., 2024). *Erysipelatoclostridium ramosum* exhibits resistance to certain antibiotics (Milosavljevic et al., 2021), while *Adlercreutzia equolifaciens*, which has anti-inflammatory properties, appears vulnerable to depletion (Oñate et al., 2023).

Cosmetics use correlated with apoptosis/cancer-related pathways, potentially reflecting bioactive ingredients like blueberry anthocyanins and chlorogenic acid. These compounds induce apoptosis, reduce oxidative stress, and regulate inflammatory cytokines (Huang et al., 2023; Wang et al., 2022). The gut-skin axis may mediate these effects (Beri, 2018), possibly through microbial metabolites with apoptotic activity (Altonsy et al., 2010; Saiki et al., 2021) or cyanobacterial compounds with cytotoxic potential (Kalaimathi et al., 2023).

iv. Health variables

Health conditions exhibited significant microbial associations. Food allergies altered beta diversity ($P = 4.2 \times 10^{-2}$), with enrichment of *Bacteroides coprocola* and *Bifidobacterium catenulatum. B. coprocola's* sensitivity to sodium nitrite suggests its overgrowth could stem from selective survival with dietary preservatives (Hrncirova et al., 2019), aligning with its association with ADHD in refined grain consumers (Wang et al., 2020). *B. catenulatum's* enrichment contrasts with its role as a common commensal in healthy adults (Ishikawa et al., 2013; Prasanna et al., 2014). While thriving in gluten-rich media (Caminero et al., 2014), its early-life association with allergy predisposition (Melli et al., 2016) suggests a dual role – protective in fibre-rich contexts but potentially pro-inflammatory in gluten-sensitive individuals. *B. coprocola's* link to Crohn's disease raises concerns about exacerbating mucosal inflammation (Hrncirova et al., 2019).

Lactose intolerant individuals showed enrichment of *Collinsella intestinalis* and *Blautia obeum*. GOS supplementation in lactose intolerant patients reduced symptoms and increased *Collinsella* abundance, suggesting enhancement of saccharolytic potential (Azcarate-Peril et al., 2021). *Collinsella* thrives with oligosaccharides, potentially aiding residual carbohydrate fermentation (Chey et al., 2020). *Blautia* shows conflicting associations – enriched in lactose mal-absorbers but identified as a risk factor for gluten-related disorders (Han et al., 2024; Li et al., 2018). Host-specific colonization patterns further complicate its relationship with lactose

intolerance (Eren et al., 2014; Xue et al., 2024). *Bacteroides stercoris* enrichment in gluten intolerance suggests a paradoxical role. Typically associated with anti-inflammatory properties (Chung et al., 2016), its enrichment in celiac disease may reflect compensatory growth in response to gluten-induced dysbiosis (Girdhar et al., 2023). *Dorea longicatena's* enrichment contrasts with its typical hydrogen cross-feeding in healthy networks (Hansen et al., 2018), where its hydrogen production may worsen symptoms like bloating. This highlights *B. stercoris's* context-dependent ecology – its pectin-degrading role may be overshadowed by inflammation-driven substrate shifts (Girdhar et al., 2023).

Alistipes putredinis enrichment in slow transit constipation (STC) (Tian et al., 2022) highlights its complex role in gut motility. Despite association with butanoate metabolism, its elevation in STC contrasts with depletion in IBS (Gryaznova et al., 2024). *A. putredinis* abundance correlates with longer transit times in healthy individuals (Asnicar et al., 2021), suggesting a shift from benign commensal to potential dysbiotic marker in pathological constipation. *Roseburia* depletion in STC (Tian et al., 2022) may worsen constipation by reducing butyrate. Laxative use reduces *A. putredinis* abundance (Liu et al., 2022), raising questions about causality in slowed transit.

Xenobiotics degradation pathways enrichment (adj. $p = 7.4 \times 10^{-3}$) highlights the gut microbiome's adaptive role in metabolizing environmental stressors. *B. coprocola* may metabolize preservatives, reducing exposure to harmful byproducts (Hrncirova et al., 2019). *B. catenulatum* may degrade allergenic compounds despite its allergy associations (Caminero et al., 2014; Melli et al., 2016). *Parabacteroides merdae*, linked to BCAA catabolism (Qiao et al., 2022), could enhance detoxification, while *A. putredinis* may mitigate oxidative stress through butyrate production (Tian et al., 2022). However, reliance on these pathways may reduce diversity or deplete commensals like *Roseburia*, potentially exacerbating conditions like constipation or food allergies.

5.6. Longitudinal analysis reveals temporal dynamics of a core yet variable microbiome

The longitudinal analysis of the urban South African population revealed pronounced inter-individual variability and subject-specific temporal dynamics in taxonomic composition, contrasting with a functionally stable microbiome dominated by conserved metabolic pathways – a pattern aligning with global studies of industrialized populations (Lloyd-Price et al., 2017; Olsson et al., 2022). It is important to interpret this in the context that second sample collections ranged from 7 – 24 months and given the small sample size ($n = 22$) we retained all samples without adjustment for this aspect to maintain statistical power. However, prior research has highlighted dynamic compositional and functional changes can occur from 6

months (Mehta et al., 2018; Han et al., 2022). Additionally, the alignment to the industrialised population research above enables the extrapolation of potentially relevant taxa and functional dynamics over time in the SAN population.

While core phyla (Firmicutes, Bacteroidota) remained stable, species-level shifts in taxa like *Blautia* and *Bacteroides thetaiotaomicron* mirrored diet- and urbanization-associated fluctuations observed in Western cohorts (Severino et al., 2024; van der Vossen et al., 2023). For example, Subject 287 exhibited marked taxonomic restructuring post-pregnancy, including weight normalization and reduced BMI, aligning with known pregnancy-related microbiome shifts (Koren et al., 2012), while Subject 199's antibiotic-driven dysbiosis reflects medication-induced volatility observed in industrialized populations (Smits et al., 2017). These individual trajectories, visualized via PCoA, highlight how life events and medical interventions drive taxonomic volatility despite functional resilience.The limited core microbiome at species level (e.g., only one genus and species at 95% prevalence) diverges from non-industrialized populations, which retain richer, stable cores (Smits et al., 2017). However, Subject 188's adoption of gas heating and vitamin supplementation – factors linked to *Alistipes putredinis* enrichment in transitional cohorts (Tian et al., 2022) – stresses urbanization's role in eroding conserved microbial patterns. This contrasts with Gacesa et al.'s (2022) identification of nine core species (e.g., *Bacteroides uniformis, Faecalibacterium prausnitzii*) across global populations, suggesting urbanization may disrupt universal microbial signatures which are also challenging to observe in smaller cohorts.

Functional redundancy, evidenced by stable KEGG pathways (82% core at 95% prevalence) despite taxonomic volatility, echoes the "functional resilience" hypothesis (Moya & Ferrer, 2016). Variable carbohydrate metabolism pathways align with Subject 276's improved bowel movements post-constipation, reflecting metabolic adaptability to urban diets rich in processed substrates (David et al., 2014). However, the cohort's high intra-individual Bray-Curtis dissimilarity exceeds long-term Western studies, potentially reflecting competing influences in diverse urban environments (Faith et al., 2013). For instance, intermittent traditional dietary practices – suggested by Subject 276's travel frequency – may transiently stabilize communities, creating hybrid ecologies where functional stability coexists with compositional volatility (Vangay et al., 2018; Ramaboli et al., 2024).

Notably, among the most temporally stable functions we observed were ATP-binding cassette (ABC) transporters. ABC transporters are central to microbial nutrient acquisition, bile acid and lipid handling, and xenobiotic efflux in the gut ecosystem – processes that are particularly relevant in urban environments characterized by processed diets, medication exposure, and household/occupational pollutants. Prior work shows microbiome-host crosstalk around bile

acids and lipid metabolism is tightly linked to insulin sensitivity and cardiometabolic risk via FXR and TGR5 signalling pathways (Sayin et al., 2013; Thomas et al., 2009), while ABC transport systems facilitate microbial survival under fluctuating nutrient and xenobiotic loads typical of urban lifestyles (Orelle et al., 2019). Thus, the persistence of ABC transporter activity likely reflects a conserved functional backbone that buffers day-to-day environmental variation in an urban professional setting.

In contrast to the transporter core, several high-abundance but temporally variable functions mapped to carbohydrate acquisition/processing and host-glycan turnover (e.g., PTS cellobiose systems; 6 phospho β glucosidase; hexosaminidase). These pathways signal flexibility in utilizing refined carbohydrates and plant polysaccharides – dietary features that can oscillate in urban professionals due to irregular meal timing, convenience foods, and hybrid traditional-Western dietary patterns. Such PTS/carbohydrate-active pathways are repeatedly linked to glycaemic control and adiposity (David et al., 2014; Kovatcheva Datchary et al., 2015), and hexosaminidase-mediated amino-sugar and glycosaminoglycan metabolism interfaces with sphingolipid/glycosphingolipid biosynthesis, pathways implicated in insulin resistance and cardiometabolic risk (Feingold & Grunfeld, 2011; Chaurasia & Summers, 2015). Importantly, in African-origin cohorts, microbial functional signatures tied to butyrate synthesis routes and LPS biosynthesis have been associated with higher cardiometabolic risk (Fei et al., 2019), underscoring why these variable carbohydrate and host-glycan functions merit attention in an urban South African context.

Additional cell-surface remodelling functions (e.g., wcaJ/colanic acid biosynthesis) and pectin/rhamnogalacturonan degradation enzymes (yteR/yesR; yesW) point to biofilm/mucus-layer interactions and the capacity to exploit plant fibres. Colanic acid and exopolysaccharide assembly can modulate biofilm formation and barrier interactions (Bharathi et al., 2023), factors tied to gut permeability and low-grade inflammation relevant for cardiometabolic health (Campos-Bayardo et al., 2025). Meanwhile, pectin/rhamnogalacturonan-degrading activities are consistent with mixed fiber intake and have been associated with SCFA production and metabolic benefits when diets are plant-forward (Makki et al., 2018). Together, these variable functions provide plausible mechanistic links between urban dietary heterogeneity and fluctuations in metabolic and inflammatory tone in professional populations.

The scarcity of persistent species contrasts with functional stability, highlighting methodological limitations: 16S-based inferences mask strain-level adaptations critical for resolving true redundancy (Heintz-Buschart & Wilmes, 2018). Sparse sampling (two samples collected with ~1 year between timepoints) further limits resolution of transient fluctuations, as seen in Subject 199's antibiotic use and Subject 188's air pollution exposure. Future work

should integrate metagenomics and metabolomics to disentangle strain-level adaptations and link taxa like *Alistipes putredinis* to functional outcomes in this understudied population.

5.7. Revealing the broader landscape

Diversity across resolution and taxonomic divergence

The SAN population's distinct microbial diversity patterns reflect both demographic and methodological influences. Compared to AWIGEN1 – a rural-urban African cohort with diverse SES facing distinct lifestyle and dietary pressures – the SAN population comprises urban professionals, likely with higher SES and homogenized urban exposures. These demographic divergences alone could explain reduced SAN phylum-level diversity relative to AWIGEN1. Methodologically, SAN's full-length 16S sequencing may under-detect low-abundance species compared to AWIGEN1/AGP's shotgun sequencing (Yue et al., 2023), though cross-platform normalization partially mitigates this (Franzosa et al., 2015). Thus, observed differences likely arise from both ecological (SES, urbanization) and technical (sequencing depth, platform) factors. The spectrum of microbiome profiles across different populations reveals how urbanization and lifestyle factors shape bacterial community structures.

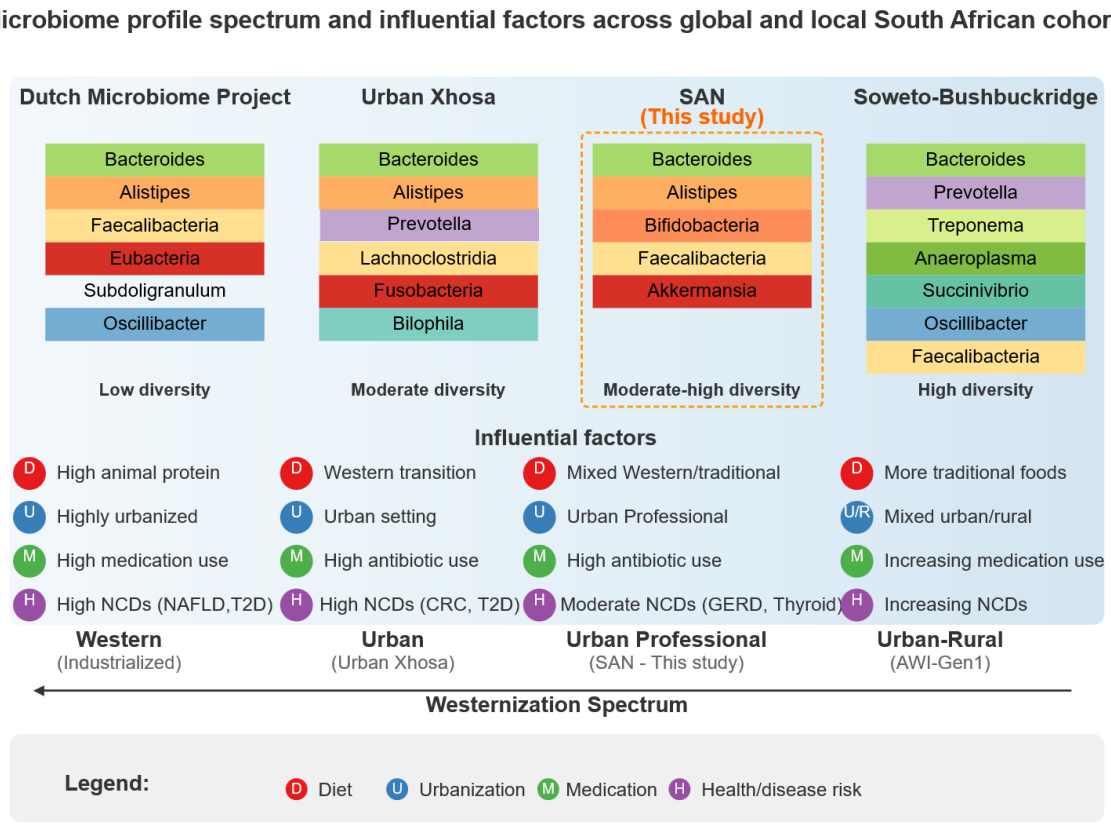


Figure 5.6 Microbiome profile spectrum and influential factors across global and local South African cohorts. Microbial compositional profiles (dominant taxa and diversity) and associated influential factors (diet, urbanization, medication use, and health/disease risk) are compared across cohorts positioned along the Western-to-Rural microbiome spectrum. The SAN population (this study) occupies an intermediate position, highlighting its unique combination of moderate-high diversity, mixed Western/traditional dietary habits, high antibiotic exposure, and moderate NCD prevalence, distinguishing it from highly Westernized (Dutch Microbiome Project) and more traditional rural cohorts (Urban-Rural AWI-Gen1). This framework highlights the SAN population's novelty as a microbiome model for understanding urbanization's complex transition dynamics and associated NCD risks in South Africa.

Figure 5.6 positions the SAN population within a westernization spectrum of microbiome profiles across South African and global cohorts. The SAN population occupies an intermediate “Urban Professional” niche with moderate-high diversity, characterized by notable presence of *Bacteroides, Alistipes, Bifidobacteria, Faecalibacteria,* and *Akkermansia*. This transitional position is potentially driven by specific influential factors: mixed Western/traditional diet, urban professional lifestyle, high antibiotic use, and moderate NCDs prevalence. The figure contextualizes how the SAN population bridges the gap between Western industrialized microbiomes (Dutch Microbiome Project) and traditional profiles (Soweto-Bushbuckridge), providing a framework for further understanding the subsequent taxonomic analyses.

At the genus level, SAN’s intermediate diversity aligns with transitional populations retaining microbial complexity despite urban homogenization pressures, as seen in African cohorts adopting hybrid diets (Smits et al., 2017). However, Smits et al. focused on transitions in non-urban populations, whereas SAN’s urban professionals may face distinct lifestyle stressors (e.g., processed diets, sedentary work) that dampen microbial richness, suggesting regional urbanization dynamics are not universally generalisable (Vangay et al., 2018; Ramaboli et al., 2024).

Species-level findings further highlight methodological and ecological constraints for exploration. SAN 16S sequencing detected a lower unique species count than AGP/AWIGEN1 (Chapter 4.7.2), though similar Shannon diversity implies comparable evenness in detected communities (Quince et al., 2017). This suggests that while SAN harbours fewer taxa overall (lower richness), the relative abundance of detected species (evenness) is similar to AGP/AWIGEN1. Beta diversity trends reinforce SAN’s uniqueness, with pronounced separation from AGP/AWIGEN1 at all taxonomic levels. This mirrors global studies where urbanization drives divergent microbiomes, yet SAN’s species-level distinctiveness (27.9% variance in PCoA) exceeds typical urban signatures, potentially reflecting unmeasured lifestyle or environmental factors (Gupta et al., 2020).

The SAN population microbiome marked by reduced phylum diversity, transitional genus profiles, and species-level distinctiveness, occupies an intermediate niche between Westernized and rural African populations. While methodological disparities (16S vs. shotgun) limit direct comparisons, the cohort's divergence at finer taxonomic resolutions suggests unique ecological drivers, such as those highlighted in the prior discussion sections (e.g. stratifying cluster-associated features) and Figure 5.6, warranting deeper exploration.

The SAN study population includes a diverse mix of ancestries, with a majority of European ancestry (~60%), introducing a non-African ancestry component that is not present in the AWIGEN1 and urban Xhosa cohorts (which comprise exclusively African ancestry participants). While prior work suggests genetic ancestry contributes minimally to gut microbiome composition (Gacesa et al., 2022; He et al., 2018; Rothschild et al., 2018), this compositional difference in ancestry may nonetheless be a contextual factor when interpreting cross-cohort differences within African studies. The SAN population's elevated Bacteroidetes and reduced Firmicutes compared to AWIGEN1 diverges from typical urbanization trends, where Firmicutes dominance is common (Sonnenburg et al., 2016). Full-length 16S's ability to resolve species like *Parabacteroides distasonis* and *Odoribacter splanchnicus* reduces "unclassified species" biases but still misses strain-level novelty captured by shotgun sequencing (Truong et al., 2017). For instance, SAN's lower Proteobacteria vs. AGP may reflect shotgun's sensitivity to low-abundance pathobionts, while full-length 16S's precision avoids primer bias for Bacteroidetes (Lloyd-Price et al., 2019; Johnson et al., 2019).

MAG analysis to bridge platforms

The MAG analysis of the SAN population's shotgun data revealed high average nucleotide identity (ANI > 95%) to existing reference genomes, indicating limited novel species recovery compared to non-Western populations (Pasolli et al., 2019). In total, n = 258 MAGs were recovered from the SAN population, of which approximately half (~ 54%) were classified as high-quality, confirming the robustness of these genomes for downstream comparative analyses (Figure 4.48). Interestingly, despite the robust MAG reconstruction and stringent quality thresholds applied, FastANI comparisons against the GTDB revealed no substantial novelty, with all genomes aligning closely (97 – 98% identity) to previously characterized species (Supplementary S4.17, Figure 4.52A).

This contrasts with studies of larger urban-rural African cohorts, where shotgun sequencing uncovered novel prokaryotic genomes, many of which fell under the Actinobacteria phylum (Actinomycetota), as well as those with unique host-disease adaptations (Maghini et al., 2025). The lack of novelty in SAN's MAGs may reflect the small cohort size compared to the > 1,800 women from Maghini et al. (2025), who were recruited from considerably more heterogeneous

environments, including both urban and rural settings, potentially capturing broader ecological niches and host-microbiome interactions. Furthermore, higher sequencing depths and methodological variations in optimising the metagenome assembly pipelines for this data employed by Maghini et al. (2025), such as deeper coverage per sample, could enhance the detection of rare and novel microbial variants that may have been missed in the SAN population due to relatively limited sequencing depth. Socioeconomic factors, including dietary diversity, exposure to antibiotics, and lifestyle differences between more industrialized versus rural populations, may also contribute significantly to these observed differences in microbial diversity and novelty.

In addition to urbanization-driven homogenization, where niche industrialized lifestyles may erode region-specific microbial diversity, these technical and socioeconomic factors collectively elucidate the importance of sampling strategy, sequencing methodology, and cohort heterogeneity when interpreting microbial novelty within populations. Although shotgun metagenomic data enabled MAG reconstruction and insight into strain-level genomic content, long-read 16S – despite its improved taxonomic resolution – cannot capture the accessory gene content critical for understanding functional adaptation within specific niches (Sczyrba et al., 2017). Thus, while SAN's MAGs validate taxonomic trends (e.g., Actinomycetota enrichment), their functional novelty remains constrained compared to less urbanized African populations.

Functional profiling and concordance

Moderate to strong correlations between SAN's PICRUSt2 (16S long-read) and AGP/AWIGEN1's HUMAnN3 (shotgun) functional predictions reflect 16S's capacity to approximate core metabolic pathways while revealing methodological biases (Franzosa et al., 2018). PICRUSt2 systematically overestimates nucleotide synthesis pathways compared to HUMAnN3, a discrepancy partially mitigated by long-read sequencing (Ye et al., 2019). This bias may artificially inflate SAN's apparent enrichment in central carbon metabolism pathways, which appeared 1-2 orders of magnitude higher than in AGP/AWIGEN1. Conversely, PICRUSt2's poor resolution of lipid and cell wall metabolism likely explains SAN's lower predicted abundances of CMP-pseudaminate biosynthesis and fatty acid initiation (Douglas et al., 2020).

Within-cohort comparisons show PICRUSt2 identified only 349 metabolic pathways versus 791 with HUMAnN3, with approximately 33.3% overlap. PICRUSt2 uniquely predicted 64 pathways absent from HUMAnN3, while HUMAnN3 uniquely detected 506 pathways. Among 285 commonly detected pathways, abundance estimates showed only moderate correlation ($r = 0.6$), with increasing divergence at lower abundances.

Cohort divergences reflect ecological differences – AGP's elevated amino acid and xenobiotic degradation pathways align with Westernized microbiomes adapted to processed diets, while AWIGEN1's starch degradation enrichment mirrors urban-transitional African populations retaining high-fibre diets (Heinken et al., 2021; Smits et al., 2017). These interpretations are constrained by 16S's inability to resolve strain-specific functions (Nayfach et al., 2016).

SAN's distinct profile – high *Bacteroides* and low Proteobacteria – reflects a hybrid ecology shaped by urban diets with potential intermittent traditional exposures (Ramaboli et al., 2024). While full-length 16S provides superior taxonomic resolution over short-read approaches, it cannot replace shotgun sequencing for functional insights. SAN's hybrid profile – taxonomically distinct yet functionally convergent – emphasizes urbanization's nuanced impacts. Future studies should integrate long-read 16S for taxonomic precision and shotgun sequencing for functional resolution, as advocated in recent critiques (McDonald et al., 2024).

5.8. Future directions for research

The findings from this study lay the groundwork for several promising research directions that could further enhance our understanding of the urban African microbiome. By addressing gaps in niche population-specific insights, clinical translation, and methodological innovation, future studies can build on these results to unravel the complex interplay between urbanization, host biology, and microbial profiles.

Population-specific investigations

While rural-urban comparisons remain valuable, the SAN population's primary strength lies in its capacity to elucidate intra-population heterogeneity among urban professionals – a critical gap given South Africa's rising NCD burden linked to urbanization (Statistics South Africa, 2023). To advance this, future studies should expand to other urban professional cohorts (e.g., university staff in Cape Town, healthcare workers in Durban) to disentangle how microenvironments – such as workplace dietary policies, pollution exposure gradients, or socioeconomic stratification within cities – drive microbiome variation. For instance, private hospital employees in Johannesburg may exhibit distinct microbial profiles compared to public-sector workers due to divergent cafeteria diets, stress levels, or antibiotic prescription practices, mirroring SES-linked dysbiosis observed in U.S. healthcare cohorts (Dai et al., 2024). Similarly, comparing professionals across industries (e.g., mining vs. academia) could clarify occupational exposures (e.g., heavy metals, sedentary routines) that modulate NCD-associated taxa like *Bacteroides* or *Prevotella* and various others (Shivani et al., 2021; Wilkins et al., 2019).

Such intra-urban profiling would complement the SAN population's findings – where *Prevotella* enrichment correlated with processed carbohydrate intake – by identifying localized drivers of

microbial shifts that precede NCD onset. For example, hypertension-linked *Collinsella* may disproportionately colonize professionals in high-stress, low-physical-activity sectors, offering actionable targets for workplace wellness programs (Clemente-Suárez et al., 2023). Importantly, harmonizing methodologies across institutions (e.g., standardized dietary surveys, metabolomic profiling) will enable meta-analyses to disentangle universal urban stressors from culturally unique factors shaping microbiome-NCD linkages in South Africa (Kim et al., 2020).

Longitudinal studies are essential to resolve the temporal dynamics hinted at by the limited SAN subset (*n* = 22). While taxonomic profiles showed significant intra-individual variation, functional pathways remained stable, suggesting a "core" metabolic resilience despite fluctuating species composition. Larger longitudinal cohorts can track how life events (e.g., migration, seasonal dietary shifts) alter this balance in rapidly urbanizing settings. Integrating metabolomics to profile short-chain fatty acids or bile acids alongside microbial data could reveal how functional stability buffers against taxonomic disruption (Muller et al., 2024). Multi-omics approaches must expand beyond 16S and shotgun metagenomics. Metatranscriptomics for example could clarify whether the *S. aureus* infection pathway enrichment in low-sugar diets reflects active pathogen transcription or residual DNA signals. Similarly, metabolomic profiling of urban populations could link functional pathways like nitrotoluene degradation to specific xenobiotic exposures (Priya et al., 2022).

Clinical applications

The urban professional South African population reveals microbiome signatures with direct translational potential. For instance, *Bacteroides stercoris* enrichment in *gluten intolerance* (Girdhar et al., 2023) and *Alistipes putredinis* in constipation (Tian et al., 2022) could serve as biomarkers for personalized diagnostics. Machine learning models trained on these taxa, alongside lifestyle metadata (e.g., refined sugar avoidance linked to *Bacteroides plebeius*), may stratify risk for conditions like IBS or food allergies, which showed subthreshold associations in this cohort. However, population-specificity remains a hurdle: while *Faecalibacterium prausnitzii* is a well-established biomarker for IBD in Western cohorts (Sokol et al., 2008), South African urban professionals exhibit distinct signatures (e.g., *Prevotella stercorea* enrichment in males), necessitating validation in African-centric cohorts to avoid misapplication of Eurocentric references.

Therapeutic strategies should leverage modifiable lifestyle-microbiome interactions. The enrichment of *Lactobacillus plantarum* in non-omnivorous diets suggests probiotic formulations could mimic plant-rich dietary benefits in urban populations with limited fibre access, akin to interventions reducing heavy metal toxicity via *Bacillus megaterium* (Simon et

al., 2023). Similarly, the association between cosmetics use and apoptosis pathways warrants trials testing whether reducing skincare product exposure alters *Cutibacterium acnes* dynamics in acne, as proposed for pollution-induced dysbiosis (Rio et al., 2024). Public health initiatives could prioritize urban gardening to boost dietary fibre intake, countering the low *Roseburia* abundance linked to processed diets, while campaigns targeting refined sugar might restore *Bacteroides plebeius*-mediated polysaccharide metabolism.

Methodological advancements

Causal inference frameworks, such as Mendelian randomization, are essential to disentangle whether microbiome alterations drive conditions like lactose intolerance (e.g., *Collinsella intestinalis* enrichment) or merely reflect immune-mediated dysregulation. While traditional microbiome studies predominantly rely on community structure assessments (e.g., 16S rRNA sequencing) or gene abundance estimations (shotgun metagenomics), these methods alone remain inherently correlative, limiting causal inference and functional interpretation regarding host health outcomes. Functional multi-omics analyses, such as metatranscriptomics and metabolomics, offer deeper insights into active microbial metabolism and host-microbial interactions, providing direct evidence linking microbial activity to host physiology (Shalon et al., 2023).

Our SAN population analysis identified numerous microbiome-host correlations, such as the enrichment of *Dialister invisus* in urban professionals with normal BMI, potentially implicating metabolic health regulation. Similarly, taxa like *Bacteroides stercoris* in gluten intolerance and *Alistipes putredinis* in constipation highlight possible microbiome-mediated interactions. However, these associations, derived solely from microbial abundance data, do not definitively indicate causal roles or distinguish taxa as active functional players versus passive biomarkers of dietary or inflammatory contexts. Integrating metatranscriptomics could clarify whether taxa such as *Dialister invisus* actively express metabolic pathways linked to BMI regulation, or whether *Bacteroides stercoris* expresses inflammatory genes in gluten intolerance contexts.

Network analyses of lifestyle-associated pathways (e.g., apoptosis, xenobiotic degradation) could pinpoint hub taxa, such as *Bacteroides coprocola* – known to metabolize dietary nitrites – as mediators of urban environmental exposures (Hrncirova et al., 2019). Metabolomics integration, particularly lipidomics, could further elucidate these relationships by identifying microbial metabolites (e.g., short-chain fatty acids, bile acids, sulfonolipids, FAHFAs) mediating observed health outcomes (Wang et al., 2023). For instance, our data noted *Blautia* spp. enrichment linked with alcohol consumption or lactose intolerance, but without metabolomics, we cannot confidently determine if these associations reflect beneficial or detrimental metabolic interactions. Incorporating metabolomics could validate whether

microbial metabolites produced by enriched taxa directly influence host physiology, thus enabling stronger causal inference.

Technologically, long-read sequencing (PacBio) and shotgun metagenomics remain critical for resolving species- and strain-level diversity in taxa such as *Prevotella*, prevalent in African microbiomes yet poorly represented in global reference databases (Quince et al., 2017). Additionally, deploying ingestible sampling devices could overcome stool sampling limitations, particularly for proximal intestinal conditions like celiac disease (Shalon et al., 2023). This approach has demonstrated substantial longitudinal and spatial microbiota variation and underlined the necessity of complementing stool-based community analyses with metabolomics and transcriptomics to accurately capture dynamic host-microbial interactions.

Portable sequencing technologies (e.g., Oxford Nanopore) hold promise for democratizing antimicrobial resistance gene monitoring in low-SES urban areas, allowing real-time assessments of antibiotic misuse impacts on taxa like *Dialister succinatiphilus*. Standardization efforts must address platform disparities by harmonizing 16S and shotgun pipelines for high-resolution species classification – such as *Bifidobacterium adolescentis* and its GABA-producing pathways – which would further enhance cross-study validation. Initiatives mirroring MIAME standards should prioritize African microbiome contexts, ensuring protocols specifically account for high *Prevotella* diversity and urbanization gradients (Brazma, 2009). Although existing protocols (e.g., STORMS checklist; Mirzayi et al., 2021) have advanced reporting standards, further methodological harmonization remains essential.

Critically, as Rio et al. (2024) emphasize, integrating exposome data – encompassing environmental factors (pollutants, chemicals, public transport use) and individual behaviours – with multi-omics analyses represents the next frontier of microbiome research. In our cohort, associations between *gas heating* and taxa like the sulphate-reducing bacterium *Desulfovibrio piger* suggest potential inflammatory implications. Yet, without integrated exposome and metabolomics data, we remain limited to speculative interpretations regarding causality. Multi-dimensional integration could clarify if such environmental exposures directly alter gut sulphur metabolism pathways, influencing microbiota composition, gene expression, metabolite production, and subsequent health outcomes.

While our cohort highlights actionable microbiome-diet-lifestyle axes, critical gaps persist. First, validation efforts must test biomarkers such as *B. stercoris* in rural-to-urban transition cohorts, distinguishing gluten-specific effects from broader dietary shifts. Second, establishing causality – such as determining if *Alistipes putredinis* drives constipation or merely thrives in slow-transit niches – demands longitudinal sampling complemented by metatranscriptomics and metabolomics. Lastly, equity considerations remain vital. Urban professionals' unique

stressors (sedentary lifestyles, processed diets) limit generalizability to broader urban landscapes and informal settlements, where pollution and malnutrition heavily influence microbial profiles. Thus, methodological advancements beyond stool sampling and Eurocentric reference databases, towards integrated multi-omics and exposome frameworks, become pivotal.

In conclusion, our observed microbiome-host correlations (e.g., involving *Prevotella stercorea*, *Dialister invisus*, *Alistipes putredinis*) provide compelling hypotheses regarding their functional roles yet remain speculative without multi-omics validation. Future microbiome research must prioritize integrative approaches: (i) metatranscriptomics: verifying active microbial gene expression, distinguishing functional metabolic roles from DNA-based presence / absence; (ii) metabolomics: identifying microbial metabolites mediating observed associations, enabling clear linkage between microbial activity and host health outcomes; and (iii) exposome integration: capturing environmental exposures (pollution, dietary preservatives, cosmetics) to elucidate pathways driving microbiome-host interactions. Adopting this integrated analytical paradigm transforms microbiome studies from descriptive observations towards robust causal inference. Ultimately, integrating multi-omics and exposome data will clarify microbial mechanisms underlying health disparities, informing targeted dietary, probiotic, and public health interventions tailored specifically to the unique, urbanizing African contexts represented in our SAN population.

6. Conclusion

6.1. Summary of findings

This PhD microbiome research project pioneers the exploration of the gut microbiome in an urban professional South African population, leveraging PacBio 16S long-read sequencing to resolve taxonomic complexity with significant precision. The study reveals a central paradox of urbanization: microbial communities exhibit taxonomic volatility – driven by processed diets, environmental exposures, and lifestyle stressors – while maintaining functional stability through conserved metabolic pathways.

Key demographic patterns emerged, including age-associated *Bifidobacterium* depletion, sex-specific adaptations (*Prevotella stercorea* in males, *Bacteroides dorei* in females), and ancestry-linked microbial resilience to urban dietary shifts. Urbanization's dual role as both disruptor and adapter was evident in the erosion of species-level core microbiomes, contrasting with rural African cohorts, and the enrichment of taxa like *Bacteroides plebeius* under refined sugar exclusion – a hallmark of dietary transition. Health-microbiome linkages,

such as *Bacteroides stercoris* enrichment in *gluten intolerance* and *Alistipes putredinis* in constipation, highlighted context-dependent trade-offs between microbial adaptation and dysbiosis. Methodologically, while PacBio 16S advanced rare-taxa resolution, disparities in functional inference (e.g., PICRUSt2 overestimation of nucleotide pathways) highlighted the irreplaceable role of shotgun metagenomics for strain-level insights. The cohort's high intra-individual variability, exceeding Western benchmarks, positions South Africa's diet, lifestyle, and SES transition as a critical model for studying urbanization's bidirectional impacts on microbial ecology and host health.

6.2. Contributions to the field

This work contributes to the South African gut microbiome landscape through the snapshot characterization of a novel niche cohort. As the first in-depth study of urban professionals in South Africa, it bridges geographic and demographic gaps, challenging the *Firmicutes*-dominance narrative with a *Bacteroidetes*-enriched, functionally resilient profile unique to South Africa's hybrid lifestyles.

Methodologically, the integration of PacBio long-read sequencing with shotgun metagenomics sets an example for resolving taxonomic complexity in underrepresented populations, while exposing limitations in 16S-based functional inference. This approach addresses the limitations of traditional 16S rRNA sequencing, which, while useful for taxonomic profiling, often falls short in functional inference due to its reliance on a single gene marker. Long-read shotgun metagenomics allows for comprehensive genomic analysis, capturing the entire microbial community's genetic material, which facilitates both taxonomic and functional insights (Bars-Cortina et al., 2024).

Recent studies have demonstrated that long-read sequencing can reconstruct complete genomes, including plasmids and mobile genetic elements, which are critical for understanding microbial ecology and evolution (Kim et al., 2024). This capability is particularly beneficial for identifying novel metabolic pathways and functional traits that may be overlooked by 16S-based methods, which are constrained by reference database biases and limited taxonomic resolution (Peterson et al., 2021). Furthermore, as sequencing costs continue to decline, the economic viability of long-read shotgun metagenomics improves, making it a more attractive option for future research endeavours (Kim et al., 2024). In light of these advancements, future studies should prioritize long-read shotgun metagenomics as the primary sequencing strategy. This shift will not only enhance the resolution of microbial diversity but also provide deeper insights into the functional potential of microbial communities, ultimately leading to a more comprehensive understanding of their roles in health and disease.

Further understanding of host-microbiome health associations amplifies their public health relevance by providing actionable insights. Microbial signatures linked to sulphur exposure (*Desulfovibrio piger*) and non-omnivorous diets (*Lactobacillus plantarum*) offer blueprints for dietary and environmental interventions to curb NCDs. The “functional resilience” framework reimagines microbial stability in transitional societies prioritizing metabolic redundancy over taxonomic conservation – a paradigm shift for urbanization studies globally.

Crucially, this study advocates for equitable research practices, challenging Eurocentric biases by exploring and highlighting African-centric taxonomic signatures that could support biomarker development along with protocols, urging the adoption of inclusive databases that represent taxa with potentially critical roles in local health contexts. By linking microbial dysbiosis to modifiable urban stressors (e.g. processed diets, air pollution), the findings may directly inform policy reforms, advocating for cleaner energy subsidies, workplace wellness programs, and urban infrastructure prioritising green spaces. This work supports positioning microbial ecology as a cornerstone of NCD prevention, potentially contributing to systemic change in how African science serves African populations. Its significance lies not only in forwarding South African microbiome characterization and research, but in advancing a roadmap for equitable, evidence-based health strategies in rapidly urbanizing Africa – a testament to the societal and scientific contribution made by this body of research.

References

3billion. (2023). *Whole Genome Sequencing cost 2023*. https://3billion.io/blog/whole-genome-sequencing-cost-2023

Aagaard, K. M. (2021). 132 Early Establishment of the Human Gut Microbiome—A Tale of Moms, Their Diets, and Babes Guts. *Journal of Animal Science*, *99*(Supplement_3), 68. https://doi.org/10.1093/jas/skab235.123

Aagaard, K. M., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The Placenta Harbors a Unique Microbiome. *Science Translational Medicine*, *6*(237), 237ra65. https://doi.org/10.1126/scitranslmed.3008599

Adewale, B. A. (2020). Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? *African Journal of Laboratory Medicine*, *9*(1), Article 1. https://doi.org/10.4102/ajlm.v9i1.1340

Adrian, M. A., Ayati, B. P., & Mangalam, A. K. (2023). A mathematical model of Bacteroides thetaiotaomicron, Methanobrevibacter smithii, and Eubacterium rectale interactions in the human gut. *Scientific Reports*, *13*(1), 21192. https://doi.org/10.1038/s41598-023-48524-4

Afolayan, A. O., Ayeni, F. A., Moissl-Eichinger, C., Gorkiewicz, G., Halwachs, B., & Högenauer, C. (2019). Impact of a Nomadic Pastoral Lifestyle on the Gut Microbiome in the Fulani Living in Nigeria. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02138

Agus, A., Clément, K., & Sokol, H. (2021). Gut microbiota-derived metabolites as central regulators in metabolic disorders. *Gut*, *70*(6), 1174–1182. https://doi.org/10.1136/gutjnl-2020-323071

Aldirawi, H., & Morales, F. G. (2023). Univariate and Multivariate Statistical Analysis of Microbiome Data: An Overview. *Applied Microbiology*, *3*(2), Article 2. https://doi.org/10.3390/applmicrobiol3020023

Aljazairy, E. A., Al-Musharaf, S., Abudawood, M., Almaarik, B., Hussain, S. D., Alnaami, A. M., Sabico, S., Al-Daghri, N. M., Clerici, M., & Aljuraiban, G. S. (2022). Influence of Adiposity on the Gut Microbiota Composition of Arab Women: A Case-Control Study. *Biology*, *11*(11), 1586. https://doi.org/10.3390/biology11111586

Allali, I., Abotsi, R. E., Tow, L. Ah., Thabane, L., Zar, H. J., Mulder, N. M., & Nicol, M. P. (2021). Human microbiota research in Africa: A systematic review reveals gaps and priorities for future research. *Microbiome*, *9*(1), 241. https://doi.org/10.1186/s40168-021-01195-7

Allali, I., Boukhatem, N., Bouguenouch, L., Hardi, H., Boudouaya, H. A., Cadenas, M. B., Ouldim, K., Amzazi, S., Azcarate-Peril, M. A., & Ghazal, H. (2018). Gut microbiome of Moroccan colorectal cancer patients. *Medical Microbiology and Immunology*, *207*(3), 211–225. https://doi.org/10.1007/s00430-018-0542-5

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, *11*(11), 1144–1146. https://doi.org/10.1038/nmeth.3103

Alou, M. T., Lagier, J.-C., & Raoult, D. (2016). Diet influence on the gut microbiota and dysbiosis related to nutritional disorders. *Human Microbiome Journal*, *1*, 3–11. https://doi.org/10.1016/j.humic.2016.09.001

Altonsy, M. O., Andrews, S. C., & Tuohy, K. M. (2010). Differential induction of apoptosis in human colonic carcinoma cells (Caco-2) by Atopobium, and commensal, probiotic and enteropathogenic bacteria: Mediation by the mitochondrial pathway. *International Journal of Food Microbiology*, *137*(2–3), 190–203. https://doi.org/10.1016/j.ijfoodmicro.2009.11.015

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amadou, T., Hosny, M., La Scola, B., & Cassir, N. (2016). "Lachnoclostridium bouchesdurhonense," a new bacterial species isolated from human gut microbiota. *New Microbes and New Infections*, *13*, 69–70. https://doi.org/10.1016/j.nmni.2016.06.015

Aminianfar, A., Saneei, P., Nouri, M., Shafiei, R., Hassanzadeh-Keshteli, A., Esmaillzadeh, A., & Adibi, P. (2021). Validity of Self-reported Height, Weight, Body Mass Index, and Waist Circumference in Iranian Adults. *International Journal of Preventive Medicine*, *12*, 75. https://doi.org/10.4103/ijpvm.IJPVM_422_18

Anderson, M. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*, 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Andrews, S. (2010). *Babraham Bioinformatics—FastQC A Quality Control tool for High Throughput Sequence Data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Anthamatten, L., von Bieberstein, P. R., Menzi, C., Zünd, J. N., Lacroix, C., de Wouters, T., & Leventhal, G. E. (2024). Stratification of human gut microbiomes by succinotype is associated with inflammatory bowel disease status. *Microbiome*, *12*(1), 186. https://doi.org/10.1186/s40168-024-01897-8

Aoun, A., Darwish, F., & Hamod, N. (2020). The Influence of the Gut Microbiome on Obesity in Adults and the Role of Probiotics, Prebiotics, and Synbiotics for Weight Loss. *Preventive Nutrition and Food Science*, *25*(2), Article 2. https://doi.org/10.3746/pnf.2020.25.2.113

Arboleya, S., Watkins, C., Stanton, C., & Ross, R. P. (2016). Gut Bifidobacteria Populations in Human Health and Aging. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.01204

Argentini, C., Lugli, G. A., Tarracchini, C., Fontana, F., Mancabelli, L., Viappiani, A., Anzalone, R., Angelini, L., Alessandri, G., Longhi, G., Bianchi, M. G., Taurino, G., Bussolati, O., Milani, C., van Sinderen, D., Turroni, F., & Ventura, M. (2024). Genomic and ecological approaches to identify the Bifidobacterium breve prototype of the healthy human gut microbiota. *Frontiers in Microbiology*, *15*, 1349391. https://doi.org/10.3389/fmicb.2024.1349391

Armstrong, A. J. S., Shaffer, M., Nusbacher, N. M., Griesmer, C., Fiorillo, S., Schneider, J. M., Preston Neff, C., Li, S. X., Fontenot, A. P., Campbell, T., Palmer, B. E., & Lozupone, C. A. (2018). An exploration of Prevotella-rich microbiomes in HIV and men who have sex with men. *Microbiome*, *6*(1), 198. https://doi.org/10.1186/s40168-018-0580-7

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., … Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature*, *473*(7346), Article 7346. https://doi.org/10.1038/nature09944

Aslam, H., Collier, F., Davis, J. A., Quinn, T. P., O'Hely, M., Pasco, J. A., Jacka, F. N., & Loughman, A. (2021). Gut Microbiome Diversity and Composition Are Associated with Habitual Dairy Intakes: A Cross-Sectional Study in Men. *The Journal of Nutrition*, *151*(11), 3400–3412. https://doi.org/10.1093/jn/nxab252

Asnicar, F., Berry, S. E., Valdes, A. M., Nguyen, L. H., Piccinno, G., Drew, D. A., Leeming, E., Gibson, R., Le Roy, C., Khatib, H. A., Francis, L., Mazidi, M., Mompeo, O., Valles-Colomer, M., Tett, A., Beghini, F., Dubois, L., Bazzani, D., Thomas, A. M., … Segata, N. (2021). Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. *Nature Medicine*, *27*(2), Article 2. https://doi.org/10.1038/s41591-020-01183-8

Asnicar, F., Manara, S., Zolfo, M., Truong, D. T., Scholz, M., Armanini, F., Ferretti, P., Gorfer, V., Pedrotti, A., Tett, A., & Segata, N. (2017). Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling. *mSystems*, *2*(1), e00164-16. https://doi.org/10.1128/mSystems.00164-16

Atukunda, P., Muhoozi, G. K. M., van den Broek, T. J., Kort, R., Diep, L. M., Kaaya, A. N., Iversen, P. O., & Westerberg, A. C. (2019). Child development, growth and microbiota: Follow-up of a randomized education trial in Uganda. *Journal of Global Health*, *9*(1), 010431. https://doi.org/10.7189/jogh.09.010431

Azcarate-Peril, M. A., Roach, J., Marsh, A., Chey, W. D., Sandborn, W. J., Ritter, A. J., Savaiano, D. A., & Klaenhammer, T. R. (2021). A double-blind, 377-subject randomized study identifies Ruminococcus, Coprococcus, Christensenella, and Collinsella as long-term potential key players in the modulation of the gut microbiome of lactose intolerant individuals by galacto-oligosaccharides. *Gut Microbes*, *13*(1), 1957536. https://doi.org/10.1080/19490976.2021.1957536

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., … Jun, W. (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

Bajaj, J. S., Ng, S. C., & Schnabl, B. (2022). Promises of microbiome-based therapies. *Journal of Hepatology*, *76*(6), 1379–1391. https://doi.org/10.1016/j.jhep.2021.12.003

Balvočiūtė, M., & Huson, D. (2017). SILVA, RDP, Greengenes, NCBI and OTT — how do these taxonomies compare? *BMC Genomics*, *18*. https://doi.org/10.1186/s12864-017-3501-4

Banerjee, J., Mishra, N., & Dhas, Y. (2015). Metagenomics: A new horizon in cancer research. *Meta Gene*, *5*, 84–89. https://doi.org/10.1016/j.mgene.2015.05.005

Baptista, D., Moreira Teixeira, L., Barata, D., Tahmasebi Birgani, Z., King, J., van Riet, S., Pasman, T., Poot, A. A., Stamatialis, D., & Rottier, R. J. (2022). 3D Lung-on-Chip Model Based on Biomimetically Microcurved Culture Membranes. *ACS Biomaterials Science & Engineering*.

Barcik, W., Boutin, R. C. T., Sokolowska, M., & Finlay, B. B. (2020). The Role of Lung and Gut Microbiota in the Pathology of Asthma. *Immunity*, *52*(2), Article 2. https://doi.org/10.1016/j.immuni.2020.01.007

Bars-Cortina, D., Ramon, E., Rius-Sansalvador, B., Guinó, E., Garcia-Serrano, A., Mach, N., Khannous-Lleiffe, O., Saus, E., Gabaldón, T., Ibáñez-Sanz, G., Rodríguez-Alonso, L., Mata, A., García-Rodríguez, A., Obón-Santacana, M., & Moreno, V. (2024). Comparison between 16S rRNA and shotgun sequencing in colorectal cancer, advanced colorectal lesions, and healthy human gut microbiota. *BMC Genomics*, *25*(1), 730. https://doi.org/10.1186/s12864-024-10621-7

Basatemur, E., Gardiner, J., Williams, C., Melhuish, E., Barnes, J., & Sutcliffe, A. (2013). Maternal prepregnancy BMI and child cognition: A longitudinal cohort study. *Pediatrics*, *131*(1), 56–63. https://doi.org/10.1542/peds.2012-0788

Behary, J., Amorim, N., Jiang, X.-T., Raposo, A., Gong, L., McGovern, E., Ibrahim, R., Chu, F., Stephens, C., & Jebeili, H. (2021). Gut microbiota impact on the peripheral immune response in non-alcoholic fatty liver disease related hepatocellular carcinoma. *Nature Communications*, *12*(1), Article 1.

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 (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berding, K., Vlckova, K., Marx, W., Schellekens, H., Stanton, C., Clarke, G., Jacka, F., Dinan, T. G., & Cryan, J. F. (2021). Diet and the Microbiota–Gut–Brain Axis: Sowing the Seeds of Good Mental Health. *Advances in Nutrition*, *12*(4), 1239–1285. https://doi.org/10.1093/advances/nmaa181

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

Beri, K. (2018). Perspective: Stabilizing the Microbiome Skin-Gut-Brain Axis with Natural Plant Botanical Ingredients in Cosmetics. *Cosmetics*, *5*(2), Article 2. https://doi.org/10.3390/cosmetics5020037

R, K. B., S, S. C., & N, S. S. (2023). "Sharing the matrix" - a cooperative strategy for survival in Salmonella enterica serovar Typhimurium. *BMC microbiology*, *23*(1), 230. https://doi.org/10.1186/s12866-023-02972-0

Bhatt Lab. (2023). *bhattlab/AWIGen2Microbiome: Repository for the AWI-Gen 2 microbiome project workflows, analysis, and results.* https://github.com/bhattlab/AWIGen2Microbiome

Bhattacharjee, S., & Lukiw, W. J. (2013). Alzheimer's disease and the microbiome. *Frontiers in Cellular Neuroscience*, *7*, 153. https://doi.org/10.3389/fncel.2013.00153

Bhusri, B., Sutheeworapong, S., Kittichotirat, W., Kusonmano, K., Thammarongtham, C., Lertampaiporn, S., Prommeenate, P., Praphanphoj, V., Kittitharaphan, W., Dulsawat, S., Paenkaew, P., & Cheevadhanarak, S. (2025). Characterization of gut microbiota on gender and age groups bias in Thai patients with autism spectrum disorder. *Scientific Reports*, *15*(1), 2587. https://doi.org/10.1038/s41598-025-86740-2

Biadgilign, S., Mgutshini, T., Haile, D., Gebremichael, B., Moges, Y., & Tilahun, K. (2017). Epidemiology of obesity and overweight in sub-Saharan Africa: A protocol for a systematic review and meta-analysis. *BMJ Open*, *7*(11), e017666. https://doi.org/10.1136/bmjopen-2017-017666

Biol-Aquino, M. A., Perdiz, C. J., Borlagdan, M., Alcantara, J. D., & Mallillin, A. (2019). Differences in the bacterial profiles of infant gut by birth process, milk diet, and choice of 16S rRNA gene target region. *Human Microbiome Journal*, *13*(September), 100062. https://doi.org/10.1016/j.humic.2019.100062

Björk, J. R., Bolte, L. A., Maltez Thomas, A., Lee, K. A., Rossi, N., Wind, T. T., Smit, L. M., Armanini, F., Asnicar, F., Blanco-Miguez, A., Board, R., Calbet-Llopart, N., Derosa, L., Dhomen, N., Brooks, K., Harland, M., Harries, M., Lorigan, P., Manghi, P., … Weersma, R. K. (2024). Longitudinal gut microbiome changes in immune checkpoint blockade-treated advanced melanoma. *Nature Medicine*, *30*(3), 785–796. https://doi.org/10.1038/s41591-024-02803-3

Blanton, L. V., Barratt, M. J., Charbonneau, M. R., Ahmed, T., & Gordon, J. I. (2016). Childhood undernutrition, the gut microbiota, and microbiota-directed therapeutics. *Science*, *352*(6293). https://doi.org/10.1126/science.aad9359

Blot, F., Marchix, J., Ejarque, M., Jimenez, S., Meunier, A., Keime, C., Trottier, C., Croyal, M., Lapp, C., Mahe, M. M., De Arcangelis, A., & Gradwohl, G. (2023). Gut Microbiota Remodeling and Intestinal Adaptation to Lipid Malabsorption After Enteroendocrine Cell Loss in Adult Mice.

Cellular and Molecular Gastroenterology and Hepatology, *15*(6), 1443–1461. https://doi.org/10.1016/j.jcmgh.2023.02.013

Boachie, M. K., Thsehla, E., Immurana, M., Kohli-Lynch, C., & Hofman, K. J. (2022). Estimating the healthcare cost of overweight and obesity in South Africa. *Global Health Action*, *15*(1), 2045092. https://doi.org/10.1080/16549716.2022.2045092

Boddy, S. L., Giovannelli, I., Sassani, M., Cooper-Knock, J., Snyder, M. P., Segal, E., Elinav, E., Barker, L. A., Shaw, P. J., & McDermott, C. J. (2021). The gut microbiome: A key player in the complexity of amyotrophic lateral sclerosis (ALS). *BMC Medicine*, *19*, 13. https://doi.org/10.1186/s12916-020-01885-3

Bokulich Nicholas A., Dillon Matthew R., Zhang Yilong, Rideout Jai Ram, Bolyen Evan, Li Huilin, Albert Paul S., & Caporaso J. Gregory. (2018). q2-longitudinal: Longitudinal and Paired-Sample Analyses of Microbiome Data. *mSystems*, *3*(6), 10.1128/msystems.00219-18. https://doi.org/10.1128/msystems.00219-18

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

Bonder, M. J., Tigchelaar, E. F., Cai, X., Trynka, G., Cenit, M. C., Hrdlickova, B., Zhong, H., Vatanen, T., Gevers, D., Wijmenga, C., Wang, Y., & Zhernakova, A. (2016). The influence of a short-term gluten-free diet on the human gut microbiome. *Genome Medicine*, *8*(1), 45. https://doi.org/10.1186/s13073-016-0295-y

Bonfili, L., Cecarini, V., Berardi, S., Scarpona, S., Suchodolski, J. S., Nasuti, C., Fiorini, D., Boarelli, M. C., Rossi, G., & Eleuteri, A. M. (2017). Microbiota modulation counteracts Alzheimer's disease progression influencing neuronal proteolysis and gut hormones plasma levels. *Scientific Reports*, *7*(1), 2426. https://doi.org/10.1038/s41598-017-02587-2

Bourdeau-Julien, I., Castonguay-Paradis, S., Rochefort, G., Perron, J., Lamarche, B., Flamand, N., Di Marzo, V., Veilleux, A., & Raymond, F. (2023). The diet rapidly and differentially affects the gut microbiota and host lipid mediators in a healthy population. *Microbiome*, *11*(1), 26. https://doi.org/10.1186/s40168-023-01469-2

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., Schulz, F., Jarett, J., Rivers, A. R., Eloe-Fadrosh, E. A., Tringe, S. G., Ivanova, N. N., Copeland, A., Clum, A., Becraft, E. D., Malmstrom, R. R., Birren, B., Podar, M., Bork, P., … Woyke, T. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, *35*(8), 725–731. https://doi.org/10.1038/nbt.3893

Bowyer, R. C. E., Jackson, M. A., Le Roy, C. I., Ni Lochlainn, M., Spector, T. D., Dowd, J. B., & Steves, C. J. (2019). Socioeconomic Status and the Gut Microbiome: A TwinsUK Cohort Study. *Microorganisms*, *7*(1), Article 1. https://doi.org/10.3390/microorganisms7010017

Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, *27*(4), 325–349. https://doi.org/10.2307/1942268

Brazma, A. (2009). Minimum Information About a Microarray Experiment (MIAME) – Successes, Failures, Challenges. *TheScientificWorldJournal*, *9*, 420–423. https://doi.org/10.1100/tsw.2009.57

Brenchley, J. M., & Douek, D. C. (2008). HIV infection and the gastrointestinal immune system. *Mucosal Immunology*, *1*(1), Article 1. https://doi.org/10.1038/mi.2007.1

Bressa, C., Bailén-Andrino, M., Pérez-Santiago, J., González-Soltero, R., Pérez, M., Montalvo-Lominchar, M. G., Maté-Muñoz, J. L., Domínguez, R., Moreno, D., & Larrosa, M. (2017). Differences in gut microbiota profile between women with active lifestyle and sedentary women. *PloS One*, *12*(2), e0171352. https://doi.org/10.1371/journal.pone.0171352

Brewster, R., Tamburini, F. B., Asiimwe, E., Oduaran, O., Hazelhurst, S., & Bhatt, A. S. (2019). Surveying Gut Microbiome Research in Africans: Toward Improved Diversity and Representation. *Trends in Microbiology*, *27*(10), 824–835. https://doi.org/10.1016/j.tim.2019.05.006

Brown, K., DeCoffe, D., Molcan, E., & Gibson, D. L. (2012). Diet-Induced Dysbiosis of the Intestinal Microbiota and the Effects on Immunity and Disease. *Nutrients*, *4*(8), Article 8. https://doi.org/10.3390/nu4081095

Bucher-Johannessen, C., Senthakumaran, T., Avershina, E., Birkeland, E., Hoff, G., Bemanian, V., Tunsjø, H., & Rounge, T. B. (2024). *Species-level verification of Phascolarctobacterium*

association to colorectal cancer (p. 2024.03.13.24304214). medRxiv. https://doi.org/10.1101/2024.03.13.24304214

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Bui, T. P. N., Mannerås-Holm, L., Puschmann, R., Wu, H., Troise, A. D., Nijsse, B., Boeren, S., Bäckhed, F., Fiedler, D., & deVos, W. M. (2021). Conversion of dietary inositol into propionate and acetate by commensal Anaerostipes associates with host health. *Nature Communications*, *12*(1), 4798. https://doi.org/10.1038/s41467-021-25081-w

Buret, A. G., & Bhargava, A. (2014). Modulatory mechanisms of enterocyte apoptosis by viral, bacterial and parasitic pathogens. *Critical Reviews in Microbiology*, *40*(1), 1–17. https://doi.org/10.3109/1040841X.2012.746952

Busato, S., Gordon, M., Chaudhari, M., Jensen, I., Akyol, T., Andersen, S., & Williams, C. (2023). Compositionality, sparsity, spurious heterogeneity, and other data-driven challenges for machine learning algorithms within plant microbiome studies. *Current Opinion in Plant Biology*, *71*, 102326. https://doi.org/10.1016/j.pbi.2022.102326

Cabezas, M. P., Fonseca, N. A., & Muñoz-Mérida, A. (2024). MIMt: A curated 16S rRNA reference database with less redundancy and higher accuracy at species-level identification. *Environmental Microbiome*, *19*(1), 88. https://doi.org/10.1186/s40793-024-00634-w

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Caminero, A., Herrán, A. R., Nistal, E., Pérez-Andrés, J., Vaquero, L., Vivas, S., Ruiz de Morales, J. M. G., Albillos, S. M., & Casqueiro, J. (2014). Diversity of the cultivable human gut microbiome involved in gluten metabolism: Isolation of microorganisms with potential interest for coeliac disease. *FEMS Microbiology Ecology*, *88*(2), 309–319. https://doi.org/10.1111/1574-6941.12295

Campbell, F., Archer, B., Laurenson-Schafer, H., Jinnai, Y., Konings, F., Batra, N., Pavlin, B., Vandemaele, K., Van Kerkhove, M. D., Jombart, T., Morgan, O., & le Polain de Waroux, O. (2021). Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Eurosurveillance*, *26*(24). https://doi.org/10.2807/1560-7917.ES.2021.26.24.2100509

Campos, P. M., Darwish, N., Shao, J., & Proszkowiec-Weglarz, M. (2022). Research Note: Choice of microbiota database affects data analysis and interpretation in chicken cecal microbiota. *Poultry Science*, *101*(8), 101971. https://doi.org/10.1016/j.psj.2022.101971

Campos-Bayardo, T. I., Román-Rojas, D., García-Sánchez, A., Cardona-Muñoz, E. G., Sánchez-Lozano, D. I., Totsuka-Sutto, S., Gómez-Hermosillo, L. F., Casillas-Moreno, J., Andrade-Sierra, J., Pazarín-Villaseñor, L., Campos-Pérez, W., Martínez-López, E., & Miranda-Díaz, A. G. (2025). The Role of TLRs in Obesity and Its Related Metabolic Disorders. *International Journal of Molecular Sciences*, *26*(5), 2229. https://doi.org/10.3390/ijms26052229

Canfora, E. E., Jocken, J. W., & Blaak, E. E. (2015). Short-chain fatty acids in control of body weight and insulin sensitivity. *Nature Reviews. Endocrinology*, *11*(10), 577–591. https://doi.org/10.1038/nrendo.2015.128

Carter, M. M., Olm, M. R., Merrill, B. D., Dahan, D., Tripathi, S., Spencer, S. P., Yu, F. B., Jain, S., Neff, N., Jha, A. R., Sonnenburg, E. D., & Sonnenburg, J. L. (2023). Ultra-deep sequencing of Hadza hunter-gatherers recovers vanishing gut microbes. *Cell*, *186*(14), 3111-3124.e13. https://doi.org/10.1016/j.cell.2023.05.046

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*. https://doi.org/10.1038/s41564-022-01184-y

Caudill, M. T., & Brayton, K. A. (2022). The Use and Limitations of the 16S rRNA Sequence for Species Classification of Anaplasma Samples. *Microorganisms*, *10*(3), 605. https://doi.org/10.3390/microorganisms10030605

Cha, T., Kim, H. H., Keum, J., Kwak, M.-J., Park, J. Y., Hoh, J. K., Kim, C.-R., Jeon, B.-H., & Park, H.-K. (2023). Gut microbiome profiling of neonates using Nanopore MinION and Illumina MiSeq sequencing. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1148466

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2022). GTDB-Tk v2: Memory friendly classification with the genome taxonomy database. *Bioinformatics (Oxford, England)*, *38*(23), 5315–5316. https://doi.org/10.1093/bioinformatics/btac672

Chaurasia, B., & Summers, S. A. (2015). Ceramides - Lipotoxic Inducers of Metabolic Disorders. *Trends in endocrinology and metabolism: TEM*, *26*(10), 538–550. https://doi.org/10.1016/j.tem.2015.07.006

Chen, L. A., & Boyle, K. (2024). The Role of the Gut Microbiome in Health and Disease in the Elderly. *Current Gastroenterology Reports*, *26*(9), 217–230. https://doi.org/10.1007/s11894-024-00932-w

Chen, Y., Li, Z., Hu, S., Zhang, J., Wu, J., Shao, N., Bo, X., Ni, M., & Ying, X. (2017). Gut metagenomes of type 2 diabetic patients have characteristic single-nucleotide polymorphism distribution in Bacteroides coprocola. *Microbiome*, *5*(1), 15. https://doi.org/10.1186/s40168-017-0232-3

Cheung, M. K., Lam, W. Y., Fung, W. Y. W., Law, P. T. W., Au, C. H., Nong, W., Kam, K. M., Kwan, H. S., & Tsui, S. K. W. (2013). Sputum Microbiota in Tuberculosis as Revealed by 16S rRNA Pyrosequencing. *PLoS ONE*, *8*(1), e54574. https://doi.org/10.1371/journal.pone.0054574

Chevallier, C., Hacquin, A.-S., & Mercier, H. (2021). COVID-19 Vaccine Hesitancy: Shortening the Last Mile. *Trends in Cognitive Sciences*, *25*(5), 331–333. https://doi.org/10.1016/j.tics.2021.02.002

Chey, W., Sandborn, W., Ritter, A. J., Foyt, H., Azcarate-Peril, M. A., & Savaiano, D. A. (2020). Galacto-Oligosaccharide RP-G28 Improves Multiple Clinical Outcomes in Lactose-Intolerant Patients. *Nutrients*, *12*(4), 1058. https://doi.org/10.3390/nu12041058

Chigurupati, H., Auddy, B., Biyani, M., Chakrabarti, S., & Stohs, S. J. (2018). Prevention of alcohol-induced DNA damage by a proprietary glycyrrhizin/D-mannitol product: A randomized, placebo-controlled, cross-over human study. *Alcohol (Fayetteville, N.Y.)*, *69*, 33–39. https://doi.org/10.1016/j.alcohol.2017.11.001

Chiu, K., Warner, G., Nowak, R. A., Flaws, J. A., & Mei, W. (2020). The Impact of Environmental Chemicals on the Gut Microbiome. *Toxicological Sciences*, *176*(2), 253–284. https://doi.org/10.1093/toxsci/kfaa065

Chu, D. M., Ma, J., Prince, A. L., Antony, K. M., Seferovic, M. D., & Aagaard, K. M. (2017). Maturation of the Infant Microbiome Community Structure and Function Across Multiple Body Sites and in Relation to Mode of Delivery Derrick. *Nature Medicine*, *23*, 314–326. https://doi.org/10.5588/ijtld.16.0716.Isoniazid

Chung, W. S. F., Walker, A. W., Louis, P., Parkhill, J., Vermeiren, J., Bosscher, D., Duncan, S. H., & Flint, H. J. (2016). Modulation of the human gut microbiota by dietary fibres occurs at the species level. *BMC Biology*, *14*(1), 3. https://doi.org/10.1186/s12915-015-0224-3

Ciuffreda, L., Rodríguez-Pérez, H., & Flores, C. (2021). Nanopore sequencing and its application to the study of microbial communities. *Computational and Structural Biotechnology Journal*, *19*, 1497–1511. https://doi.org/10.1016/j.csbj.2021.02.020

Claassen-Weitz, S., Gardner-Lubbe, S., Nicol, P., Botha, G., Mounaud, S., Shankar, J., Nierman, W. C., Mulder, N., Budree, S., Zar, H. J., Nicol, M. P., & Kaba, M. (2018). HIV-exposure, early life feeding practices and delivery mode impacts on faecal bacterial profiles in a South African birth cohort. *Scientific Reports*, *8*, 5078. https://doi.org/10.1038/s41598-018-22244-6

Claesson, M. J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P., & O'Toole, P. W. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Research*, *38*(22), e200. https://doi.org/10.1093/nar/gkq873

Claus, H. (2014). Microbial Degradation of 2,4,6-Trinitrotoluene In Vitro and in Natural Environments. In *Environmental Science and Engineering (Subseries: Environmental Science)* (Vol. 112, pp. 15–38). https://doi.org/10.1007/978-3-319-01083-0_2

Clemente-Suárez, V. J., Beltrán-Velasco, A. I., Redondo-Flórez, L., Martín-Rodríguez, A., & Tornero-Aguilera, J. F. (2023). Global Impacts of Western Diet and Its Effects on Metabolism and Health: A Narrative Review. *Nutrients*, *15*(12), 2749. https://doi.org/10.3390/nu15122749

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

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Complete Genomics. (2024, September 16). *Next-Generation Sequencing Costs: The Sub $100 Genome*. https://www.completegenomics.com/next-generation-sequencing-costs/

Costea, P. I., Coelho, L. P., Sunagawa, S., Munch, R., Huerta-Cepas, J., Forslund, K., Hildebrand, F., Kushugulova, A., Zeller, G., & Bork, P. (2017). Subspecies in the global human gut microbiome. *Molecular Systems Biology*, *13*(12), 960. https://doi.org/10.15252/msb.20177589

Craven, L., Rahman, A., Nair Parvathy, S., Beaton, M., Silverman, J., Qumosani, K., Hramiak, I., Hegele, R., Joy, T., Meddings, J., Urquhart, B., Harvie, R., McKenzie, C., Summers, K., Reid, G., Burton, J. P., & Silverman, M. (2020). Allogenic Fecal Microbiota Transplantation in Patients With Nonalcoholic Fatty Liver Disease Improves Abnormal Small Intestinal Permeability: A Randomized Control Trial. *Official Journal of the American College of Gastroenterology | ACG*, *115*(7), 1055–1065. https://doi.org/10.14309/ajg.0000000000000661

CSIR. (2019). *CSIR Annual Report 2018/19*.

Cui, Z., Zhou, Y., Li, H., Zhang, Y., Zhang, S., Tang, S., & Guo, X. (2012). Complex sputum microbial composition in patients with pulmonary tuberculosis. *BMC Microbiology*, *12*(1), 276. https://doi.org/10.1186/1471-2180-12-276

Cullen, J. M. A., Shahzad, S., & Dhillon, J. (2023). A systematic review on the effects of exercise on gut microbial diversity, taxonomic composition, and microbial metabolites: Identifying research gaps and future directions. *Frontiers in Physiology*, *14*. https://doi.org/10.3389/fphys.2023.1292673

Cumin, C., Huang, Y.-L., Everest-Dass, A., & Jacob, F. (2021). Deciphering the Importance of Glycosphingolipids on Cellular and Molecular Mechanisms Associated with Epithelial-to-Mesenchymal Transition in Cancer. *Biomolecules*, *11*(1), 62. https://doi.org/10.3390/biom11010062

Da Silva, H. E., Teterina, A., Comelli, E. M., Taibi, A., Arendt, B. M., Fischer, S. E., Lou, W., & Allard, J. P. (2018). Nonalcoholic fatty liver disease is associated with dysbiosis independent of body mass index and insulin resistance. *Scientific Reports*, *8*(1), Article 1.

Dai, D. L. Y., Petersen, C., & Turvey, S. E. (2024). Reduce, reinforce, and replenish: Safeguarding the early-life microbiota to reduce intergenerational health disparities. *Frontiers in Public Health*, *12*. https://doi.org/10.3389/fpubh.2024.1455503

Dang, A. T., & Marsland, B. J. (2019). Microbes, metabolites, and the gut–lung axis. *Mucosal Immunology*, *12*(4), Article 4.

Daniluk, U., Alifier, M., & Kaczmarski, M. (2012). Probiotic-induced apoptosis and its potential relevance to mucosal inflammation of gastrointestinal tract. *Advances in Medical Sciences*, *57*(2), 175–182. https://doi.org/10.2478/v10039-012-0025-7

Das, M. K., Bishwal, S. C., Das, A., Dabral, D., Badireddy, V. K., Pandit, B., Varghese, G. M., & Nanda, R. K. (2015). Deregulated tyrosine-phenylalanine metabolism in pulmonary tuberculosis patients. *Journal of Proteome Research*, *14*(4), 1947–1956. https://doi.org/10.1021/acs.jproteome.5b00016

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

de Almeida, N. E. C., Esteves, F. G., Dos Santos-Pinto, J. R. A., Peres de Paula, C., da Cunha, A. F., Malavazi, I., Palma, M. S., & Rodrigues-Filho, E. (2020). Digestion of Intact Gluten Proteins by Bifidobacterium Species: Reduction of Cytotoxicity and Proinflammatory Responses. *Journal of Agricultural and Food Chemistry*, *68*(15), 4485–4492. https://doi.org/10.1021/acs.jafc.0c01421

de Goffau, M. C., Jallow, A. T., Sanyang, C., Prentice, A. M., Meagher, N., Price, D. J., Revill, P. A., Parkhill, J., Pereira, D. I. A., & Wagner, J. (2022). Gut microbiomes from Gambian infants reveal the development of a non-industrialized Prevotella-based trophic network. *Nature Microbiology*, *7*(1), 132–144. https://doi.org/10.1038/s41564-021-01023-6

de Groot, P., Nikolic, T., Pellegrini, S., Sordi, V., Imangaliyev, S., Rampanelli, E., Hanssen, N., Attaye, I., Bakker, G., Duinkerken, G., Joosten, A., Prodan, A., Levin, E., Levels, H., Potter van Loon, B., van Bon, A., Brouwer, C., van Dam, S., Simsek, S., … Nieuwdorp, M. (2021). Faecal microbiota transplantation halts progression of human new-onset type 1 diabetes in a randomised controlled trial. *Gut*, *70*(1), 92–105. https://doi.org/10.1136/gutjnl-2020-322630

de la Cuesta-Zuluaga, J., Kelley, S. T., Chen, Y., Escobar, J. S., Mueller, N. T., Ley, R. E., McDonald, D., Huang, S., Swafford, A. D., Knight, R., & Thackray, V. G. (2019). Age- and Sex-Dependent Patterns of Gut Microbial Diversity in Human Adults. *mSystems*, *4*(4), 10.1128/msystems.00261-19. https://doi.org/10.1128/msystems.00261-19

de Vos, W. M. (2017). Microbe Profile: Akkermansia muciniphila: a conserved intestinal symbiont that acts as the gatekeeper of our mucosa. *Microbiology (Reading, England)*, *163*(5), 646–648. https://doi.org/10.1099/mic.0.000444

Dekaboruah, E., Suryavanshi, M. V., Chettri, D., & Verma, A. K. (2020). Human microbiome: An academic update on human body site specific surveillance and its possible role. *Archives of Microbiology*, *202*(8), Article 8. https://doi.org/10.1007/s00203-020-01931-x

Del Castillo-Izquierdo, Á., Mayneris-Perxachs, J., & Fernández-Real, J. M. (2022). Bidirectional relationships between the gut microbiome and sexual traits. *American Journal of Physiology. Cell Physiology*, *322*(6), C1223–C1229. https://doi.org/10.1152/ajpcell.00116.2022

Depommier, C., Everard, A., Druart, C., Plovier, H., Van Hul, M., Vieira-Silva, S., Falony, G., Raes, J., Maiter, D., Delzenne, N. M., de Barsy, M., Loumaye, A., Hermans, M. P., Thissen, J.-P., de Vos, W. M., & Cani, P. D. (2019). Supplementation with Akkermansia muciniphila in overweight and obese human volunteers: A proof-of-concept exploratory study. *Nature Medicine*, *25*(7), 1096–1103. https://doi.org/10.1038/s41591-019-0495-2

Derrien, M., Vaughan, E. E., Plugge, C. M., & de Vos, W. M. (2004). Akkermansia muciniphila gen. Nov., sp. Nov., a human intestinal mucin-degrading bacterium. *International Journal of Systematic and Evolutionary Microbiology*, *54*(Pt 5), 1469–1476. https://doi.org/10.1099/ijs.0.02873-0

Dethlefsen, L., & Relman, D. A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proceedings of the National Academy of Sciences*, *108*(Supplement_1), 4554–4561. https://doi.org/10.1073/pnas.1000087107

Dexter, F., Bayman, E. O., & Dexter, E. U. (2017). Monte Carlo Simulations Comparing Fisher Exact Test and Unequal Variances t Test for Analysis of Differences Between Groups in Brief Hospital Lengths of Stay. *Anesthesia & Analgesia*, *125*(6), 2141. https://doi.org/10.1213/ANE.0000000000002428

Díaz-López, A., Bulló, M., Martínez-González, M. A., Corella, D., Estruch, R., Fitó, M., Gómez-Gracia, E., Fiol, M., García de la Corte, F. J., Ros, E., Babio, N., Serra-Majem, L., Pintó, X., Muñoz, M. Á., Francés, F., Buil-Cosiales, P., & Salas-Salvadó, J. (2016). Dairy product consumption and risk of type 2 diabetes in an elderly Spanish Mediterranean population at high cardiovascular risk. *European Journal of Nutrition*, *55*(1), 349–360. https://doi.org/10.1007/s00394-015-0855-8

Dimidi, E., Cox, S. R., Rossi, M., & Whelan, K. (2019). Fermented Foods: Definitions and Characteristics, Impact on the Gut Microbiota and Effects on Gastrointestinal Health and Disease. *Nutrients*, *11*(8), E1806. https://doi.org/10.3390/nu11081806

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 of the United States of America*, *107*(26), 11971–11975. https://doi.org/10.1073/pnas.1002601107

Doroszkiewicz, J., Groblewska, M., & Mroczko, B. (2021). The Role of Gut Microbiota and Gut–Brain Interplay in Selected Diseases of the Central Nervous System. *International Journal of Molecular Sciences*, *22*(18), Article 18. https://doi.org/10.3390/ijms221810028

Douglas, G. M. (2021, April 2). *Full pipeline script*. GitHub. https://github.com/picrust/picrust2/wiki/Full-pipeline-script

Douglas, G. M., Beiko, R. G., & Langille, M. G. I. (2018). Predicting the Functional Potential of the Microbiome from Marker Genes Using PICRUSt. *Methods in Molecular Biology (Clifton, N.J.)*, *1849*, 169–177. https://doi.org/10.1007/978-1-4939-8728-3_11

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*, *38*(6), 685–688. https://doi.org/10.1038/s41587-020-0548-6

Douillard, F. P., & de Vos, W. M. (2019). Biotechnology of health-promoting bacteria. *Biotechnology Advances*, *37*(6), 107369. https://doi.org/10.1016/j.biotechadv.2019.03.008

Duan, R., Zhang, C., Li, G., Li, J., & Duan, L. (2024). Antibiotic Exposure and Risk of New-Onset Inflammatory Bowel Disease: A Systematic Review and Dose-Response Meta-Analysis. *Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association*, S1542-3565(24)00218-0. https://doi.org/10.1016/j.cgh.2024.02.010

Duranti, S., Ruiz, L., Lugli, G. A., Tames, H., Milani, C., Mancabelli, L., Mancino, W., Longhi, G., Carnevali, L., Sgoifo, A., Margolles, A., Ventura, M., Ruas-Madiedo, P., & Turroni, F. (2020). Bifidobacterium adolescentis as a key member of the human gut microbiota in the production of GABA. *Scientific Reports*, *10*(1), 14112. https://doi.org/10.1038/s41598-020-70986-z

Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D., & De Cesare, A. (2021). Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Scientific Reports*, *11*(1), 3030. https://doi.org/10.1038/s41598-021-82726-y

Dwiyanto, J., Hussain, M. H., Reidpath, D., Ong, K. S., Qasim, A., Lee, S. W. H., Lee, S. M., Foo, S. C., Chong, C. W., & Rahman, S. (2021). Ethnicity influences the gut microbiota of individuals sharing a geographical location: A cross-sectional study from a middle-income country. *Scientific Reports*, *11*(1), 2618. https://doi.org/10.1038/s41598-021-82311-3

Earl, J. P., Adappa, N. D., Krol, J., Bhat, A. S., Balashov, S., Ehrlich, R. L., Palmer, J. N., Workman, A. D., Blasetti, M., Sen, B., Hammond, J., Cohen, N. A., Ehrlich, G. D., & Mell, J. C. (2018). Species-level bacterial community profiling of the healthy sinonasal microbiome using Pacific Biosciences sequencing of full-length 16S rRNA genes 06 Biological Sciences 0604 Genetics 06 Biological Sciences 0605 Microbiology. *Microbiome*, *6*(1). https://doi.org/10.1186/s40168-018-0569-2

Echegaray, N., Yilmaz, B., Sharma, H., Kumar, M., Pateiro, M., Ozogul, F., & Lorenzo, J. M. (2023). A novel approach to Lactiplantibacillus plantarum: From probiotic properties to the omics insights. *Microbiological Research*, *268*, 127289. https://doi.org/10.1016/j.micres.2022.127289

Ecklu-Mensah, G., Choo-Kang, C., Maseng, M. G., Donato, S., Bovet, P., Viswanathan, B., Bedu-Addo, K., Plange-Rhule, J., Oti Boateng, P., Forrester, T. E., Williams, M., Lambert, E. V., Rae, D., Sinyanya, N., Luke, A., Layden, B. T., O'Keefe, S., Gilbert, J. A., & Dugas, L. R. (2023). Gut microbiota and fecal short chain fatty acids differ with adiposity and country of origin: the METS-microbiome study. *Nature communications*, *14*(1), 5160. https://doi.org/10.1038/s41467-023-40874-x

Eren, A. M., Borisy, G. G., Huse, S. M., & Mark Welch, J. L. (2014). Oligotyping analysis of the human oral microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(28), E2875-2884. https://doi.org/10.1073/pnas.1409644111

Espín, J. C., González-Sarrías, A., & Tomás-Barberán, F. A. (2017). The gut microbiota: A key factor in the therapeutic effects of (poly)phenols. *Biochemical Pharmacology*, *139*, 82–93. https://doi.org/10.1016/j.bcp.2017.04.033

Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J. P., Druart, C., Bindels, L. B., Guiot, Y., Derrien, M., Muccioli, G. G., Delzenne, N. M., Vos, W. M. de, & Cani, P. D. (2013). Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proceedings of the National Academy of Sciences*, *110*(22), 9066–9071. https://doi.org/10.1073/pnas.1219451110

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Ezeji, J. C., Sarikonda, D. K., Hopperton, A., Erkkila, H. L., Cohen, D. E., Martinez, S. P., Cominelli, F., Kuwahara, T., Dichosa, A. E. K., Good, C. E., Jacobs, M. R., Khoretonenko, M., Veloo, A., & Rodriguez-Palacios, A. (2021). Parabacteroides distasonis: Intriguing aerotolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health. *Gut Microbes*, *13*(1), 1922241. https://doi.org/10.1080/19490976.2021.1922241

FACTS. (2019, June 26). *Crafting a new era of fermented foods and beverages | FACTS in Regulatory & Nutrition*. https://www.factssa.com/news/crafting-a-new-era-of-fermented-foods-and-beverages/

Fairweather-Tait, S. J., Bao, Y., Broadley, M. R., Collings, R., Ford, D., Hesketh, J. E., & Hurst, R. (2011). Selenium in human health and disease. *Antioxidants & Redox Signaling*, *14*(7), 1337–1383. https://doi.org/10.1089/ars.2010.3275

Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., Clemente, J. C., Knight, R., Heath, A. C., Leibel, R. L., Rosenbaum, M., & Gordon, J. I. (2013). The long-term stability of the human gut microbiota. *Science (New York, N.Y.)*, *341*(6141), 1237439. https://doi.org/10.1126/science.1237439

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

Fang, P., Kazmi, S. A., Jameson, K. G., & Hsiao, E. Y. (2020). The Microbiome as a Modifier of Neurodegenerative Disease Risk. *Cell Host & Microbe*, *28*(2), 201–222. https://doi.org/10.1016/j.chom.2020.06.008

Farhana, L., Antaki, F., Murshed, F., Mahmud, H., Judd, S. L., Nangia-Makker, P., Levi, E., Yu, Y., & Majumdar, A. P. (2018). Gut microbiome profiling and colorectal cancer in African Americans and Caucasian Americans. *World Journal of Gastrointestinal Pathophysiology*, *9*(2), 47–58. https://doi.org/10.4291/wjgp.v9.i2.47

FDA. (2022, November 30). *FDA Approves First Fecal Microbiota Product*. FDA; FDA. https://www.fda.gov/news-events/press-announcements/fda-approves-first-fecal-microbiota-product

Fei N, Bernabé BP, Lie L, Baghdan D, Bedu-Addo K, et al. (2019). The human microbiota is associated with cardiometabolic risk across the epidemiologic transition. *PLOS ONE* 14(7): e0215262. https://doi.org/10.1371/journal.pone.0215262

Fei N, Choo-Kang C, Reutrakul S, Crowley SJ, Rae D, et al. (2021). Gut microbiota alterations in response to sleep length among African-origin adults. *PLOS ONE* 16(9): e0255323. https://doi.org/10.1371/journal.pone.0255323

Feingold, K. R., & Grunfeld, C. (2011). The role of HDL in innate immunity. *Journal of lipid research*, *52*(1), 1-3. https://doi.org/10.1194/jlr.E012138

Feng, J., Peng, J., Hsiao, Y.-C., Liu, C.-W., Yang, Y., Zhao, H., Teitelbaum, T., Wang, X., & Lu, K. (2024). Non/Low-Caloric Artificial Sweeteners and Gut Microbiome: From Perturbed Species to Mechanisms. *Metabolites*, *14*(10), 544. https://doi.org/10.3390/metabo14100544

Feng, Q., Liu, Z., Zhong, S., Li, R., Xia, H., Jie, Z., Wen, B., Chen, X., Yan, W., Fan, Y., Guo, Z., Meng, N., Chen, J., Yu, X., Zhang, Z., Kristiansen, K., Wang, J., Xu, X., He, K., & Li, G. (2016). Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep22525

Feng, Y., Jaratlerdsiri, W., Patrick, S. M., Lyons, R. J., Haynes, A.-M., Collins, C. C., Stricker, P. D., Bornman, M. S. R., & Hayes, V. M. (2019). Metagenomic analysis reveals a rich bacterial content in high-risk prostate tumors from African men. *The Prostate*, *79*(15), 1731–1738. https://doi.org/10.1002/pros.23897

Feraco, A., Armani, A., Gorini, S., Camajani, E., Quattrini, C., Filardi, T., Karav, S., Strollo, R., Caprio, M., & Lombardo, M. (2024). Gender Differences in Dietary Patterns and Eating Behaviours in Individuals with Obesity. *Nutrients*, *16*(23), Article 23. https://doi.org/10.3390/nu16234226

Ferdous, T., Jiang, L., Dinu, I., Groizeleau, J., Kozyrskyj, A. L., Greenwood, C. M. T., & Arrieta, M.-C. (2022). The rise to power of the microbiome: Power and sample size calculation for microbiome studies. *Mucosal Immunology*, *15*(6), 1060–1070. https://doi.org/10.1038/s41385-022-00548-1

Ferreira, A. M., Moreira, F., Guimaraes, T., Spadão, F., Ramos, J. F., Batista, M. V., Filho, J. S., Costa, S. F., & Rocha, V. (2018). Epidemiology, risk factors and outcomes of multi-drug-resistant bloodstream infections in haematopoietic stem cell transplant recipients: Importance of previous gut colonization. *Journal of Hospital Infection*, *100*(1), 83–91. https://doi.org/10.1016/j.jhin.2018.03.004

Filippo, C. D., Cavalieri, D., Paola, M. D., Ramazzotti, M., Poullet, J. B., Massart, S., Collini, S., Pieraccini, G., & Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(33), 14691–14696. https://doi.org/10.1073/pnas.1005963107

Flygel, T. T., Sovershaeva, E., Claassen-Weitz, S., Hjerde, E., Mwaikono, K. S., Odland, J. Ø., Ferrand, R. A., Mchugh, G., Gutteberg, T. J., Nicol, M. P., Cavanagh, J. P., & Flægstad, T. (2020). Composition of Gut Microbiota of Children and Adolescents With Perinatal Human Immunodeficiency Virus Infection Taking Antiretroviral Therapy in Zimbabwe. *The Journal of Infectious Diseases*, *221*(3), 483–492. https://doi.org/10.1093/infdis/jiz473

Fouladi, F., Bailey, M. J., Patterson, W. B., Sioda, M., Blakley, I. C., Fodor, A. A., Jones, R. B., Chen, Z., Kim, J. S., Lurmann, F., Martino, C., Knight, R., Gilliland, F. D., & Alderete, T. L. (2020). Air pollution exposure is associated with the gut microbiome as revealed by shotgun metagenomic sequencing. *Environment International*, *138*, 105604. https://doi.org/10.1016/j.envint.2020.105604

Franzosa, E. A., Huang, K., Meadow, J. F., Gevers, D., Lemon, K. P., Bohannan, B. J. M., & Huttenhower, C. (2015). Identifying personal microbiomes using metagenomic codes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(22), E2930-2938. https://doi.org/10.1073/pnas.1423854112

Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., Lipson, K. S., Knight, R., Caporaso, J. G., Segata, N., & Huttenhower, C. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*, *15*(11), 962–968. https://doi.org/10.1038/s41592-018-0176-y

Frati, F., Salvatori, C., Incorvaia, C., Bellucci, A., Di Cara, G., Marcucci, F., & Esposito, S. (2018). The Role of the Microbiome in Asthma: The Gut–Lung Axis. *International Journal of Molecular Sciences*, *20*(1), 123. https://doi.org/10.3390/ijms20010123

Fu, Y., Gu, Z., Cao, H., Zuo, C., Huang, Y., Song, Y., Jiang, Y., & Wang, F. (2024). The role of the gut microbiota in neurodegenerative diseases targeting metabolism. *Frontiers in Neuroscience*, *18*. https://doi.org/10.3389/fnins.2024.1432659

Fujisaka, S., Watanabe, Y., & Tobe, K. (2023). *The gut microbiome: A core regulator of metabolism*. https://doi.org/10.1530/JOE-22-0111

Fujita, K., Matsushita, M., De Velasco, M. A., Hatano, K., Minami, T., Nonomura, N., & Uemura, H. (2023). The Gut-Prostate Axis: A New Perspective of Prostate Cancer Biology through the Gut Microbiome. *Cancers*, *15*(5), 1375. https://doi.org/10.3390/cancers15051375

Fung, T. C., Hsiao, E. Y., & Olsen, C. A. (2017). Interactions between the microbiota, immune and nervous systems in health and disease | Nature Neuroscience. *Nature Neuroscience*, *20*, 145–155. https://doi.org/10.1038/nn.4476

Gacesa, R., Kurilshikov, A., Vich Vila, A., Sinha, T., Klaassen, M. a. Y., Bolte, L. A., Andreu-Sánchez, S., Chen, L., Collij, V., Hu, S., Dekens, J. a. M., Lenters, V. C., Björk, J. R., Swarte, J. C., Swertz, M. A., Jansen, B. H., Gelderloos-Arends, J., Jankipersadsing, S., Hofker, M., … Weersma, R. K. (2022). Environmental factors shaping the gut microbiome in a Dutch population. *Nature*, *604*(7907), Article 7907. https://doi.org/10.1038/s41586-022-04567-7

Gand, M., Navickaite, I., Bartsch, L.-J., Grützke, J., Overballe-Petersen, S., Rasmussen, A., Otani, S., Michelacci, V., Matamoros, B. R., González-Zorn, B., Brouwer, M. S. M., Di Marcantonio, L., Bloemen, B., Vanneste, K., Roosens, N. H. C. J., AbuOun, M., & De Keersmaecker, S. C. J. (2024). Towards facilitated interpretation of shotgun metagenomics long-read sequencing data analyzed with KMA for the detection of bacterial pathogens and their antimicrobial resistance genes. *Frontiers in Microbiology*, *15*. https://doi.org/10.3389/fmicb.2024.1336532

Garcia-Gonzalez, N., Comas, J. C., Harris, H. M. B., Strain, C., Stanton, C., Hill, C., Corsetti, A., & Gahan, C. G. M. (2022). Impact of Food Origin Lactiplantibacillus plantarum Strains on the Human Intestinal Microbiota in an in vitro System. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.832513

Gehrig, J. L., Portik, D. M., Driscoll, M. D., Jackson, E., Chakraborty, S., Gratalo, D., Ashby, M., & Valladares, R. (2022). Finding the right fit: Evaluation of short-read and long-read sequencing approaches to maximize the utility of clinical microbiome data. *Microbial Genomics*, *8*(3), 000794. https://doi.org/10.1099/mgen.0.000794

Ghosh, T. S., Shanahan, F., & O'Toole, P. W. (2022). Toward an improved definition of a healthy microbiome for healthy aging. *Nature Aging*, *2*(11), Article 11. https://doi.org/10.1038/s43587-022-00306-9

Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, *18*, 9–19. https://doi.org/10.1016/j.csbj.2019.11.002

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica*, *25*(2), 141–151. https://doi.org/10.11613/BM.2015.015

Gille, D., Schmid, A., Walther, B., & Vergères, G. (2018). Fermented Food and Non-Communicable Chronic Diseases: A Review. *Nutrients*, *10*(4), E448. https://doi.org/10.3390/nu10040448

Girdhar, K., Dogru, Y. D., Huang, Q., Yang, Y., Tolstikov, V., Raisingani, A., Chrudinova, M., Oh, J., Kelley, K., Ludvigsson, J. F., Kiebish, M. A., Palm, N. W., Ludvigsson, J., & Altindis, E. (2023). Dynamics of the gut microbiome, IgA response, and plasma metabolome in the development of pediatric celiac disease. *Microbiome*, *11*, 9. https://doi.org/10.1186/s40168-022-01429-2

GlobeNewswire. (2024, October 12). *Whole Genome Sequencing Market Analysis and Forecast 2024-2034: Cost of Sequencing a Human Genome has Decreased from about $3 Billion at the end of the Human Genome Project in 2003 to $600 in 2024*. GlobeNewswire News Room. https://www.globenewswire.com/news-release/2024/12/10/2994337/28124/en/Whole-Genome-Sequencing-Market-Analysis-and-Forecast-2024-2034-Cost-of-Sequencing-a-Human-Genome-has-Decreased-from-about-3-Billion-at-the-end-of-the-Human-Genome-Project-in-2003-t.html

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.02224

Golen, T. (2021, November 1). *What are postbiotics?* Harvard Health. https://www.health.harvard.edu/nutrition/what-are-postbiotics

Gomez, A., Petrzelkova, K. J., Burns, M. B., Yeoman, C. J., Amato, K. R., Vlckova, K., Modry, D., Todd, A., Jost Robinson, C. A., Remis, M. J., Torralba, M. G., Morton, E., Umaña, J. D., Carbonero, F., Gaskins, H. R., Nelson, K. E., Wilson, B. A., Stumpf, R. M., White, B. A., … Blekhman, R. (2016). Gut Microbiome of Coexisting BaAka Pygmies and Bantu Reflects Gradients of

Traditional Subsistence Patterns. *Cell Reports*, *14*(9), 2142–2153. https://doi.org/10.1016/j.celrep.2016.02.013

Gomez-Arango, L. F., Barrett, H. L., Wilkinson, S. A., Callaway, L. K., McIntyre, H. D., Morrison, M., & Dekker Nitert, M. (2018). Low dietary fiber intake increases Collinsella abundance in the gut microbiota of overweight and obese pregnant women. *Gut Microbes*, *9*(3), 189–201. https://doi.org/10.1080/19490976.2017.1406584

Gómez-Gallego, C., García-Mantrana, I., Martínez-Costa, C., Salminen, S., Isolauri, E., & Collado, M. C. (2019). The Microbiota and Malnutrition: Impact of Nutritional Status During Early Life. *Annual Review of Nutrition*, *39*(Volume 39, 2019), 267–290. https://doi.org/10.1146/annurev-nutr-082117-051716

González-Zancada, N., Redondo-Useros, N., Díaz, L. E., Gómez-Martínez, S., Marcos, A., & Nova, E. (2020). Association of Moderate Beer Consumption with the Gut Microbiota and SCFA of Healthy Adults. *Molecules*, *25*(20). https://doi.org/10.3390/molecules25204772

Gopalakrishnan, V., Spencer, C. N., Nezi, L., Reuben, A., Andrews, M. C., Karpinets, T. V., Prieto, P. A., Vicente, D., Hoffman, K., Wei, S. C., Cogdill, A. P., Zhao, L., Hudgens, C. W., Hutchinson, D. S., Manzo, T., Petaccia de Macedo, M., Cotechini, T., Kumar, T., Chen, W. S., … Wargo, J. A. (2018). Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science (New York, N.Y.)*, *359*(6371), 97–103. https://doi.org/10.1126/science.aan4236

Gorvitovskaia, A., Holmes, S. P., & Huse, S. M. (2016). Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle. *Microbiome*, *4*(1), 15. https://doi.org/10.1186/s40168-016-0160-7

Gray, G. E., Bekker, L.-G., Laher, F., Malahleha, M., Allen, M., Moodie, Z., Grunenberg, N., Huang, Y., Grove, D., Prigmore, B., Kee, J. J., Benkeser, D., Hural, J., Innes, C., Lazarus, E., Meintjes, G., Naicker, N., Kalonji, D., Nchabeleng, M., … Corey, L. (2021). Vaccine Efficacy of ALVAC-HIV and Bivalent Subtype C gp120–MF59 in Adults. *New England Journal of Medicine*, *384*(12), 1089–1100. https://doi.org/10.1056/NEJMoa2031499

Grice, E. A., & Segre, J. A. (2012). The human microbiome: Our second genome. *Annual Review of Genomics and Human Genetics*, *13*, 151–170. https://doi.org/10.1146/annurev-genom-090711-163814

Grüning, B., Dale, R., Sjödin, A., Chapman, B. A., Rowe, J., Tomkins-Tinch, C. H., Valieris, R., & Köster, J. (2018). Bioconda: Sustainable and comprehensive software distribution for the life sciences. *Nature Methods*, *15*(7), 475–476. https://doi.org/10.1038/s41592-018-0046-7

Gryaznova, M., Smirnova, Y., Burakova, I., Morozova, P., Lagutina, S., Chizhkov, P., Korneeva, O., & Syromyatnikov, M. (2024). Fecal Microbiota Characteristics in Constipation-Predominant and Mixed-Type Irritable Bowel Syndrome. *Microorganisms*, *12*(7), 1414. https://doi.org/10.3390/microorganisms12071414

Guohong-Liu, null, Qingxi-Zhao, null, & Hongyun-Wei, null. (2019). Characteristics of intestinal bacteria with fatty liver diseases and cirrhosis. *Annals of Hepatology*, *18*(6), 796–803. https://doi.org/10.1016/j.aohep.2019.06.020

Gupta, N., El-Gawaad, N. S. A., Mallasiy, L. O., Gupta, H., Yadav, V. K., Alghamdi, S., & Qusty, N. F. (2024). Microbial dysbiosis and the aging process: A review on the potential age-deceleration role of Lactiplantibacillus plantarum. *Frontiers in Microbiology*, *15*. https://doi.org/10.3389/fmicb.2024.1260793

Gupta, V. K., Kim, M., Bakshi, U., Cunningham, K. Y., Davis, J. M., Lazaridis, K. N., Nelson, H., Chia, N., & Sung, J. (2020). A predictive index for health status using species-level gut microbiome profiling. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-18476-8

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Han, N., Peng, X., Zhang, T., Qiang, Y., Li, X., & Zhang, W. (2024). Temporal dynamics and species-level complexity of Prevotella spp. in the human gut microbiota: Implications for enterotypes and health. *Frontiers in Microbiology*, *15*. https://doi.org/10.3389/fmicb.2024.1414000

Han, N., Zhang, T., Qiang, Y. *et al.* Time-scale analysis of the long-term variability of human gut microbiota characteristics in Chinese individuals. *Commun Biol* **5**, 1414 (2022). https://doi.org/10.1038/s42003-022-04359-9

Hansen, L. B. S., Roager, H. M., Søndertoft, N. B., Gøbel, R. J., Kristensen, M., Vallès-Colomer, M., Vieira-Silva, S., Ibrügger, S., Lind, M. V., Mærkedahl, R. B., Bahl, M. I., Madsen, M. L., Havelund, J., Falony, G., Tetens, I., Nielsen, T., Allin, K. H., Frandsen, H. L., Hartmann, B., …

Pedersen, O. (2018). A low-gluten diet induces changes in the intestinal microbiome of healthy Danish adults. *Nature Communications*, *9*(1), 4630. https://doi.org/10.1038/s41467-018-07019-x

Hansen, M. E. B., Rubel, M. A., Bailey, A. G., Ranciaro, A., Thompson, S. R., Campbell, M. C., Beggs, W., Dave, J. R., Mokone, G. G., Mpoloka, S. W., Nyambo, T., Abnet, C., Chanock, S. J., Bushman, F. D., & Tishkoff, S. A. (2019). Population structure of human gut bacteria in a diverse cohort from rural Tanzania and Botswana. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-018-1616-9

Haro, C., Rangel-Zúñiga, O. A., Alcalá-Díaz, J. F., Gómez-Delgado, F., Pérez-Martínez, P., Delgado-Lista, J., Quintana-Navarro, G. M., Landa, B. B., Navas-Cortés, J. A., Tena-Sempere, M., Clemente, J. C., López-Miranda, J., Pérez-Jiménez, F., & Camargo, A. (2016). Intestinal Microbiota Is Influenced by Gender and Body Mass Index. *PLOS ONE*, *11*(5), e0154090. https://doi.org/10.1371/journal.pone.0154090

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., Duda, S. N., & REDCap Consortium. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, *95*, 103208. https://doi.org/10.1016/j.jbi.2019.103208

Hawkins, K. G., Casolaro, C., Brown, J. A., Edwards, D. A., & Wikswo, J. P. (2020). The Microbiome and the Gut-Liver-Brain Axis for Central Nervous System Clinical Pharmacology: Challenges in Specifying and Integrating *In Vitro* and *In Silico* Models. *Clinical Pharmacology & Therapeutics*, *108*(5), Article 5. https://doi.org/10.1002/cpt.1870

He, Y., Wu, W., Zheng, H.-M., Li, P., McDonald, D., Sheng, H.-F., Chen, M.-X., Chen, Z.-H., Ji, G.-Y., Zheng, Z.-D.-X., Mujagond, P., Chen, X.-J., Rong, Z.-H., Chen, P., Lyu, L.-Y., Wang, X., Wu, C.-B., Yu, N., Xu, Y.-J., … Zhou, H.-W. (2018). Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nature Medicine*, *24*(10), 1532–1535. https://doi.org/10.1038/s41591-018-0164-x

Hehemann, J.-H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., & Michel, G. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature*, *464*(7290), 908–912. https://doi.org/10.1038/nature08937

Heinken, A., Basile, A., Hertel, J., Thinnes, C., & Thiele, I. (2021). Genome-Scale Metabolic Modeling of the Human Microbiome in the Era of Personalized Medicine. *Annual Review of Microbiology*, *75*, 199–222. https://doi.org/10.1146/annurev-micro-060221-012134

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

Helliwell, J., Layard, R., & Sachs, J. (2019). World Happiness Report 2019. *Happiness and Subjective Well-Being*. https://www.wellbeingintlstudiesrepository.org/hw_happiness/4

Henneke, L., Schlicht, K., Andreani, N. A., Hollstein, T., Demetrowitsch, T., Knappe, C., Hartmann, K., Jensen-Kroll, J., Rohmann, N., Pohlschneider, D., Geisler, C., Schulte, D. M., Settgast, U., Türk, K., Zimmermann, J., Kaleta, C., Baines, J. F., Shearer, J., Shah, S., … Laudes, M. (2022). A dietary carbohydrate—Gut Parasutterella—Human fatty acid biosynthesis metabolic axis in obesity and type 2 diabetes. *Gut Microbes*, *14*(1), 2057778. https://doi.org/10.1080/19490976.2022.2057778

Hiippala, K., Barreto, G., Burrello, C., Diaz-Basabe, A., Suutarinen, M., Kainulainen, V., Bowers, J. R., Lemmer, D., Engelthaler, D. M., Eklund, K. K., Facciotti, F., & Satokari, R. (2020). Novel Odoribacter splanchnicus Strain and Its Outer Membrane Vesicles Exert Immunoregulatory Effects in vitro. *Frontiers in Microbiology*, *11*, 575455. https://doi.org/10.3389/fmicb.2020.575455

Hildebrandt, M. A., Hoffmann, C., Sherrill–Mix, S. A., Keilbaugh, S. A., Hamady, M., Chen, Y.-Y., Knight, R., Ahima, R. S., Bushman, F., & Wu, G. D. (2009). High-Fat Diet Determines the Composition of the Murine Gut Microbiome Independently of Obesity. *Gastroenterology*, *137*(5), 1716-1724.e2. https://doi.org/10.1053/j.gastro.2009.08.042

Hippe, B., Zwielehner, J., Liszt, K., Lassl, C., Unger, F., & Haslberger, A. G. (2011). Quantification of butyryl CoA:acetate CoA-transferase genes reveals different butyrate production capacity in individuals according to diet and age. *FEMS Microbiology Letters*, *316*(2), 130–135. https://doi.org/10.1111/j.1574-6968.2010.02197.x

Ho, S. S., Urban, A. E., & Mills, R. E. (2020). Structural variation in the sequencing era. *Nature Reviews Genetics*, *21*(3), Article 3. https://doi.org/10.1038/s41576-019-0180-9

Holingue, C., Volk, H., Crocetti, D., Gottlieb, B., Spira, A. P., & Mostofsky, S. H. (2021). Links between parent-reported measures of poor sleep and executive function in childhood autism and

attention deficit hyperactivity disorder. *Sleep Health*, *7*(3), 375–383. https://doi.org/10.1016/j.sleh.2020.12.006
Hrncirova, L., Hudcovic, T., Sukova, E., Machova, V., Trckova, E., Krejsek, J., & Hrncir, T. (2019). Human gut microbes are susceptible to antimicrobial food additives in vitro. *Folia Microbiologica*, *64*(4), 497–508. https://doi.org/10.1007/s12223-018-00674-z
Hu, H.-J., Park, S.-G., Jang, H. B., Choi, M.-G., Park, K.-H., Kang, J. H., Park, S. I., Lee, H.-J., & Cho, S.-H. (2015). Obesity Alters the Microbial Community Profile in Korean Adolescents. *PLOS ONE*, *10*(7), e0134333. https://doi.org/10.1371/journal.pone.0134333
Huang, S., Li, M., Jin, D., & Xie, Y. (2023). *Indian Journal of Pharmaceutical Sciences*. https://www.ijpsonline.com/
Hughes, R., Alkan, Z., Keim, N., & Kable, M. (2017). *Impact of sequence variant detection and bacterial DNA extraction methods on the measurement of microbial community composition in human stool*. 212134. https://doi.org/10.1101/212134
Humphreys, C. (2020). 19—Intestinal Permeability. In J. E. Pizzorno & M. T. Murray (Eds.), *Textbook of Natural Medicine (Fifth Edition)* (pp. 166-177.e4). Churchill Livingstone. https://doi.org/10.1016/B978-0-323-43044-9.00019-4
Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*, 119. https://doi.org/10.1186/1471-2105-11-119
Inoue, T., Funatsu, Y., Ohnishi, M., Isogawa, M., Kawashima, K., Tanaka, M., Moriya, K., Kawaratani, H., Momoda, R., & Iio, E. (2022). Bile acid dysmetabolism in the gut-microbiota-liver axis under hepatitis C virus infection. *Liver International*, *42*(1), Article 1.
Isenring, J., Stevens, M. J. A., Jans, C., Lacroix, C., & Geirnaert, A. (2022). Identification of Valerate as Carrying Capacity Modulator by Analyzing Lactiplantibacillus plantarum Colonization of Colonic Microbiota in vitro. *Frontiers in Microbiology*, *13*, 910609. https://doi.org/10.3389/fmicb.2022.910609
Ishikawa, E., Matsuki, T., Kubota, H., Makino, H., Sakai, T., Oishi, K., Kushiro, A., Fujimoto, J., Watanabe, K., Watanuki, M., & Tanaka, R. (2013). Ethnic diversity of gut microbiota: Species characterization of *Bacteroides fragilis* group and genus *Bifidobacterium* in healthy Belgian adults, and comparison with data from Japanese subjects. *Journal of Bioscience and Bioengineering*, *116*(2), 265–270. https://doi.org/10.1016/j.jbiosc.2013.02.010
Jae Hyun Kim, Ji-Young Jeon, Yong-Jin Im, Na Ha, Jeon-Kyung Kim, Seol Ju Moon, & Min-Gul Kim. (2023). *Long-term taxonomic and functional stability of the gut microbiome from human fecal samples | Scientific Reports*. *13*(114). https://doi.org/10.1038/s41598-022-27033-w
Jaeggi, T., Kortman, G. A. M., Moretti, D., Chassard, C., Holding, P., Dostal, A., Boekhorst, J., Timmerman, H. M., Swinkels, D. W., Tjalsma, H., Njenga, J., Mwangi, A., Kvalsvig, J., Lacroix, C., & Zimmermann, M. B. (2015). Iron fortification adversely affects the gut microbiome, increases pathogen abundance and induces intestinal inflammation in Kenyan infants. *Gut*, *64*(5), 731–742. https://doi.org/10.1136/gutjnl-2014-307720
Jaffar, S., & Gill, G. (2017). The crisis of diabetes in sub-Saharan Africa. *The Lancet Diabetes & Endocrinology*, *5*(8), 574–575. https://doi.org/10.1016/S2213-8587(17)30219-X
Jang, L.-G., Choi, G., Kim, S.-W., Kim, B.-Y., Lee, S., & Park, H. (2019). The combination of sport and sport-specific diet is associated with characteristics of gut microbiota: An observational study. *Journal of the International Society of Sports Nutrition*, *16*(1), 21. https://doi.org/10.1186/s12970-019-0290-y
Jeffery, I. B., Lynch, D. B., & O'Toole, P. W. (2016). Composition and temporal stability of the gut microbiota in older persons. *The ISME Journal*, *10*(1), 170–182. https://doi.org/10.1038/ismej.2015.88
Jha, A. R., Shmalberg, J., Tanprasertsuk, J., Perry, L., Massey, D., & Honaker, R. W. (2020). Characterization of gut microbiomes of household pets in the United States using a direct-to-consumer approach. *PLOS ONE*, *15*(2), e0227289. https://doi.org/10.1371/journal.pone.0227289
Jiménez, E., Marín, M. L., Martín, R., Odriozola, J. M., Olivares, M., Xaus, J., Fernández, L., & Rodríguez, J. M. (2008). Is meconium from healthy newborns actually sterile? *Research in Microbiology*, *159*(3), 187–193. https://doi.org/10.1016/j.resmic.2007.12.007
Johnson, A. J., Vangay, P., Al-Ghalith, G. A., Hillmann, B. M., Ward, T. L., Shields-Cutler, R. R., Kim, A. D., Shmagel, A. K., Syed, A. N., Walter, J., Menon, R., Koecher, K., & Knights, D. (2019). Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. *Cell Host and Microbe*, *25*(6), 789-802.e5. https://doi.org/10.1016/j.chom.2019.05.005

Johnson, E. L., Heaver, S. L., Waters, J. L., Kim, B. I., Bretin, A., Goodman, A. L., Gewirtz, A. T., Worgall, T. S., & Ley, R. E. (2020). Sphingolipids produced by gut bacteria enter host metabolic pathways impacting ceramide levels. *Nature Communications*, *11*(1), 2471. https://doi.org/10.1038/s41467-020-16274-w

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

Johnson, L., Schopp, L., Waggie, F., & Frantz, J. (2021). Assessment of risk factors for Non-Communicable diseases among a cohort of community health workers in Western Cape, South Africa. *Malawi Medical Journal*, *33*(3), 196–203. https://doi.org/10.4314/mmj.v33i3.7

Jorgensen, J. A., Choo-Kang, C., Wang, L., Issa, L., Gilbert, J. A., Ecklu-Mensah, G., Luke, A., Bedu-Addo, K., Forrester, T., Bovet, P., Lambert, E. V., Rae, D., Argos, M., Kelly, T. N., Sargis, R. M., Dugas, L. R., Dai, Y., & Layden, B. T. (2024). Toxic Metals Impact Gut Microbiota and Metabolic Risk in Five African-Origin Populations. *medRxiv : the preprint server for health sciences*, 2024.10.07.24315016. https://doi.org/10.1101/2024.10.07.24315016

Kaakoush, N. O. (2020). Sutterella Species, IgA-degrading Bacteria in Ulcerative Colitis. *Trends in Microbiology*, *28*(7), 519–522. https://doi.org/10.1016/j.tim.2020.02.018

Kabwe, M. H., Vikram, S., Mulaudzi, K., Jansson, J. K., & Makhalanyane, T. P. (2020). The gut mycobiota of rural and urban individuals is shaped by geography. *BMC Microbiology*, *20*(1), 257. https://doi.org/10.1186/s12866-020-01907-3

Kalaimathi, K., Shine, K., Rajiv Gandhi, G., Vijayakumar, S., Ayyanar, M., Amalraj, S., Varatharaju, G., Thiruvengadam, M., Antony Ceasar, S., Prabhu, S., & Jose, J. (2023). Cyanobacterial metabolites as novel potential suppressors of breast cancer: A comparative *in silico* pharmacological assessment. *Intelligent Pharmacy*, *1*(3), 133–144. https://doi.org/10.1016/j.ipha.2023.06.002

Kalantar-Zadeh, K., Berean, K. J., Burgell, R. E., Muir, J. G., & Gibson, P. R. (2019). Intestinal gases: Influence on gut disorders and the role of dietary manipulations. *Nature Reviews. Gastroenterology & Hepatology*, *16*(12), 733–747. https://doi.org/10.1038/s41575-019-0193-z

Kamil, R. Z., Murdiati, A., Juffrie, M., Nakayama, J., & Rahayu, E. S. (2021). Gut Microbiota and Short-Chain Fatty Acid Profile between Normal and Moderate Malnutrition Children in Yogyakarta, Indonesia. *Microorganisms*, *9*(1), Article 1. https://doi.org/10.3390/microorganisms9010127

Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, *3*, e1165. https://doi.org/10.7717/peerj.1165

Kang, D.-W., Adams, J. B., Coleman, D. M., Pollard, E. L., Maldonado, J., McDonough-Means, S., Caporaso, J. G., & Krajmalnik-Brown, R. (2019). Long-term benefit of Microbiota Transfer Therapy on autism symptoms and gut microbiota. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-42183-0

Kargar, S., Roshan, M., Ghoreishi, S., Akhlaghi, A., Kanani, M., & Hosseini Ghaffari, M. (2020). Extended colostrum feeding for 2 weeks improves growth performance and reduces the susceptibility to diarrhea and pneumonia in neonatal Holstein dairy calves. *Journal of Dairy Science*, *103*. https://doi.org/10.3168/jds.2020-18355

Karim, Md. R., Iqbal, S., Mohammad, S., Lee, J. H., Jung, D., Mathiyalagan, R., Yang, D.-C., Yang, D. U., & Kang, S. C. (2023). *A review on* Impact of dietary interventions, drugs, and traditional herbal supplements on the gut microbiome. *Microbiological Research*, *271*, 127346. https://doi.org/10.1016/j.micres.2023.127346

Kassambara, A. (2023). *ggpubr: "ggplot2" Based Publication Ready Plots* (Version 0.6.0) [Computer software]. https://cran.r-project.org/web/packages/ggpubr/index.html

Kates, A. E., Jarrett, O., Skarlupka, J. H., Sethi, A., Duster, M., Watson, L., Suen, G., Poulsen, K., & Safdar, N. (2020). Household Pet Ownership and the Microbial Diversity of the Human Gut Microbiota. *Frontiers in Cellular and Infection Microbiology*, *10*, 73. https://doi.org/10.3389/fcimb.2020.00073

Kato, K., Odamaki, T., Mitsuyama, E., Sugahara, H., Xiao, J.-Z., & Osawa, R. (2017). Age-Related Changes in the Composition of Gut Bifidobacterium Species. *Current Microbiology*, *74*(8), 987–995. https://doi.org/10.1007/s00284-017-1272-4

Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Kennedy, D. O. (2016). B Vitamins and the Brain: Mechanisms, Dose and Efficacy--A Review. *Nutrients*, *8*(2), 68. https://doi.org/10.3390/nu8020068

Kers, J. G., & Saccenti, E. (2022). The Power of Microbiome Studies: Some Considerations on Which Alpha and Beta Metrics to Use and How to Report Results. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.796025

Khaledi, M., Poureslamfar, B., Alsaab, H. O., Tafaghodi, S., Hjazi, A., Singh, R., Alawadi, A. H., Alsaalamy, A., Qasim, Q. A., & Sameni, F. (2024). The role of gut microbiota in human metabolism and inflammatory diseases: A focus on elderly individuals. *Annals of Microbiology*, *74*(1), 1. https://doi.org/10.1186/s13213-023-01744-5

Khanna, S., Vazquez-Baeza, Y., González, A., Weiss, S., Schmidt, B., Muñiz-Pedrogo, D. A., Rainey, J. F., Kammer, P., Nelson, H., Sadowsky, M., Khoruts, A., Farrugia, S. L., Knight, R., Pardi, D. S., & Kashyap, P. C. (2017). Changes in microbial ecology after fecal microbiota transplantation for recurrent C. difficile infection affected by underlying inflammatory bowel disease. *Microbiome*, *5*(1), 55. https://doi.org/10.1186/s40168-017-0269-3

Kho, Z. Y., & Lal, S. K. (2018). The Human Gut Microbiome – A Potential Controller of Wellness and Disease. *Frontiers in Microbiology*, *9*. https://www.frontiersin.org/articles/10.3389/fmicb.2018.01835

Kieliszek, M., & Błażejak, S. (2016). Current Knowledge on the Importance of Selenium in Food for Living Organisms: A Review. *Molecules (Basel, Switzerland)*, *21*(5), 609. https://doi.org/10.3390/molecules21050609

Kim, C., Pongpanich, M., & Porntaveetus, T. (2024). Unraveling metagenomics through long-read sequencing: A comprehensive review. *Journal of Translational Medicine*, *22*(1), 111. https://doi.org/10.1186/s12967-024-04917-1

Kim, H.-Y. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics*, *42*(2), 152–155. https://doi.org/10.5395/rde.2017.42.2.152

Kim, J. E., & Kim, H. S. (2019). Microbiome of the Skin and Gut in Atopic Dermatitis (AD): Understanding the Pathophysiology and Finding Novel Management Strategies. *Journal of Clinical Medicine*, *8*(4), 444. https://doi.org/10.3390/jcm8040444

Kim, M.-S., Kim, Y., Choi, H., Kim, W., Park, S., Lee, D., Kim, D. K., Kim, H. J., Choi, H., Hyun, D.-W., Lee, J.-Y., Choi, E. Y., Lee, D.-S., Bae, J.-W., & Mook-Jung, I. (2020a). Transfer of a healthy microbiota reduces amyloid and tau pathology in an Alzheimer's disease animal model. *Gut*, *69*(2), 283–294. https://doi.org/10.1136/gutjnl-2018-317431

Kim, M.-S., Kim, Y., Choi, H., Kim, W., Park, S., Lee, D., Kim, D. K., Kim, H. J., Choi, H., Hyun, D.-W., Lee, J.-Y., Choi, E. Y., Lee, D.-S., Bae, J.-W., & Mook-Jung, I. (2020b). Transfer of a healthy microbiota reduces amyloid and tau pathology in an Alzheimer's disease animal model. *Gut*, *69*(2), 283–294. https://doi.org/10.1136/gutjnl-2018-317431

Kim, Y. S., Unno, T., Kim, B.-Y., & Park, M.-S. (2020). Sex Differences in Gut Microbiota. *The World Journal of Men's Health*, *38*(1), 48–60. https://doi.org/10.5534/wjmh.190009

Klatt, N. R., Broedlow, C., Osborn, J. M., Gustin, A. T., Dross, S., O'Connor, M. A., Coronado, E., Barnette, P., Hensley-McBain, T., Zevin, A. S., Fuller, D. H., & Manuzak, J. A. (2021). Effects of persistent modulation of intestinal microbiota on SIV/HIV vaccination in rhesus macaques. *Npj Vaccines*, *6*(1). https://doi.org/10.1038/s41541-021-00298-4

Knipe, M., & Krygsman, S. (2024). The opportunity cost of household transport expenditure in South Africa. *Journal of Transport and Supply Chain Management; Vol 18 (2024)DO - 10.4102/Jtscm.V18i0.1081*. https://jtscm.co.za/index.php/jtscm/article/view/1081/1857

Koay, W., Siems, L., & Persaud, D. (2018). The Microbiome and HIV Persistence: Implications for Viral Remission and Cure. *Current Opinion in HIV and AIDS*, *13*(1), 61–68. https://doi.org/10.1097/COH.0000000000000434

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 of the United States of America*, *108*(SUPPL. 1), 4578–4585. https://doi.org/10.1073/pnas.1000081107

Komodromou, I., Andreou, E., Vlahoyiannis, A., Christofidou, M., Felekkis, K., Pieri, M., & Giannaki, C. D. (2024). Exploring the Dynamic Relationship between the Gut Microbiome and Body Composition across the Human Lifespan: A Systematic Review. *Nutrients*, *16*(5), 660. https://doi.org/10.3390/nu16050660

Kontomina, E., Garefalaki, V., Fylaktakidou, K. C., Evmorfidou, D., Eleftheraki, A., Avramidou, M., Udoh, K., Panopoulou, M., Felföldi, T., Márialigeti, K., Fakis, G., & Boukouvala, S. (2022). A taxonomically representative strain collection to explore xenobiotic and secondary metabolism in bacteria. *PLoS ONE*, *17*(7), e0271125. https://doi.org/10.1371/journal.pone.0271125

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kootte, R. S., Levin, E., Salojärvi, J., Smits, L. P., Hartstra, A. V., Udayappan, S. D., Hermes, G., Bouter, K. E., Koopen, A. M., Holst, J. J., Knop, F. K., Blaak, E. E., Zhao, J., Smidt, H., Harms, A. C., Hankemeijer, T., Bergman, J. J. G. H. M., Romijn, H. A., Schaap, F. G., … Nieuwdorp, M. (2017). Improvement of Insulin Sensitivity after Lean Donor Feces in Metabolic Syndrome Is Driven by Baseline Intestinal Microbiota Composition. *Cell Metabolism*, *26*(4), 611-619.e6. https://doi.org/10.1016/j.cmet.2017.09.008

Koppel, N., Maini Rekdal, V., & Balskus, E. P. (2017). Chemical transformation of xenobiotics by the human gut microbiota. *Science (New York, N.Y.)*, *356*(6344), eaag2770. https://doi.org/10.1126/science.aag2770

Koren, O., Goodrich, J. K., Cullender, T. C., Spor, A., Laitinen, K., Kling Bäckhed, H., Gonzalez, A., Werner, J. J., Angenent, L. T., Knight, R., Bäckhed, F., Isolauri, E., Salminen, S., & Ley, R. E. (2012). Host Remodeling of the Gut Microbiome and Metabolic Changes during Pregnancy. *Cell*, *150*(3), 470–480. https://doi.org/10.1016/j.cell.2012.07.008

Kovatcheva-Datchary, P., Nilsson, A., Akrami, R., Lee, Y. S., De Vadder, F., Arora, T., Hallen, A., Martens, E., Björck, I., & Bäckhed, F. (2015). Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. *Cell metabolism*, *22*(6), 971-982. https://doi.org/10.1016/j.cmet.2015.10.001

Kozlov, A. M., Zhang, J., Yilmaz, P., Glöckner, F. O., & Stamatakis, A. (2016). Phylogeny-aware identification and correction of taxonomically mislabeled sequences. *Nucleic Acids Research*, *44*(11), 5022–5033. https://doi.org/10.1093/nar/gkw396

Kraft, F., & Kurth, I. (2019). Long-read sequencing in human genetics. *Medizinische Genetik*, *31*(2), 198–204. https://doi.org/10.1007/s11825-019-0249-z

Krautkramer, K. A., Fan, J., & Bäckhed, F. (2021). Gut microbial metabolites as multi-kingdom intermediates. *Nature Reviews Microbiology*, *19*(2), Article 2. https://doi.org/10.1038/s41579-020-0438-4

Krishnamoorthy, S., Coetzee, V., Kruger, J., Potgieter, H., & Buys, E. M. (2020). Dysbiosis Signatures of Fecal Microbiota in South African Infants with Respiratory, Gastrointestinal, and Other Diseases. *The Journal of Pediatrics*, *218*, 106-113.e3. https://doi.org/10.1016/j.jpeds.2019.11.029

Kruger, H. S., De Lange-Loots, Z., Kruger, I. M., & Pieters, M. (2022). The Metabolic Profiles of Metabolically Healthy Obese and Metabolically Unhealthy Obese South African Adults over 10 Years. *International Journal of Environmental Research and Public Health*, *19*(9), Article 9. https://doi.org/10.3390/ijerph19095061

Kuntz, T. M., & Gilbert, J. A. (2017). Introducing the Microbiome into Precision Medicine. *Trends in Pharmacological Sciences*, *38*(1), 81–91. https://doi.org/10.1016/j.tips.2016.10.001

Kushkevych, I., Dordević, D., & Vítězová, M. (2019). Toxicity of hydrogen sulfide toward sulfate-reducing bacteria Desulfovibrio piger Vib-7. *Archives of Microbiology*, *201*(3), 389–397. https://doi.org/10.1007/s00203-019-01625-z

Kwak, S., Usyk, M., Beggs, D., Choi, H., Ahdoot, D., Wu, F., Maceda, L., Li, H., Im, E.-O., Han, H.-R., Lee, E., Wu, A. H., Hayes, R. B., & Ahn, J. (2024). Sociobiome—Individual and neighborhood socioeconomic status influence the gut microbiome in a multi-ethnic population in the US. *Npj Biofilms and Microbiomes*, *10*(1), 1–10. https://doi.org/10.1038/s41522-024-00491-y

Ladeira, R., Tap, J., & Derrien, M. (2023). Exploring Bifidobacterium species community and functional variations with human gut microbiome structure and health beyond infancy. *Microbiome Research Reports*, *2*(2), 9. https://doi.org/10.20517/mrr.2023.01

Lagkouvardos, I., Pukall, R., Abt, B., Foesel, B. U., Meier-Kolthoff, J. P., Kumar, N., Bresciani, A., Martínez, I., Just, S., Ziegler, C., Brugiroux, S., Garzetti, D., Wenning, M., Bui, T. P. N., Wang, J., Hugenholtz, F., Plugge, C. M., Peterson, D. A., Hornef, M. W., … Clavel, T. (2016). The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nature Microbiology*, *1*(10), 1–15. https://doi.org/10.1038/nmicrobiol.2016.131

Lam, K. N., Alexander, M., & Turnbaugh, P. J. (2019). Precision Medicine Goes Microscopic: Engineering the Microbiome to Improve Drug Outcomes. *Cell Host & Microbe*, *26*(1), 22–34. https://doi.org/10.1016/j.chom.2019.06.011

Lampeter, T., Love, C., Tang, T. T., Marella, A. S., Lee, H. Y., Oganyan, A., Moffat, D., Kareem, A., Rusling, M., Massmann, A., Orr, M., Bongiorno, C., & Yuan, L.-L. (2023). Risk of bias

assessment tool for systematic review and meta-analysis of the gut microbiome. *Gut Microbiome*, *4*, e13. https://doi.org/10.1017/gmb.2023.12
Lang, D., Zhang, S., Ren, P., Liang, F., Sun, Z., Meng, G., Tan, Y., Li, X., Lai, Q., Han, L., Wang, D., Hu, F., Wang, W., & Liu, S. (2020). Comparison of the two up-to-date sequencing technologies for genome assembly: HiFi reads of Pacific Biosciences Sequel II system and ultralong reads of Oxford Nanopore. *GigaScience*, *9*(12), giaa123. https://doi.org/10.1093/gigascience/giaa123
Lang, S., & Schnabl, B. (2020). Microbiota and Fatty Liver Disease—The Known, the Unknown, and the Future. *Cell Host & Microbe*, *28*(2), 233–244. https://doi.org/10.1016/j.chom.2020.07.007
Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923
Lau, A. S.-Y., Yanagisawa, N., Hor, Y.-Y., Lew, L.-C., Ong, J.-S., Chuah, L.-O., Lee, Y.-Y., Choi, S.-B., Rashid, F., Wahid, N., Sugahara, H., Xiao, J.-Z., & Liong, M.-T. (2018). *Bifidobacterium longum BB536 alleviated upper respiratory illnesses and modulated gut microbiota profiles in Malaysian pre-school children*. https://doi.org/10.3920/BM2017.0063
Lau, R. I., Zhang, F., Liu, Q., Su, Q., Chan, F. K., & Ng, S. C. (2022). Gut microbiota in COVID-19: Key microbial changes, potential mechanisms and clinical applications. *Nature Reviews Gastroenterology & Hepatology*, 1–15.
Laudadio, I., Fulci, V., Palone, F., Stronati, L., Cucchiara, S., & Carissimi, C. (2018). Quantitative Assessment of Shotgun Metagenomics and 16S rDNA Amplicon Sequencing in the Study of Human Gut Microbiome. *OMICS: A Journal of Integrative Biology*, *22*(4), 248–254. https://doi.org/10.1089/omi.2018.0013
Laursen, M. F. (2021). Gut Microbiota Development: Influence of Diet from Infancy to Toddlerhood. *Annals of Nutrition and Metabolism*, *77*(Suppl. 3), 21–34. https://doi.org/10.1159/000517912
Lay, C., Rigottier-Gois, L., Holmstrøm, K., Rajilic, M., Vaughan, E. E., de Vos, W. M., Collins, M. D., Thiel, R., Namsolleck, P., Blaut, M., & Doré, J. (2005). Colonic microbiota signatures across five northern European countries. *Applied and Environmental Microbiology*, *71*(7), 4153–4155. https://doi.org/10.1128/AEM.71.7.4153-4155.2005
Lebba, V., Totino, V., Gagliardi, A., Santangelo, F., Cacciotti, F., Trancassini, M., Mancini, C., Cicerone, C., Corazziari, E., Pantanella, F., & Schippa, S. (2016). Eubiosis and dysbiosis: The two sides of the microbiota. *The New Microbiologica*, *39*(1), 1–12.
LeBlanc, K. E., Baer-Sinnott, S., Lancaster, K. J., Campos, H., Lau, K. H. K., Tucker, K. L., Kushi, L. H., & Willett, W. C. (2024). Perspective: Beyond the Mediterranean Diet—Exploring Latin American, Asian, and African Heritage Diets as Cultural Models of Healthy Eating. *Advances in Nutrition*, *15*(5), 100221. https://doi.org/10.1016/j.advnut.2024.100221
Lee, J.-E., Ha, J. S., Park, H.-Y., & Lee, E. (2020). Alteration of gut microbiota composition by short-term low-dose alcohol intake is restored by fermented rice liquor in mice. *Food Research International*, *128*, 108800. https://doi.org/10.1016/j.foodres.2019.108800
Lee, J.-Y., Na, Y.-A., Kim, E., Lee, H.-S., & Kim, P. (2016). *The Actinobacterium Corynebacterium glutamicum, an Industrial Workhorse*. *26*(5), 807–822. https://doi.org/10.4014/jmb.1601.01053
Lee, K. A., Luong, M. K., Shaw, H., Nathan, P., Bataille, V., & Spector, T. D. (2021). The gut microbiome: What the oncologist ought to know. *British Journal of Cancer*, *125*(9), Article 9. https://doi.org/10.1038/s41416-021-01467-x
Lee, M.-T., Le, H. H., & Johnson, E. L. (2021). Dietary sphinganine is selectively assimilated by members of the mammalian gut microbiome. *Journal of Lipid Research*, *62*, 100034. https://doi.org/10.1194/jlr.RA120000950
Lee, S. Y., Kim, J. H., Lee, D. Y., & Hur, S. J. (2023). Characterization of gut microbiota in mouse models of aging and sarcopenia. *Microbiological Research*, *275*, 127462. https://doi.org/10.1016/j.micres.2023.127462
Lee, S. Y., & Sung, J. H. (2018). Gut–liver on a chip toward an in vitro model of hepatic steatosis. *Biotechnology and Bioengineering*, *115*(11), Article 11.
Leggat, D. J., Cohen, K. W., Willis, J. R., Fulp, W. J., deCamp, A. C., Kalyuzhniy, O., Cottrell, C. A., Menis, S., Finak, G., Ballweber-Fleming, L., Srikanth, A., Plyler, J. R., Schiffner, T., Liguori, A., Rahaman, F., Lombardo, A., Philiponis, V., Whaley, R. E., Seese, A., … Schief, W. R. (2022). Vaccination induces HIV broadly neutralizing antibody precursors in humans. *Science*, *378*(6623), Article 6623. https://doi.org/10.1126/science.add6502
Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
Li, J., Fu, R., Yang, Y., Horz, H.-P., Guan, Y., Lu, Y., Lou, H., Tian, L., Zheng, S., Liu, H., Wang, S., & Xu, S. (2018). A metagenomic approach to dissect the genetic composition of enterotypes in

Han Chinese and two Muslim groups. *Systematic and Applied Microbiology*, *41*(1), 1–12. https://doi.org/10.1016/j.syapm.2017.09.006

Li, N., Chen, H., Cheng, Y., Xu, F., Ruan, G., Ying, S., Tang, W., Chen, L., Chen, M., Lv, L., Ping, Y., Chen, D., & Wei, Y. (2021). Fecal Microbiota Transplantation Relieves Gastrointestinal and Autism Symptoms by Improving the Gut Microbiota in an Open-Label Study. *Frontiers in Cellular and Infection Microbiology*, *11*, 759435. https://doi.org/10.3389/fcimb.2021.759435

Li, P., Ji, B., Luo, H., Sundh, D., Lorentzon, M., & Nielsen, J. (2022). One-year supplementation with Lactobacillus reuteri ATCC PTA 6475 counteracts a degradation of gut microbiota in older women with low bone mineral density. *Npj Biofilms and Microbiomes*, *8*(1), Article 1. https://doi.org/10.1038/s41522-022-00348-2

Li, S., He, Y., Liu, J., Chen, K., Yang, Y., Tao, K., Yang, J., Luo, K., & Ma, X. (2024). An umbrella review of socioeconomic status and cancer. *Nature Communications*, *15*(1), 9993. https://doi.org/10.1038/s41467-024-54444-2

Li, Y., Wang, D. D., Satija, A., Ivey, K. L., Li, J., Wilkinson, J. E., Li, R., Baden, M., Chan, A. T., Huttenhower, C., Rimm, E. B., Hu, F. B., & Sun, Q. (2021). Plant-Based Diet Index and Metabolic Risk in Men: Exploring the Role of the Gut Microbiome. *The Journal of Nutrition*, *151*(9), 2780–2789. https://doi.org/10.1093/jn/nxab175

Liang, Q., Zhang, M., Hu, Y., Zhang, W., Zhu, P., Chen, Y., Xue, P., Li, Q., & Wang, K. (2020). Gut Microbiome Contributes to Liver Fibrosis Impact on T Cell Receptor Immune Repertoire. *Frontiers in Microbiology*, *11*. https://www.frontiersin.org/articles/10.3389/fmicb.2020.571847

Lim, S., Moon, J. H., Shin, C. M., Jeong, D., & Kim, B. (2020). Effect of Lactobacillus sakei, a Probiotic Derived from Kimchi, on Body Fat in Koreans with Obesity: A Randomized Controlled Study. *Endocrinology and Metabolism*, *35*(2), 425–434. https://doi.org/10.3803/EnM.2020.35.2.425

Lima, A. R., Herrera, D. R., Francisco, P. A., Pereira, A. C., Lemos, J., Abranches, J., & Gomes, B. P. F. A. (2021). Detection of Streptococcus mutans in symptomatic and asymptomatic infected root canals. *Clinical Oral Investigations*, *25*(6), 3535–3542. https://doi.org/10.1007/s00784-020-03676-9

Lin, H., & Peddada, S. D. (2020). Analysis of compositions of microbiomes with bias correction. *Nature Communications*, *11*(1), 3514. https://doi.org/10.1038/s41467-020-17041-7

Lin, X., Waring, K., Ghezzi, H., Tropini, C., Tyson, J., & Ziels, R. M. (2024). High accuracy meets high throughput for near full-length 16S ribosomal RNA amplicon sequencing on the Nanopore platform. *PNAS Nexus*, *3*(10), pgae411. https://doi.org/10.1093/pnasnexus/pgae411

Liu, A., Garrett, S., Hong, W., & Zhang, J. (2024). Staphylococcus aureus Infections and Human Intestinal Microbiota. *Pathogens*, *13*(4), 276. https://doi.org/10.3390/pathogens13040276

Liu, Q., Mak, J. W. Y., Su, Q., Yeoh, Y. K., Lui, G. C.-Y., Ng, S. S. S., Zhang, F., Li, A. Y. L., Lu, W., Hui, D. S.-C., Chan, P. K., Chan, F. K. L., & Ng, S. C. (2021). Gut microbiota dynamics in a prospective cohort of patients with post-acute COVID-19 syndrome. *Gut*. https://doi.org/10.1136/gutjnl-2021-325989

Liu, W., Fang, X., Zhou, Y., Dou, L., & Dou, T. (2022). Machine learning-based investigation of the relationship between gut microbiome and obesity status. *Microbes and Infection*, *24*(2), 104892. https://doi.org/10.1016/j.micinf.2021.104892

Liu, X., Dai, M., Ma, Y., Zhao, N., Wang, Z., Yu, Y., Xu, Y., Zhang, H., Xiang, L., Tian, H., Shui, G., Zhang, F., & Wang, J. (2022). Reconstruction and Dynamics of the Human Intestinal Microbiome Observed *In Situ*. *Engineering*, *15*, 89–101. https://doi.org/10.1016/j.eng.2021.03.015

Liu, X., Li, X., Xia, B., Jin, X., Zou, Q., Zeng, Z., Zhao, W., Yan, S., Li, L., Yuan, S., Zhao, S., Dai, X., Yin, F., Cadenas, E., Liu, R. H., Zhao, B., Hou, M., Liu, Z., & Liu, X. (2021a). High-fiber diet mitigates maternal obesity-induced cognitive and social dysfunction in the offspring via gut-brain axis. *Cell Metabolism*, *33*(5), 923-938.e6. https://doi.org/10.1016/j.cmet.2021.02.002

Liu, X., Li, X., Xia, B., Jin, X., Zou, Q., Zeng, Z., Zhao, W., Yan, S., Li, L., Yuan, S., Zhao, S., Dai, X., Yin, F., Cadenas, E., Liu, R. H., Zhao, B., Hou, M., Liu, Z., & Liu, X. (2021b). High-fiber diet mitigates maternal obesity-induced cognitive and social dysfunction in the offspring via gut-brain axis. *Cell Metabolism*, *33*(5), 923-938.e6. https://doi.org/10.1016/j.cmet.2021.02.002

Liu, X., Liu, Y., Liu, J., Zhang, H., Shan, C., Guo, Y., Gong, X., Cui, M., Li, X., & Tang, M. (2024). Correlation between the gut microbiome and neurodegenerative diseases: A review of metagenomics evidence. *Neural Regeneration Research*, *19*(4), 833–845. https://doi.org/10.4103/1673-5374.382223

Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A. B., Brady, A., Creasy, H. H., McCracken, C., Giglio, M. G., McDonald, D., Franzosa, E. A., Knight, R., White, O., &

Huttenhower, C. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*, *550*(7674), 61–66. https://doi.org/10.1038/nature23889

Lobstein, C. T., & Brinsden, H. (2017). *We are grateful for the compilers of online databases and information on obesity especially the World Health Organization's Global Health Observatory, the NCD Risk Factor Collaboration based at Imperial College London and OECD. Please see the Data Sources section of this report in Appendix 1 for more details.* 242.

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

Lozupone, C. A., Li, M., Campbell, T. B., Flores, S. C., Linderman, D., Gebert, M. J., Knight, R., Fontenot, A. P., & Palmer, B. E. (2013). Alterations in the gut microbiota associated with HIV-1 infection. *Cell Host & Microbe*, *14*(3), 329–339. https://doi.org/10.1016/j.chom.2013.08.006

Lozupone, C. A., Stombaugh, J., Gonzalez, A., Ackermann, G., Wendel, D., Vázquez-Baeza, Y., Jansson, J. K., Gordon, J. I., & Knight, R. (2013). Meta-analyses of studies of the human microbiota. *Genome Research*, *23*(10), 1704–1714. https://doi.org/10.1101/gr.151803.112

Lozupone, C., & Knight, R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, *71*(12), 8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005

Luczynski, P., McVey Neufeld, K.-A., Oriach, C. S., Clarke, G., Dinan, T. G., & Cryan, J. F. (2016). Growing up in a Bubble: Using Germ-Free Animals to Assess the Influence of the Gut Microbiota on Brain and Behavior. *The International Journal of Neuropsychopharmacology*, *19*(8), pyw020. https://doi.org/10.1093/ijnp/pyw020

Luies, L., Reenen, M. van, Ronacher, K., Walzl, G., & Loots, D. T. (2017). Predicting tuberculosis treatment outcome using metabolomics. *Biomarkers in Medicine*, *11*(12), 1057–1067. https://doi.org/10.2217/bmm-2017-0133

Lundberg, J. O., Moretti, C., Benjamin, N., & Weitzberg, E. (2021). Symbiotic bacteria enhance exercise performance. *British Journal of Sports Medicine*, *55*(5), 243. https://doi.org/10.1136/bjsports-2020-102094

Lv, W., Liu, C., Ye, C., Sun, J., Tan, X., Zhang, C., Qu, Q., Shi, D., & Guo, S. (2017). Structural modulation of gut microbiota during alleviation of antibiotic-associated diarrhea with herbal formula. *International Journal of Biological Macromolecules*, *105*, 1622–1629. https://doi.org/10.1016/j.ijbiomac.2017.02.060

Lyu, M., Wang, Y., Fan, G., Wang, X., Xu, S., & Zhu, Y. (2017). Balancing Herbal Medicine and Functional Food for Prevention and Treatment of Cardiometabolic Diseases through Modulating Gut Microbiota. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.02146

Ma, J., Lyu, Y., Liu, X., Jia, X., Cui, F., Wu, X., Deng, S., & Yue, C. (2022). Engineered probiotics. *Microbial Cell Factories*, *21*(1), 72. https://doi.org/10.1186/s12934-022-01799-0

Ma, S., Wang, N., Zhang, P., Wu, W., & Fu, L. (2021). Fecal microbiota transplantation mitigates bone loss by improving gut microbiome composition and gut barrier function in aged rats. *PeerJ*, *9*, e12293. https://doi.org/10.7717/peerj.12293

Macrotrends. (2025). *South Africa Life Expectancy 1950-2025*. https://www.macrotrends.net/global-metrics/countries/ZAF/south-africa/life-expectancy

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

Mahdavinia, M., Rasmussen, H. E., Engen, P., Van den Berg, J. P., Davis, E., Engen, K., Green, S. J., Naqib, A., Botha, M., Gray, C., Lunjani, N., Hlela, C., Basera, W., Hobane, L., Watkins, A., Tobin, M. C., Landay, A., Keshavarzian, A., & Levin, M. E. (2017). Atopic dermatitis and food sensitization in South African toddlers: Role of fiber and gut microbiota. *Annals of Allergy, Asthma & Immunology: Official Publication of the American College of Allergy, Asthma, & Immunology*, *118*(6), 742-743.e3. https://doi.org/10.1016/j.anai.2017.04.011

Makki, K., Deehan, E. C., Walter, J., & Bäckhed, F. (2018). The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell host & microbe*, *23*(6), 705-715. https://doi.org/10.1016/j.chom.2018.05.012

Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., Tickle, T. L., Weingart, G., Ren, B., Schwager, E. H., Chatterjee, S., Thompson, K. N., Wilkinson, J. E., Subramanian, A., Lu, Y., Waldron, L., Paulson, J. N., Franzosa, E. A., Bravo, H. C., & Huttenhower, C. (2021).

Multivariable association discovery in population-scale meta-omics studies. *PLoS Computational Biology*, *17*(11), e1009442. https://doi.org/10.1371/journal.pcbi.1009442

Mandal, S., Godfrey, K. M., McDonald, D., Treuren, W. V., Bjørnholt, J. V., Midtvedt, T., Moen, B., Rudi, K., Knight, R., Brantsæter, A. L., Peddada, S. D., & Eggesbø, M. (2016). Fat and vitamin intakes during pregnancy have stronger relations with a pro-inflammatory maternal microbiota than does carbohydrate intake. *Microbiome*, *4*(1), 55. https://doi.org/10.1186/s40168-016-0200-3

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491

Manor, O., Dai, C. L., Kornilov, S. A., Smith, B., Price, N. D., Lovejoy, J. C., Gibbons, S. M., & Magis, A. T. (2020). Health and disease markers correlate with gut microbiome composition across thousands of people. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-18871-1

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), Article 1. https://doi.org/10.14806/ej.17.1.200

McDonald, D., Hyde, E., Debelius, J. W., Morton, J. T., Gonzalez, A., Ackermann, G., Aksenov, A. A., Behsaz, B., Brennan, C., Chen, Y., DeRight Goldasich, L., Dorrestein, P. C., Dunn, R. R., Fahimipour, A. K., Gaffney, J., Gilbert, J. A., Gogul, G., Green, J. L., Hugenholtz, P., … Knight, R. (2018). American Gut: An Open Platform for Citizen Science Microbiome Research. *mSystems*, *3*(3), e00031-18. https://doi.org/10.1128/mSystems.00031-18

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

McGrattan, A., van Aller ,Carla, Narytnyk ,Alla, Reidpath ,Daniel, Keage ,Hannah, Mohan ,Devi, Su ,Tin Tin, Stephan ,Blossom, Robinson ,Louise, & and Siervo, M. (2022). Nutritional interventions for the prevention of cognitive impairment and dementia in developing economies in East-Asia: A systematic review and meta-analysis. *Critical Reviews in Food Science and Nutrition*, *62*(7), 1838–1855. https://doi.org/10.1080/10408398.2020.1848785

Mehta, R.S., Abu-Ali, G.S., Drew, D.A. *et al.* Stability of the human faecal microbiome in a cohort of adult men. *Nat Microbiol* **3**, 347–355 (2018). https://doi.org/10.1038/s41564-017-0096-0

Melli, L. C. F. L., do Carmo-Rodrigues, M. S., Araújo-Filho, H. B., Solé, D., & de Morais, M. B. (2016). Intestinal microbiota and allergic diseases: A systematic review. *Allergologia Et Immunopathologia*, *44*(2), 177–188. https://doi.org/10.1016/j.aller.2015.01.013

Mena-Vázquez, N., Godoy-Navarrete, F. J., Lisbona-Montañez, J. M., Redondo-Rodriguez, R., Manrique-Arija, S., Rioja, J., Mucientes, A., Ruiz-Limón, P., Garcia-Studer, A., Ortiz-Márquez, F., Oliver-Martos, B., Cano-García, L., & Fernández-Nebro, A. (2023). Inflammatory Biomarkers in the Diagnosis and Prognosis of Rheumatoid Arthritis-Associated Interstitial Lung Disease. *International Journal of Molecular Sciences*, *24*(7), 6800. https://doi.org/10.3390/ijms24076800

Meng, J., Tao, J., Abu, Y., Sussman, D. A., Girotra, M., Franceschi, D., & Roy, S. (2022). HIV-Positive Patients on Antiretroviral Therapy Have an Altered Mucosal Intestinal but Not Oral Microbiome. *Microbiology Spectrum*, e02472-22. https://doi.org/10.1128/spectrum.02472-22

Messer, J. S., & Chang, E. B. (2018). Chapter 36—Microbial Physiology of the Digestive Tract and Its Role in Inflammatory Bowel Diseases. In H. M. Said (Ed.), *Physiology of the Gastrointestinal Tract (Sixth Edition)* (pp. 795–810). Academic Press. https://doi.org/10.1016/B978-0-12-809954-4.00036-0

Millington, A. (2019). *South Africa has just been ranked the unhealthiest country on earth: Indigo Wellbeing Index*. BusinessInsider. https://www.businessinsider.co.za/most-unhealthy-countries-in-the-world-ranked-2019-3

Mills, S., Lane, J. A., Smith, G. J., Grimaldi, K. A., Ross, R. P., & Stanton, C. (2019). Precision Nutrition and the Microbiome Part II: Potential Opportunities and Pathways to Commercialisation. *Nutrients*, *11*(7), E1468. https://doi.org/10.3390/nu11071468

Mills, S., Yang, B., Smith, G. J., Stanton, C., & Ross, R. P. (2023). Efficacy of Bifidobacterium longum alone or in multi-strain probiotic formulations during early life and beyond. *Gut Microbes*, *15*(1), 2186098. https://doi.org/10.1080/19490976.2023.2186098

Milosavljevic, M. N., Kostic, M., Milovanovic, J., Zaric, R. Z., Stojadinovic, M., Jankovic, S. M., & Stefanovic, S. M. (2021). Antimicrobial treatment of Erysipelatoclostridium ramosum invasive

infections: A systematic review. *Revista Do Instituto De Medicina Tropical De Sao Paulo*, *63*, e30. https://doi.org/10.1590/S1678-9946202163030

Mirzayi, C., Renson, A., Genomic Standards Consortium, Massive Analysis and Quality Control Society, Furlanello, C., Sansone, S.-A., Zohra, F., Elsafoury, S., Geistlinger, L., Kasselman, L. J., Eckenrode, K., van de Wijgert, J., Loughman, A., Marques, F. Z., MacIntyre, D. A., Arumugam, M., Azhar, R., Beghini, F., Bergstrom, K., … Waldron, L. (2021). Reporting guidelines for human microbiome research: The STORMS checklist. *Nature Medicine*. https://doi.org/10.1038/s41591-021-01552-x

Mobeen, F., Sharma, V., & Prakash, T. (2018). Enterotype Variations of the Healthy Human Gut Microbiome in Different Geographical Regions. *Bioinformation*, *14*(9), 560–573. https://doi.org/10.6026/97320630014560

Montalban-Arques, A., Katkeviciute, E., Busenhart, P., Bircher, A., Wirbel, J., Zeller, G., Morsy, Y., Borsig, L., Glaus Garzon, J. F., Müller, A., Arnold, I. C., Artola-Boran, M., Krauthammer, M., Sintsova, A., Zamboni, N., Leventhal, G. E., Berchtold, L., de Wouters, T., Rogler, G., … Scharl, M. (2021). Commensal *Clostridiales* strains mediate effective anti-cancer immune response against solid tumors. *Cell Host & Microbe*, *29*(10), 1573-1588.e7. https://doi.org/10.1016/j.chom.2021.08.001

Morita, E., Yokoyama, H., Imai, D., Takeda, R., Ota, A., Kawai, E., Hisada, T., Emoto, M., Suzuki, Y., & Okazaki, K. (2019). Aerobic Exercise Training with Brisk Walking Increases Intestinal Bacteroides in Healthy Elderly Women. *Nutrients*, *11*(4), 868. https://doi.org/10.3390/nu11040868

Mosca, A., Abreu Y Abreu, A. T., Gwee, K. A., Ianiro, G., Tack, J., Nguyen, T. V. H., & Hill, C. (2022). The clinical evidence for postbiotics as microbial therapeutics. *Gut Microbes*, *14*(1), 2117508. https://doi.org/10.1080/19490976.2022.2117508

Moya, A., & Ferrer, M. (2016). Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. *Trends in Microbiology*, *24*(5), 402–413. https://doi.org/10.1016/j.tim.2016.02.002

Mu, C., Yang, Y., Luo, Z., & Zhu, W. (2017). Temporal microbiota changes of high-protein diet intake in a rat model. *Anaerobe*, *47*, 218–225. https://doi.org/10.1016/j.anaerobe.2017.06.003

Mukaka, M. M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal : The Journal of Medical Association of Malawi*, *24*(3), 69. https://pmc.ncbi.nlm.nih.gov/articles/PMC3576830/

Mukhopadhya, I., Hansen, R., Nicholl, C. E., Alhaidan, Y. A., Thomson, J. M., Berry, S. H., Pattinson, C., Stead, D. A., Russell, R. K., El-Omar, E. M., & Hold, G. L. (2011). A comprehensive evaluation of colonic mucosal isolates of Sutterella wadsworthensis from inflammatory bowel disease. *PloS One*, *6*(10), e27076. https://doi.org/10.1371/journal.pone.0027076

Muller, E., Shiryan, I., & Borenstein, E. (2024). Multi-omic integration of microbiome data for identifying disease-associated modules. *Nature Communications*, *15*(1), 2621. https://doi.org/10.1038/s41467-024-46888-3

Munoz, S., Guzman-Rodriguez, M., Sun, J., Zhang, Y.-G., Noordhof, C., He, S.-M., Allen-Vercoe, E., Claud, E. C., & Petrof, E. O. (2016). Rebooting the microbiome. *Gut Microbes*, *7*(4), 353–363. https://doi.org/10.1080/19490976.2016.1188248

Musso, G., Gambino, R., & Cassader, M. (2010). Obesity, diabetes, and gut microbiota: The hygiene hypothesis expanded? *Diabetes Care*, *33*(10), 2277–2284. https://doi.org/10.2337/dc10-0556

Mutyambizi, C., Pavlova, M., Hongoro, C., & Groot, W. (2020). Inequalities and factors associated with adherence to diabetes self-care practices amongst patients at two public hospitals in Gauteng, South Africa. *BMC Endocrine Disorders*, *20*(1), 15. https://doi.org/10.1186/s12902-020-0492-y

Myer, P. R., McDaneld, T. G., Kuehn, L. A., Dedonder, K. D., Apley, M. D., Capik, S. F., Lubbers, B. V., Harhay, G. P., Harhay, D. M., Keele, J. W., Henniger, M. T., Clemmons, B. A., & Smith, T. P. L. (2020). Classification of 16S rRNA reads is improved using a niche-specific database constructed by near-full length sequencing. *PloS One*, *15*(7), e0235498. https://doi.org/10.1371/journal.pone.0235498

Nagao-Kitamoto, H., Leslie, J. L., Kitamoto, S., Jin, C., Thomsson, K. A., Gillilland, M. G., Kuffa, P., Goto, Y., Jenq, R. R., Ishii, C., Hirayama, A., Seekatz, A. M., Martens, E. C., Eaton, K. A., Kao, J. Y., Fukuda, S., Higgins, P. D. R., Karlsson, N. G., Young, V. B., & Kamada, N. (2020). Interleukin-22-mediated host glycosylation prevents Clostridioides difficile infection by modulating the metabolic activity of the gut microbiota. *Nature Medicine*, *26*(4), 608–617. https://doi.org/10.1038/s41591-020-0764-0

Nagata, N., Takeuchi, T., Masuoka, H., Aoki, R., Ishikane, M., Iwamoto, N., Sugiyama, M., Suda, W., Nakanishi, Y., Terada-Hirashima, J., Kimura, M., Nishijima, T., Inooka, H., Miyoshi-Akiyama, T., Kojima, Y., Shimokawa, C., Hisaeda, H., Zhang, F., Yeoh, Y. K., … Ohno, H. (2022). Human

Gut Microbiota and Its Metabolites Impact Immune Responses in COVID-19 and Its Complications. *Gastroenterology*, *164*(2), 272. https://doi.org/10.1053/j.gastro.2022.09.024

Navarre, D. A., Zhu, M., & Hellmann, H. (2022). Plant Antioxidants Affect Human and Gut Health, and Their Biosynthesis Is Influenced by Environment and Reactive Oxygen Species. *Oxygen*, *2*(3), Article 3. https://doi.org/10.3390/oxygen2030025

Nayfach, S., Rodriguez-Mueller, B., Garud, N., & Pollard, K. S. (2016). An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Research*, *26*(11), 1612–1625. https://doi.org/10.1101/gr.201863.115

NCDRisC. (2016). Trends in adult body-mass index in 200 countries from 1975 to 2014: A pooled analysis of 1698 population-based measurement studies with 19·2 million participants. *The Lancet*, *387*(10026), 1377–1396. https://doi.org/10.1016/S0140-6736(16)30054-X

Ndongo, S., Cadoret, F., Dubourg, G., Delerce, J., Fournier, P.-E., Raoult, D., & Lagier, J.-C. (2017). '*Collinsella phocaeensis*' sp. Nov., '*Clostridium merdae*' sp. Nov., '*Sutterella massiliensis*' sp. Nov., '*Sutturella timonensis*' sp. Nov., '*Enorma phocaeensis*' sp. Nov., '*Mailhella massiliensis*' gen. Nov., sp. Nov., '*Mordavella massiliensis*' gen. Nov., sp. Nov. And '*Massiliprevotella massiliensis*' gen. Nov., sp. Nov., 9 new species isolated from fresh stool samples of healthy French patients. *New Microbes and New Infections*, *17*, 89–95. https://doi.org/10.1016/j.nmni.2017.02.005

Nearing, J. T., Douglas, G. M., Hayes, M. G., MacDonald, J., Desai, D. K., Allward, N., Jones, C. M. A., Wright, R. J., Dhanani, A. S., Comeau, A. M., & Langille, M. G. I. (2022). Microbiome differential abundance methods produce different results across 38 datasets. *Nature Communications*, *13*(1), 342. https://doi.org/10.1038/s41467-022-28034-z

Nguyen, S., Tran, H., Long, J., Shrubsole, M., Cai, H., Yang, Y., Nguyen, L., Nguyen, G., Nguyen, C., Ta, T., Wu, J., Cai, Q., Zheng, W., Tran, T., & Shu, X. (2024). Gut Microbiome of Patients With Breast Cancer in Vietnam. *JCO Global Oncology*, *10*, e2300234. https://doi.org/10.1200/GO.23.00234

Nichols, E. (2023). *FDA 21 CFR Part 11—7 Tips to Ensure Compliance*. https://www.greenlight.guru/blog/tips-to-comply-with-fda-21-cfr-part-11

Nieuwoudt, S., Manderson, L., & Norris, S. A. (2018). Infant feeding practices in Soweto, South Africa: Implications for healthcare providers. *South African Medical Journal*, *108*(9), 756–762. https://doi.org/10.7196/SAMJ.2018.v108i9.13358

Nilson, R., Penumutchu, S., Pagano, F. S., & Belenky, P. (2024). Metabolic changes associated with polysaccharide utilization reduce susceptibility to some β-lactams in Bacteroides thetaiotaomicron. *mSphere*, *9*(8), e00103-24. https://doi.org/10.1128/msphere.00103-24

Noble, E. E., Hsu, T. M., & Kanoski, S. E. (2017). Gut to Brain Dysbiosis: Mechanisms Linking Western Diet Consumption, the Microbiome, and Cognitive Impairment. *Frontiers in Behavioral Neuroscience*, *11*. https://www.frontiersin.org/articles/10.3389/fnbeh.2017.00009

Ocvirk, S., Wilson, A. S., Posma, J. M., Li, J. V., Koller, K. R., Day, G. M., Flanagan, C. A., Otto, J. E., Sacco, P. E., Sacco, F. D., Sapp, F. R., Wilson, A. S., Newton, K., Brouard, F., DeLany, J. P., Behnning, M., Appolonia, C. N., Soni, D., Bhatti, F., … O'Keefe, S. J. D. (2020). A prospective cohort analysis of gut microbial co-metabolism in Alaska Native and rural African people at high and low risk of colorectal cancer. *The American Journal of Clinical Nutrition*, *111*(2), 406–419. https://doi.org/10.1093/ajcn/nqz301

Odamaki, T., Kato, K., Sugahara, H., Hashikura, N., Takahashi, S., Xiao, J., Abe, F., & Osawa, R. (2016). Age-related changes in gut microbiota composition from newborn to centenarian: A cross-sectional study. *BMC Microbiology*, *16*(1), 90. https://doi.org/10.1186/s12866-016-0708-5

Oduaran, O. H., Tamburini, F. B., Gómez-Olivé, F. X., Norris, S. A., Tollman, S. M., Wade, A. N., Wagner, R. G., & Bhatt, A. S. (2020). *Gut Microbiome Profiling of a Rural and Urban South African population Reveals Biomarkers of a Population in Lifestyle Transition*. https://doi.org/10.1101/2020.02.27.964023

Ogata, Y., Suda, W., Ikeyama, N., Hattori, M., Ohkuma, M., & Sakamoto, M. (2019). Complete Genome Sequence of Phascolarctobacterium faecium JCM 30894, a Succinate-Utilizing Bacterium Isolated from Human Feces. *Microbiology Resource Announcements*, *8*(3), e01487-18. https://doi.org/10.1128/MRA.01487-18

Oh, B. S., Choi, W. J., Kim, J.-S., Ryu, S. W., Yu, S. Y., Lee, J.-S., Park, S.-H., Kang, S. W., Lee, J., Jung, W. Y., Kim, Y.-M., Jeong, J.-H., & Lee, J. H. (2021). Cell-Free Supernatant of Odoribacter splanchnicus Isolated From Human Feces Exhibits Anti-colorectal Cancer Activity. *Frontiers in Microbiology*, *12*, 736343. https://doi.org/10.3389/fmicb.2021.736343

O'Hara, A., & Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Reports*, *7*, 688–693. https://doi.org/10.1038/sj.embor.7400731

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., … Weedon, J. (2022). *vegan: Community Ecology Package* (Version 2.6-4) [Computer software]. https://cran.r-project.org/web/packages/vegan/index.html

Oldenburg, C. E., Sié, A., Coulibaly, B., Ouermi, L., Dah, C., Tapsoba, C., Bärnighausen, T., Lebas, E., Arzika, A. M., Cummings, S., Zhong, L., Lietman, T. M., Keenan, J. D., & Doan, T. (2019). Indirect Effect of Azithromycin Use on the Intestinal Microbiome Diversity of Untreated Children: A Randomized Trial. *Open Forum Infectious Diseases*, *6*(3), ofz061. https://doi.org/10.1093/ofid/ofz061

Oldenburg, C. E., Sié, A., Coulibaly, B., Ouermi, L., Dah, C., Tapsoba, C., Bärnighausen, T., Ray, K. J., Zhong, L., Cummings, S., Lebas, E., Lietman, T. M., Keenan, J. D., & Doan, T. (2018). Effect of Commonly Used Pediatric Antibiotics on Gut Microbial Diversity in Preschool Children in Burkina Faso: A Randomized Clinical Trial. *Open Forum Infectious Diseases*, *5*(11), ofy289. https://doi.org/10.1093/ofid/ofy289

Olsson, L. M., Boulund, F., Nilsson, S., Khan, M. T., Gummesson, A., Fagerberg, L., Engstrand, L., Perkins, R., Uhlén, M., Bergström, G., Tremaroli, V., & Bäckhed, F. (2022). Dynamics of the normal gut microbiota: A longitudinal one-year population study in Sweden. *Cell Host & Microbe*, *30*(5), 726-739.e3. https://doi.org/10.1016/j.chom.2022.03.002

Oñate, F. P., Chamignon, C., Burz, S. D., Lapaque, N., Monnoye, M., Philippe, C., Bredel, M., Chêne, L., Farin, W., Paillarse, J.-M., Boursier, J., Ratziu, V., Mousset, P.-Y., Doré, J., Gérard, P., & Blottière, H. M. (2023). Adlercreutzia equolifaciens Is an Anti-Inflammatory Commensal Bacterium with Decreased Abundance in Gut Microbiota of Patients with Metabolic Liver Disease. *International Journal of Molecular Sciences*, *24*(15), 12232. https://doi.org/10.3390/ijms241512232

ONT. (2024). *Welcome to Oxford Nanopore Technologies*. Oxford Nanopore Technologies. https://nanoporetech.com/

Orelle, C., Mathieu, K., & Jault, J. M. (2019). Multidrug ABC transporters in bacteria. *Research in microbiology*, *170*(8), 381-391. https://doi.org/10.1016/j.resmic.2019.06.001

Ou, J., Carbonero, F., Zoetendal, E. G., DeLany, J. P., Wang, M., Newton, K., Gaskins, H. R., & O'Keefe, S. J. (2013). Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans1234. *The American Journal of Clinical Nutrition*, *98*(1), 111–120. https://doi.org/10.3945/ajcn.112.056689

Ozato, N., Saito, S., Yamaguchi, T., Katashima, M., Tokuda, I., Sawada, K., Katsuragi, Y., Kakuta, M., Imoto, S., Ihara, K., & Nakaji, S. (2019). Blautia genus associated with visceral fat accumulation in adults 20–76 years of age. *Npj Biofilms and Microbiomes*, *5*(1), 28. https://doi.org/10.1038/s41522-019-0101-x

PacBio. (2024). *PacBio Revio | Long-read sequencing at scale*. https://www.pacb.com/revio/

Pacific Bioscience. (2023). *Preparing multiplexed amplicon libraries using PacBio barcoded M13 primers and SMRTbell® prep kit 3.0*.

Paek, J., Shin, Y., Kim, J.-S., Kim, H., Kook, J.-K., Paek, W. K., & Chang, Y.-H. (2017). Description of *Absiella argi* gen. Nov., sp. Nov., and transfer of *Eubacterium dolichum* and *Eubacterium tortuosum* to the genus *Absiella as Absiella dolichum* comb. Nov. And *Absiella tortuosum* comb. Nov. *Anaerobe*, *48*, 70–75. https://doi.org/10.1016/j.anaerobe.2017.07.006

Palmas, V., Pisanu, S., Madau, V., Casula, E., Deledda, A., Cusano, R., Uva, P., Vascellari, S., Loviselli, A., Manzin, A., & Velluzzi, F. (2021). Gut microbiota markers associated with obesity and overweight in Italian adults. *Scientific Reports*, *11*(1), 5532. https://doi.org/10.1038/s41598-021-84928-w

Pan, S., Zhu, C., Zhao, X.-M., & Coelho, L. P. (2022). A deep siamese neural network improves metagenome-assembled genomes in microbiome datasets across different environments. *Nature Communications*, *13*(1), 2326. https://doi.org/10.1038/s41467-022-29843-y

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M. C., Rice, B. L., DuLong, C., Morgan, X. C., Golden, C. D., Quince, C., Huttenhower, C., & Segata, N. (2019). Extensive Unexplored Human Microbiome Diversity

Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*, *176*(3), 649. https://doi.org/10.1016/j.cell.2019.01.001
Patangia, D. V., Anthony Ryan, C., Dempsey, E., Paul Ross, R., & Stanton, C. (2022). Impact of antibiotics on the human microbiome and consequences for host health. *MicrobiologyOpen*, *11*(1), e1260. https://doi.org/10.1002/mbo3.1260
Pei, T., Zhu, D., Yang, S., Hu, R., Wang, F., Zhang, J., Yan, S., Ju, L., He, Z., Han, Z., He, J., Yan, Y., Wang, M., Xiao, W., & Ma, Y. (2022). Bacteroides plebeius improves muscle wasting in chronic kidney disease by modulating the gut-renal muscle axis. *Journal of Cellular and Molecular Medicine*, *26*(24), 6066–6078. https://doi.org/10.1111/jcmm.17626
Peng, C., Gathagan, R. J., Covell, D. J., Medellin, C., Stieber, A., Robinson, J. L., Zhang, B., Pitkin, R. M., Olufemi, M. F., Luk, K. C., Trojanowski, J. Q., & Lee, V. M.-Y. (2018). Cellular milieu imparts distinct pathological α-synuclein strains in α-synucleinopathies. *Nature*, *557*(7706), 558–563. https://doi.org/10.1038/s41586-018-0104-4
Pereira, L. (2014). *The Future of South Africa's Food System: What is research telling us?*
Pérez Martínez, G., Bäuerl, C., & Collado, M. C. (2014). Understanding gut microbiota in elderly's health will enable intervention through probiotics. *Beneficial Microbes*, *5*, 1–12. https://doi.org/10.3920/BM2013.0079
Peters, B. A., Yi, S. S., Beasley, J. M., Cobbs, E. N., Choi, H. S., Beggs, D. B., Hayes, R. B., & Ahn, J. (2020). US nativity and dietary acculturation impact the gut microbiome in a diverse US population. *The ISME Journal*, *14*(7), 1639–1650. https://doi.org/10.1038/s41396-020-0630-6
Peterson, D., Bonham, K. S., Rowland, S., Pattanayak, C. W., Consortium, R., & Klepac-Ceraj, V. (2021). Comparative Analysis of 16S rRNA Gene and Metagenome Sequencing in Pediatric Gut Microbiomes. *Frontiers in Microbiology*, *12*, 670336. https://doi.org/10.3389/fmicb.2021.670336
Petrov, V. A., Saltykova, I. V., Zhukova, I. A., Alifirova, V. M., Zhukova, N. G., Dorofeeva, Y. B., Tyakht, A. V., Kovarsky, B. A., Alekseev, D. G., Kostryukova, E. S., Mironova, Y. S., Izhboldina, O. P., Nikitina, M. A., Perevozchikova, T. V., Fait, E. A., Babenko, V. V., Vakhitova, M. T., Govorun, V. M., & Sazonov, A. E. (2017). Analysis of Gut Microbiota in Patients with Parkinson's Disease. *Bulletin of Experimental Biology and Medicine*, *162*(6), 734–737. https://doi.org/10.1007/s10517-017-3700-7
Philips, C. A., Pande, A., Shasthry, S. M., Jamwal, K. D., Khillan, V., Chandel, S. S., Kumar, G., Sharma, M. K., Maiwall, R., Jindal, A., Choudhary, A., Hussain, M. S., Sharma, S., & Sarin, S. K. (2017). Healthy Donor Fecal Microbiota Transplantation in Steroid-Ineligible Severe Alcoholic Hepatitis: A Pilot Study. *Clinical Gastroenterology and Hepatology*, *15*(4), 600–602. https://doi.org/10.1016/j.cgh.2016.10.029
Philips, C. A., Phadke, N., Ganesan, K., Ranade, S., & Augustine, P. (2018). Corticosteroids, nutrition, pentoxifylline, or fecal microbiota transplantation for severe alcoholic hepatitis. *Indian Journal of Gastroenterology*, *37*(3), 215–225. https://doi.org/10.1007/s12664-018-0859-4
Phuong-Nguyen, K., McGee, S. L., Aston-Mourney, K., Mcneill, B. A., Mahmood, M. Q., & Rivera, L. R. (2024). Yoyo Dieting, Post-Obesity Weight Loss, and Their Relationship with Gut Health. *Nutrients*, *16*(18). https://doi.org/10.3390/nu16183170
Piao, H., Hawley, E., Kopf, S., Descenzo, R., Sealock, S., Henick-Kling, T., & Hess, M. (2015). Insights into the bacterial community and its temporal succession during the fermentation of wine grapes. *Frontiers in Microbiology*, *6*, 809. https://doi.org/10.3389/fmicb.2015.00809
Pimentel, M., Mathur, R., & Chang, C. (2013). Gas and the Microbiome. *Current Gastroenterology Reports*, *15*(12), 356. https://doi.org/10.1007/s11894-013-0356-y
Pinart, M., Dötsch, A., Schlicht, K., Laudes, M., Bouwman, J., Forslund, S. K., Pischon, T., & Nimptsch, K. (2021). Gut Microbiome Composition in Obese and Non-Obese Persons: A Systematic Review and Meta-Analysis. *Nutrients*, *14*(1), 12. https://doi.org/10.3390/nu14010012
Piscotta, F. J., Hoffmann, H.-H., Choi, Y. J., Small, G. I., Ashbrook, A. W., Koirala, B., Campbell, E. A., Darst, S. A., Rice, C. M., & Brady, S. F. (2021). Metabolites with SARS-CoV-2 Inhibitory Activity Identified from Human Microbiome Commensals. *mSphere*. https://doi.org/10.1128/mSphere.00711-21
Portela, N. D., Galván, C., Sanmarco, L. M., Bergero, G., Aoki, M. P., Cano, R. C., & Pesoa, S. A. (2023). Omega-3-Supplemented Fat Diet Drives Immune Metabolic Response in Visceral Adipose Tissue by Modulating Gut Microbiota in a Mouse Model of Obesity. *Nutrients*, *15*(6), 1404. https://doi.org/10.3390/nu15061404
Portincasa, P., Khalil, M., Graziani, A., Frühbeck, G., Baffy, G., Garruti, G., Di Ciaula, A., & Bonfrate, L. (2024). Gut microbes in metabolic disturbances. Promising role for therapeutic manipulations? *European Journal of Internal Medicine*, *119*, 13–30. https://doi.org/10.1016/j.ejim.2023.10.002

Posit team. (2024). *RStudio: Integrated Development Environment for R* (Version 2023.12.1.402) [Computer software]. Posit Software, PBC, Boston, MA. https://www.posit.co/

Prasanna, P. H. P., Grandison, A. S., & Charalampopoulos, D. (2014). Bifidobacteria in milk products: An overview of physiological and biochemical properties, exopolysaccharide production, selection criteria of milk products and health benefits. *Food Research International*, *55*, 247–262. https://doi.org/10.1016/j.foodres.2013.11.013

Price, C. E., & O'Toole, G. A. (2021). The Gut-Lung Axis in Cystic Fibrosis. *Journal of Bacteriology*, *203*(20), Article 20. https://doi.org/10.1128/JB.00311-21

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*, *5*(3), e9490. https://doi.org/10.1371/journal.pone.0009490

Priya, A., Dutta, K., & Daverey, A. (2022). A comprehensive biotechnological and molecular insight into plastic degradation by microbial community. *Journal of Chemical Technology & Biotechnology*, *97*(2), 381–390. https://doi.org/10.1002/jctb.6675

Pu, Y., Sun, Z., Zhang, H., Huang, Q., Wang, Z., Mei, Z., Wang, P., Kong, M., Yang, W., Lin, C., Zhou, X., Lin, S., Huang, Q., Huang, L., Sun, L., Yuan, C., Xu, Q., Tang, H., Wang, X., & Zheng, Y. (2024). Gut microbial features and circulating metabolomic signatures of frailty in older adults. *Nature Aging*, *4*(9), 1249–1262. https://doi.org/10.1038/s43587-024-00678-0

Pugh, S. J., Richardson, G. A., Hutcheon, J. A., Himes, K. P., Brooks, M. M., Day, N. L., & Bodnar, L. M. (2015). Maternal Obesity and Excessive Gestational Weight Gain Are Associated with Components of Child Cognition123. *The Journal of Nutrition*, *145*(11), 2562–2569. https://doi.org/10.3945/jn.115.215525

Qiao, S., Liu, C., Sun, L., Wang, T., Dai, H., Wang, K., Bao, L., Li, H., Wang, W., Liu, S.-J., & Liu, H. (2022). Gut Parabacteroides merdae protects against cardiovascular damage by enhancing branched-chain amino acid catabolism. *Nature Metabolism*, *4*(10), 1271–1286. https://doi.org/10.1038/s42255-022-00649-y

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., Peng, Y., Zhang, D., Jie, Z., Wu, W., Qin, Y., Xue, W., Li, J., Han, L., Lu, D., … Wang, J. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, *490*(7418), 55–60. https://doi.org/10.1038/nature11450

Qu, C., Lu, J., Chen, Y., Li, J., Xu, X., & Li, F. (2024). Unravelling the role of gut microbiota in acute pancreatitis: Integrating Mendelian randomization with a nested case-control study. *Frontiers in Microbiology*, *15*, 1401056. https://doi.org/10.3389/fmicb.2024.1401056

QuantStack. (2024). *QuantStack*. https://quantstack.net/

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

Racanelli, V., & Rehermann, B. (2006). The liver as an immunological organ. *Hepatology*, *43*(S1), Article S1.

Rachek, S., Nikpoor, N., Gómez del Pulgar, E. M., Gonzaga, A., Sanz, Y., & Tompkins, T. A. (2021). Complete Genome Sequence of Phascolarctobacterium faecium G 104, Isolated from the Stools of a Healthy Lean Donor. *Microbiology Resource Announcements*, *10*(4), e01054-20. https://doi.org/10.1128/MRA.01054-20

Ramaboli, M. C., Ocvirk, S., Khan Mirzaei, M., Eberhart, B. L., Valdivia-Garcia, M., Metwaly, A., Neuhaus, K., Barker, G., Ru, J., Nesengani, L. T., Mahdi-Joest, D., Wilson, A. S., Joni, S. K., Layman, D. C., Zheng, J., Mandal, R., Chen, Q., Perez, M. R., Fortuin, S., … O'Keefe, S. J. D. (2024). Diet changes due to urbanization in South Africa are linked to microbiome and metabolome signatures of Westernization and colorectal cancer. *Nature Communications*, *15*(1), 3379. https://doi.org/10.1038/s41467-024-46265-0

Rausch, P., Rühlemann, M., Hermes, B. M., Doms, S., Dagan, T., Dierking, K., Domin, H., Fraune, S., von Frieling, J., Hentschel, U., Heinsen, F.-A., Höppner, M., Jahn, M. T., Jaspers, C., Kissoyan, K. A. B., Langfeldt, D., Rehman, A., Reusch, T. B. H., Roeder, T., … Baines, J. F. (2019). Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome*, *7*(1), 133. https://doi.org/10.1186/s40168-019-0743-1

Ravikrishnan, A., Wijaya, I., Png, E., Chng, K. R., Ho, E. X. P., Ng, A. H. Q., Mohamed Naim, A. N., Gounot, J.-S., Guan, S. P., Hanqing, J. L., Guan, L., Li, C., Koh, J. Y., de Sessions, P. F., Koh, W.-P., Feng, L., Ng, T. P., Larbi, A., Maier, A. B., … Nagarajan, N. (2024). Gut metagenomes of Asian octogenarians reveal metabolic potential expansion and distinct microbial species associated with aging phenotypes. *Nature Communications*, *15*(1), 7751. https://doi.org/10.1038/s41467-024-52097-9

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E., Benenson, M., Gurunathan, S., Tartaglia, J., McNeil, J. G., Francis, D. P., Stablein, D., Birx, D. L., Chunsuttiwat, S., Khamboonruang, C., … Kim, J. H. (2009). Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand. *New England Journal of Medicine*, *361*(23), 2209–2220. https://doi.org/10.1056/NEJMoa0908492

Reveles, K. R., Ryan, C. N., Chan, L., Cosimi, R. A., & Haynes, W. L. (2018). Proton pump inhibitor use associated with changes in gut microbiota composition. *Gut*, *67*(7), 1369–1370. https://doi.org/10.1136/gutjnl-2017-315306

Reynoso-García, J., Miranda-Santiago, A. E., Meléndez-Vázquez, N. M., Acosta-Pagán, K., Sánchez-Rosado, M., Díaz-Rivera, J., Rosado-Quiñones, A. M., Acevedo-Márquez, L., Cruz-Roldán, L., Tosado-Rodríguez, E. L., Figueroa-Gispert, M. D. M., & Godoy-Vitorino, F. (2022). A complete guide to human microbiomes: Body niches, transmission, development, dysbiosis, and restoration. *Frontiers in Systems Biology*, *2*. https://www.frontiersin.org/articles/10.3389/fsysb.2022.951403

Rideout, J. R., Chase, J. H., Bolyen, E., Ackermann, G., González, A., Knight, R., & Caporaso, J. G. (2016). Keemei: Cloud-based validation of tabular bioinformatics file formats in Google Sheets. *GigaScience*, *5*, 27. https://doi.org/10.1186/s13742-016-0133-6

Ridlon, J. M., Kang, D.-J., Hylemon, P. B., & Bajaj, J. S. (2015). Gut microbiota, cirrhosis and alcohol regulate bile acid metabolism in the gut. *Digestive Diseases (Basel, Switzerland)*, *33*(3), 338–345. https://doi.org/10.1159/000371678

Rio, P., Gasbarrini, A., Gambassi, G., & Cianci, R. (2024). Pollutants, microbiota and immune system: Frenemies within the gut. *Frontiers in Public Health*, *12*, 1285186. https://doi.org/10.3389/fpubh.2024.1285186

Roager, H. M., & Licht, T. R. (2018). Microbial tryptophan catabolites in health and disease. *Nature Communications*, *9*(1), 3294. https://doi.org/10.1038/s41467-018-05470-4

Roberts, R. J., Carneiro, M. O., & Schatz, M. C. (2013). The advantages of SMRT sequencing. *Genome Biology*, *14*(6). https://doi.org/10.1186/gb-2013-14-6-405

Rohrhofer, J., Zwirzitz, B., Selberherr, E., & Untersmayr, E. (2021). The Impact of Dietary Sphingolipids on Intestinal Microbiota and Gastrointestinal Immune Homeostasis. *Frontiers in Immunology*, *12*, 635704. https://doi.org/10.3389/fimmu.2021.635704

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

Ruiz, L., Delgado, S., Ruas-Madiedo, P., Sánchez, B., & Margolles, A. (2017). Bifidobacteria and Their Molecular Communication with the Immune System. *Frontiers in Microbiology*, *8*, 2345. https://doi.org/10.3389/fmicb.2017.02345

Russell, W. R., Gratz, S. W., Duncan, S. H., Holtrop, G., Ince, J., Scobbie, L., Duncan, G., Johnstone, A. M., Lobley, G. E., Wallace, R. J., Duthie, G. G., & Flint, H. J. (2011). High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health. *The American Journal of Clinical Nutrition*, *93*(5), 1062–1072. https://doi.org/10.3945/ajcn.110.002188

SADAG. (2019). https://www.sadag.org/

Saiki, H., Okano, Y., Yasuma, T., Toda, M., Takeshita, A., Abdel-Hamid, A. M., Fridman D'Alessandro, V., Tsuruga, T., D'Alessandro-Gabazza, C. N., Katayama, K., Sugimoto, M., Fujimoto, H., Yamanaka, K., Kobayashi, T., Cann, I., & Gabazza, E. C. (2021). A Microbiome-Derived Peptide Induces Apoptosis of Cells from Different Tissues. *Cells*, *10*(11). https://doi.org/10.3390/cells10112885

Sakuma, K., Hayashi, S., Yasaka, Y., Nishijima, H., Funabashi, H., Hayashi, M., Matsuoka, H., & Saito, M. (2013). Analysis of odor compounds in feces of mice that were exposed to various stresses during breeding. *Experimental Animals*, *62*(2), 101–107. https://doi.org/10.1538/expanim.62.101

Salazar-Jaramillo Laura, de la Cuesta-Zuluaga Jacobo, Chica Luis A., Cadavid María, Ley Ruth E., Reyes Alejandro, & Escobar Juan S. (2024). Gut microbiome diversity within Clostridia is negatively associated with human obesity. *mSystems*, *9*(8), e00627-24. https://doi.org/10.1128/msystems.00627-24

Sanders, M. E., Benson, A., Lebeer, S., Merenstein, D. J., & Klaenhammer, T. R. (2018). Shared mechanisms among probiotic taxa: Implications for general probiotic claims. *Current Opinion in Biotechnology*, *49*, 207–216. https://doi.org/10.1016/j.copbio.2017.09.007

Santos-Marcos, J. A., Mora-Ortiz, M., Tena-Sempere, M., Lopez-Miranda, J., & Camargo, A. (2023). Interaction between gut microbiota and sex hormones and their relation to sexual dimorphism in metabolic diseases. *Biology of Sex Differences*, *14*(1), 4. https://doi.org/10.1186/s13293-023-00490-2

Sarker, G., & Peleg-Raibstein, D. (2018). Maternal Overnutrition Induces Long-Term Cognitive Deficits across Several Generations. *Nutrients*, *11*(1), E7. https://doi.org/10.3390/nu11010007

Saseendran, A., Setia, L., Chhabria, V., Chakraborty, D., & Barman Roy, A. (2019). *Impact of Noise in Dataset on Machine Learning Algorithms*. https://doi.org/10.13140/RG.2.2.25669.91369

Sayin, S. I., Wahlström, A., Felin, J., Jäntti, S., Marschall, H. U., Bamberg, K., Angelin, B., Hyötyläinen, T., Orešič, M., & Bäckhed, F. (2013). Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. *Cell metabolism*, *17*(2), 225–235. https://doi.org/10.1016/j.cmet.2013.01.003

Scheiman, J., Luber, J. M., Chavkin, T. A., MacDonald, T., Tung, A., Pham, L.-D., Wibowo, M. C., Wurth, R. C., Punthambaker, S., Tierney, B. T., Yang, Z., Hattab, M. W., Avila-Pacheco, J., Clish, C. B., Lessard, S., Church, G. M., & Kostic, A. D. (2019). Meta-omics analysis of elite athletes identifies a performance-enhancing microbe that functions via lactate metabolism. *Nature Medicine*, *25*(7), 1104–1109. https://doi.org/10.1038/s41591-019-0485-4

Scheperjans, F., Aho, V., Pereira, P. A. B., Koskinen, K., Paulin, L., Pekkonen, E., Haapaniemi, E., Kaakkola, S., Eerola-Rautio, J., Pohja, M., Kinnunen, E., Murros, K., & Auvinen, P. (2015). Gut microbiota are related to Parkinson's disease and clinical phenotype. *Movement Disorders: Official Journal of the Movement Disorder Society*, *30*(3), 350–358. https://doi.org/10.1002/mds.26069

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Horn, D. J. V., & Weber, C. F. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. https://doi.org/10.1128/AEM.01541-09

Schneider, K. M., Mohs, A., Gui, W., Galvez, E. J., Candels, L. S., Hoenicke, L., Muthukumarasamy, U., Holland, C. H., Elfers, C., & Kilic, K. (2022). Imbalanced gut microbiota fuels hepatocellular carcinoma development by shaping the hepatic inflammatory microenvironment. *Nature Communications*, *13*(1), Article 1.

Schnorr, S. L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., Turroni, S., Biagi, E., Peano, C., Severgnini, M., Fiori, J., Gotti, R., De Bellis, G., Luiselli, D., Brigidi, P., Mabulla, A., Marlowe, F., Henry, A. G., & Crittenden, A. N. (2014). Gut microbiome of the Hadza hunter-gatherers. *Nature Communications*, *5*(1), 3654. https://doi.org/10.1038/ncomms4654

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia and Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., Robbertse, B., Sharma, S., Soussov, V., Sullivan, J. P., Sun, L., Turner, S., & Karsch-Mizrachi, I. (2020). NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database: The Journal of Biological Databases and Curation*, *2020*, baaa062. https://doi.org/10.1093/database/baaa062

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

Sczyrba, A., Hofmann, P., Belmann, P., Koslicki, D., Janssen, S., Dröge, J., Gregor, I., Majda, S., Fiedler, J., Dahms, E., Bremges, A., Fritz, A., Garrido-Oter, R., Jørgensen, T. S., Shapiro, N., Blood, P. D., Gurevich, A., Bai, Y., Turaev, D., … McHardy, A. C. (2017). Critical Assessment of Metagenome Interpretation—A benchmark of metagenomics software. *Nature Methods*, *14*(11), 1063–1071. https://doi.org/10.1038/nmeth.4458

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, *12*(6), R60. https://doi.org/10.1186/gb-2011-12-6-r60

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., & Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, *9*(8), Article 8. https://doi.org/10.1038/nmeth.2066

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

Seong, H. J., Baek, Y., Lee, S., & Jin, H.-J. (2024). Gut microbiome and metabolic pathways linked to sleep quality. *Frontiers in Microbiology*, *15*. https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2024.1418773

Sepich-Poore, G. D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J. A., & Knight, R. (2021). The microbiome and human cancer. *Science*, *371*(6536), Article 6536. https://doi.org/10.1126/science.abc4552

Severino, A., Tohumcu, E., Tamai, L., Dargenio, P., Porcari, S., Rondinella, D., Venturini, I., Maida, M., Gasbarrini, A., Cammarota, G., & Ianiro, G. (2024). The microbiome-driven impact of western diet in the development of noncommunicable chronic disorders. *Best Practice & Research Clinical Gastroenterology*, *72*, 101923. https://doi.org/10.1016/j.bpg.2024.101923

Shalon, D., Culver, R. N., Grembi, J. A., Folz, J., Treit, P. V., Shi, H., Rosenberger, F. A., Dethlefsen, L., Meng, X., Yaffe, E., Aranda-Díaz, A., Geyer, P. E., Mueller-Reif, J. B., Spencer, S., Patterson, A. D., Triadafilopoulos, G., Holmes, S. P., Mann, M., Fiehn, O., … Huang, K. C. (2023). Profiling the human intestinal environment under physiological conditions. *Nature*, *617*(7961), Article 7961. https://doi.org/10.1038/s41586-023-05989-7

Shanahan, F., Ghosh, T. S., & O'Toole, P. W. (2021). The Healthy Microbiome—What Is the Definition of a Healthy Gut Microbiome? *Gastroenterology*, *160*(2), 483–494. https://doi.org/10.1053/j.gastro.2020.09.057

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sherwin, E., Dinan, T. G., & Cryan, J. F. (2018). Recent developments in understanding the role of the gut microbiota in brain health and disease. *Annals of the New York Academy of Sciences*, *1420*(1), Article 1. https://doi.org/10.1111/nyas.13416

Shi, H., ter Horst, R., Nielen, S., Bloemendaal, M., Jaeger, M., Joosten, I., Koenen, H., Joosten, L. A. B., Schweren, L. J. S., Vasquez, A. A., Netea, M. G., & Buitelaar, J. (2022). The gut microbiome as mediator between diet and its impact on immune function. *Scientific Reports*, *12*(1), 5149. https://doi.org/10.1038/s41598-022-08544-y

Shinde, Y., & Deokar, G. (2024). Regulation of Gut Microbiota by Herbal Medicines. *Current Drug Metabolism*, *25*(2), 110–127. https://doi.org/10.2174/0113892002287336240328083220

Shivani, S., Chattopadhyay, A., & Chuang, E. Y. (2021). Targeting the gut microbiome for non-communicable diseases: Present and future. *Annals of Translational Medicine*, *9*(5), 361. https://doi.org/10.21037/atm-21-144

Shuai, M., Zuo, L.-S.-Y., Miao, Z., Gou, W., Xu, F., Jiang, Z., Ling, C.-W., Fu, Y., Xiong, F., Chen, Y.-M., & Zheng, J.-S. (2021). Multi-omics analyses reveal relationships among dairy consumption, gut microbiota and cardiometabolic health. *EBioMedicine*, *66*, 103284. https://doi.org/10.1016/j.ebiom.2021.103284

Siddiqui, R., Makhlouf, Z., Alharbi, A. M., Alfahemi, H., & Khan, N. A. (2022). The Gut Microbiome and Female Health. *Biology*, *11*(11), Article 11. https://doi.org/10.3390/biology11111683

Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., & Banfield, J. F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, *3*(7), 836–843. https://doi.org/10.1038/s41564-018-0171-1

Simon, A., Colom, J., Mazhar, S., Khokhlova, E., Deaton, J., & Rea, K. (2023). Bacillus megaterium Renuspore® as a potential probiotic for gut health and detoxification of unwanted dietary contaminants. *Frontiers in Microbiology*, *14*. https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1125616

Simpson, E. H. (1949). Measurement of Diversity. *Nature*, *163*(4148), 688–688. https://doi.org/10.1038/163688a0

Singer, M., Koedooder, R., Bos, M., Poort, L., Schoenmakers, S., Savelkoul, P., Laven, J., de Jonge, J., Morré, S., & Budding, A. (2021). The profiling of microbiota in vaginal swab samples using 16S rRNA gene sequencing and IS-pro analysis. *BMC Microbiology*, *21*. https://doi.org/10.1186/s12866-021-02149-7

Smith, L., Grabovac, I., Veronese, N., Soysal, P., Isik, A. T., Stubbs, B., Yang, L., & Jackson, S. E. (2019). Sleep Quality, Duration, and Associated Sexual Function at Older Age: Findings from the English Longitudinal Study of Ageing. *The Journal of Sexual Medicine*, *16*(3), 427–433. https://doi.org/10.1016/j.jsxm.2019.01.005

Smits, S. A., Leach, J., Sonnenburg, E. D., Gonzalez, C. G., Lichtman, J. S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J. E., Dominguez-Bello, M. G., & Sonnenburg, J. L. (2017). Seasonal Cycling in the Gut Microbiome of the Hadza Hunter-Gatherers of Tanzania. *Science (New York, N.Y.)*, *357*(6353), 802–806. https://doi.org/10.1126/science.aan4834

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L. G., Gratadoux, J.-J., Blugeon, S., Bridonneau, C., Furet, J.-P., Corthier, G., Grangette, C., Vasquez, N., Pochart, P., Trugnan, G., Thomas, G., Blottière, H. M., Doré, J., Marteau, P., Seksik, P., & Langella, P. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(43), 16731–16736. https://doi.org/10.1073/pnas.0804812105

Sonnenburg, E. D., Smits, S. A., Tikhonov, M., Higginbottom, S. K., Wingreen, N. S., & Sonnenburg, J. L. (2016). Diet-induced extinctions in the gut microbiota compound over generations. *Nature*, *529*(7585), 212–215. https://doi.org/10.1038/nature16504

Sonnenburg, J. L., & Bäckhed, F. (2016). Diet–microbiota interactions as moderators of human metabolism. *Nature*, *535*(7610), 56–64. https://doi.org/10.1038/nature18846

Stanislawski, M. A., Dabelea, D., Lange, L. A., Wagner, B. D., & Lozupone, C. A. (2019). Gut microbiota phenotypes of obesity. *Npj Biofilms and Microbiomes*, *5*(1), 1–9. https://doi.org/10.1038/s41522-019-0091-8

Statista. (2021). *Leading causes of death in Africa 2019 | Statista*. https://www.statista.com/statistics/1029337/top-causes-of-death-africa/

Statistics South Africa. (2019). *General Household Survey 2018*.

Statistics South Africa. (2023, October 17). *Rising Non-Communicable Diseases: A Looming Health Crisis | Statistics South Africa*. https://www.statssa.gov.za/?p=16729

Stommel, M., & Schoenborn, C. A. (2009). Accuracy and usefulness of BMI measures based on self-reported weight and height: Findings from the NHANES & NHIS 2001-2006. *BMC Public Health*, *9*, 421. https://doi.org/10.1186/1471-2458-9-421

Sugino, K. Y., Paneth, N., & Comstock, S. S. (2019). Michigan cohorts to determine associations of maternal pre-pregnancy body mass index with pregnancy and infant gastrointestinal microbial communities: Late pregnancy and early infancy. *PLOS ONE*, *14*(3), e0213733. https://doi.org/10.1371/journal.pone.0213733

Sun, M.-F., Zhu, Y.-L., Zhou, Z.-L., Jia, X.-B., Xu, Y.-D., Yang, Q., Cui, C., & Shen, Y.-Q. (2018). Neuroprotective effects of fecal microbiota transplantation on MPTP-induced Parkinson's disease mice: Gut microbiota, glial reaction and TLR4/TNF-α signaling pathway. *Brain, Behavior, and Immunity*, *70*, 48–60. https://doi.org/10.1016/j.bbi.2018.02.005

Sun, S., Jones, R. B., & Fodor, A. A. (2020). Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories. *Microbiome*, *8*(1), 46. https://doi.org/10.1186/s40168-020-00815-y

Sun, Y., Sommerville, N. R., Liu, J. Y. H., Ngan, M. P., Poon, D., Ponomarev, E. D., Lu, Z., Kung, J. S. C., & Rudd, J. A. (2020). Intra-gastrointestinal amyloid-β1–42 oligomers perturb enteric function and induce Alzheimer's disease pathology. *The Journal of Physiology*, *598*(19), Article 19. https://doi.org/10.1113/JP279919

Sun, Y., Ye, F., Huang, Q., Du, F., Song, T., Yuan, H., Liu, X., & Yao, D. (2023). Linking ecological niches to bacterial community structure and assembly in polluted urban aquatic ecosystems. *Frontiers in Microbiology*, *14*, 1288304. https://doi.org/10.3389/fmicb.2023.1288304

Suppli, M. P., Bagger, J. I., Lelouvier, B., Broha, A., Demant, M., Kønig, M. J., Strandberg, C., Lund, A., Vilsbøll, T., & Knop, F. K. (2021). Hepatic microbiome in healthy lean and obese humans. *JHEP Reports*, *3*(4), 100299. https://doi.org/10.1016/j.jhepr.2021.100299

Tamanai-Shacoori, Z., Smida, I., Bousarghin, L., Loreal, O., Meuric, V., Fong, S. B., Bonnaure-Mallet, M., & Jolivet-Gougeon, A. (2017). Roseburia spp.: A marker of health? *Future Microbiology*, *12*, 157–170. https://doi.org/10.2217/fmb-2016-0130

Tamburini, F. B., Maghini, D., Oduaran, O. H., Brewster, R., Hulley, M. R., Sahibdeen, V., Norris, S. A., Tollman, S., Kahn, K., Wagner, R. G., Wade, A. N., Wafawanaka, F., Gómez-Olivé, F. X., Twine, R., Lombard, Z., Hazelhurst, S., & Bhatt, A. S. (2022). Short- and long-read metagenomics of urban and rural South African gut microbiomes reveal a transitional composition and undescribed taxa. *Nature Communications*, *13*(1), 926. https://doi.org/10.1038/s41467-021-27917-x

Tang, W., Liu, J., Ma, Y., Wei, Y., Liu, J., & Wang, H. (2021). Impairment of Intestinal Barrier Function Induced by Early Weaning via Autophagy and Apoptosis Associated With Gut Microbiome and Metabolites. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.804870

Tedersoo, L., Tooming-Klunderud, A., & Anslan, S. (2018). PacBio metabarcoding of Fungi and other eukaryotes: Errors, biases and perspectives. *New Phytologist*, *217*(3), 1370–1385. https://doi.org/10.1111/nph.14776

Thomas, C., Gioiello, A., Noriega, L., Strehle, A., Oury, J., Rizzo, G., Macchiarulo, A., Yamamoto, H., Mataki, C., Pruzanski, M., Pellicciari, R., Auwerx, J., & Schoonjans, K. (2009). TGR5-mediated bile acid sensing controls glucose homeostasis. *Cell metabolism*, *10*(3), 167–177. https://doi.org/10.1016/j.cmet.2009.08.001

Tian, H., Ye, C., Yang, B., Cui, J., Zheng, Z., Wu, C., Zhou, S., Lv, X., Qin, N., Qin, H., Li, N., & Chen, Q. (2022). Gut Metagenome as a Potential Diagnostic and Predictive Biomarker in Slow Transit Constipation. *Frontiers in Medicine*, *8*. https://doi.org/10.3389/fmed.2021.777961

Tierney, B. T., Yang, Z., Luber, J. M., Beaudin, M., Wibowo, M. C., Baek, C., Mehlenbacher, E., Patel, C. J., & Kostic, A. D. (2019). The Landscape of Genetic Content in the Gut and Oral Human Microbiome. *Cell Host & Microbe*, *26*(2), Article 2. https://doi.org/10.1016/j.chom.2019.07.008

Trapecar, M., Wogram, E., Svoboda, D., Communal, C., Omer, A., Lungjangwa, T., Sphabmixay, P., Velazquez, J., Schneider, K., Wright, C. W., Mildrum, S., Hendricks, A., Levine, S., Muffat, J., Lee, M. J., Lauffenburger, D. A., Trumper, D., Jaenisch, R., & Griffith, L. G. (2021). Human physiomimetic model integrating microphysiological systems of the gut, liver, and brain for studies of neurodegenerative diseases. *Science Advances*, *7*(5), Article 5. https://doi.org/10.1126/sciadv.abd1707

Treiber, M. L., Taft, D. H., Korf, I., Mills, D. A., & Lemay, D. G. (2020). Pre- and post-sequencing recommendations for functional annotation of human fecal metagenomes. *BMC Bioinformatics*, *21*(1), 74. https://doi.org/10.1186/s12859-020-3416-y

Tremlett, H., Bauer, K. C., Appel-Cresswell, S., Finlay, B. B., & Waubant, E. (2017). The gut microbiome in human neurological disease: A review. *Annals of Neurology*, *81*(3), 369–382. https://doi.org/10.1002/ana.24901

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., & Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods*, *12*(10), Article 10. https://doi.org/10.1038/nmeth.3589

Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C., & Segata, N. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Research*, *27*(4), 626–638. https://doi.org/10.1101/gr.216242.116

Tuomela, J., Kaprio, J., Sipilä, P. N., Silventoinen, K., Wang, X., Ollikainen, M., & Piirtola, M. (2019). Accuracy of self-reported anthropometric measures—Findings from the Finnish Twin Study. *Obesity Research & Clinical Practice*, *13*(6), 522–528. https://doi.org/10.1016/j.orcp.2019.10.006

Turnbaugh, P. J., Bäckhed, F., Fulton, L., & Gordon, J. I. (2008). Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host & Microbe*, *3*(4), 213–223. https://doi.org/10.1016/j.chom.2008.02.015

UNAIDs. (2023). *Global HIV & AIDS statistics—Fact sheet*. https://www.unaids.org/en/resources/fact-sheet

UNICEF. (2022, March 20). *Baseline report on diet and physical activity | UNICEF South Africa*. https://www.unicef.org/southafrica/reports/baseline-report-diet-and-physical-activity

University of the Witwatersrand. (2022). *Facts and figures*. https://www.wits.ac.za/about-wits/facts-and-figures/

Valeri, F., & Endres, K. (2021). How biological sex of the host shapes its gut microbiota. *Frontiers in Neuroendocrinology*, *61*, 100912. https://doi.org/10.1016/j.yfrne.2021.100912

Vallès, Y., & Francino, M. P. (2018). Air Pollution, Early Life Microbiome, and Development. *Current Environmental Health Reports*, *5*(4), 512–521. https://doi.org/10.1007/s40572-018-0215-y

van der Vossen, E. W. J., Davids, M., Bresser, L. R. F., Galenkamp, H., van den Born, B.-J. H., Zwinderman, A. H., Levin, E., Nieuwdorp, M., & de Goffau, M. C. (2023). Gut microbiome transitions across generations in different ethnicities in an urban setting—The HELIUS study. *Microbiome*, *11*(1), 99. https://doi.org/10.1186/s40168-023-01488-z

van Zyl, S., van Rooyen, F. C., Joubert, G., Kruger, W. H., & Walsh, C. M. (2020). A Comparison of the Socio-Behavioral-Metabolic Risk Profiles and Associated Factors for Chronic Diseases of Lifestyle in Urban and Rural Communities in Central South Africa. *Frontiers in Public Health*, *8*. https://www.frontiersin.org/articles/10.3389/fpubh.2020.570676

Vangay, P., Johnson, A. J., Ward, T. L., Al-Ghalith, G. A., Shields-Cutler, R. R., Hillmann, B. M., Lucas, S. K., Beura, L. K., Thompson, E. A., Till, L. M., Batres, R., Paw, B., Pergament, S. L., Saenyakul, P., Xiong, M., Kim, A. D., Kim, G., Masopust, D., Martens, E. C., … Knights, D. (2018). US Immigration Westernizes the Human Gut Microbiome. *Cell*, *175*(4), 962-972.e10. https://doi.org/10.1016/j.cell.2018.10.029

Varghese, B., Ling, Z., & Ren, X. (2022). Reconstructing the pulmonary niche with stem cells: A lung story. *Stem Cell Research & Therapy*, *13*(1), Article 1.

Varghese, S., Rao, S., Khattak, A., Zamir, F., & Chaari, A. (2024). Physical Exercise and the Gut Microbiome: A Bidirectional Relationship Influencing Health and Performance. *Nutrients*, *16*(21), Article 21. https://doi.org/10.3390/nu16213663

Veloo, A. C. M., Baas, W. H., Haan, F. J., Coco, J., & Rossen, J. W. (2019). Prevalence of antimicrobial resistance genes in *Bacteroides* spp. And *Prevotella* spp. Dutch clinical isolates. *Clinical Microbiology and Infection*, *25*(9), 1156.e9-1156.e13. https://doi.org/10.1016/j.cmi.2019.02.017

Vento, S., Dzudzor, B., Cainelli, F., & Tachi, K. (2018). Liver cirrhosis in sub-Saharan Africa: Neglected, yet important. *The Lancet Global Health*, *6*(10), e1060–e1061. https://doi.org/10.1016/S2214-109X(18)30344-9

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

Viaud, S., Saccheri, F., Mignot, G., Yamazaki, T., Daillère, R., Hannani, D., Enot, D. P., Pfirschke, C., Engblom, C., Pittet, M. J., Schlitzer, A., Ginhoux, F., Apetoh, L., Chachaty, E., Woerther, P.-L., Eberl, G., Bérard, M., Ecobichon, C., Clermont, D., … Zitvogel, L. (2013). The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. *Science (New York, N.Y.)*, *342*(6161), 971–976. https://doi.org/10.1126/science.1240537

Vidaurrazaga, M. M., McKinley, G. F., & Camins, B. C. (2020). Postpartum Fusobacterium gonidiaformans bacteremia. *Anaerobe*, *62*, 102168. https://doi.org/10.1016/j.anaerobe.2020.102168

Vilella, D. A. (2019). *Next-Generation-Sequencing.v1.2.5 @albertvilella*. Google Docs. https://docs.google.com/spreadsheets/d/1GMMfhyLK0-q8XkIo3YxlWaZA5vVMuhU1kg41g4xLkXc/htmlview?hl=en_GB&usp=embed_facebook

Villarreal-Soto, S. A., Beaufort, S., Bouajila, J., Souchard, J.-P., & Taillandier, P. (2018). Understanding Kombucha Tea Fermentation: A Review. *Journal of Food Science*, *83*(3), 580–588. https://doi.org/10.1111/1750-3841.14068

Vinogradova, E., Mukhanbetzhanov, N., Nurgaziyev, M., Jarmukhanov, Z., Aipova, R., Sailybayeva, A., Bekbossynova, M., Kozhakhmetov, S., & Kushugulova, A. (2024). Impact of urbanization on gut microbiome mosaics across geographic and dietary contexts. *mSystems*, *9*(10), e00585-24. https://doi.org/10.1128/msystems.00585-24

Vogt, N. M., Kerby, R. L., Dill-McFarland, K. A., Harding, S. J., Merluzzi, A. P., Johnson, S. C., Carlsson, C. M., Asthana, S., Zetterberg, H., Blennow, K., Bendlin, B. B., & Rey, F. E. (2017). Gut microbiome alterations in Alzheimer's disease. *Scientific Reports*, *7*(1), 13537. https://doi.org/10.1038/s41598-017-13601-y

Vujkovic-Cvijin, I., Sklar, J., Jiang, L., Natarajan, L., Knight, R., & Belkaid, Y. (2020). Host variables confound gut microbiota studies of human disease. *Nature*, *587*(7834), 448–454. https://doi.org/10.1038/s41586-020-2881-9

Vyas, T. (2017). EFFECT OF CHRONIC NAIL BITING AND NON-NAIL BITING HABIT ON THE ORAL CARRIAGE OF ENTEROBACTERIACEAE. *Journal of Advanced Medical and Dental Sciences Research*, *5*(5). https://doi.org/10.21276/jamdsr.2017.5.5.14

Wallenborn, J. T., Gunier, R. B., Pappas, D. J., Chevrier, J., & Eskenazi, B. (2022). Breastmilk, Stool, and Meconium: Bacterial Communities in South Africa. *Microbial Ecology*, *83*(1), 246–251. https://doi.org/10.1007/s00248-021-01758-z

Wandai, M. E., Aagaard-Hansen, J., Manda, S. O. M., & Norris, S. A. (2020). Transitions between body mass index categories, South Africa. *Bulletin of the World Health Organization*, *98*, 878-885l. https://api.semanticscholar.org/CorpusID:227343133

Wanders, R., Waterham, H., & Ferdinandusse, S. (2016). Metabolic Interplay between Peroxisomes and Other Subcellular Organelles Including Mitochondria and the Endoplasmic Reticulum. *Frontiers in Cell and Developmental Biology*, *3*. https://doi.org/10.3389/fcell.2015.00083

Wang, L., Pan, X., Jiang, L., Chu, Y., Gao, S., Jiang, X., Zhang, Y., Chen, Y., Luo, S., & Peng, C. (2022). The Biological Activity Mechanism of Chlorogenic Acid and Its Applications in Food Industry: A Review. *Frontiers in Nutrition*, *9*, 943911. https://doi.org/10.3389/fnut.2022.943911

Wang, L.-J., Yang, C.-Y., Chou, W.-J., Lee, M.-J., Chou, M.-C., Kuo, H.-C., Yeh, Y.-M., Lee, S.-Y., Huang, L.-H., & Li, S.-C. (2020). Gut microbiota and dietary patterns in children with attention-deficit/hyperactivity disorder. *European Child & Adolescent Psychiatry*, *29*(3), 287–297. https://doi.org/10.1007/s00787-019-01352-2

Wang, S., Ju, D., & Zeng, X. (2024). Mechanisms and Clinical Implications of Human Gut Microbiota-Drug Interactions in the Precision Medicine Era. *Biomedicines*, *12*(1), 194. https://doi.org/10.3390/biomedicines12010194

Wang, T., Shi, Z., Ren, H., Xu, M., Lu, J., Yang, F., Ye, C., Wu, K., Chen, M., Xu, X., Liu, D., Kong, L., Zheng, R., Zheng, J., Li, M., Xu, Y., Zhao, Z., Chen, Y., Yang, H., … Wang, W. (2024). Divergent age-associated and metabolism-associated gut microbiome signatures modulate cardiovascular disease risk. *Nature Medicine*, *30*(6), 1722–1731. https://doi.org/10.1038/s41591-024-03038-y

Wang, X., Li, L., Bian, C., Bai, M., Yu, H., Gao, H., Zhao, J., Zhang, C., & Zhao, R. (2023). Alterations and correlations of gut microbiota, fecal, and serum metabolome characteristics in a rat model of alcohol use disorder. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.1068825

Wang, X., Zhang, P., & Zhang, X. (2021). Probiotics Regulate Gut Microbiota: An Effective Method to Improve Immunity. *Molecules*, *26*(19), 6076. https://doi.org/10.3390/molecules26196076

Wang, Z., Emmerich, A., Pillon, N. J., Moore, T., Hemerich, D., Cornelis, M. C., Mazzaferro, E., Broos, S., Ahluwalia, T. S., Bartz, T. M., Bentley, A. R., Bielak, L. F., Chong, M., Chu, A. Y., Berry, D., Dorajoo, R., Dueker, N. D., Kasbohm, E., Feenstra, B., … Hoed, M. den. (2022). Genome-wide association analyses of physical activity and sedentary behavior provide insights into underlying mechanisms and roles in disease prevention. *Nature Genetics*, *54*(9), 1332–1344. https://doi.org/10.1038/s41588-022-01165-1

Wastyk, H. C., Fragiadakis, G. K., Perelman, D., Dahan, D., Merrill, B. D., Yu, F. B., Topf, M., Gonzalez, C. G., Treuren, W. V., Han, S., Robinson, J. L., Elias, J. E., Sonnenburg, E. D., Gardner, C. D., & Sonnenburg, J. L. (2021). Gut-microbiota-targeted diets modulate human immune status. *Cell*. https://doi.org/10.1016/j.cell.2021.06.019

Watanabe, R. M. (2011). *Inherited destiny? Genetics and gestational diabetes mellitus*.

Wegierska, A. E., Charitos, I. A., Topi, S., Potenza, M. A., Montagnani, M., & Santacroce, L. (2022). The Connection Between Physical Exercise and Gut Microbiota: Implications for Competitive Sports Athletes. *Sports Medicine (Auckland, N.z.)*, *52*(10), 2355–2369. https://doi.org/10.1007/s40279-022-01696-x

Weingarden, A., González, A., Vázquez-Baeza, Y., Weiss, S., Humphry, G., Berg-Lyons, D., Knights, D., Unno, T., Bobr, A., Kang, J., Khoruts, A., Knight, R., & Sadowsky, M. J. (2015). Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent Clostridium difficile infection. *Microbiome*, *3*, 10. https://doi.org/10.1186/s40168-015-0070-0

Weinstock. (2019). *Human gut microbiome and mock communities (ID 552603)—BioProject—NCBI*. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA552603

Weiss, A. S., Niedermeier, L. S., von Strempel, A., Burrichter, A. G., Ring, D., Meng, C., Kleigrewe, K., Lincetto, C., Hübner, J., & Stecher, B. (2023). Nutritional and host environments determine community ecology and keystone species in a synthetic gut bacterial community. *Nature Communications*, *14*(1), 4780. https://doi.org/10.1038/s41467-023-40372-0

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R., & Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, *5*(1), 27. https://doi.org/10.1186/s40168-017-0237-y

Westfall, S., Lomis, N., Kahouli, I., Dia, S. Y., Singh, S. P., & Prakash, S. (2017). Microbiome, probiotics and neurodegenerative diseases: Deciphering the gut brain axis. *Cellular and Molecular Life Sciences*, *74*(20), Article 20. https://doi.org/10.1007/s00018-017-2550-9

Wexler Hannah M. (2007). Bacteroides: The Good, the Bad, and the Nitty-Gritty. *Clinical Microbiology Reviews*, *20*(4), 593–621. https://doi.org/10.1128/cmr.00008-07

Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, *30*(3), 279–338. https://doi.org/10.2307/1943563

Wieërs, G., Belkhir, L., Enaud, R., Leclercq, S., Philippart de Foy, J.-M., Dequenne, I., de Timary, P., & Cani, P. D. (2019). How Probiotics Affect the Microbiota. *Frontiers in Cellular and Infection Microbiology*, *9*, 454. https://doi.org/10.3389/fcimb.2019.00454

Wilkins, L. J., Monga, M., & Miller, A. W. (2019). Defining Dysbiosis for a Cluster of Chronic Diseases. *Scientific Reports*, *9*(1), 12918. https://doi.org/10.1038/s41598-019-49452-y

Wilmanski, T., Diener, C., Rappaport, N., Patwardhan, S., Wiedrick, J., Lapidus, J., Earls, J. C., Zimmer, A., Glusman, G., Robinson, M., Yurkovich, J. T., Kado, D. M., Cauley, J. A., Zmuda, J., Lane, N. E., Magis, A. T., Lovejoy, J. C., Hood, L., Gibbons, S. M., … Price, N. D. (2021). Gut microbiome pattern reflects healthy ageing and predicts survival in humans. *Nature Metabolism*, *3*(2), 274–286. https://doi.org/10.1038/s42255-021-00348-0

Winglee, K., Eloe-Fadrosh, E., Gupta, S., Guo, H., Fraser, C., & Bishai, W. (2014). Aerosol Mycobacterium tuberculosis Infection Causes Rapid Loss of Diversity in Gut Microbiota. *PLOS ONE*, *9*(5), Article 5. https://doi.org/10.1371/journal.pone.0097048

Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J. S., Voigt, A. Y., Palleja, A., Ponnudurai, R. P., Sunagawa, S., Coelho, L. P., Schrotz-King, P., Vogtmann, E., Habermann, N., Niméus, E., Thomas, A. M., Manghi, P., Gandini, S., … Zeller, G. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature Medicine*, *25*(4), 679–689. https://doi.org/10.1038/s41591-019-0406-6

Wolf, A. R., Wesener, D. A., Cheng, J., Houston-Ludlam, A. N., Beller, Z. W., Hibberd, M. C., Giannone, R. J., Peters, S. L., Hettich, R. L., Leyn, S. A., Rodionov, D. A., Osterman, A. L., & Gordon, J. I. (2019). Bioremediation of a Common Product of Food Processing by a Human Gut Bacterium. *Cell Host & Microbe*, *26*(4), 463-477.e8. https://doi.org/10.1016/j.chom.2019.09.001

Won, S.-M., Chen, S., Lee, S. Y., Lee, K. E., Park, K. W., & Yoon, J.-H. (2020). Lactobacillus sakei ADM14 Induces Anti-Obesity Effects and Changes in Gut Microbiome in High-Fat Diet-Induced Obese Mice. *Nutrients*, *12*(12), 3703. https://doi.org/10.3390/nu12123703

Wood, L. F., Brown, B. P., Lennard, K., Karaoz, U., Havyarimana, E., Passmore, J.-A. S., Hesseling, A. C., Edlefsen, P. T., Kuhn, L., Mulder, N., Brodie, E. L., Sodora, D. L., & Jaspan, H. B. (2018). Feeding-Related Gut Microbial Composition Associates With Peripheral T-Cell Activation and Mucosal Gene Expression in African Infants. *Clinical Infectious Diseases*, *67*(8), 1237–1246. https://doi.org/10.1093/cid/ciy265

World Health Organization. (2016). WORLD HEALTH STATISTICS - MONITORING HEALTH FOR THE SDGs. *World Health Organization*, 1.121. https://doi.org/10.1017/CBO9781107415324.004

World Health Organization. (2018). *Global action plan for the prevention and control of NCDs 2013–2020*. https://www.who.int/publications-detail-redirect/9789241506236

World Health Organization. (2024). *Tuberculosis (TB)*. https://www.who.int/news-room/fact-sheets/detail/tuberculosis

Wu, F., Guo, X., Zhang, J., Zhang, M., Ou, Z., & Peng, Y. (2017). Phascolarctobacterium faecium abundant colonization in human gastrointestinal tract. *Experimental and Therapeutic Medicine*, *14*(4), 3122–3126. https://doi.org/10.3892/etm.2017.4878

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

Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics (Oxford, England)*, *32*(4), 605–607. https://doi.org/10.1093/bioinformatics/btv638

Xavier, R. J., & Podolsky, D. K. (2007). Unravelling the pathogenesis of inflammatory bowel disease. *Nature*, *448*(7152), 427–434. https://doi.org/10.1038/nature06005

Xiao, L., Zhou, T., Zuo, Z., Sun, N., & Zhao, F. (2024). Spatiotemporal patterns of the pregnancy microbiome and links to reproductive disorders. *Science Bulletin*, *69*(9), 1275–1285. https://doi.org/10.1016/j.scib.2024.02.001

Xiao, M., Zhang, C., Duan, H., Narbad, A., Zhao, J., Chen, W., Zhai, Q., Yu, L., & Tian, F. (2024). Cross-feeding of bifidobacteria promotes intestinal homeostasis: A lifelong perspective on the host health. *Npj Biofilms and Microbiomes*, *10*(1), 1–15. https://doi.org/10.1038/s41522-024-00524-6

Xu, L., Sun, L., Zhang, S., Wang, S., & Lu, M. (2019). High-resolution profiling of gut bacterial communities in an invasive beetle using PacBio smrt sequencing system. *Insects*, *10*(8). https://doi.org/10.3390/insects10080248

Xue, M.-Y., Xie, Y.-Y., Zang, X.-W., Zhong, Y.-F., Ma, X.-J., Sun, H.-Z., & Liu, J.-X. (2024). Deciphering functional groups of rumen microbiome and their underlying potentially causal relationships in shaping host traits. *iMeta*, *3*(4), e225. https://doi.org/10.1002/imt2.225

Yakobi, S. H., & Nwodo, U. U. (2025). Microbiological methodologies: Comparative evaluation of microbial community and enhanced antibiotic susceptibility testing. *Electronic Journal of Biotechnology*. https://doi.org/10.1016/j.ejbt.2025.01.001

Yan, H., Qin, Q., Chen, J., Yan, S., Li, T., Gao, X., Yang, Y., Li, A., & Ding, S. (2022). Gut Microbiome Alterations in Patients With Visceral Obesity Based on Quantitative Computed Tomography. *Frontiers in Cellular and Infection Microbiology*, *11*. https://doi.org/10.3389/fcimb.2021.823262

Yan, H., Qin, Q., Yan, S., Chen, J., Yang, Y., Li, T., Gao, X., & Ding, S. (2022). Comparison Of The Gut Microbiota In Different Age Groups In China. *Frontiers in Cellular and Infection Microbiology*, *12*. https://doi.org/10.3389/fcimb.2022.877914

Yan, P., Sun, Y., Luo, J., Liu, X., Wu, J., & Miao, Y. (2023). Integrating the serum proteomic and fecal metaproteomic to analyze the impacts of overweight/obesity on IBD: A pilot investigation. *Clinical Proteomics*, *20*(1), 6. https://doi.org/10.1186/s12014-023-09396-y

Yang. (2023). *105S rRNA full length amplicon of diarrhoea patients—SRA - NCBI*. https://www.ncbi.nlm.nih.gov/sra/?term=SRR24512451

Yang, F., Xie, X., Li, X., Liao, H., & Zou, B. (2022). Analysis of Psychological and Gut Microbiome Characteristics in Patients With Non-erosive Reflux Disease. *Frontiers in Psychiatry*, *12*. https://www.frontiersin.org/article/10.3389/fpsyt.2021.741049

Yang, J., Hirai, Y., Iida, K., Ito, S., Trumm, M., Terada, S., Sakai, R., Tsuchiya, T., Tabata, O., & Kamei, K. (2020). *Integrated gut–liver-on-a-chip platform as an in vitro human model of non-alcoholic fatty liver disease* (p. 2020.06.10.141606). bioRxiv. https://doi.org/10.1101/2020.06.10.141606

Yang, M., Bose, S., Lim, S., Seo, J., Shin, J., Lee, D., Chung, W. H., Song, E. J., Nam, Y. D., & Kim, H. (2020). Beneficial effects of newly isolated akkermansia muciniphila strains from the human gut on obesity and metabolic dysregulation. *Microorganisms*, *8*(9), 1–26. https://doi.org/10.3390/microorganisms8091413

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

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

Yenerall, J., You, W., & Hill, J. (2015). *A Multidimensional View of Food Environment Impact on Consumption and Food Security* (2015 AAEA & WAEA Joint Annual Meeting, July 26-28, San Francisco, California 205538). Agricultural and Applied Economics Association. https://econpapers.repec.org/paper/agsaaea15/205538.htm

Yeoh, Y. K., Chen, Z., Wong, M. C. S., Hui, M., Yu, J., Ng, S. C., Sung, J. J. Y., Chan, F. K. L., & Chan, P. K. S. (2020). Southern Chinese populations harbour non-nucleatum Fusobacteria possessing homologues of the colorectal cancer-associated FadA virulence factor. *Gut*, *69*(11), 1998–2007. https://doi.org/10.1136/gutjnl-2019-319635

Yeoh, Y. K., Sun, Y., Ip, L. Y. T., Wang, L., Chan, F. K. L., Miao, Y., & Ng, S. C. (2022). Prevotella species in the human gut is primarily comprised of Prevotella copri, Prevotella stercorea and related lineages. *Scientific Reports*, *12*(1), 9055. https://doi.org/10.1038/s41598-022-12721-4

Yeoh, Y. K., Zuo, T., Lui, G. C.-Y., Zhang, F., Liu, Q., Li, A. Y., Chung, A. C., Cheung, C. P., Tso, E. Y., Fung, K. S., Chan, V., Ling, L., Joynt, G., Hui, D. S.-C., Chow, K. M., Ng, S. S. S., Li, T. C.-M., Ng, R. W., Yip, T. C., … Ng, S. C. (2021). Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut*. https://doi.org/10.1136/gutjnl-2020-323020

Yin, X.-F., Ye, T., Chen, H.-L., Liu, J., Mu, X.-F., Li, H., Wang, J., Hu, Y.-J., Cao, H., & Kang, W.-Q. (2024). The microbiome compositional and functional differences between rectal mucosa and feces. *Microbiology Spectrum*, *12*(8), e03549-23. https://doi.org/10.1128/spectrum.03549-23

Yoshida, N., Emoto, T., Yamashita, T., Watanabe, H., Hayashi, T., Tabata, T., Hoshi, N., Hatano, N., Ozawa, G., Sasaki, N., Mizoguchi, T., Amin, H. Z., Hirota, Y., Ogawa, W., Yamada, T., & Hirata, K.-I. (2018). Bacteroides vulgatus and Bacteroides dorei Reduce Gut Microbial Lipopolysaccharide Production and Inhibit Atherosclerosis. *Circulation*, *138*(22), 2486–2498. https://doi.org/10.1161/CIRCULATIONAHA.118.033714

Younossi, Z. M., Golabi, P., Paik, J. M., Henry, A., Van Dongen, C., & Henry, L. (2023). The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): A systematic review. *Hepatology (Baltimore, Md.)*, *77*(4), 1335–1347. https://doi.org/10.1097/HEP.0000000000000004

Yu, C., Liu, S., Chen, L., Shen, J., Niu, Y., Wang, T., Zhang, W., & Fu, L. (2019). Effect of exercise and butyrate supplementation on microbiota composition and lipid metabolism. *The Journal of Endocrinology*, *243*(2), 125–135. https://doi.org/10.1530/JOE-19-0122

Yu, Z., Shen, X., Wang, A., Hu, C., & Chen, J. (2023). The gut microbiome: A line of defense against tuberculosis development. *Frontiers in Cellular and Infection Microbiology*, *13*. https://doi.org/10.3389/fcimb.2023.1149679

Yue, Y., Read, T. D., Fedirko, V., Satten, G. A., & Hu, Y.-J. (2023). Integrative analysis of microbial 16S gene and shotgun metagenomic sequencing data improves statistical efficiency. *bioRxiv: The Preprint Server for Biology*, 2023.06.27.546795. https://doi.org/10.1101/2023.06.27.546795

Yun, E. J., Yu, S., Park, N. J., Cho, Y., Han, N. R., Jin, Y.-S., & Kim, K. H. (2021). Metabolic and enzymatic elucidation of cooperative degradation of red seaweed agarose by two human gut bacteria. *Scientific Reports*, *11*(1), 13955. https://doi.org/10.1038/s41598-021-92872-y

Zampolli, J., De Giani, A., Rossi, M., Finazzi, M., & Di Gennaro, P. (2024). Who inhabits the built environment? A microbiological point of view on the principal bacteria colonizing our urban areas. *Frontiers in Microbiology*, *15*. https://doi.org/10.3389/fmicb.2024.1380953

Zeevi, D., Korem, T., Godneva, A., Bar, N., Kurilshikov, A., Lotan-Pompan, M., Weinberger, A., Fu, J., Wijmenga, C., Zhernakova, A., & Segal, E. (2019). Structural variation in the gut microbiome associates with host health. *Nature*, *568*(7750), 43–48. https://doi.org/10.1038/s41586-019-1065-y

Zhang, B., Wang, H.-H. E., Bai, Y.-M., Tsai, S.-J., Su, T.-P., Chen, T.-J., Wang, Y.-P., & Chen, M.-H. (2022). Bidirectional association between inflammatory bowel disease and depression among patients and their unaffected siblings. *Journal of Gastroenterology and Hepatology*, *37*(7), 1307–1315. https://doi.org/10.1111/jgh.15855

Zhang, L., Zheng, Z., Huang, H., Fu, Y., Chen, T., Liu, C., Yi, Q., Lin, C., Zeng, Y., Ou, Q., & Zeng, Y. (2024). Multi-omics reveals deoxycholic acid modulates bile acid metabolism via the gut microbiota to antagonize carbon tetrachloride-induced chronic liver injury. *Gut Microbes*, *16*(1), 2323236. https://doi.org/10.1080/19490976.2024.2323236

Zhang, X., Wang, H., Peng, S., Kang, J., Xie, Z., Tang, R., Xing, Y., He, Y., Yuan, H., Xie, C., & Liu, Y. (2022). Effect of microplastics on nasal and intestinal microbiota of the high-exposure population. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.1005535

Zhao, Q., Chen, Y., Huang, W., Zhou, H., & Zhang, W. (2023). Drug-microbiota interactions: An emerging priority for precision medicine. *Signal Transduction and Targeted Therapy*, *8*(1), 1–27. https://doi.org/10.1038/s41392-023-01619-w

Zhao, Z., Zhong, L., Zhou, P., Deng, Y., Liu, G., Li, P., Zeng, J., Zhang, Y., Tang, X., & Zhang, M. (2024). Impact of Dietary Fatty Acid Composition on the Intestinal Microbiota and Fecal Metabolism of Rats Fed a High-Fructose/High-Fat Diet. *Nutrients*, *16*(21), 3774. https://doi.org/10.3390/nu16213774

Zhu, X., Li, Y., Jiang, Y., Zhang, J., Duan, R., Liu, L., Liu, C., Xu, X., Yu, L., Wang, Q., Xiong, F., Ni, C., Xu, L., & He, Q. (2021). Prediction of Gut Microbial Community Structure and Function in Polycystic Ovary Syndrome With High Low-Density Lipoprotein Cholesterol. *Frontiers in Cellular and Infection Microbiology*, *11*, 665406. https://doi.org/10.3389/fcimb.2021.665406

Zhu, X., Xu, P., Zhu, R., Gao, W., Yin, W., Lan, P., Zhu, L., & Jiao, N. (2024). Multi-kingdom microbial signatures in excess body weight colorectal cancer based on global metagenomic analysis. *Communications Biology*, *7*(1), 1–14. https://doi.org/10.1038/s42003-023-05714-0

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

Żółkiewicz, J., Marzec, A., Ruszczyński, M., & Feleszko, W. (2020). Postbiotics—A Step Beyond Pre- and Probiotics. *Nutrients*, *12*(8), Article 8. https://doi.org/10.3390/nu12082189

Zuo, T., Zhang, F., Lui, G. C. Y., Yeoh, Y. K., Li, A. Y. L., Zhan, H., Wan, Y., Chung, A. C. K., Cheung, C. P., Chen, N., Lai, C. K. C., Chen, Z., Tso, E. Y. K., Fung, K. S. C., Chan, V., Ling, L., Joynt, G., Hui, D. S. C., Chan, F. K. L., … Ng, S. C. (2020). Alterations in Gut Microbiota of Patients With COVID-19 During Time of Hospitalization. *Gastroenterology*, *159*(3), 944-955.e8. https://doi.org/10.1053/j.gastro.2020.05.048

Zymo Research. (2023). *ZymoBIOMICS DNA Miniprep Kit*. Zymo Research International. https://zymoresearch.eu/products/zymobiomics-dna-miniprep-kit