

DIETARY GENOMICS AS A COMPLEMENTARY TOOL IN CHILDHOOD
MALNUTRITION INTERVENTION

by

Ammara Aqeel

Department of Molecular Genetics and Microbiology
Duke University

Defense Date: March 18th, 2025

Approved:

Lawrence A. David, Supervisor

Allison E. Ashley-Koch, Co-Chair

Martin D. Smith

Kelly D. Brownell

Jason Arnold

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of
Philosophy in the Department of Molecular Genetics and Microbiology in The Graduate School of
Duke University
2025

ABSTRACT

DIETARY GENOMICS AS A COMPLEMENTARY TOOL IN CHILDHOOD
MALNUTRITION INTERVENTION

by

Ammara Aqeel

Department of Molecular Genetics and Microbiology
Duke University

Defense Date: March 18th, 2025

Approved:

Lawrence A. David, Supervisor

Allison E. Ashley-Koch, Co-Chair

Martin D. Smith

Kelly D. Brownell

Jason Arnold

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Molecular Genetics and Microbiology in The Graduate
School of
Duke University
2025

Copyright by
Ammara Aqeel
2025
All rights reserved

Abstract

Malnutrition remains a critical global health challenge, especially among children. Developing effective treatment and prevention policies require accurate and large-scale dietary intake data yet traditional methods of dietary assessments rely on self-report methods and suffer from limitations of recall bias, reporting errors, and logistical constraints— particularly in resource-limited settings. Emerging omics-based approaches offer objective alternatives, with dietary genomics, specifically DNA metabarcoding, providing a non-invasive means to detect and quantify food intake from residual DNA in fecal samples. While recent studies have demonstrated feasibility for identifying dietary components and developing epidemiological metrics, the application of dietary genomics in clinical interventions for malnutrition remains unexplored. This dissertation evaluates dietary genomics as a scalable tool for monitoring dietary intervention uptake and impact in malnutrition across diverse contexts. In Chapter 2, we investigate whether food provisioning improves dietary quality in children with obesity (Durham, USA) in a randomized clinical trial using both traditional and genomic dietary assessment. Our results show comprehensive dietary improvements with food provisioning compared to usual care, with dietary genomics confirming consumption of provided foods and corroborating dietary shifts. These findings support expanding “Food-is-Medicine” interventions to pediatric populations and provide a roadmap for how dietary genomics can be integrated into clinical obesity management. In Chapter 3, we evaluate dietary genomics as a compliance biomarker in stool samples from malnourished infants from rural Pakistan receiving ready-to-use therapeutic and supplementary foods (RUTF/RUSF). Our results show significant detection of chickpea DNA (primary RUSF ingredient) coinciding with RUSF administration. Genomic analysis also uncovered unexpected dietary behaviors, such as widespread tea consumption, which may impede nutritional recovery. These findings establish dietary genomics as a novel method for compliance monitoring and highlight its potential for informing public health interventions. Collectively, the results in this thesis advance dietary genomics as a tool for enhancing interventions against childhood malnutrition and facilitating the development of effective context-specific nutritional guidelines and policies.

Contents

Abstract.....	iv
List of Tables.....	vii
List of Figures.....	viii
Acknowledgements.....	x
1. Introduction	1
1.1 Malnutrition as a global health problem.....	1
1.2 Prevention and treatment strategies.....	5
1.3 Methods of dietary assessment and challenges	8
1.4 Dietary genomics as an addition to the malnutrition toolbox.....	13
1.5 Contributions of this work	16
1.5.1 Collaboration and Authorship	16
2. Grocery intervention and DNA-based assessment to improve diet quality in pediatric obesity: a pilot randomized controlled study.....	18
2.1 Introduction	18
2.2 Results.....	20
2.2.1 Descriptive statistics.....	20
2.2.2 Primary outcome: change in diet quality from baseline to endpoint.....	23
2.2.3 Secondary outcomes	29
2.3 Discussion	35
2.4 Methods.....	37
2.4.1 Recruitment.....	38
2.4.2 Study Design.....	39
2.4.3 Treatment.....	40
2.4.4 Stool collection and DNA extraction.....	41
2.4.5 Primary outcome	42

2.4.6 Secondary outcomes	43
2.4.7 Statistical analyses.....	45
2.4.8 Data availability	49
3. Stool Genomics for Enhanced Monitoring of Nutrition Interventions in Childhood	
Malnutrition.....	50
3.1 Introduction	50
3.2 Results.....	52
3.3 Discussion	57
3.4 Methods.....	59
3.4.1 Data availability	63
4. Conclusion.....	64
4.1 Summary.....	64
4.2 Ongoing and future research.....	66
Appendix A	70
Appendix B.....	83
Appendix C.....	87
Appendix D.....	92
Bibliography.....	94
Biography.....	104

List of Tables

Table 1: Demographic descriptive statistics at baseline.	21
Table 2: Description of DSQ, weight, and food security outcomes stratified by treatment group for week 4 (endpoint) and week 8 (follow-up).....	23
Table 3: Estimated coefficients of intervention on DSQ, weight, and food-security outcomes in linear regression.....	28
Table 4: Factor loadings from principal component analysis (PCA) of dietary intake patterns and gut microbiome profiles.	31
Table 5: Estimated coefficients from linear mixed effects models with random intercept adjusted for appropriate covariates on dietary and microbiome patterns.	33
Table 6: Demographic descriptive characteristics at baseline between subjects who were included in the analysis and those who were excluded due to lost to follow-up or discontinued.....	70
Table 7: Summary of study deviation notes for participants lost to follow-up or discontinued.....	71
Table 8: Intervention and usual care participants experience with stool sampling. Assessed from participants who completed the study.	72
Table 9: Descriptive statistics of DSQ predicted intake of calcium (DSQcalc) (mg) in both treatment groups at baseline, week 4, and week 8.	72
Table 10: Top taxa identified as drivers of PC1 and PC2 in FoodSeq PCAs.	79
Table 11: Top taxa identified as drivers of PC1 and PC2 in dietary principal components analysis... ..	83
Table 12: Coding scheme used for analyzing participant interviews.....	88
Table 13: Representative quotes from major codes ($\geq 16.7\%$ of participants) across 12 coded interviews.	89

List of Figures

Figure 1: CONSORT diagram and study design.	20
Figure 2: High concordance between plant taxa identified in FoodSeq and foods corresponding to them ordered in groceries.....	30
Figure 3: Sampling design of infant subset from SEEM cohort included in this study and detection of Chickpea DNA from stool samples aligning with administration of RUSF.	52
Figure 4: Broad dietary patterns captured by FoodSeq including plant dietary composition, most frequently consumed taxa and potential dairy intake changes.....	54
Figure 5: Tea detected consistently across all timepoints and dietary diversity increases with age.	56
Figure 6: Total order price and item price distribution across all 64 grocery orders.....	73
Figure 7: Counts of food item types from 15 different food categories across all grocery orders.	74
Figure 8: Dietary intake distribution of participants throughout the study, compared to USDA 2020-2025 dietary guidelines (66).	75
Figure 9: Frequency across samples and total abundance of top 25 plant taxa detected in enrolled participants using FoodSeq.	76
Figure 10: PCA and scree plots of FoodSeq and Microbiome PCA	77
Figure 11: FoodSeq dietary diversity (pFR) and microbiome alpha diversity across treatment groups and relationship between pFR and PC1 from integrated cohort FoodSeq PCA.....	78
Figure 12: Food security status shifts through course of study.....	79
Figure 13: Major dietary taxa DNA detection patterns do not align with RUSF administration.	83
Figure 14: Dietary DNA Composition in Stool Samples: Heatmap of Vertebrate Taxa Abundances and Presence Counts of Common Food Taxa in Healthy Controls.....	85
Figure 15: Health status does not impact dietary composition and chickpea DNA detection is not impacted by socioeconomic or other dietary factors.....	86
Figure 16: Example of a personalized genomic dietary data return report provided to 17 participants (parents and children) in the CHOICES study.	87
Figure 17: Preliminary findings from 12 coded interviews using coding scheme.....	89
Figure 18: Genomic dietary diversity (pFR) and its relationship with deforestation and elevation across 19 villages in western Honduras.	92
Figure 19: Principal Component Analysis (PCA) of plant dietary composition across samples, with biplots indicating the relative contributions of the top 10 dietary components.	93

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Lawrence David, whose open-door policy, unwavering support, and trust in my independence shaped my Ph.D. journey. He gave me the freedom to explore, the opportunities to grow, and the autonomy to drive my research in directions I was passionate about. Beyond science, I am grateful for our heartfelt conversations about identity, home, belonging, and purpose.

I am also deeply grateful to my committee members, Allison Ashley-Koch, Martin Smith, Kelly Brownell, Raphael Valdivia, and Jason Arnold, for their guidance and encouragement throughout this journey. I always looked forward to the chance to reflect on my progress and gain fresh insights from their thoughtful questions.

Being part of the David Lab has been a privilege, and I extend my heartfelt thanks to all past and present lab members who have made it such a vibrant, inspiring, and supportive environment. In particular, I want to thank Brianna Petrone, whose invaluable mentorship and example shaped me as both a scientist and a person. Jeffrey Letourneau first welcomed me into the lab and had unwavering confidence in my abilities; Benjamin Neubert has been a constant friend through thick and thin, and I will always remember the turbulent trip we overcame together during our conference travel to Ireland; Teresa McDonald redefined my perspective of what is possible, and Jun Zeng enriched our work with his deep and insightful questions. Michelle Kirtley provided critical feedback on my manuscripts, making them ever more thorough. Sharon Jiang, the heart of the lab, was instrumental in making this work possible and was someone I could always rely on, for her keen experimental expertise, her support, and her fascinating book recommendations. To the newer members of the lab, Mengyi Dong, Dorothy Superdock, Aurian Gonzalez, Jacob Hiatt, Kylie Reed, Nolan Ives, Caroline Rivera, Anna Bauer, Ashish Subramanian, and Nicole Duncan, it has been a pleasure getting to know you and witnessing the lab evolve in exciting new directions as you joined.

Most importantly, I owe my success to the unwavering support of my friends and family. My deepest gratitude goes to my husband, Altamash Rafiq, who kept me fueled with chai and delicious

food, patiently listened to my endless rants and wild ideas, and stood by my side every step of the way. My parents have been instrumental in my education, making innumerable sacrifices to help me reach this stage. My sisters ensured this journey was never without adventure, adding humor and excitement throughout the way. I am also grateful to Agastya Sharma and Aruna Menon for our fun discussions, weekend dinners, and art nights, you brought so much joy and balance to this journey.

1. Introduction

Despite international commitments such as the United Nations Sustainable Development Goals, malnutrition remains a critical global health challenge with particularly severe impacts on children in low- and middle-income countries (LMICs). Development and implementation of appropriate treatment interventions and prevention policies requires understanding dietary patterns in the affected/at-risk populations. However, traditional dietary assessment methods rely on self-reports, suffering from well-established limitations of recall bias, reporting errors, and logistical constraints, particularly in resource-limited settings. Emerging omics-based approaches offer the potential to develop objective biomarkers and measures of diet. Among these, dietary genomics— specifically DNA metabarcoding— offers a promising, non-invasive method to detect and quantify dietary intake through residual food DNA from fecal samples. Initial studies have demonstrated the feasibility of this approach for not only identifying plant and animal components of human diets but also developing summary metrics with potential for epidemiological application. However, the utility of dietary genomics in interventions for improving clinical dietary outcomes remains to be explored. In this dissertation, I evaluate dietary genomics as a tool for assessing the uptake and impact of dietary interventions for malnutrition across diverse contexts. In this chapter, I outline the global burden of malnutrition, its causes, consequences, and treatment strategies, alongside challenges in traditional dietary assessment. I further examine emerging omics-based alternatives and the potential of dietary genomics to enhance malnutrition treatment interventions.

1.1 Malnutrition as a global health problem

In 2015, the United Nations introduced the Sustainable Development Goals (SDGs), a set of 17 interconnected objectives supported by specific long-term targets and indicators to address global challenges and foster a more sustainable future by 2030 (1). The second goal, “Zero Hunger,” set an ambitious target to eradicate malnutrition in all its forms by 2030 (2). Despite international commitments, however, malnutrition persists as a major global challenge (3).

A broad term referring to deficiencies, excesses, or imbalances in nutrient intake, malnutrition can take three different forms: undernutrition, micronutrient deficiencies and overweight and obesity (3). Undernutrition itself includes several categories: wasting (low weight-for-height; indicative of acute weight loss), stunting (low height-for-age; signaling chronic undernutrition), and underweight (low weight-for-age, signaling a combination of wasting and stunting) (3). Micronutrient deficiencies refer to the insufficient intake of essential vitamins and minerals such as vitamin A, iron, and iodine etc. needed for production of key biomolecules necessary for health and development (3) and are of particular concern globally as vulnerable groups like children and pregnant women are disproportionately impacted (3). Finally, overweight and obesity are characterized by an excessive accumulation of body fat, increasing the risk of noncommunicable diseases, including heart disease, stroke, diabetes, and certain cancers (3). Notably, these different forms of malnutrition are not mutually exclusive and can coexist within the same population, household, or even individual, an increasingly prevalent global phenomenon termed the “Double Burden of Malnutrition” due to a nutrition transition towards ultra-processed and high energy-density foods (3–5).

In 2022, malnutrition impacted an estimated 390 million adults globally in the form of underweight, 2.5 billion adults as overweight, and 890 million through obesity (3). Among children and adolescents between 5-19 years, 390 million were estimated to be overweight, 160 million affected by obesity, and 190 million suffering from thinness (body mass index (BMI) for age < 2 standard deviations below reference median) (3). In children under 5 years old, an estimated 149 million were impacted by stunting while 37 million were classified overweight or with obesity (3). Most alarmingly, undernutrition was associated with approximately half of all deaths in children under 5 years old (3).

This high prevalence of malnutrition in children is particularly concerning due to its profound and often irreversible effects on health and development (6). Nutrition starts in the womb and early childhood, specifically the “first 1,000” days from conception to the second birthday represent a critical window for physical growth and neurodevelopment during which nutritional deficits can result in permanent impairments such as compromised cognitive function, delayed psychomotor development,

and deficits in emotional and behavioral regulation (6,7). Furthermore, undernutrition significantly weakens immune competence, increasing susceptibility to infectious diseases such as diarrhea, pneumonia, and malaria along with exacerbating the severity and duration of these conditions (8). Overweight and obesity on the other hand, lead to an elevated risk of noncommunicable diseases in adulthood, including type 2 diabetes, cardiovascular disorders, and metabolic syndrome (9). Beyond health outcomes, childhood malnutrition adversely affects educational achievement, reduces economic productivity, and contributes to the intergenerational transmission of poverty (6).

The causes of childhood malnutrition are multifactorial, and the global burden is not uniform. LMICs (Gross national income per capita \leq \$1,145 - \$14,005 (10)) account for 91% of stunted and 92% of wasted children worldwide, with South Asia and Sub-Saharan Africa bearing the highest proportions of stunting (6). Inadequate exclusive breastfeeding for the first six months and insufficiently varied complementary foods contribute to stunting, wasting, micronutrient deficiencies, and noncommunicable diseases (6). Globally, only 49% of newborns receive breastmilk within the first hour of life, and just 44% of infants under six months are exclusively breastfed. Breastfeeding is further undermined by limited public health support, inadequate counselling, and aggressive marketing of substitutes (6). Among children aged 6–23 months, 72% do not consume a minimally diverse diet: while over 75% receive grains and breastmilk, 46% are not given fruits or vegetables, and 60% miss out on nutrient-dense foods like eggs, fish, or meat (6), with the lowest dietary diversities again reported in South Asia (18%) and sub-Saharan Africa (22%) (6). Wealth disparities exacerbate these gaps: only 18% of children from the poorest households meet minimum dietary diversity versus 38% in richer households although the low percentages even in richer households indicate that factors beyond income— such as knowledge, desirability and convenience— are critical (6).

For women, economic constraints, intra-household dynamics, and cultural norms can pose barriers to proper nutrition during pregnancy and post-partum, leading to increased risk of maternal anemia through inadequate iodine, iron, and calcium intake, hypertensive disorders, hemorrhage, maternal and newborn mortality, preterm delivery, intrauterine growth restriction, reduced breastmilk

quality, and low birthweight (6). Over 20 million infants are born with low birthweight each year, more than half being in South Asia (6). High costs, especially of nutrient-rich foods, further block dietary diversity, and households with fewer resources often resort to lower-quality, cheaper options (6). An estimated 385 million children live in extreme poverty, lacking access to a healthy diet, basic nutrition services, and safe water and sanitation. (6).

As a result of a globally pervasive nutrition transition entailing shifts in diet, consumption, and energy expenditure patterns linked to economic growth, globalization, and urbanization (11), 45% of all children impacted by overweight globally now live in Asia, with 78% of these children living in lower middle-income and upper middle-income countries (37% and 41%, respectively) (6). In LMICs, urban migration, rising incomes, increasing bread-winners per household, infrastructure development, and liberalized global trade policies have fueled private sector investment in the food industry (4), and contributed to heightened demand for inexpensive, ready-to-eat or ready-to-heat products leading to an “ultra-processed food revolution”, characterized by consumption of convenient and relatively affordable ultra-processed foods high in refined carbohydrates, fats, sugars, and salt (4). In many LMICs, ultra-processed food purchases have soared with the expansion of retail food outlets and the consolidation of the food supply chain under agribusinesses, manufacturers, and service companies (4). Euromonitor International data highlight rising sales of non-essential snack items and sugary beverages in Chile, South Africa, the Philippines, and Malaysia, with India and China among the top five markets for sugary drink producers (4). Research into infant feeding in low- and middle-income contexts further indicates that commercial snack foods often replace nutritious alternatives for young children, influenced by children’s taste preferences which can outweigh cost and other decision factors (6). Moreover, extensive marketing and labeling of commercial products promote early introduction of foods (before six months of age) and encourage consumption of items high in sugar or artificial flavors accompanied by unsubstantiated health claims (6). Overall, this rapid transition in food environments and dietary habits has led to rising rates of overweight, obesity, and non-communicable diseases in parallel with preexisting undernutrition in the population (4,11).

1.2 Prevention and treatment strategies

Strategies for the management of childhood malnutrition are multifaceted, ranging from maternal nutritional deficiency mitigation, immunization and micronutrient supplementation, exclusive breastfeeding support during the first six months of life and complementary feeding support afterwards, sanitation interventions (12), social safety net programs (13), and lifestyle and behavior interventions (14).

At the earliest stages, strategies focus on preventing the development of malnutrition during pregnancy. Balanced protein-energy supplementation for pregnant women, providing less than 25% of total energy from protein, can reduce intrauterine growth restriction/low birthweight (12) and prenatal supplementation with multiple micronutrients, including iron and folic acid, addresses essential micronutrient deficiencies (12,13). Post-birth interventions encourage exclusive breastfeeding given demonstrated improvements in child growth, immunity against respiratory and gastrointestinal infections, and survival along with environmental, economic, and social benefits (12,13). Interventional strategies at this stage focus on maternal nutritional counselling to provide safe and nutritious complementary foods to infants by encouraging healthy feeding practices when the household is food secure and provision of complementary foods when the household is food insecure along with fortification and supplementation of complementary foods with micronutrients or increased caloric density (12,13). Sanitation and access to clean water are similarly critical to lower the risk of infections such as diarrhea that can aggravate malnutrition (12). In cases where undernutrition is already present, highly fortified and energy-dense lipid-based nutrient supplements (e.g., ready-to-use therapeutic foods) and high-concentration micronutrient formulations are employed to treat severe acute malnutrition (SAM; weight-for-height < -3 standard deviations of the WHO growth standard, a mid-upper arm circumference (MUAC) < 115 mm, or the presence of nutritional edema) (12,15). Moderate acute malnutrition (MAM; weight-for-height between -2 and -3 standard deviations or MUAC between 115–125 mm) is managed with supplementary feeding programs to prevent

deterioration (12,15). Both conditions are addressed through a framework of Community-Based Management of Acute Malnutrition (CMAM) instead of direct in-patient care at a facility (12).

Social safety net programs such as income support, food transfers, subsidies, and school feeding programs are another strategy to mitigate the malnutrition risk due to food insecurity and poverty (13). Conditional cash transfer programs can specifically target mothers, incentivizing health and nutrition education while providing financial support, fostering greater nutritional awareness and adoption of best practices (13). Various forms of income support, food transfer, and subsidy programs have been implemented globally with demonstrated positive impacts on food insecurity and education and healthcare access (13). In the United States, the Supplemental Nutrition Assistance Program (SNAP), established in 1964 through the Food Stamp Act (16), supplements grocery budgets for low-income families, enabling the purchase of nutrient-dense foods such as fruits, vegetables, dairy, meat, bread, and cereals (17,18). Additionally, the Women, Infants, and Children (WIC) program provides access to nutritious foods, breastfeeding support, and nutrition education for pregnant and new mothers and young children facing nutritional risk (19). Similarly, Mexico's *Oportunidades* program utilizes conditional cash transfers to improve education and health outcomes in low-income families (20), while Egypt's *Takaful* and *Karama* programs provides both conditional and unconditional cash transfers to enhance food security and alleviate poverty (21). School feeding programs further reinforce these efforts by providing snacks, meals, or take-home rations to a collective 368 million children across 150 countries (13). In high-income countries, these initiatives promote healthy dietary intake among children, who can obtain up to 50% of their daily caloric requirements from school meals (22), while also reducing exposure to ultra-processed foods high in energy density, salt, and sugar (13). In LMICs, such programs serve as an additional safeguard against undernutrition by ensuring stable and reliable food access (13).

In terms of managing childhood obesity, a combination of behavioral and clinical interventions are implemented to reduce adiposity, prevent obesity-related chronic diseases, and address psychosocial and physical complications (14). Behavioral strategies encompass dietary

modifications, physical activity promotion, sedentary behavior and screen time reduction, and sleep hygiene improvements, primarily delivered through school-based, preschool-based, community-based, and home-based programs (14,23). In cases where behavioral approaches are insufficient, clinical interventions, including anti-obesity pharmacotherapy and metabolic or bariatric surgery, may be considered (14,23). Dietary interventions for obesity management may involve nutritional education alone or in combination with moderate energy restriction, structured meal plans, or adherence to dietary patterns aligned with national guidelines (14). In the US for example, dietary intervention can be provided through intensive health behavior and lifestyle treatment (IHBLT) which is a family-centered approach coordinated by a medical clinic/home (24). Multiple stakeholders are engaged, ranging from the child and the family to pediatric health care providers such as registered dietitians, nurses, social workers, exercise specialists and psychologists as well the broader community and schools to provide face-to-face nutrition and physical activity education and support over a course of 26 contact hours over 3-12 months (24). These strategies emphasize increased fruit and vegetable intake while reducing consumption of energy-dense, nutrient-poor foods and sugar-sweetened beverages (14). Studies in children and adolescents (aged 2–20 years) have demonstrated modest reductions in total energy intake, lower sugar-sweetened beverage consumption, and higher fruit and vegetable intake following such interventions (14). Physical activity interventions typically incorporate structured exercise programs or education, fostering a supportive, enjoyable, and non-judgmental environment that enhances motor skills, confidence, and peer socialization (14). Sedentary behavior interventions focus on parental involvement, home environment modifications (e.g., limiting access to electronic games in bedrooms), and digital device monitoring (14). Psychological interventions, either as adjuncts to other approaches or standalone treatments, target factors such as body image concerns, negative mood, and stimulus control (14). Cognitive behavioral therapy (CBT) is the most widely used method, effectively improving dietary habits, psychosocial health, self-esteem, quality of life, and anthropometric measures, including BMI and waist circumference (14). Pharmacological treatments may be indicated for adolescents in whom behavioral interventions alone have been insufficient (14).

Several glucagon-like peptide-1 receptor agonists and combination therapies are now approved for chronic obesity treatment in adolescents aged 12–18 years. In addition to liraglutide (3 mg daily), which was initially the only approved option (14), semaglutide (Wegovy) received FDA approval in 2022 for once-weekly use, demonstrating significant efficacy in reducing body mass index (25). Additionally, the combination therapy phentermine/topiramate (Qsymia) was approved in 2022 for weight management in adolescents, providing an alternative mechanism of action (26). Setmelanotide (Imcivree) is also available for individuals with rare genetic obesity disorders (27). For severe adolescent obesity, metabolic and bariatric surgery remains the most effective and durable intervention, with Roux-en-Y gastric bypass and vertical sleeve gastrectomy achieving BMI reductions of approximately 25–40% over 1–9 years (14). Beyond weight loss, these procedures yield significant improvements in obesity-related comorbidities, cardiometabolic risk factors, musculoskeletal pain, and functional mobility (14). In terms of setting, school-based studies have reported the most favorable outcomes for combined dietary intake and physical activity interventions with secondary utilization of the home setting (e.g. parental outreach) improving the chances of success of the intervention (23). Fewer studies have been conducted in the pre-school and community-based settings and offer modest to weak support for the efficacy of combined diet and physical activity interventions (23). Exclusive home-based interventions are rare in the literature and have not provided conclusive results (23).

1.3 Methods of dietary assessment and challenges

Development of effective malnutrition interventions depend on dietary intake assessments within target populations for evaluating nutritional adequacy, informing agricultural, nutritional, and social policies, and tracking progress toward global nutrition targets (28,29). High-quality diet data shape programs for examining consumer food choices and their relationships with agricultural production and the food environment as well as with health impacts (29). Dietary intake data also help identify targets for food fortification and contribute to the development of population-specific dietary guidelines, food labeling regulations, and public health initiatives (29). Dietary assessment also

constitutes an integral component of nutrition interventions for measuring adherence to behavioral interventions (30,31), facilitate behavior modification (31), and evaluate intervention efficacy and outcomes (30,31).

Dietary assessment methods predominantly rely on self-reporting and can be categorized into retrospective and prospective approaches (28,32,33). Among retrospective methods, 24-hour dietary recalls (24 hr-recalls) require individuals to recall all foods and beverages consumed over the past 24 hours (28,34). These can be either self-administered or interviewer-administered, conducted using traditional paper-based formats or technology-enhanced tools such as web-based systems (e.g., ASA24 (35)), mobile applications, cameras, and digital assistants (32). This method captures detailed information, including food products, brand names, portion sizes, ingredients, and preparation methods (32). To enhance recall accuracy, 24 hr-recalls commonly employ the multi-pass method, which systematically guides participants through multiple rounds of questioning to improve specificity and completeness (34). Another commonly used retrospective tool is the food frequency questionnaire (FFQ) or diet history, where individuals report their habitual dietary intake over a predefined period, ranging from a week to a year (28). Respondents select foods from a predetermined list, and portion size options may be included to create quantitative, semi-quantitative, or non-quantitative instruments (32). FFQs are widely used due to their ease of administration and cost-effectiveness compared to 24-hr recalls. However, as the food list is predefined, FFQs require careful adaptation to ensure cultural and population relevance (32). Prospective methods, such as food diaries, require participants to document foods and beverages consumed in real time over a specified period, estimating portion sizes and weights using measurement aids such as scales, measuring cups, reference images or food models to improve accuracy (32).

However, both retrospective and prospective approaches suffer from limitations inherent to self-reporting that affect the accuracy and reliability of intake data (28,32). Retrospective methods are prone to memory error such as recall bias (under-reporting or over-reporting), errors in estimating portion sizes, and omissions in memory (28). The accuracy of 24-hr recalls depends on the respondent's

recall ability and the interviewer's proficiency in probing appropriately, which can lead to underreporting or overreporting of intake; additionally, 24-hr recall must be conducted over multiple days to obtain a reliable estimate of usual dietary patterns (32). The primary limitation of FFQs is their lower resolution in capturing specific food types and quantities, as well as the requirement for sufficient literacy levels to complete the instrument accurately (32). Prospective methods, while reducing recall bias, impose high participant burden— for instance, requiring individuals to weigh foods before each meal— which can lead to altered eating behaviors as participants modify their dietary intake to simplify the recording process (28).

24-hr recalls with multiple passes, FFQs, and food diaries also constitute the primary methods for dietary assessment in children (32), posing another layer of challenges (36–38). Compared to adults, children have limited literacy, comprehension, and attention spans, making self-reporting difficult (38). Additionally, dietary habits and food intake patterns evolve more rapidly over shorter age intervals, further complicating assessment (32). For infants and toddlers, self-reporting is not possible. Consequently, parents, primary caretakers or other adults often act as surrogate or supplementary reporters (32,37). However, parental knowledge of a child's dietary intake is often incomplete or biased, particularly as children grow older (39,40) and begin consuming food outside the direct supervision of the primary reporter (e.g., at school or family events) (32,37). Beyond these methodological concerns, dietary assessments must also be adapted to the cultural, socio-economic, and environmental context of each country (28). Factors such as seasonality, literacy levels, food traditions, supply chains, and food availability all influence the feasibility and accuracy of dietary data collection (28). Given that LMICs bear the highest burden of malnutrition, substantial modifications to dietary assessment approaches are necessary for their specific geographic and economic contexts (6). Adapting dietary assessment tools is a complex and non-trivial task, with successful modifications— such as the validation of an FFQ for a new cultural setting— meriting dedicated studies (41). While 24-hr recalls and FFQs are preferred over food diaries in LMICs due to the latter's higher participant burden (32), several logistical and financial challenges persist (28). These include the high costs of training and

employing nutrition-trained field workers, time-intensive data collection, limited research infrastructure, and the lack of country-specific food composition databases and automated data collection platforms in local languages (28).

Acknowledging these limitations, the field has shifted towards methodological advancements through smart technology, biological biomarkers and omics-based approaches (33,42). Smart technology innovations include image-based food capture pre- and post-intake with automated estimation of food identity and portion size through machine learning and artificial intelligence facilitated by reference objects of known dimensions for precise quantification (28,33). Image capture can be performed using smartphone cameras, standalone digital cameras, or wearable cameras that continuously take photographs independent of eating events (33). In LMICs, image-capture based dietary assessment using mobile or smartphone cameras has shown promise in overcoming literacy barriers while providing an additional advantage of local data storage, without requiring an active internet connection (28). However, large-scale implementation remains challenging due to the need for robust analytical methods capable of deriving accurate nutrient estimates from images given the variability in food preparation, portion sizes, and individual consumption behaviors, posing significant challenges for computer vision algorithms before analysis can be automated in free-living individuals (33). Another emerging approach involves wearable sensor-based devices designed to detect hand-to-mouth movements and physiological eating actions such as biting, chewing, and swallowing (33). These devices may include wrist-mounted sensors that track food movement toward the mouth, as well as sensors positioned on the neck or jaw to monitor eating behaviors (33). However, testing has been limited to controlled environments with small sample sizes, and further validation in real-world settings with larger study populations is required before wider implemented (33).

In terms of biological approaches, advancements have been made in the development of objective and readily administrable biomarkers of diet (33,42). Recovery biomarkers, exhibiting direct correlations with intake (33) include examples such as doubly-labeled water which is used to estimate total energy intake, urinary markers of sodium and potassium intake or urinary nitrogen which reflects

protein consumption (33). However, recovery biomarkers are available only for a few dietary components and thus, broader applicability remains limited (33). Another category, concentration biomarkers, is obtained from biological samples such as plasma but does not directly correspond to dietary intake due to variations in metabolism and individual characteristics, including obesity (33). These biomarkers are particularly useful for examining tissue concentrations in relation to health outcomes and can be detected in various matrices, including hair, adipose tissue, plasma, and red blood cell membranes (33). Predictive biomarkers have gained visibility in recent years, particularly for assessing intake of sugars, phytosterols/stanols, polyphenols, isoflavones, and whole grains (33). Urinary sucrose and fructose excretion over 24 hours, as well as the carbon stable isotope ratio ($^{13}\text{C}/^{12}\text{C}$), have been proposed as biomarkers of sugar intake (33). Alkylresorcinols (ARs), phenolic lipids found predominantly in the bran fraction of wheat and rye and with lower concentrations in barley, rice, and oats, have been suggested as biomarkers for whole grain consumption (33). A novel non-invasive approach, skin reflectance spectroscopy, has been introduced to objectively assess fruit and vegetable intake by measuring carotenoid pigments that accumulate in the skin and contribute to its yellow hue, although the method is more expensive than the other biomarkers (33). Metabolomics has also emerged as a powerful tool in nutrition research, enabling the detection of dietary metabolites in saliva, urine, blood, and biopsies using mass spectrometry (MS) and nuclear magnetic resonance (NMR) techniques (33,42). This approach has broad applications, from identifying diet-disease relationships to discovering new dietary biomarkers and elucidating mechanisms of action in nutritional interventions (33). Notably, metabolomics has linked branched-chain amino acids (BCAAs) to type 2 diabetes risk and has facilitated the identification of dietary biomarkers for broccoli, tea, coffee, and citrus fruits (33). However, due to the complex processes of nutrient absorption and metabolism, blood metabolomics offer only partial insights into dietary intake (43). Recent efforts have sought to address this constraint through development of food-specific databases of mass spectrometry spectra to enable detection of food signatures in fecal, urinary, or blood metabolomes (43). However, while these reference spectra pinpoint food-derived compounds with high precision, they offer minimal

information regarding overall food abundance (43). Similarly, no metabolomic biomarker has been identified that captures the breadth of food items covered by self-report based dietary assessment in a single measurement (42). Therefore, an ongoing need exists for the discovery and validation of novel biomarkers capable of assessing a broad range of dietary components and patterns (33,42).

1.4 Dietary genomics as an addition to the malnutrition toolbox

Dietary genomics, specifically through DNA metabarcoding of human stool samples, is one approach that offers the potential for such broad biomarkers of diet in a single, non-invasive measure (44–46). Originally developed in wildlife ecology as a tool for mapping predator-prey relationships, dietary DNA metabarcoding detects residual fragments of dietary DNA from consumed food within fecal samples (47–49). This process comprises five key steps: (1) extracting DNA from fecal samples, (2) amplifying taxonomically informative regions using specific DNA barcodes, (3) sequencing the amplified DNA through next-generation sequencing (NGS) technologies, (4) analyzing the sequenced data via specialized bioinformatics pipelines, and (5) identifying taxa present in the sample based on reference sequence databases (47). Given the universality of DNA as a conserved biomolecule across all living organisms, DNA metabarcoding can be applied globally to any sample from which DNA can be obtained (48) and as this approach transitions into human nutrition research, it presents new opportunities to refine dietary assessments and enhance our understanding of diet-health relationships at an unprecedented level of taxonomic resolution.

Initially, DNA metabarcoding employed targeted PCR amplification using specific primers for target prey species or taxonomic groups such as sand shrimp feeding on stone flounder, whale sharks consuming larval crabs, or collembola preying on nematodes (48). The technique further evolved into using broad or group-specific primers to amplify a range of dietary DNA, followed by identification of individual taxa through cloning and sequencing (48). This approach enabled analyses such as assessment of extinct mammalian diet from coprolites and the stomach contents of ancient humans, as well as contemporary ecological research on predator-prey relationships across a diversity of habits (48). The advent of NGS technologies significantly streamlined dietary DNA analysis, replacing

traditional cloning-based sequencing methods and enabling the rapid generation of large datasets (48), spurring studies on the diets of species ranging from worms, insects, and mollusks to bats, penguins and Australian fur seals (48). Concurrently, the expansion of comprehensive sequence databases through extensive DNA metabarcoding of both animal and plant species facilitated accurate taxonomic identification of the vast sequence data produced and substantially broadened the application of the technique (48), revealing otherwise undetected key dietary components across multiple species such as jellyfish consumption by albatross species and deep-sea fish predation by shearwaters (49).

In recent years, DNA metabarcoding techniques have been adapted for dietary assessment in humans. Reese et al. conducted the first pilot study, evaluating the effectiveness of DNA metabarcoding in characterizing the plant-based component of human diets by analyzing fecal samples from 11 individuals participating in a controlled feeding trial with either a plant-based or animal product-based diet (44). Reese et al. employed a widely used plant DNA metabarcoding primer targeting the chloroplast *trnL* (UAA)-P6 region (50), a marker favored for dietary studies due to its short amplicon length, conserved primer-binding sites, sufficient interspecific variability and demonstrated success in detecting plant DNA in human fecal samples (50) and diet composition analyses in wild herbivores (51). The study identified chloroplast DNA corresponding to both cooked and raw preparations of common dietary items such as grains, vegetables, fruits, and herbs with a recall of 0.86 and a precision of 0.55 when compared to written dietary records and effectively distinguished between baseline and intervention phase diet samples (44). However, the study also highlighted methodological limitations, including a high PCR failure rate and difficulty of differentiating certain plant species at the sequence level (44). Building on this foundation, Schneider et al. demonstrated the application of DNA metabarcoding in forensic analysis of human stomach contents post-mortem by assessing both plant and vertebrate dietary components in 48 stomach content samples collected during medicolegal autopsies and identifying 124 unique plant sequences and 34 unique vertebrate sequences (45).

The most comprehensive study to date, conducted by Petrone et al., advanced DNA metabarcoding as a tool for quantifying dietary plant diversity in humans (46). Using the chloroplast trnL-P6 marker, they analyzed 1,029 fecal samples from 324 participants across two interventional feeding studies and three observational cohorts (46). Petrone et al. introduced plant metabarcoding richness (pMR), defined as the number of plant taxa detected per sample, and demonstrated its correlation with dietary intake data from controlled feeding interventions and food frequency questionnaires (46). In adolescents, where validated dietary surveys were unavailable (52), trnL metabarcoding identified 111 plant taxa, with 86 consumed by multiple individuals and four— wheat, chocolate, corn, and members of the potato family— consumed by over 70% of participants (46). Adolescent pMR was associated with age and household income, replicating trends from prior epidemiological studies and demonstrated the potential of metabarcoding to generate dietary data in contexts where traditional methods are limited (46). To address amplification challenges noted by Reese et al., Petrone et al. implemented several protocol refinement, including adoption of a high-fidelity polymerase to reduce PCR errors, implementation of a two-step amplification process to minimize barcode bias, increase in annealing temperature to 63°C for specificity, and quadrupling template volume for efficiency (46). These optimizations improved PCR success to 92%, increased sequencing yield, and enabled the detection of additional plant taxa per sample, outperforming the response rates of validated FFQs (76 – 82%) (46). Beyond laboratory improvements, the study demonstrated enhanced taxonomic assignment by expanding their dietary plant reference database from 185 to 791 sequences, covering 468 plant species and improving species-, genus-, and family-level resolution to 83.2%, 92.6%, and 99.3%, respectively (46) and transitioning from clustering operational taxonomic units (OTUs) to exact amplicon sequence variants (ASVs) (46). Collectively, these enhancements reduced unassigned reads to 0.9%, with a median of 0.2% per sample, significantly increasing analytical accuracy and robustness (46).

Collectively, these studies underscore the potential of dietary genomics to provide high-resolution taxonomic identification, objective dietary assessment, and quantifiable dietary patterns and

ecological metrics. However, no study has yet explored the application of dietary genomics as a complementary tool for enhancing dietary assessment in malnutrition interventions. Such applications could yield critical insights by enabling objective diet tracking and measurement of intervention adherence. Given the universal nature of DNA, this approach could be particularly valuable given the disproportionate burden of malnutrition in populations that are challenging to assess using traditional dietary methods, including children, individuals in low- and middle-income countries, and those in contexts of high poverty and limited educational attainment.

1.5 Contributions of this work

In this thesis, I aim to establish human dietary assessment through dietary genomics (hereafter “FoodSeq”) as a complementary tool for assessing the uptake and impact of dietary interventions for different forms of malnutrition in different geographical and demographic contexts.

In Chapter 2, I evaluate a “food-is-medicine” intervention implemented as a grocery provisioning program for improving dietary quality among children with obesity in Durham, North Carolina, USA. Using FoodSeq, I assess whether genomic dietary analysis enhances the measurement of intervention adherence and validates dietary outcomes by addressing the well-documented limitations of traditional self-reported dietary assessments.

In Chapter 3, I investigate the feasibility of dietary genomics as an objective method for monitoring compliance with a ready-to-use supplementary food (RUSF) intervention among undernourished infants in *Matiani*, Sindh, Pakistan. This study examines whether DNA-based dietary assessment can serve as a scalable tool to support nutrition programs in resource-limited settings, where conventional dietary assessment methods face significant challenges.

1.5.1 Collaboration and Authorship

Material in Chapter 2 has been adapted from a previously published manuscript (53), co-authored with Melissa C. Kay, Jun Zeng, Brianna L. Petrone, Chengxin Yang, Tracy Truong,

Covington B. Brown, Sharon Jiang, Veronica M. Carrion, Stephanie Bryant, Michelle C. Kirtley, Cody D. Neshteruk, Sarah C. Armstrong, Lawrence A. David.

Material in Chapter 3 is currently being prepared for publication and has been co-authored with Najeeha Iqbal, Sheraz Ahmed, Sanam Iram Soomro, Michael Barratt, Teresa McDonald, Nolan Ives, Olivia Osborne, Sharon Jiang, Fayaz Umrani, Kazi Ahsan, Jeffrey Gordon, Asad Ali, Lawrence David.

2. Grocery intervention and DNA-based assessment to improve diet quality in pediatric obesity: a pilot randomized controlled study

2.1 Introduction

Childhood obesity affects 20% of US children and adolescents between ages 2 and 19 years, representing a nearly 40% increase over the past two decades (54). Left untreated, children with obesity can develop preventable chronic diseases such as type 2 diabetes, hypertension, and cardiovascular disease (9). Childhood obesity frequently co-occurs with household food insecurity, defined as “limited or uncertain availability or capacity to obtain and access nutritionally adequate and safe food” (55,56). Estimated to impact 2.3 million US households (57), food insecurity alone can disrupt dietary patterns, resulting in stunted childhood growth and development (58). The co-occurrence of obesity and food insecurity not only exacerbates children's health outcomes (55) but also impairs efforts to treat obesity itself (59).

Treating obesity in the context of food insecurity requires addressing both issues simultaneously (55). Interventions for children with obesity have historically focused on educational approaches to improve dietary habits (60). However, for households already struggling with inconsistent food access and quality, nutritional counseling that advises consumption of more fruits and vegetables and less energy-dense foods can be challenging to implement and insensitive to the barriers that patients face (55). The latest guidance emphasizes an intensive health behavior and lifestyle treatment (IHBLT) approach, consisting of 26 or more hours of tailored, family-based, face-to-face support covering nutrition, behavior changes, and physical activity over a 3- to 12-month period (24). However, studies have suggested that cost, rather than lack of nutritional awareness, may drive purchases of unhealthy foods in low-income households (61,62). Thus, interventions focusing on improving access to healthy foods may improve diet quality and obesity outcomes more effectively.

At the same time, measuring the efficacy and compliance of dietary interventions through self-reported dietary assessment remains challenging (38). Children often lack the literacy and attention

span needed to reliably respond to dietary surveys (38). Surrogate reporting by parents has known biases and is limited by incomplete knowledge of children's diet (38). Stool genomics offers an objective alternative. The recently developed technique FoodSeq characterizes dietary intake of individual plant taxa by sequencing residual food DNA in the human stool (46). The universal language of DNA standardizes dietary data comparisons across cultural, geographic, and demographic differences. FoodSeq can further assess dietary diversity by enumerating unique plant taxa in each sample, generating a plant FoodSeq Richness (pFR) score. This score correlates with validated measures of diet quality and has replicated established epidemiological findings in a pediatric obesity cohort (46). Gut microbiome assessment through stool genomics can also provide an early readout of diet-driven health changes. Although the definition of a “healthy” microbiome remains elusive (63), microbiome compositional and functional diversity and the presence of fiber-degrading bacteria that produce anti-inflammatory short-chain fatty acids such as butyrate are important markers of gut health, are known to be depleted in obesity (63), and can be rapidly modulated through dietary changes (64).

We conducted a pilot clinical trial (Childhood Obesity Microbiome Study [CHOICE]) to determine whether increased access to healthy foods improves diet quality in children with obesity. Children were randomized to either 4 weeks of usual care or usual care supplemented with a food-provisioning intervention involving weekly dietitian-guided online grocery shopping worth \$25 per person (up to \$100 per family). As self-reported survey-based dietary assessment remains the field standard, the primary outcome was a change in children's diet quality assessed as increased alignment of Dietary Screener Questionnaire (DSQ, self-administered, National Cancer Institute) (65) derived dietary intake values with US Department of Agriculture (USDA) dietary guidelines (66). In order to address limitations of self-reported pediatric dietary data, however, we implemented FoodSeq as an orthogonal, genomics-based dietary assessment and conducted microbiome analysis to assess potential early health impacts as secondary outcomes. FoodSeq analyses included within-cohort comparisons and a post hoc comparison to a larger pediatric obesity cohort (i.e., Pediatric Obesity Microbiome and Metabolism Study [POMMS], n = 195, ages 10–18 years) (46,52) from similar households recruited

from the same clinic. Changes in weight and food-insecurity status from baseline to end of treatment and follow-up were included as additional secondary outcomes.

2.2 Results

2.2.1 Descriptive statistics

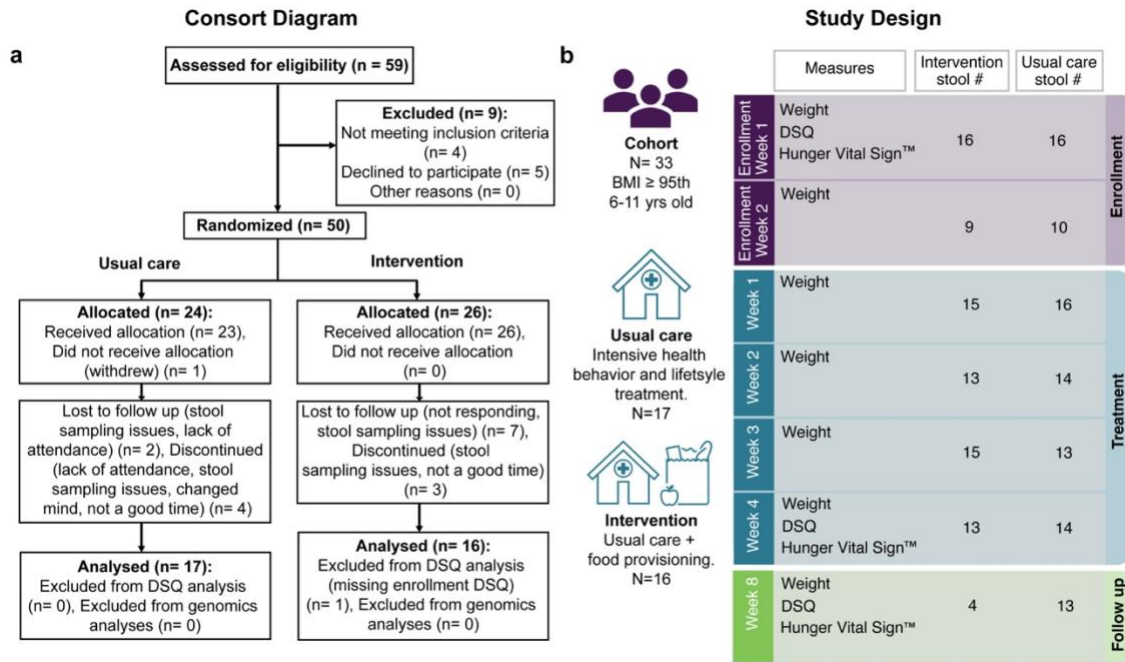


Figure 1: CONSORT diagram and study design.

(a) Derivation of study population and short form reasoning for exclusions. (b) Children with obesity were randomized into either intervention or usual care. The study consisted of 10 total weeks, with two enrollment weeks, four treatment weeks, and a follow-up at week 8 post-enrollment. Measures included Dietary Screener Questionnaire, Hunger Vital Sign™, and weight. Number of stool samples received from intervention and usual care groups respectively are shown for each week.

A total of 59 participants were assessed for eligibility at the Healthy Lifestyles Clinic (Fig. 1a), with 24 and 26 participants randomized to usual care and intervention, respectively. A total of 17 usual care and 16 intervention participants completed the study (Fig. 1a). Descriptive statistics of participants who completed the study versus those lost to follow-up or who discontinued were provided in Table 6 (Appendix A). Participant loss or discontinuation was primarily attributed to participant non-responsiveness followed by problems with study compliance and stool sample collection (Appendix A: Table 7). Stool sampling was largely consistent across all weeks other than Week 8 for intervention,

which showed a significant drop-off (Fig. 1b). A greater proportion of intervention participants reported negative or mixed experiences with stool sampling compared with those in usual care (Appendix A: Table 8), and, within the intervention, participants who did not provide Week 8 stool samples reported negative/mixed experiences more frequently (Appendix A: Table 8).

There were no significant differences in baseline demographic or dietary characteristics between the groups (Table 1; Appendix A: Table 6). Participants' mean age was 8.9 (SD 1.5) years, with mean BMI of 29.5 (SD 4.8) kg/m², and they were almost evenly split between female and male (48.5% vs. 51.5%, respectively). Most participants self-reported race as either “other” (45.5%) or “Black or African American” (36.4%). Roughly one-half of the participants self-reported ethnicity as “Hispanic or Latino” (54.5%). Most parents reported education level as “some college or below” (67.7%). Baseline predicted calcium intake of participants aligned with dietary guidelines (66) (Appendix A: Table 9). Over one-third of participants reported food insecurity at baseline (36.4%). For intervention, all households except one reported four or more members, meeting the maximum grocery support threshold (\$100). The one household with three members, however, was also assigned a support of \$100 due to a protocol oversight.

Table 1: Demographic descriptive statistics at baseline.

	Usual care (17)	Intervention (16)	Total (33)
Child's race^a			
Black or African American	8 (47.1%)	4 (25.0%)	12 (36.4%)
White	1 (5.9%)	3 (18.8%)	4 (12.1%)
Other	6 (35.3%)	9 (56.3%)	15 (45.5%)
Two or more races	2 (11.8%)	0 (0%)	2 (6.1%)
Child's ethnicity			
Hispanic or Latino	8 (47.1%)	10 (62.5%)	18 (54.5%)
Not Hispanic or Latino	9 (52.9%)	5 (31.3%)	14 (42.4%)
Prefer not to answer	0 (0%)	1 (6.3%)	1 (3.0%)
Child's sex			
Female	7 (41.2%)	9 (56.3%)	16 (48.5%)
Male	10 (58.8%)	7 (43.8%)	17 (51.5%)
Child's age (years)			
Mean (SD)	8.8 (± 1.8)	9.0 (± 1.3)	8.9 (± 1.5)
Median [Q1, Q3]	8 [8, 11]	9 [8, 10]	9 [8, 10]
[Min, Max]	[6, 11]	[6, 11]	[6, 11]
Child's BMI at baseline			
Mean (SD)	29.7 (± 4.5)	29.2 (± 5.2)	29.5 (± 4.8)

Median [Q1, Q3]	29.5 [26.5, 34.7]	28 [24.7, 32.8]	28.8 [26, 34.4]
[Min, Max]	[21.7, 37.2]	[22.3, 38.4]	[21.7, 38.4]
Child's weight at baseline (pounds)			
Mean (SD)	131.2 (\pm 39.5)	139.1 (\pm 36.4)	134.9 (\pm 37.7)
Median [Q1, Q3]	120.8 [110, 156]	137 [116, 155.7]	121.2 [110.9, 157.8]
[Min, Max]	[78, 196]	[73.4, 217.5]	[73.4, 217.5]
Unknown	0 (0%)	1 (6.3%)	1 (3.0%)
Having other family member(s) with obesity			
Yes	5 (29.4%)	3 (18.8%)	8 (24.2%)
No	9 (52.9%)	6 (37.5%)	15 (45.5%)
Unknown	3 (17.6%)	7 (43.8%)	10 (30.3%)
Parent's education^b			
Some college or below	11 (64.7%)	12 (75.0%)	23 (69.7%)
Undergraduate or higher	6 (35.3%)	4 (25.5%)	10 (30.3%)
Prefer not to answer	0 (0%)	0 (0%)	0 (0%)
Parent's race^a			
Black or African American	8 (47.1%)	3 (18.8%)	11 (33.3%)
White	2 (11.8%)	3 (18.8%)	5 (15.2%)
Other	4 (23.5%)	8 (50.0%)	12 (36.4%)
Two or more races	1 (5.9%)	0 (0%)	1 (3.0%)
Unknown	2 (11.8%)	2 (12.5%)	4 (12.1%)
Parent's ethnicity			
Hispanic or Latino	8 (47.1%)	11 (68.8%)	19 (57.6%)
Not Hispanic or Latino	7 (41.2%)	4 (25.0%)	11 (33.3%)
Prefer not to answer	0 (0%)	0 (0%)	0 (0%)
Unknown	2 (11.8%)	1 (6.3%)	3 (9.1%)
Family's food insecurity status at baseline			
Secure	11 (64.7%)	10 (62.5%)	21 (63.6%)
Insecure	6 (35.3%)	6 (37.5%)	12 (36.4%)
<i>Note:</i> There were no statistically significant study group differences in any of the variables shown.			
Abbreviations: Max, maximum; Min, minimum; Q, quartile.			
a "American Indian/Alaska Native," "Asian," and "Native Hawaiian or Other Pacific Islander" races were merged into "Other." Different multiracial categories were merged into "Two or more races."			
b Education levels from associate's degree and below were merged into "Some college or below." Other levels except for "Prefer not to answer" were merged into "Undergraduate or higher."			

The intervention received 64 grocery orders with a median total order price of \$96.45 [IQR: \$93.13, \$100.04] and median item price across all orders of \$3.28 [IQR: \$1.98, \$5.00] (Appendix A: Fig. 6). Vegetables (n = 432) were the most frequently ordered food category, followed by fruits (n = 279)

and whole grains (n = 196; Appendix A: Fig. 7), whereas water (n = 1), alternative protein (n = 3), and sauce (n = 5) were the most infrequently ordered (Appendix A: Fig. 7). There were 23 orders for alternative dairy and 14 for animal dairy (e.g., cheese, yogurt; Appendix A: Fig. 7). In terms of specific foods, wheat products (n = 102), chicken (n = 63), tomatoes (n = 58), eggs (n = 54), and strawberries (n = 49) were the topmost frequently ordered items (Appendix A: Fig. 7).

2.2.2 Primary outcome: change in diet quality from baseline to endpoint

Both usual care and intervention demonstrated a decrease in DSQ predicted intake of total added sugars (DSQsug) and DSQ predicted intake of added sugars from sugar-sweetened beverages (DSQssb) over the treatment period (Table 2). No other DSQ components changed within usual care (Table 2). Within intervention, all DSQ components assessed changed significantly, with increases in DSQ predicted intake of fruits/vegetables, including legumes and excluding French fries (DSQfvlnf; marginally significant), DSQ predicted intake of whole grains (DSQwhgr), and DSQ predicted intake of fiber (DSQfiber) and decreases in DSQ predicted intake of dairy (DSQdairy) and DSQ predicted intake of calcium (DSQcalc; Table 2).

Table 2: Description of DSQ, weight, and food security outcomes stratified by treatment group for week 4 (endpoint) and week 8 (follow-up).

	Usual care (17)	<i>p</i> value ^a	Intervention (16)	<i>p</i> value ^a	Total (33)	<i>p</i> value ^b
Change from baseline to week 4						
Child weight (lbs)						
Mean (SD)	0.2 (± 8.8)	0.650	-0.6 (± 2.6)	0.609	-0.1 (± 6.9)	0.364
Median [Q1, Q3]	1.4 [-3.8, 5.2]		-1.0 [-2.9, 1.2]		0.6 [-3.4, 3.3]	
[Min, Max]	[-19.0, 15.7]		[-3.9, 4]		[-19.0, 15.7]	
Unknown	3 (17.6%)		6 (37.5%)		9 (27.3%)	
DSQ predicted intake of fruits/ vegetables including legumes and excluding French fries (DSQfvlnf) (cup eq)						
Mean (SD)	-0.09 (± 0.80)	0.890	0.24 (± 0.53)	0.048	0.07 (± 0.69)	0.089

Median [Q1, Q3]	-0.02 [-0.48, 0.23]		0.28 [0.11, 0.52]		0.13 [-0.20, 0.31]	
[Min, Max]	[-1.97, 1.58]		[-1.15, 1.19]		[-1.97, 1.58]	
Unknown	0 (0%)		1 (6.3%)		1 (3%)	
DSQ predicted intake of dairy (DSQdairy) (cup eq)						
Mean (SD)	0.28 (\pm 1.39)	0.776	-1.092 (\pm 1.164)	<0.001	-0.36 (\pm 1.45)	0.006
Median [Q1, Q3]	0.00 [-0.32, 0.94]		-0.36 [-2.23, -0.17]		-0.17 [-1.07, 0.00]	
[Min, Max]	[-2.35, 2.65]		[-3.06, 0.11]		[-3.06, 2.65]	
Unknown	0 (0%)		1 (6.3%)		1 (3%)	
DSQ predicted intake of whole grains (DSQwhgr) (oz eq)						
Mean (SD)	0.00 (\pm 0.21)	0.963	0.19 (\pm 0.26)	0.018	0.09 (\pm 0.25)	0.018
Median [Q1, Q3]	0.04 [-0.12, 0.09]		0.18 [0.09, 0.34]		0.09 [-0.04, 0.24]	
[Min, Max]	[-0.43, 0.35]		[-0.36, 0.75]		[-0.43, 0.75]	
Unknown	0 (0%)		1 (6.3%)		1 (3%)	
DSQ predicted intake of total added sugars (DSQsug) (tsp eq)						
Mean (SD)	-3.65 (\pm 5.61)	0.003	-4.80 (\pm 7.61)	0.003	-4.19 (\pm 6.54)	0.941
Median [Q1, Q3]	-2.28 [-4.27, -0.35]		-1.25 [-5.21, -0.44]		-1.70 [-4.78, -0.34]	
[Min, Max]	[-20.00, 1.94]		[-22.27, 0.81]		[-22.27, 1.94]	
Unknown	0 (0%)		1 (6.3%)		1 (3%)	
DSQ predicted intake of added sugars from sugar-sweetened beverages (DSQssb) (tsp eq)						
Mean (SD)	-2.38 (\pm 4.95)	0.003	-3.63 (\pm 7.40)	0.030	-2.97 (\pm 6.15)	0.622
Median [Q1, Q3]	-0.51 [-1.87, 0.00]		-0.61 [-1.75, 0.15]		-0.56 [-1.88, 0]	
[Min, Max]	[-19.46, 0.20]		[-21.50, 0.61]		[-21.50, 0.61]	

Unknown	0 (0%)		1 (6.3%)		1 (3%)	
DSQ predicted intake of fiber (DSQfiber) (gm)						
Mean (SD)	-0.80 (\pm 2.77)	0.243	1.61 (\pm 1.85)	0.004	0.33 (\pm 2.64)	0.005
Median [Q1, Q3]	-0.61 [-1.12, 0.53]		1.46 [0.10, 2.71]		0.13 [-0.97, 2.13]	
[Min, Max]	[-8.55, 3.97]		[-0.99, 5.95]		[-8.55, 5.95]	
Unknown	0 (0%)		1 (6.3%)		1 (3%)	
DSQ predicted intake of calcium (DSQcalc) (mg)						
Mean (SD)	58 (\pm 351)	0.747	-252 (\pm 283)	0.003	-88 (\pm 353)	0.024
Median [Q1, Q3]	-6 [-54, 265]		-84 [-518, -17]		-49 [-268, 32]	
[Min, Max]	[-667, 665]		[-692, 35]		[-692, 665]	
Unknown	0 (0%)		1 (6.3%)		1 (3%)	
Family's food insecurity status at week 4						
secured	10 (58.8%)		13 (81.3%)		23 (69.7%)	0.259
insecure	7 (41.2%)		3 (18.8%)		10 (30.3%)	
Change in food insecurity status at week 4 from baseline		1.000		0.248		
secured to insecure	1 (5.9%)		0 (0%)		1 (3%)	0.161
unchanged	16 (94.1%)		13 (81.3%)		29 (87.9%)	
insecure to secured	0 (0%)		3 (18.8%)		3 (9.1%)	
Change from baseline to week 8						
Child weight (lbs)						
Mean (SD)	0.1 (\pm 8.1)	0.660	-1.0 (\pm 2.4)	0.313	-0.2 (\pm 6.9)	0.265
Median [Q1, Q3]	2.3 [-2.6, 4.8]		-1.2 [-2.6, 0.5]		0.9 [-3.0, 3.2]	
[Min, Max]	[-18.0, 11.6]		[-4.0, 2.4]		[-18.0, 11.6]	
Unknown	3 (17.6%)		10 (62.5%)		13 (39.4%)	
DSQ predicted intake of fruits/ vegetables including legumes and excluding French fries (DSQfvlnf) (cup eq)						
Mean (SD)	0.00 (\pm 0.82)	0.644	-0.09 (\pm 0.33)	0.542	-0.04 (\pm 0.64)	0.421

Median [Q1, Q3]	0.15 [-0.42, 0.46]		-0.10 [-0.15, 0.13]		0.05 [-0.17, 0.20]	
[Min, Max]	[-2.31, 1.12]		[-1.08, 0.26]		[-2.31, 1.12]	
Unknown	0 (0%)		2 (12.5%)		2 (6.1%)	
DSQ predicted intake of dairy (DSQdairy) (cup eq)						
Mean (SD)	0.04 (\pm 1.24)	0.940	-0.91 (\pm 1.16)	0.027	-0.39 (\pm 1.28)	0.110
Median [Q1, Q3]	-0.01 [-0.31, 0.65]		-0.27 [-1.96, -0.10]		-0.19 [-0.56, 0.16]	
[Min, Max]	[-2.42, 2.56]		[-2.91, 0.34]		[-2.91, 2.56]	
Unknown	1 (5.9%)		3 (18.8%)		4 (12.1%)	
DSQ predicted intake of whole grains (DSQwhgr) (oz eq)						
Mean (SD)	0.08 (\pm 0.27)	0.353	0.10 (\pm 0.20)	0.135	0.09 (\pm 0.241)	0.953
Median [Q1, Q3]	0.05 [-0.03, 0.24]		0.05 [-0.04, 0.16]		0.05 [-0.04, 0.23]	
[Min, Max]	[-0.33, 0.60]		[-0.17, 0.56]		[-0.33, 0.60]	
Unknown	0 (0%)		2 (12.5%)		2 (6.1%)	
DSQ predicted intake of total added sugars (DSQsug) (tsp eq)						
Mean (SD)	-3.71 (\pm 6.31)	0.009	-3.01 (\pm 5.75)	0.017	-3.39 (\pm 5.97)	0.739
Median [Q1, Q3]	-2.61 [-5.57, -0.07]		-1.74 [-2.77, -0.32]		-1.76 [-4.48, -0.14]	
[Min, Max]	[-21.93, 4.66]		[-20.87, 1.73]		[-21.93, 4.66]	
Unknown	0 (0%)		2 (12.5%)		2 (6.1%)	
DSQ predicted intake of added sugars from sugar-sweetened beverages (DSQssb) (tsp eq)						
Mean (SD)	-2.29 (\pm 5.10)	0.010	-2.56 (\pm 5.69)	0.025	-2.41 (\pm 5.28)	0.936
Median [Q1, Q3]	-0.26 [-1.94, 0.00]		-0.83 [-1.24, 0]		-0.67 [-1.42, 0]	

[Min, Max]	[-19.81, 0.70]		[-19.94, 1.15]		[-19.94, 1.15]	
Unknown	0 (0%)		2 (12.5%)		2 (6.1%)	
DSQ predicted intake of fiber (DSQfiber) (gm)						
Mean (SD)	-0.64 (\pm 3.15)	0.706	0.585 (\pm 1.574)	0.168	-0.09 (\pm 2.60)	0.268
Median [Q1, Q3]	0.05 [-1.94, 1.07]		0.61 [0.10, 1.52]		0.38 [-1.69, 1.52]	
[Min, Max]	[-9.63, 3.96]		[-2.76, 3.34]		[-9.63, 3.96]	
Unknown	1 (5.9%)		3 (18.8%)		4 (12.1%)	
DSQ predicted intake of calcium (DSQcalc) (mg)						
Mean (SD)	3 (\pm 306)	0.940	-227 (\pm 278)	0.017	-100 (\pm 311)	0.110
Median [Q1, Q3]	-9 [-112, 133]		-87 [-526, -24]		-40 [-172, 27]	
[Min, Max]	[-554, 684]		[-641, 94]		[-641, 684]	
Unknown	1 (5.9%)		3 (18.8%)		4 (12.1%)	
Family's food insecurity status at week 8						
secured	12 (70.6%)		9 (56.3%)		21 (63.6%)	0.798
insecure	5 (29.4%)		6 (37.5%)		11 (33.3%)	
Unknown	0 (0%)		1 (6.3%)		1 (3%)	
Change in food insecurity status at week 8 from baseline		1.000		1.000		
secured to insecure	1 (5.9%)		2 (12.5%)		3 (9.1%)	0.835
unchanged	14 (82.4%)		11 (68.8%)		25 (75.8%)	
insecure to secured	2 (11.8%)		2 (12.5%)		4 (12.1%)	
Unknown	0 (0%)		1 (6.3%)		1 (3%)	
<p>Abbreviations: DSQ, Dietary Screener Questionnaire; DSQcalc, DSQ predicted intake of calcium; DSQdairy, DSQ predicted intake of dairy; DSQfiber, DSQ predicted intake of fiber; DSQfvlnf, DSQ predicted intake of fruits/vegetables, including legumes and excluding French fries; DSQssb, DSQ predicted intake of added sugars from sugar-sweetened beverages; DSQsug, DSQ predicted intake of total added sugars; DSQwhgr, DSQ predicted intake of whole grains; eq, equivalents; Max, maximum; Min, minimum; Q, quartile.</p> <p>^a Comparison of outcomes within group. P values from Wilcoxon signed rank test for DSQ and weight outcomes and McNemar test for food-insecurity outcomes (see Appendix A: Fig. 12 for figure depiction of food-insecurity outcomes). Significant p values are bold.</p> <p>^b Comparison of outcomes between usual care and intervention. P values from Wilcoxon rank sum test for DSQ and weight outcomes and Fisher exact test for food-insecurity outcomes. Significant p values are bold.</p>						

Comparing the impact of intervention over usual care, the intervention was again associated with an increase in DSQwhgr score ($\beta = 0.20$, 95% confidence interval [CI]: 0.05 to 0.34; $p = 0.013$), an increase in DSQfiber score ($\beta = 2.52$, 95% CI: 1.28 to 3.76; $p < 0.001$), a decrease in DSQdairy score ($\beta = -1.31$, 95% CI: -2.02 to -0.60 ; $p = 0.001$), and a decrease in DSQcalc score ($\beta = -318.20$, 95% CI: -504.16 to -132.25 ; $p = 0.002$; Table 3), consistent with our microbiome-promoting dietary protocol of high-fiber, low-dairy foods. Overall dietary intake shifts in intervention improved alignment with USDA dietary guidelines (Appendix A: Figure 8).

Table 3: Estimated coefficients of intervention on DSQ, weight, and food-security outcomes in linear regression.

Outcome: Baseline to Week 4	β (95% CI)	<i>p</i> value
Change in Weight ^a (lbs)	0.28 (-5.60, 6.16)	0.927
Change in DSQfvlnf (fruit and vegetables, no French fries) ^a (cup eq)	0.38 (0.00, 0.76)	0.062
Change in DSQdairy (dairy) ^a (cup eq)	-1.31 (-2.02, -0.60)	0.001
Change in DSQwhgr (whole grains) ^a (oz eq)	0.20 (0.05, 0.34)	0.013
Change in DSQsug (total added sugars) ^a (tsp eq)	-0.70 (-1.86, 0.46)	0.246
Change in DSQssb (added sugars from beverages) ^a (tsp eq)	0.35 (-0.13, 0.82)	0.163
Change in DSQfib (fiber) ^a (gm)	2.52 (1.28, 3.76)	<0.001
Change in DSQcalc (calcium) ^a (mg)	-318.20 (-504.16, -132.25)	0.002
Food insecurity at week 4 ^b	0.33 (0.07, 1.61)	0.170
Outcome: Baseline to Week 8	β (95% CI)	<i>p</i> value
Change in Weight ^a (lbs)	-1.15 (-8.07, 5.77)	0.749
Change in DSQfvlnf (fruit and vegetables, no French fries) ^a (cup eq)	-0.02 (-0.45, 0.40)	0.918
Change in DSQdairy (dairy) ^a (cup eq)	-0.78 (-1.47, -0.09)	0.036
Change in DSQwhgr (whole grains) ^a (oz eq)	0.03 (-0.15, 0.20)	0.778
Change in DSQsug (total added sugars) ^a (tsp eq)	1.32 (-0.95, 3.58)	0.265
Change in DSQssb (added sugars from beverages) ^a (tsp eq)	1.01 (-1.35, 3.37)	0.411
Change in DSQfib (fiber) ^a (gm)	1.24 (-0.45, 2.93)	0.162

Change in DSQcalc (calcium) ^a (mg)	-212.75 (-371.34, -54.17)	0.014
Food insecurity at week 8 ^b	1.65 (0.29, 9.33)	0.570
<p>Abbreviations: DSQ, Dietary Screener Questionnaire; DSQcalc, DSQ predicted intake of calcium; DSQdairy, DSQ predicted intake of dairy; DSQfiber, DSQ predicted intake of fiber; DSQfvlnf, DSQ predicted intake of fruits/vegetables, including legumes and excluding French fries; DSQssb, DSQ predicted intake of added sugars from sugar-sweetened beverages; DSQsug, DSQ predicted intake of total added sugars; DSQwhgr, DSQ predicted intake of whole grains; eq, equivalents.</p> <p>^a Linear regression adjusted for the outcome at baseline. No other covariates included. Estimated β coefficient reported. Significant p values are bold.</p> <p>^b Logistic regression. No other covariates included. Estimated odds ratio reported. Significant p values are bold.</p>		

2.2.3 Secondary outcomes

2.2.3.1 FoodSeq

FoodSeq identified 112 unique trnL-P6 sequence variants. Taxonomic assignments for these variants mapped to 45 plant families, 80 plant genera, and 63 plant species. Four taxa (wheat, corn, cacao bean, and soybean/guar bean) were detected in over 50% of the samples (Appendix A: Fig. 9). There was high concordance between plant taxa detected in FoodSeq and foods corresponding with these taxa in grocery orders (adjusted R² = 0.65; p < 0.0001; Fig. 2). Notable exceptions included cacao, corn, and guar bean/soybean (detected more than ordered) and fruits, vegetables, and legumes food categories (ordered more than detected; Fig. 2).

Comparison of grocery foods and corresponding FoodSeq plant taxa

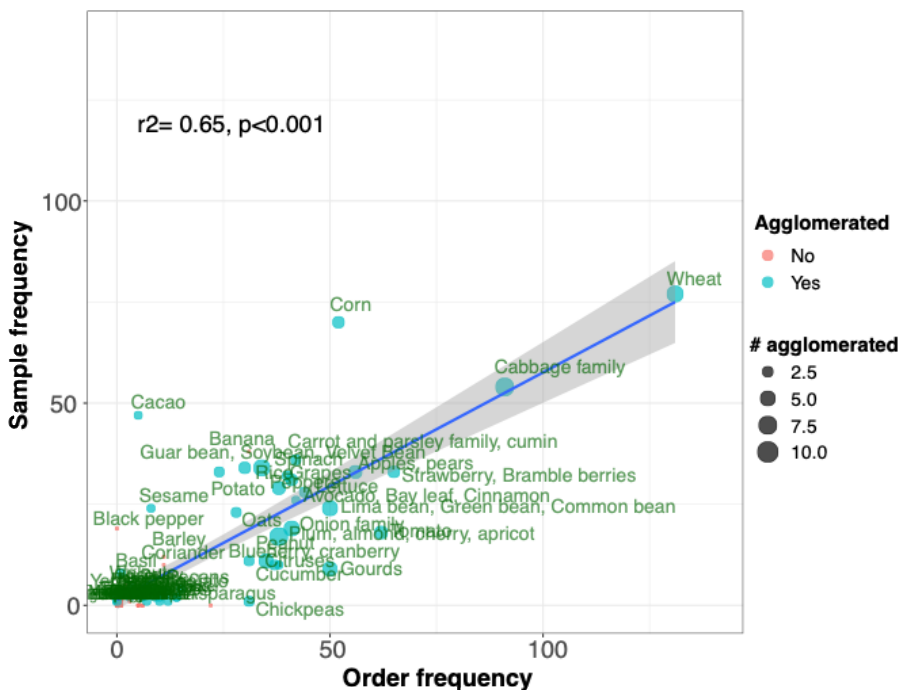


Figure 2: High concordance between plant taxa identified in FoodSeq and foods corresponding to them ordered in groceries.

Linear regression (adjusted $r^2 = 0.65$, $p < 0.001$). Foods mapping multiple FoodSeq taxa were counted for each applicable taxon and taxa mapping to multiple foods counted for each applicable food (“agglomerated”, Cyan dots, number of items agglomerated indicated by size of the dot). Plant taxa with only one distinct corresponding food item represented by salmon dots. Foods found more frequently through FoodSeq but not ordered as much in the groceries were more likely to be processed or added as an ingredient to packaged foods (e.g. Chocolate, Corn, “Guar bean, soybean, velvet bean”). Foods ordered more frequently but not detected as often through FoodSeq were more likely to be Fruits, vegetables or legumes species (e.g. Tomato, Gourds, Chickpeas).

Within enrolled participants, two major dietary patterns were identified (Appendix A: Fig. 10a,b), explaining 16.5% of dietary intake variation (PC1 = 9.3% and PC2 = 7.2%). However, recognizable dietary relationships could not be resolved from the directionality of factor loadings driving these dietary patterns (Table 4), potentially due to our limited sample size, and no significant differences were seen from baseline to end of treatment for these mixed dietary patterns involving either usual care or intervention (Table 5).

Table 4: Factor loadings from principal component analysis (PCA) of dietary intake patterns and gut microbiome profiles.

Biomarker	Sample Set			
FoodSeq	Enrolled participants	Factor loading	PC1	PC2
		Spinach	2.03	-1.83
		Cabbage family	1.62	-1.34
		Rice	-	-1.26
		Barley	0.79	-0.83
		Cabbage family.1	0.55	-0.79
		Guar bean, Soybean	-1.43	-0.78
		Carrot and parsley family, cumin	2.37	-0.44
		Lettuce	2.29	0.37
		Blueberry	0.85	0.67
		Oats	0.95	0.94
		Grapes	1.01	0.95
		Apples, pears	-1.10	1.55
		Strawberry, Bramble berries	1.62	3.43
		Avocado, Bay leaf, Cinnamon	1.26	-
		Cacao	-2.38	-
			Name	PC1
	Integrated with POMMS	Cacao	1.88	-2.79
		Guar bean, Soybean	1.31	-1.59
		Potato, goji berry	1.40	-1.07
		Wheats	1.14	-0.97
		Corn	2.03	-0.77
		Black pepper	1.03	0.32
		Grapes	1.22	0.64
		Spinach	0.68	0.81
		Carrot and parsley family, cumin	2.43	0.88
		Cabbage family	0.49	1.01
		Cabbage family.1	0.31	1.66
		Strawberry, Bramble berries	1.52	1.77
		Lettuce	1.42	2.09
		Capsicum	1.22	-
		Banana	1.21	-
		Name	PC1	PC2
16S rRNA	Enrolled participants	Faecalibacterium prausnitzii	-	-2.47
		Faecalibacterium	-	-1.96
		Faecalibacterium prausnitzii.1	0.85	-1.78
		Eubacterium coprostanoligenes group	-	-1.72

	UCG-002	-	-1.63
	Bacteroides uniformis	1.18	-1.16
	Bifidobacterium	-1.25	-1.08
	Prevotella copri	-3.37	-0.86
	Dialister	-1.62	0.46
	Ruminococcus torques group	1.52	0.58
	Ruminococcus gnavus group	1.46	0.84
	Bifidobacterium	2.02	-
	Subdoligranulum	-1.64	-
	Prevotella_9	-2.06	-
	Prevotella_9	-2.43	-
<p><i>Note:</i> Top 15 factor loadings for each PC were selected after implementation of a threshold of 0.3 (lower than this threshold is denoted by a dash: “-”). Values under PC1 and PC2 indicate directionality of that loading on the respective PC. Factor loadings were sorted by PC2. Positive values are depicted in green and negative in red.</p> <p>Abbreviations: PCA, principal component analysis; POMMS, Pediatric Obesity Microbiome and Metabolism Study; UCG, uncultured genus-level group.</p>			

Assuming improved robustness and interpretability of PCs with increased sample size, we integrated samples with a larger dataset of children and adolescents with obesity (i.e., POMMS; see Methods and Appendix: Fig. 10c) drawn from similar households and recruited from the same clinic (52). This integration identified three major dietary patterns explaining 22.9% of dietary intake variation (PC1 = 8.5%, PC2 = 7.7%, and PC3 = 6.7%; Appendix: Fig. 10d). PC1 (hereafter referred to as “total food pattern” for simplicity) was driven by “cacao,” “guar bean, soybean,” “potato, goji berry,” “wheats,” “corn,” “black pepper,” “grapes,” “spinach,” “carrot and parsley family, cumin,” “cabbage family” (two variants), “strawberry, brambleberries,” “lettuce,” “capsicum,” and “banana,” all in the positive direction. For PC2, the positive end was driven by “black pepper,” “grapes,” “spinach,” “carrot and parsley family, cumin,” “cabbage family” (two variants), “strawberry, brambleberries,” and “lettuce” and the negative end by “cacao,” “guar bean, soybean,” “potato, goji berry,” “wheats,” and “corn” (hereafter referred to as “prudent to Western pattern” for simplicity; designation of leafy greens as “prudent” and snacks and energy-dense foods as “Western” in factor analysis results of previous literature) (67). Further analyses did not include PC3.

Using the larger POMMS enrollment sample set from participants with obesity (n = 195) as the reference, significant dietary shifts were observed at the enrollment and treatment time points for usual care and intervention (Table 4). Usual care enrollment was strongly shifted toward the prudent end of the prudent to Western pattern ($\beta = 7.25$, 95% CI: 3.44 to 11.12; $p < 0.001$). However, during the treatment period, it shifted toward the Western end, although still distinct from the reference ($\beta = 4.79$, 95% CI: 1.19 to 8.38; $p = 0.013$; Table 5). In contrast, the intervention's potential deviation toward the prudent end at enrollment ($\beta = 4.47$, 95% CI: -0.03 to 8.96; $p = 0.059$) became significant during the treatment period ($\beta = 8.64$, 95% CI: 5.18 to 12.14; $p < 0.001$; Table 5). No changes were identified on the total food pattern for either group when compared with the POMMS enrollment reference (Table 5).

Table 5: Estimated coefficients from linear mixed effects models with random intercept adjusted for appropriate covariates on dietary and microbiome patterns.

Biomarker	Sample set	Comparison	Group	PC	beta (95% CI)	p value	
FoodSeq	Enrolled participants	a Change from enrollment to treatment	Usual care	1	0.69 (-1.16, 2.74)	0.500	
				2	-1.12 (-3.06, 0.87)	0.269	
			Intervention	1	1.22 (-0.68, 3.33)	0.240	
				2	1.83 (-0.66, 4.37)	0.163	
	External Integration	b Difference between POMMS enrollment and CHOICE usual care	Enrollment	1	1.01 (-3.36, 5.40)	0.658	
				1	1.09 (-3.04, 5.22)	0.620	
			Enrollment	2	7.25 (3.44, 11.12)	<0.001	
				Treatment	2	4.79 (1.19, 8.38)	0.013
			b Difference between POMMS enrollment and CHOICE intervention	Enrollment	1	-2.28 (-7.48, 2.91)	0.404
					1	-0.33 (-4.31, 2.84)	0.876
				Enrollment	2	4.47 (-0.03, 8.96)	0.059
					Treatment	2	8.64 (5.18, 12.14)
16S rRNA	Enrolled participants	b Change from enrollment to treatment	Usual care	1	-0.57 (-1.94, 0.80)	0.422	
				2	1.05 (-0.86, 2.96)	0.288	
			Intervention	1	0.60 (-1.03, 2.20)	0.471	
				2	-1.90 (-4.01, 0.31)	0.088	
<i>Note.</i> All comparisons included age, race and ethnicity, and sex as covariates. Significant <i>p</i> values are bolded.							

Abbreviations: CHOICE, Childhood Obesity Microbiome Study; PC, principal component; POMMS, Pediatric Obesity Microbiome and Metabolism Study.

^a Additionally included sequencing batch.

^b Additionally included PCR amplification batch.

The total food pattern was strongly associated with pFR (dietary diversity) in the combined sample set ($\beta = 1.03$; $p < 0.001$; Appendix A: Fig. 11). There was no shift in pFR from enrollment to treatment for either usual care or intervention (Appendix A: Fig. 11).

2.2.3.2 16S rRNA microbiome composition

Two PCs explained 15.9% of microbiome composition variation (PC1 = 9.0% and PC2 = 6.9%; Appendix: Fig. 10e,f). Intervention showed a potential negative shift on PC2 during the treatment period compared with enrollment ($\beta = -1.9$, 95% CI: -4.01 to 0.31 ; $p = 0.088$; Table 5), aligning with an increase in bacterial taxa such as *Faecalibacterium* and *F. prausnitzii* strains, *Eubacterium coprostanoligenes* group, an *Oscillospiraceae* uncultured genus-level group-002 (UCG-002) strain, *Bacteroides uniformis*, *Bifidobacterium*, and *Prevotella copri* (Table 4). Bacterial taxa driving variation along PC1 (Table 4) did not differ between intervention enrollment and treatment periods (Table 5). Neither PC1 nor PC2 was different between usual care enrollment and treatment periods (Table 5). Microbiome alpha-diversity did not change between enrollment and treatment across four different metrics of diversity for either group.

2.2.3.3 Diet quality at follow-up, change in weight, and change in food-insecurity status

DSQ scores that decreased as a result of treatment were maintained at follow-up (i.e., DSQ_{sug} and DSQ_{ssb} in usual care and DSQ_{sug}, DSQ_{ssb}, DSQ_{dairy}, and DSQ_{calc} in intervention; Tables 2 and 3). Increases in DSQ scores as a result of treatment reverted to baseline by the follow-up time point (Table 2).

Neither group showed within-group changes in weight from baseline to Week 4 (Table 2), and there was no impact of intervention over usual care (Table 3). Within intervention, three participants changed from food insecure to secure, whereas one participant changed from food secure to insecure

in usual care by Week 4 (Table 2; Appendix A: Fig. 12). However, the intervention overall did not predict food-insecurity status at Week 4 (Table 3; Appendix A: Fig. 12). Additionally, the intervention was not predictive of changes in weight from baseline to follow-up or of food-insecurity status at Week 8 (Table 3; Appendix A: Fig. 12).

2.3 Discussion

Childhood obesity and household food-insecurity frequently co-occur and disproportionately affect ethnic minority populations in the United States (57,68). We tested whether increased access to healthy foods through food provisioning could improve diet quality, weight, food-insecurity, and microbiome outcomes in an ethnically representative pilot cohort of children with obesity from a food-insecure context (36.4% of our cohort vs. the US national average of 10.2% of households reporting food insecurity in 2021) (57). Dietitian-guided food provisioning, in addition to IHBLT, shifted all DSQ scores in line with the protocol and was superior to usual care in increasing whole grain and fiber intake and reducing dairy intake. Furthermore, these dietary shifts did not require restriction or more complexity in diet variety as shown by our genomic measure of dietary diversity (i.e., pFR), potentially translating into accessible behavior changes for patients and higher compliance with recommendations. Treatment-driven reductions in dietary components were also durable post treatment, but increases in dietary components were lost, suggesting a dependence on ongoing food provisioning. Similarly, weight, food-insecurity status, and microbiome outcomes did not improve. Our results build on the success of the “Food is Medicine” approach (medically tailored food prescription to improve patient health) (66,69) in adult cohorts (69), demonstrating the viability of this approach in a pediatric setting.

To our knowledge, this is the first prospective study to incorporate objective genomic dietary assessment to mitigate biases from self-reported dietary assessments (38). The strong concordance between ordered grocery foods and corresponding plant taxa identified suggested that participants consumed dietitian-recommended foods and demonstrated FoodSeq's potential as a measure of dietary compliance. FoodSeq also empirically captured variation along a “prudent to Western” dietary pattern, representing an axis of energy density, with fruit and vegetable taxa (70) on the prudent end and staple

crops (e.g., corn, wheat, potato) (71) and plants that tend to be highly processed such as cacao and soybean/guar bean (71–73) on the Western end. Although fruits can be consumed in processed forms, the co-occurrence with vegetable taxa suggested minimal processing. In contrast, the co-occurrence of soybean with cacao, the primary ingredient of chocolate, suggested consumption of this taxa in more processed forms (e.g., vegetable oil (72) or soy lecithin (74)). Alternatively, these taxa could represent guar bean (our DNA marker does not resolve differences between soybean and guar bean), the gum of which is a common food additive in processed foods (73).

Although the robust decrease in dairy consumption aligned with our microbiome-friendly protocol, the concurrent sustained decrease in calcium contrasted with recommended guidelines of 1000 mg/day of calcium for children aged 6 to 8 years and 1300 mg/day for children aged 9 to 11 years (66). Although developing a calcium deficiency during the 4-week study duration was unlikely, dairy remains a primary source of calcium for growing children. In our study specifically, continued treatment at the Healthy Lifestyles Clinic after study completion would address any calcium-intake concerns. However, in general practice, long-term dairy reduction could lead to calcium depletion, causing decreased peak bone mass, dental and growth problems, and neuromuscular symptoms (75). Given that our protocol aimed for beneficial impacts on the microbiome but our microbiome results did not indicate such changes, we would not recommend dairy reduction in future microbiome-conscious dietary interventions.

The study's limitations include a modest sample size and brief duration, which may have hindered detectability of weight, food-insecurity, and microbiome changes despite significant dietary impacts (particularly on dairy consumption). For example, the American Academy of Pediatrics recommends at least 26 h of IHBLT over a period of 3 to 12 months (24) for obesity treatment, a period that is considerably longer than our study. The Hunger Vital Sign (which assesses food insecurity over the past year) was also not adjusted for the study's duration, leading to potential repeating of enrollment responses from participants at Weeks 4 and 8. Alternatively, this degree of food provisioning on a short time scale might not alleviate the psychological burden of food insecurity.

Similarly, our current dietary protocol might be insufficient to produce a microbiome shift compared with the drastic dietary changes previously shown to induce rapid microbiome changes (64). The brief duration of our study also precluded assessment of the long-term viability of dietary changes or the scalability of \$100 grocery provisions per week at a population level. However, the focus of our study was to assess the efficacy of direct food provisioning in improving diet, weight, microbiome, and food-insecurity outcomes instead of aiming to model a food-benefit program. Future studies could scale these components in cohort size and study duration to improve detectability of outcomes and identify factors for long-term sustainability. Although there was a possibility of increased caloric intake due to excess leftover household finances in the intervention, this was unlikely given the DSQ score and FoodSeq shifts demonstrated in this group toward low-energy-density, fiber-rich foods, which are, in fact, known to lead to reductions in caloric intake (70). Future studies could directly investigate household purchasing behavior shifts due to food provisioning. The scarcity of stool samples for Week 8 from the intervention group could not be definitively explained because we did not collect participant reasons for sample non-submission. Future studies could investigate participant study experiences and stool sampling feasibility in greater detail. Finally, FoodSeq analyses do not provide data on animal components of diet due to the trnL-P6 marker's specificity to plants, although similar markers for animal-based foods are being adapted for dietary assessment in humans (45).

Collectively, our results highlight the potential of a dietitian-guided food-provisioning intervention for improving diet quality in children with obesity. Concurrently, we answer the call for objective biomarkers of diet (42) through FoodSeq and provide a road map for using genomic tools to assess the impact of dietary interventions.

2.4 Methods

Methods and procedures received Institutional Review Board approval at Duke University (protocol #00107438). This study was registered at ClinicalTrials.gov (NCT04770311). A completed Consolidated Standards of Reporting Trials (CONSORT) checklist and extended methods are provided in online Supporting Information.

2.4.1 Recruitment

Children aged 6 to 11 years ($n = 59$) with BMI (weight in kilograms divided by height in meters squared) ≥ 95 th percentile for age and sex and whose caregivers could receive home-delivered groceries or had access to a nearby Walmart for pickup were recruited by a clinical research coordinator before, during, or after their visit to the Duke Children's Healthy Lifestyles Clinic (a comprehensive child obesity treatment program) in Durham, North Carolina (52) between April and November 2021. Flyers were posted in the clinic for study promotion. Exclusion criteria were as follows: 1) obesity from underlying conditions; 2) current or planned use of weight loss medications, including steroids, antipsychotics, or antidepressants, in the next 4 weeks; 3) initiation of stimulant medication in the past 3 months; 4) current or recent antibiotic use in the past 4 weeks; 5) participation in the pharmacotherapy or weight loss surgery track at the Healthy Lifestyles Clinic; or 6) weight loss exceeding 5% in the past 6 months. Child weight >400 lb was not an exclusion criterion as a weight of 400 lb was considered implausible for a child between 6-11 years old. The inability to receive deliveries or pickup groceries was unlikely to have excluded participants with extreme food insecurity as participants were only recruited from the Healthy Lifestyles Clinic (requiring some form of transport to get to clinic) and none of the caregivers indicated an inability to pick up groceries due to transportation barriers. In fact, caregivers heavily favored pickup over deliveries due to concerns of theft.

Caregivers consented on behalf of children (in English/Spanish) electronically on Research Electronic Data Capture (REDCap, a secure, web-based software platform) (76) in a private space in the Children's Healthy Lifestyles Clinic; children provided verbal assent if they were < 12 years old and written assent if they were ≥ 12 years old; however, as only children up to 11 years old were recruited, no written assent was obtained. In case caregivers or potential participants expressed discomfort staying in the clinic longer than their intended visit or required more time, they had the option to consent afterwards over phone. Participants could also ask study-related questions during the enrollment visit or later through phone or email to be answered by the study team.

As a pilot study, sample size was primarily informed by recruitment practicalities. A sample size of 40 was determined to provide 82% power ($\alpha = 0.05$) to detect a two-unit diet quality change between groups. A 2-unit diet quality change as chosen as the desired effect size given the meaningful change it could represent across DSQ dietary components (e.g. a 2-unit increase in wholegrains corresponds to +2 ounces of wholegrains or a 2-unit decrease in added sugars corresponds to -2 teaspoons of added sugars) in bringing children closer to recommended daily guidelines (66). Exploratory inferential tests were conducted and will provide data on feasibility and effect size variability estimates for a larger trial.

2.4.2 Study Design

Participants were randomized (simple randomization) concurrently with enrollment in REDCap (76) to either usual care consisting of 4 weeks of IHBLT or intervention, in which, in addition to IHBLT, participants received weekly groceries worth \$25 per person (up to \$100 per family) chosen under guidance from a registered dietitian (RD) for 4 weeks. Participants and study personnel were not blinded given the logistics of food provisioning and the measures collected. A total of 24 participants were randomized to usual care, and 26 were randomized to the intervention group (Fig. 1a). Descriptive statistics for participants who were lost to follow-up or discontinued are provided in Table 6 (Appendix A).

The study consisted of 2 consecutive enrollment weeks followed by 4 consecutive treatment weeks and a follow-up in the fourth week after the end of treatment (Fig. 1b). Week 8 (follow-up) measures were collected between June 2021 and February 2022. Baseline measures included the DSQ (self-administered) (65), demographics, food-insecurity status (Hunger Vital Sign) (77), height, weight, blood pressure, lipid panel, glucose measurements, insulin levels, and liver enzymes.

Caregivers reported child weight, measured using study-provided digital scales (manufacturer: SmarTake, Amazon Standard Identification Number: B08CSBWHC7; accuracy of 0.2 lb/0.09 kg; weight measurement range between 11 lb – 400 lb (5 kg – 180 kg); were not connected to an app), during weekly phone check-ins (due to COVID-19 restrictions) to be recorded in REDCap (76) by

study staff. All caregiver reported values were inspected for credibility by study staff. The DSQ and Hunger Vital Sign were repeated at Week 4 and Week 8 (Fig. 1b). Stool samples were collected weekly (Fig. 1b); caregivers could consent to sending in their own stool samples as well. Participants in both groups received \$10 in compensation for the enrollment visit and \$10 per received stool sample (in addition to any treatment-related compensation).

2.4.3 Treatment

The usual care participants received weekly IHBLT (24) visits at the Duke Healthy Lifestyles Clinic (HLC) which is a specialized weight management program for children and teenagers aged 2-18 years old that has provided care to over 15,000 children to date and accommodates 800 new patients every year (78). Treatment choices, including lifestyle changes, diet modifications, medications, or surgery, are decided through a collaborative process involving the child, their parents, and healthcare providers, considering the child's age, severity of obesity, associated health conditions, and family preferences (78). The clinic's dietary advice is provided through a motivational interview counselling style and emphasizes reducing intake of sugary drinks and processed foods while promoting the consumption of vegetables, whole grains, and healthy fats, tailored to individual and family needs (78). Children enrolled in the clinic continue treatment until they either reach a BMI < 95th percentile for age and gender or until they become 23 years old. Although these participants initially did not receive monetary compensation, a protocol modification offering \$50 per clinic visit was implemented to promote equity with the intervention after one participant withdrew upon assignment to usual care instead of the intervention. Only one other participant had been randomized to usual care before this protocol modification and this participant was lost to follow-up. All remaining usual care participants were recruited after this modification. As such, a descriptive statistics analysis comparing participants before and after this modification was not conducted. All analyzed usual care participants were enrolled after this modification and received the same compensation.

For the intervention, participants' primary caregivers had weekly video calls with the RD to order groceries worth \$25 per person (up to \$100 per family) from Walmart for home delivery or

pickup. Caregivers could flexibly choose from foods aligning with the CHILD-1 diet (79), an evidence-based adaptation of the Dietary Approaches to Stop Hypertension (DASH) (80) diet, which was designed for children with obesity. Specifically, the diet promoted increased fiber intake from vegetables, fruits, whole grains, nuts and seeds, and fermented foods. Processed meat, processed carbohydrates/sugary foods, foods high in sodium (e.g., frozen convenience meals), and fast food were limited. Restricting animal dairy was the only modification to the CHILD-1 diet in our study. This restriction, combined with the higher fiber emphasis of the CHILD-1 diet, was chosen for microbiome-promoting effects (81), particularly given the emerging role of dairy fats and proteins in gut inflammation and dysbiosis (64,82,83). Caregivers could substitute vegan alternatives for dairy where requested. The RD did not actively prescribe dairy alternatives to compensate for participant calcium requirements by age because a calcium deficiency was not expected to emerge within the 4-week duration of this intervention, and any emergent calcium-intake concerns would be addressed through continued treatment at the Healthy Lifestyles Clinic after study completion. Caregivers were encouraged to order groceries at their assigned support level, however, they could order less than the maximum amount if desired. At the start of each grocery ordering call, the RD reminded caregivers of encouraged and discouraged foods and prompted them with what meals they had planned for breakfast, lunch, dinner etc. The RD used caregiver responses to add resulting relevant ingredients to the online cart, switching items to align with the dietary protocol as needed (e.g. whole wheat bread instead of white bread). Recommendations were personalized to family food preferences. The RD also suggested new diet compliant foods if caregivers requested recommendations. However, on rare occasions when a caregiver insisted on items not on the approved list (e.g. cheese)— despite being offered alternatives and discussions about dietary guidelines— the dietitian acquiesced, striving to keep these instances infrequent (Appendix A: Fig. 7).

2.4.4 Stool collection and DNA extraction

Caregivers received Microsetta Stool Collection kits (84) (stool-catching paper, collection tube with 95% ethanol as preservative, and instructions) with prepaid FedEx return labels for stool

collection. Caregivers obtained nickel-sized stool samples from participants (often collected by children themselves) and labeled tubes with participant identifier, date of collection, and week number. Upon receipt of stool samples from caregivers, samples were stored at 4°C until removal of ethanol and subsequent storage at -80°C (within 4 weeks from collection). Prior to DNA extraction, each stool sample was homogenized to obtain aliquots representative of the whole sample. The first 23 samples were manually homogenized in the collection tube using tongue depressors. Homogenates were divided into 1 mL aliquots, centrifuged at 10 kRPM for 3 min, and the ethanol was aspirated out. Aliquots were then weighed and stored at -80°C. DNA was extracted from one aliquot per sample (mass < 250 mg) using the DNeasy PowerSoil Pro Kit (QIAGEN, Hilden, Germany). Samples did not require randomization as extraction was conducted in tubes instead of a plate. After the first 23 samples, the homogenization protocol was amended for efficiency. Samples were transferred to a 50 ml conical tube and centrifuged at 3000 RCF for 15 minutes. After aspirating the ethanol, the sample was weighed and stored at -80°C. Thawed samples were homogenized by converting to 10% (weight/volume) Phosphate Buffer Solution (PBS) fecal slurries and aliquoted out in 200 ml volumes for DNA extraction. The whole sample slurry and aliquots were then stored at -80°C. Samples were randomized and extracted using the MagAttract PowerSoil kit (QIAGEN, Hilden, Germany). As samples from each homogenization process were sequenced in separate runs, any differences resulting from homogenization were accounted for by including sequencing batch as a covariate where appropriate in the statistical analyses. A subset of 24 samples was processed using the PowerSoil Pro kit due to reagent limitations. While the choice of extraction kit impacts extracted DNA concentrations, subsequent equimolar pooling of samples during library preparation mitigated these differences.

2.4.5 Primary outcome

Changes in dietary intake from baseline (Enrollment Week 1) to end of treatment (Week 4) were measured within and between treatment groups using caregiver responses (with occasional clarification from children) to the DSQ (self-administered) (65). The DSQ is a 26-item dietary recall

inventory (filled out on paper or online) developed using the 2009-2010 National Health and Nutrition Examination Survey (NHANES) 24-h dietary recall data (65). The DSQ assesses intake frequency of fiber (in grams); calcium (in milligrams); added sugars (teaspoon equivalents); whole grains (ounce equivalents); dairy (cup equivalents); vegetables, fruits, and legumes (cup equivalents); red meat (times per day); and processed meat (times per day) over the past 30 days using up to 11 frequency categories ranging from “never” to “6 or more times per day” (85). Individual item responses are converted using National Cancer Institute-provided scoring algorithms to portion size estimates using NHANES 24-h dietary recall validated age- and sex-adjusted regression models (86). We interpreted increases or decreases in specific dietary components that aligned more closely with USDA dietary guidelines (66) as indicators of improved diet quality. High-quality diets typically score higher in fruits, vegetables, whole grains, and other beneficial components and lower in added sugars and red meat. The DSQ was ideal for our cohort because it has been deemed an acceptable tool for measuring key obesity-related dietary behaviors in children (87) and can yield quantified intake estimates, unlike most dietary screeners (65).

2.4.6 Secondary outcomes

2.4.6.1 Diet quality at follow-up, change in weight, and change in food-insecurity status

Changes in DSQ scores from baseline (Enrollment Week 1) to follow-up (Week 8) were assessed within and between treatment groups to assess longevity of dietary outcomes. Changes in weight from baseline to endpoint (Week 4) and baseline to follow-up were compared within and between the treatment groups.

Household food-insecurity status, reported by caregivers using the Hunger Vital Sign (77), was recorded for baseline, endpoint, and follow-up and compared within and between groups for the latter two time points. The Hunger Vital Sign, developed in 2010 by the Children’s Health Watch (77), consists of two validated questions based on the US Household Food Security Survey Module (88) to screen households at risk of food insecurity (77). Participants are asked to rate the following statements as “never true”, “sometimes true”, or “often true”: 1) “Within the past 12 months we worried whether

our food would run out before we got money to buy more” and 2) “Within the past 12 months the food we bought just didn’t last and we didn’t have money to get more”. Participants are classified as food insecure if the response to either one or both statements is “sometimes true” or often true” (77). The Hunger Vital Sign is validated for food-insecurity assessment in children (89).

2.4.6.2 FoodSeq

Residual dietary plant DNA in stool samples was assessed following a previously described procedure (46) to objectively evaluate treatment-driven dietary pattern changes. TrnL-P6 DNA from extracted stool DNA was first amplified in 10 µl primary PCR reactions using KAPA HiFi HotStart PCR Kit (KAPA Biosystems) using previously listed reaction components and volumes using the following cycling conditions: initial denaturation at 95 °C for 3 min, succeeded by 35 cycles of 98 °C for 20 s, 63 °C for 15 s, and 72 °C for 15 s. PCR reaction batches further included a no template control and a positive control using DNA from an *Ilex* species (collected from the wild in Durham, North Carolina, USA and identified in our data as *Ilex paraguariensis*) which were used to assess whether the whole batch could be advanced to the secondary round of PCR reactions. Primary PCR products were diluted 1:100 and transferred to secondary 10-cycle PCR reactions, conducted in 50 µl reactions using previously described reaction components and volumes to tag samples with unique 8-bp barcode combinations and Illumina adapters. When pooling resulting cleaned amplicons, a maximum volume cap was set at 20 µl to 25 µl for low concentration samples. Resulting libraries were concentrated and quantified by qPCR and fluorimeter. Libraries were run on the Illumina MiniSeq platform in three batches with paired-end sequencing using 150-cycle High or 300-cycle Mid Illumina Kits as appropriate. Libraries for sequencing contained 30% PhiX (Illumina) spike-ins to add nucleotide diversity. Sequencing results were demultiplexed and filtered and trimmed using a QIIME2 bioinformatic pipeline (<https://qiime2.org/>) (46), and species identities were assigned using a custom dietary plants reference database (46) and the `assignSpecies` function in DADA2 version 1.22.0 (<https://benjjneb.github.io/dada2/>). During taxonomic assignment using `assignSpecies` from DADA2, ASVs with multiple matches were assigned to a representative last common ancestor taxon

of all matched taxa. Further quality control was performed by a) analyzing the number of reads at each step of the pipeline, b) verifying the percentage of phi-X spike in, c) merging amplicon sequence variants (ASV) such that ASVs that were complete subsets of each other were merged into one, and d) comparing read counts between samples, positive, and no template controls so that contaminants could be identified. Contaminants were also screened using decontam v1.8.09 and subsequently removed.

While it was possible that changed dietary habits could produce variation in bowel habits and thus impact FoodSeq results, this was unlikely to be a significant concern. Participants were allowed a week to provide each corresponding stool sample, making no bowel movements in this duration unlikely. Similarly, textural variances were accounted for by homogenizing and normalizing stool samples by weight before DNA extraction and normalizing by DNA concentration before sequencing.

2.4.6.3 Microbiome 16S ribosomal RNA sequencing

Microbiota composition was determined using 16S ribosomal RNA (rRNA) gene amplicon sequencing following previously described procedures (90). DNA was amplified using custom bar-coded 16S rRNA primers targeting the 16S V4 rRNA gene (91) using published procedures (92). The bar-coded library was sequenced on the Illumina MiniSeq platform with paired-end 150-base pair (bp) reads. Primer trimming was conducted using Trimmomatic version 0.39 (<https://github.com/usadellab/Trimmomatic>) without quality filtering (33). QIIME2 was used for demultiplexing (90), and sequence variant identification and taxonomy assignment were conducted using a DADA2 analysis pipeline (90) in conjunction with version 138.1 of the Silva database (93).

2.4.7 Statistical analyses

All analyses were conducted in R (version 4.1.3, R Project for Statistical Computing). Changes in the DSQ and weight were assessed using Wilcoxon rank sum test (nonparametric) due to small samples sizes in each group. A χ^2 test was used to compare food-insecurity outcomes. Impact of the intervention on DSQ and weight outcomes over usual care was assessed through linear regression and, for food-insecurity status, using logistic regression. Models were adjusted for baseline, and diagnostic plots (e.g., quantile-quantile [QQ] plot) were used to assess normality of the residuals (residuals

generally followed the reference line with minor deviations near the tails). Means, medians, and interquartile ranges (IQR) of outcomes were reported together to enable comprehensive understanding of the data.

Sequencing data were handled as “phyloseq” objects with the “Phyloseq” R package (version 1.38.0, R Project for Statistical Computing). Data preprocessing included removing samples from adults (n= 62), participants who did not complete the study (n= 3), duplicates (i.e. samples submitted for the same week timepoint, n= 8), from week timepoints not being assessed in the study (e.g. week 5 or 7, n= 4) and samples with illegible labels where the week timepoint could not be confirmed by the participant (n= 1, from intervention enrollment week 1 bringing total analyzed samples for this timepoint in the intervention group to 15), leading to the 180 samples analyzed in this study. Data processing proceeded to filtering unidentified and low representation taxa, and center-log ratio-transforming (Clr) (microbiome package, version 1.24.0) read counts such that the read counts of a taxon in a sample were described relative to the geometric mean of the read counts all the taxa in the sample to account for compositionality (94). For FoodSeq, taxa that could not be identified at least to the super kingdom level were filtered out after transformation. This filtered set was also used for pFR calculation. For 16S rRNA, samples with fewer than 5,000 read counts were filtered out, along with microbial taxa appearing less than 3 times across at least 10 percent of samples following previously published protocols¹¹ before Clr transformation. Enrollment Week 1 and 2 samples were labeled “Enrollment,” Weeks 1 to 4 as “Treatment,” and Week 8 as “Follow-up.”

FoodSeq-detected plant taxa from the intervention group were compared against foods ordered in the groceries using linear regression. Frequency of distinct food ingredients were compared with the frequencies at which FoodSeq taxa corresponding to those ingredients were detected across stool samples. Foods mapping multiple FoodSeq taxa were counted for each applicable taxon, and taxa mapping to multiple foods counted for each applicable food (indicated as agglomerated in Fig. 2). Items that were “unavailable” or “canceled” were removed prior to analysis.

Diet and microbiome composition were analyzed using principal component analysis (PCA, “prcomp” function, parameters: centered, not scaled; full list of plant taxa mapping to PCA factor loadings in Appendix A: Table 10). Principal Component Analysis (PCA) was utilized as it is a well-established technique for dimensionality reduction and analysis of sequencing data (especially for addressing the challenges of sparsity and compositionality) (94) as well as for empirical derivation of dietary patterns from dietary recall data (95). PCA scree plots were used to identify number of Principal Components (PCs) describing most of the variation in the data (and as such presenting major dietary patterns). PCA factor loadings were identified using a threshold of $\geq |0.3|$ to obtain taxa that were primary contributors to a given PC (96). Factor loadings represent the coefficients of the linear combinations of original features used to construct PCs and identify the contribution of original features to these PCs. Factor loadings were relabeled with representative short-form common plant names for interpretability. In interpreting PCA results, the proximity of points on a PCA plot (Appendix A: Fig. 10) indicates the similarity between observations: points that are close together represent observations with similar characteristics (in the case of FoodSeq and 16S RNA sequencing, similar dietary and microbiome compositions respectively) and vice versa for points distant from each other (97). Additionally, the direction of factor loadings within the PCA reveals the nature of correlations between variables; factor loadings that point in the same direction suggest a positive correlation, meaning these variables tend to increase or decrease together (97). Conversely, factor loadings in opposite directions indicate a negative correlation, where an increase in one variable is associated with a decrease in another (97).

pFR was calculated as the number of unique taxa with at least one read count in each sample, as previously described for plant Metabarcoding Richness (46). Microbiome-diversity was calculated using the “estimate_richness” function (methods: “Observed,” “Shannon,” “InvSimpson,” “Fisher”) on the phyloseq object. Linear mixed-effects models (“lmer” function from “lme4” [version 1.1.34] and “lmerTest” [version 3.1.3, R Project for Statistical Computing]) were used to evaluate dietary pattern, microbiome, and pFR shifts as a function of the treatment time point, accounting for repeated

measures and relevant covariates. Because outliers and dataset-specific characteristics can disproportionately affect PCA on small datasets (98), we used the “merge_phyloseq” function and the “collapseNoMismatch” function of the “dada2” R package (version 1.22.0, R Project for Statistical Computing) to integrate FoodSeq data from the current cohort (i.e., CHOICE) with an external cohort (i.e., POMMS) (18) in a post hoc approach to increase sample size and stabilize PCA results, improving principal component (PC) robustness and interpretability (98). The POMMS cohort was chosen for the integrated cohort comparison as these participants were also recruited from the Duke Healthy Lifestyles clinic in Durham, North Carolina and had similar demographic and household backgrounds as CHOICE (52). To integrate the datasets, unfiltered and untransformed FoodSeq data from the 180 CHOICE samples included in this study was first combined with unprocessed POMMS FoodSeq data (all samples, n= 384) using the 'merge_phyloseq' function and the 'collapseNoMismatch' function of the 'dada2' R package (version 1.22.0). This step alone did not impact any statistical analyses as it simply combined the two datasets into one. All further preprocessing (such as taxa filtration) and data transformation on this combined dataset proceeded this initial step. PCA was conducted on this combined dataset and only enrollment samples from POMMS participants in the intervention group (n = 195, BMI (kg/m²) ≥ 95th percentile for age and sex) (52) were used as the reference to which CHOICE treatment groups at Enrollment and Treatment timepoints were compared to ensure appropriate comparability given participants obesity status and the timepoint preceding commencement of any study interventions (IHBLT, Low-carb diet, Drugs, Surgery) (52). The follow-up time point was not included in linear mixed-effects models for either the enrolled cohort or the POMMS integration due to insufficient samples available from the intervention group (Fig. 1b).

The following model calls were used for the linear mixed effects models:

CHOICE FoodSeq:

lmer(PC1 ~ *timepoint* + *seq_date* + *age* + *race* + *ethnicity* + *gender* + (1 | *id*), *data* = *subset*)

lmer(PC2 ~ *timepoint* + *seq_date* + *age* + *race* + *ethnicity* + *gender* + (1 | *id*), *data* = *subset*)

$lmer(pFR \sim timepoint + seq_date + age + race + ethnicity + gender + (1 | id), data = subset)$ where subset is the dataset filtered down to the appropriate treatment group and timepoints and seq_date is the sequencing batch.

CHOICE 16S rRNA:

$lmer(PC1 \sim timepoint + pcr_batch_16s + age + race + ethnicity + gender + (1 | id), data = subset)$

$lmer(PC2 \sim timepoint + pcr_batch_16s + age + race + ethnicity + gender + (1 | id), data = subset)$

$lmer(\alpha\text{-diversity measure} \sim timepoint + pcr_batch_16s + age + race + ethnicity + gender + (1 | id), data = subset)$

where subset is the dataset filtered down to the appropriate treatment group and timepoints and pcr_batch_16s is the PCR amplification batch.

Cross-cohort FoodSeq:

$lmer(PC1 \sim cohort + pcr_batch + age + race + ethnicity + gender + (1 | id), data = subset)$

$lmer(PC2 \sim cohort + pcr_batch + age + race + ethnicity + gender + (1 | id), data = subset)$ where subset is the dataset filtered down to the appropriate treatment and timepoint groups and pcr_batch is the PCR amplification batch.

2.4.8 Data availability

Deidentified data, data dictionaries, and code notebooks for reproducing manuscript results from processed trnL-P6 and 16S ribosomal RNA data can be provided upon request and will be shared in compliance with relevant study agreements, ethical standards, and regulations. The bioinformatic pipeline and reference database for analyzing raw trnL-P6 sequencing data, as well as an example dataset and tutorial, are available on Zenodo (<https://zenodo.org/record/8004348>).

3. Stool Genomics for Enhanced Monitoring of Nutrition Interventions in Childhood Malnutrition

3.1 Introduction

Childhood malnutrition – wasting, stunting or underweight – remains a major global health problem, affecting over 194 million children in 2022 (3) and accounting for almost half of all deaths in children under five across the globe (3). Malnutrition can be driven by a combination of socio-economic factors, environmental and biologic factors such as wealth (99), food access (100), hygiene & sanitation (101), gut microbial dysbiosis (102) and intestinal diseases like environmental enteric dysfunction (EED) (103). Low- and middle-income countries (LMICs) bear a disproportionate share of this burden, facing long-term consequences such as increased childhood mortality and morbidity, impaired cognitive development, reduced adult work capacity, and heightened risk of metabolic diseases later in life (3,104).

Global treatment regimens over the past decade have transitioned towards Community-Based Management of Acute Malnutrition (CMAM), moving away from inpatient care that often requires extensive hospital stays and significant human and financial resources (105,106). The CMAM model deploys community health workers (CHWs) to screen and enroll children with malnutrition in an outpatient treatment regimen. For severe acute malnutrition (SAM; weight-for-height z-score (WHZ) below -3 standard deviations (SD) from the median of the WHO growth standards, mid-upper arm circumference (MUAC) <115 mm, or the presence of nutritional edema (15)), this typically includes a weekly weight-adjusted prescription of ready-to-use therapeutic food (RUTF) rations and systematic antibiotic treatment (105–107). In contrast, moderate acute malnutrition (MAM; WHZ between -2 and -3 SD or a MUAC between 115–125 mm, without nutritional edema (15)) is generally managed with ready-to-use supplemental foods (RUSF) at lower doses and prescribes an antibiotic regimen only in the case of underlying concomitant infection (108). RUTFs, formulated as a single-dose foil package of an energy-nutrient dense paste, can be eaten directly from the packet without preparation (reducing contamination risk) and are shelf-stable without refrigeration for up to two years (109).

Adherence to RUTF/RUSF therapy is crucial, as insufficient intake diminishes the likelihood of adequate weight gain (110). However, socio-economic factors such as household size, wealth, and food insecurity, as well as dietary practices, can significantly impact treatment. For example, studies have documented the sharing of RUTFs among household members in food-insecure settings or large families (111), with some even reporting instances where families sold RUTFs to supplement their income and address food shortages (112). Furthermore, adherence to RUTF/RUSF therapy alone may not guarantee recovery if other dietary practices interfere, such as the consumption of age-inappropriate foods like tea, which can suppress appetite and inhibit iron absorption (113). However, monitoring compliance with RUTF therapy within the CMAM framework presents a substantial challenge, requiring regular home-visits by CHWs and/or collection of empty RUTF packets by caregivers (105,106,110,114,115). This approach demands considerable human resources and extensive training to conduct effective home-based monitoring, not to mention the additional effort to adapt food assessment methods to the local context for accurately tracking food intake patterns (116). Novel approaches are needed that can provide objective, lower-resource alternatives for monitoring RUTF/RUSF intake as well as capturing demographic, socioeconomic and/or dietary impacts on compliance. One such approach is presented by stool genomics, specifically FoodSeq, an emerging technique that characterizes dietary intake by sequencing residual food DNA in human stool (46).

Here, we applied FoodSeq (trnL-P6 chloroplast marker for dietary plants (46) and 12SV5 mitochondrial marker for dietary vertebrates (45,117,118)) to a subset of a previously studied infant cohort from *Matiari*, Pakistan (114). A total of 150 stool samples were included from 60 infants with moderate acute malnutrition (wasting as defined by weight-for-height Z score [WHZ] ≤ -2 ; 3.1 – 15.2 months of age) and 30 healthy infants (WHZ > 0 , HAZ > -1.0 ; 12.1 – 15.2 months of age) (114) (Fig. 3a). Caregivers were provided breast-feeding and complementary feeding education at home for four weeks from enrollment (infants 3–6–months-of-age) (Fig. 3a). A locally produced RUSF *Acha Mum* consisting of chickpea, edible oil, dried milk, and fortified with a vitamin and mineral mix (114,115) was administered to wasted children of weaning age (mean age of 9.89 ± 0.99 months) for an average

duration of 57.8 ± 6.3 days (Fig. 3a). Compliance was monitored through weekly home visits by study staff (114).

3.2 Results

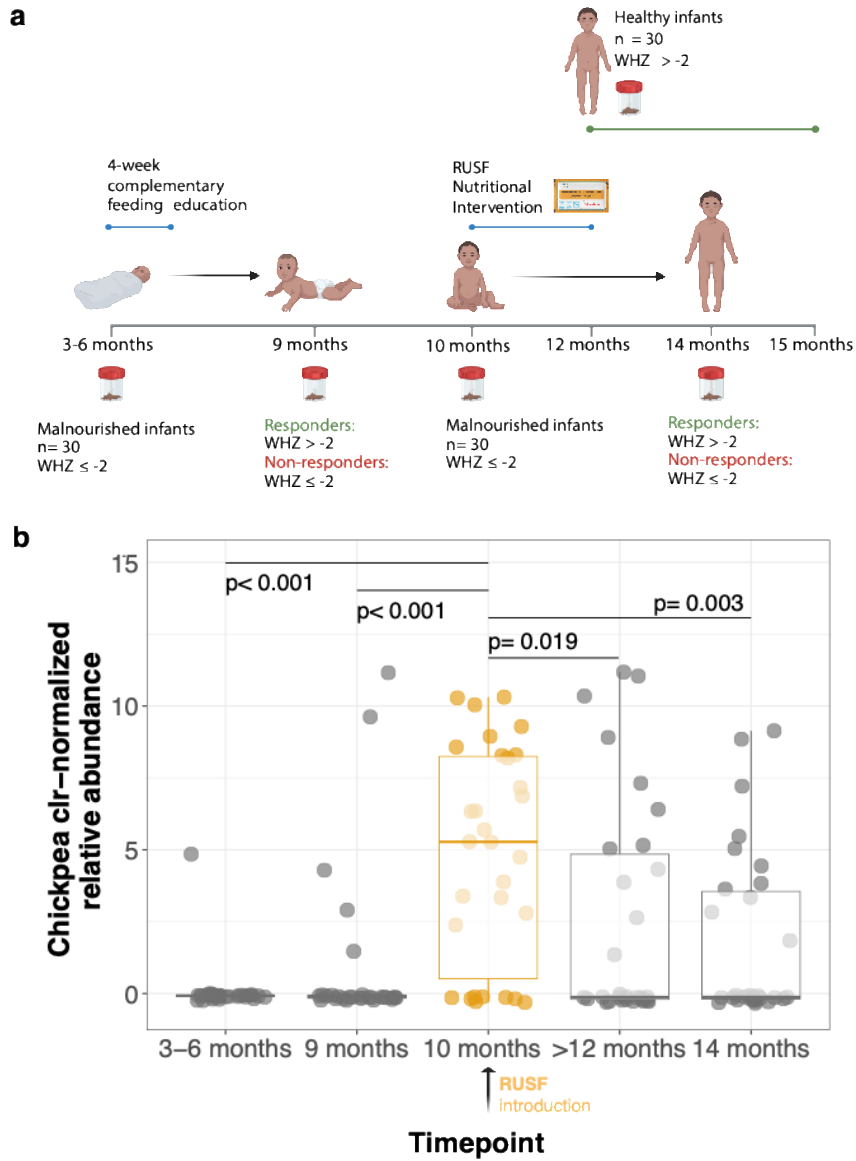


Figure 3: Sampling design of infant subset from SEEM cohort included in this study and detection of Chickpea DNA from stool samples aligning with administration of RUSF.

(a) A cohort of 350 malnourished infants was followed longitudinally in the SEEM study through ages 3-6 months to 15 months (114). From this cohort, our study obtained stool samples from a subset of 30 infants with moderate acute malnutrition (MAM; weight-for-height z-scores (WHZ) \leq -2) at 3-6 months of age when caregivers received an educational intervention covering WHO recommended breast feeding and complementary feeding practice and at approximately 9 months and another subset

of 30 malnourished infants at approximately 10 months of age when caregivers were provided a locally-produced, chickpea-based RUSF (*Acha Mum*) for children and subsequently at approximately 14 months of age. Thirty stool samples from healthy infants (WHZ > 0, HAZ > -1.0) between the ages of 12.1–15.2 months were also included in the analysis. **(b)** Clr-normalized relative abundances of chickpea (*Cicer arietinum*) DNA amplified from stool samples compared across all timepoints. The 10-month timepoint when RUSF was administered is highlighted in orange. Box plots show median and interquartile ranges, while individual dots represent samples from each subject. *Wilcoxon signed-rank test* for 3-6 months vs 9 months and 10 months vs 14 months and *Wilcoxon rank-sum test* for all other timepoint comparisons were conducted. Significance results of statistical testing are marked.

In the malnourished cohort, FoodSeq detected a significant spike in chickpea DNA (*Cicer arietinum*; primary plant ingredient of *Acha Mum*) center-log-ratio normalized relative abundances (hereafter “Clr value”) at 10 months (median= 5.27, IQR[7.75]; present in 22/30 total samples), coinciding with *Acha Mum* administration (Fig. 3b). These DNA levels were significantly higher than observed at pre-administration time points (median= -0.07, IQR[0.06] at 3-6 months, $p < 0.001$ and median= -0.13, IQR[0.10], $p < 0.001$ at 9 months; *Wilcoxon rank-sum test*), natural chickpea consumption in healthy control samples (median= -0.13, IQR[5.06], $p = 0.019$; *Wilcoxon rank-sum test*) and post-administration in malnourished infant samples at 14 months (median= -0.12, IQR[3.75], $p = 0.003$; *Wilcoxon signed-rank*) (Fig. 3b). Major dietary taxa responsible for variation in plant dietary composition were identified through Principal Components Analysis (PCA): Wheats; Tea; Banana; Cowpea Family; Chickpea; Carrot, parsley, cumin family; Pea; Wild rice; Pepper family; Coriander, Cilantro (Fig. 4a, Appendix B: Table 11) and other than chickpea, were not found to display a Clr value spike coinciding with RUSF administration similar to that of chickpea (Appendix B: Fig. 13). PCA was not performed for vertebrate dietary composition due to low dataset complexity and multiple samples being dominated by a single taxon (Appendix B: Fig. 14a). Instead, vertebrate taxa present in more than 30% samples (comparable to chickpea detection) (Fig. 4b) were analyzed for comparative Clr values across timepoints. Water buffalo was the only vertebrate taxon satisfying this criterion and the primary cattle species identified in this cohort (Fig. 4b); Clr values of water buffalo DNA increased consistently from 3-6 months to 14 months (median= -0.68, IQR [0.36] to median= 3.26, IQR[5.79], $p = 0.010$; *Wilcoxon rank-sum test*) (Fig. 4c), potentially reflecting a transition towards cattle dairy from breast milk as infants grew older. However, while dried milk was an animal-sourced ingredient in *Acha*

Mum, water buffalo Clr values did not spike at the 10-month RUSF administration timepoint similar to chickpea (Fig. 4c). Other species detected that could serve as a source of milk (cow, goat) also did not demonstrate a similar Clr value increase at 10 months (Fig. 4c). Rather Clr values for cow DNA displayed a consistent decrease compared with 3-6 months and 9 months timepoints (Fig. 4c). Goat DNA did not demonstrate a significant shift between any timepoint (Fig. 4c).

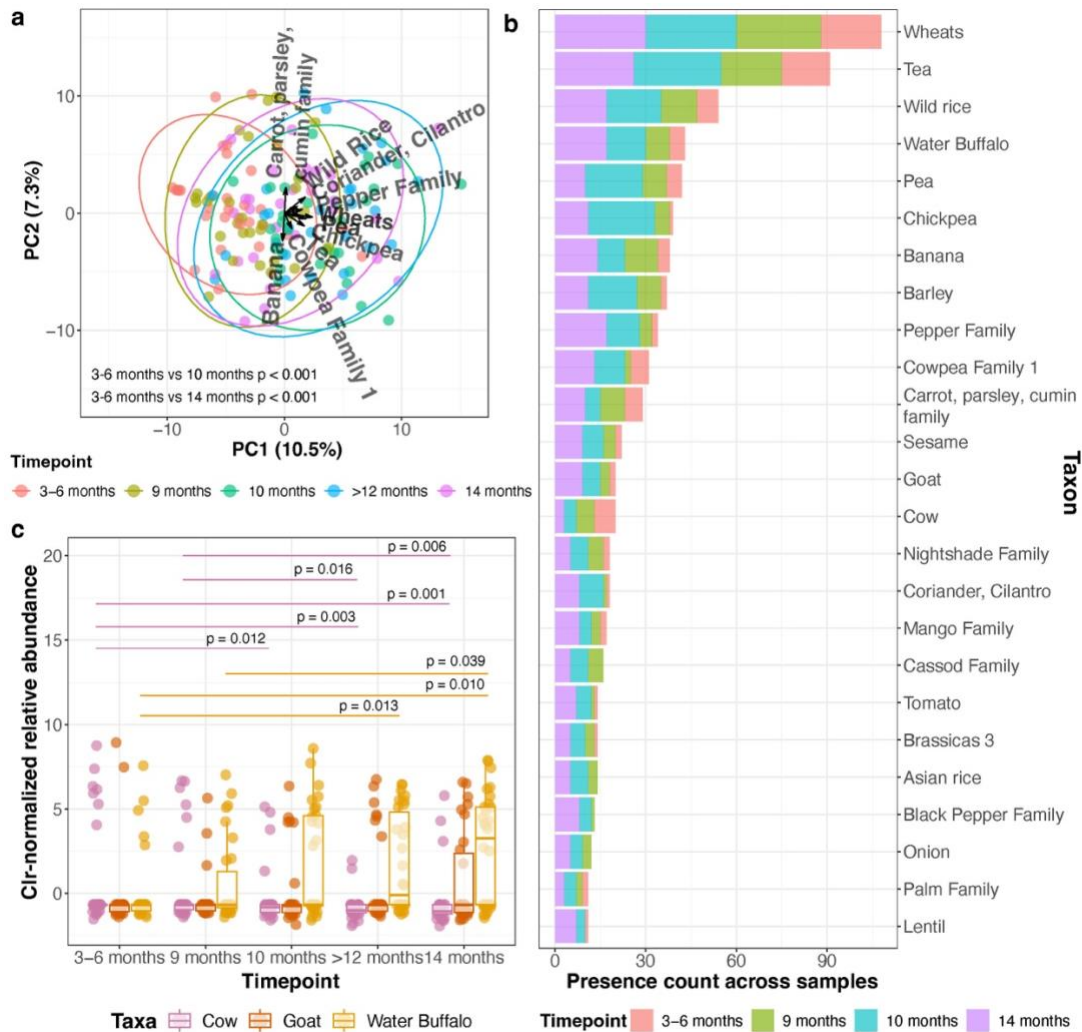


Figure 4: Broad dietary patterns captured by FoodSeq including plant dietary composition, most frequently consumed taxa and potential dairy intake changes.

(a) Principal Component Analysis (PCA) of plant dietary composition with biplot indicating relative contributions of top 10 dietary components. Points colored by timepoint (bottom of figure), timepoint clusters denoted by ellipses and compared along PC1 values using linear mixed effect models with gender and PCR batch as fixed effects and sample ID as random effects. Significant differences as marked. **(b)** Bar plot showing the presence counts of top 25 food (plant and vertebrate combined) taxa detected in participant stool samples across all timepoints except >12 months (healthy controls, see

Appendix B: Fig. 14b for top 25 taxa in healthy controls). **(c)** Clr-normalized relative abundance values of water buffalo, cow, and goat DNA detected in stool samples compared across all timepoints. Box plots show median and interquartile ranges, while individual dots represent samples from each subject. Wilcoxon signed-rank test for 3-6 months vs 9 months and 10 months vs 14 months and Wilcoxon rank-sum test for all other timepoint comparisons were conducted. Significance results of statistical testing are marked.

To understand other dietary factors that might influence treatment outcomes, we examined broader patterns captured by FoodSeq. Overall, FoodSeq detected 89 unique plant and 16 unique vertebrate taxa (Fig. 4b). Wheat was found across >50% of samples, banana (31.6% samples) was the most frequently detected fruit, peas the most frequently detected legume (35 % samples), and water buffalo the most frequently detected animal food source (35.8% samples) (Fig. 4b). Likely environmental contaminants (dog, Indian mouse) were detected at low frequencies (Appendix B: Fig. 14a), potentially indicating sanitation and/or contamination issues in food preparation. Most notably, we observed widespread non-adherence to WHO complementary feeding recommendations (119) with regard to consumption of tea, an age-inappropriate food for infants (also detected across >50% samples) (Fig. 5a). Tea was present consistently across all timepoints (even as young as 3-6 months), increased significantly in median Clr value from 9 to 10 months (median= 3.63, IQR[8.31] to median= 7.63, IQR[4.09], $p = 0.006$; *Wilcoxon rank-sum test*) at which point all but one sample were tea positive followed by a subsequent decreased in median Clr value by 14 months (median= 5.76, IQR[3.59], $p = 0.019$; *Wilcoxon signed-rank test*) (Fig. 5a). In terms of overall patterns, we examined how age-related dietary expansion was reflected in measures of dietary diversity and detected a strong positive correlation between age and both a traditionally calculated Dietary Diversity Score (DDS) and plant FoodSeq Richness (pFR), a genomic measure of dietary diversity calculated by quantifying the unique plant taxa detected in each sample (46) ($\beta = 0.42$, $p < 0.0001$ and $\beta = 0.72$, $p < 0.0001$ respectively) (Fig. 5b). DDS and pFR were not directly related in our cohort, suggesting capture of distinct aspects of dietary diversity. Age also significantly impacted dietary composition patterns as approximated by PCA principal components. Dietary composition was significantly different at 10 months and 14 months when compared with 3-6 months ($p < 0.001$; linear mixed effects model with gender and PCR batch as fixed effects and participant ID as random effects; healthy controls excluded from

comparison) (Fig. 4a). In fact, the effect of age on dietary composition overshadowed any differences resulting from malnourished vs healthy status as any impact of health status on dietary composition was lost when accounting for age (Appendix B: Fig. 15a).

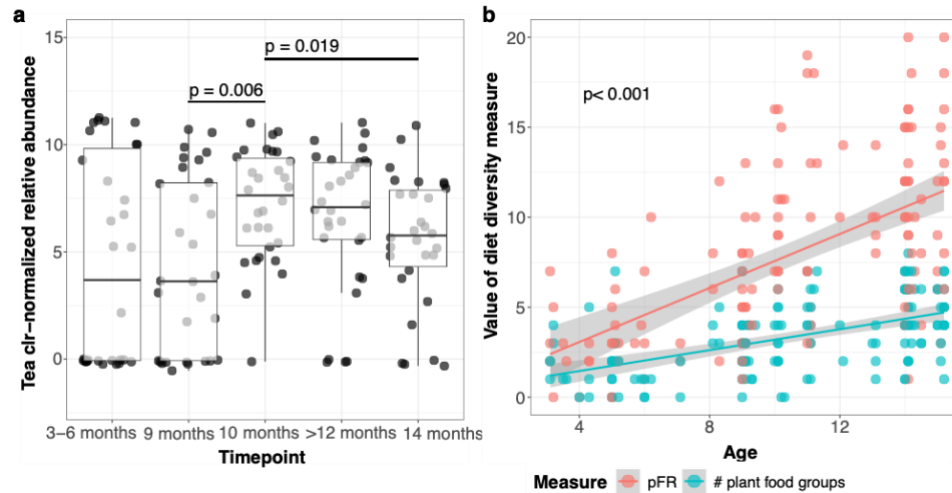


Figure 5: Tea detected consistently across all timepoints and dietary diversity increases with age.

(a) Clr-normalized relative abundances of Tea (*Camellia sinensis*) DNA amplified from stool samples compared across all timepoints. Box plots show median and interquartile ranges, while individual dots represent samples from each subject. *Wilcoxon signed-rank test* for 3-6 months vs 9 months and 10 months vs 14 months and *Wilcoxon rank-sum test* for all other timepoint comparisons were conducted. Significance results of statistical testing are marked. **(b)** Age-dependent increase in dietary diversity measured by traditional dietary diversity score (DDS, salmon) and plant FoodSeq Richness (pFR, cyan). Both metrics show significant positive correlation with age ($p < 0.0001$; Linear mixed effects model with participant ID as random effect and PCR batch and gender as fixed effects). Shaded areas represent 95% confidence intervals.

In terms of the potential impact of dietary and socioeconomic factors on RUSF compliance specifically, chickpea DNA levels at the 10-month timepoint did not partition by wealth index, food insecurity or family size (Appendix B: Fig. 15b) and were not associated with DDS, pFR, or major dietary patterns in this cohort as approximated by plant dietary principal components (Appendix B: Fig. 15c), suggesting that RUSF in this setting was not affected by local socio-economic factors or dietary practices and that the absence of chickpea DNA in some 10-month samples (Fig. 3b) might be more likely due to methodological limitations such as a gap between RUSF administration and stool sampling rather than participant non-compliance.

3.3 Discussion

Collectively, these findings highlight the potential of stool genomics to augment current Community-Based Management of Acute Malnutrition (CMAM) practices by detecting consumption of chickpea-based RUSF and empirically revealing unexpected dietary behaviors that might impede successful treatment outcomes. The most striking example is the widespread presence of tea, suggesting its role in complementary feeding is far more extensive than previously documented. To date, only one prior study has reported tea consumption in over 50% of infants (113). While tea consumption contradicts complementary feeding guidelines, it did not appear to negatively impact RUSF consumption, mirroring the lack of socioeconomic or dietary influences on chickpea DNA detection. From a programmatic perspective, chickpea-based RUSF appears well-accepted for routine outpatient care, with no major structural barriers to adherence, though measures to reduce inappropriate feeding practices such as tea provision may still be warranted.

By providing an objective, culturally adaptable method for assessing dietary intake, this technique offers significant methodological advantages over traditional CMAM compliance monitoring strategies. FoodSeq addressed the unique challenges of dietary assessment in rural *Matiari*, Pakistan by offering direct, objective detection of both standard local foods (e.g., wheat and water buffalo) and intervention-specific ingredients (i.e., chickpea from *Acha Mum*). For example, the most frequently detected taxon wheat is a primary food staple in Pakistan and an ingredient of commonly consumed complementary foods such as *roti* (flat bread), *daliya* (porridge), and *suji* (wheat derivative) (113). Rice is the key ingredient used in other popular complementary food dishes such as *kitchri* (rice and pulses) and *kheer* (rice with milk) (113). Banana has previously been reported as the most frequently used fruit for complementary feeding (113). The detection of water buffalo as the primary cattle species over cow is consistent with the species being native to South Asia and its widespread use for milk and meat (120), and aligns with 24-hour recall results from this study documenting the use of water buffalo milk diluted with water for infant complementary feeding (data not shown). This further underscores FoodSeq's ability to disentangle culturally specific taxonomic sources of commonly consumed foods.

By thus utilizing the universal language of DNA, FoodSeq captures geographic and cultural dietary patterns with minimal protocol modification, unlike traditional dietary assessment techniques that require substantial regional adaptations (29,116), and offers an objective complement to self-reporting or region-specific questionnaires.

In terms of limitations, despite dried milk being one of the primary ingredients of *Acha Mum*, milk-producing taxa did not exhibit Clr value patterns similar to chickpea. Several factors could explain this discrepancy. It is possible that the process of producing dried milk degrades DNA beyond detectable levels, that the quantity of milk used in RUSF production is too low for reliable detection, or that the 12SV5 marker has lower detection efficiency than the trnL-P6 marker. These findings suggest that this detection approach may need further optimization for the animal-derived ingredients of RUSF/RUTF. Furthermore, chickpea DNA was detected in only 22/30 samples from infants who received RUSF, yielding a sensitivity of 0.73 given that all samples at the 10-month timepoint had consumed RUSF (specificity could not be calculated as there were no true negatives in this dataset). Given the retrospective nature of our study where the alignment of stool sampling with RUSF administration was not controlled, this timing gap between intake and stool collection likely contributed to non-detection of chickpea apart from any inherent detection limits of the FoodSeq approach. Future optimization efforts should focus on establishing the detection window for RUSF/RUTF DNA in stool to help align stool collection protocols with feeding schedules for optimal sampling frequency and sensitivity. Given the potential of FoodSeq DNA abundances to be informative of portion size for some plant taxa (46), future studies should also aim to quantify the relationship between RUSF/RUTF intake and residual DNA levels as well as investigating how different RUSF/RUTF formulations affect DNA detectability across diverse geographical, cultural, and biological contexts. Although next-generation sequencing approaches like FoodSeq may be cost-prohibitive in low-resource settings, the compatibility of stool-derived dietary DNA with simple ethanol-based preservation and room-temperature shipping (53) (which circumvents cold-chain constraints) along with the development of targeted PCR assays for specific DNA sequences of

interest, leveraging increasingly accessible field PCR setups (121), could reduce costs and expedite results. Finally, stool collection itself can present challenges in certain settings due to cultural sensitivities, logistical constraints, or participant compliance. Potential strategies to mitigate these challenges include community engagement to improve acceptance and integrating stool collection into existing healthcare workflows.

Overall, our study demonstrates the potential of stool genomics, specifically FoodSeq, as an innovative approach to monitoring RUSF/RUTF compliance in low-resource settings. Future trials and interventions could deploy FoodSeq to monitor compliance with other nutritional interventions or to detect inappropriate feeding practices that undermine standard therapies, especially in challenging or remote environments. Standardizing the method to measure specific ingredients (e.g., by spike-ins of marker foods or synthetic DNA.) might allow for high-specificity detection and more precise quantitative assessment. As these tools evolve, FoodSeq-based insights can be integrated into broader strategies to combat undernutrition, monitor novel therapeutic foods, and inform policies aimed at improving child feeding practices. Ultimately, coupling stool genomics with flexible study designs could expedite targeted interventions for malnourished children by identifying non-responsive populations, revealing deleterious feeding habits, and guiding adaptive, context-specific nutrition programs.

3.4 Methods

This study included a subset of 90 participants from the Study of Environmental Enteropathy (SEEM; NCT03588013; 365 malnourished with WHZ ≤ -2 , 51 healthy with WHZ > 0 , HAZ > -1.0), a prospective longitudinal study conducted in rural *Matiari*, Pakistan (March 2016-March 2019) (114) (Fig. 3; [Created in BioRender. Aqeel, A. (2022) <https://BioRender.com/f56s058Wti>]). SEEM followed infants from birth to 24 months through a multi-national collaboration between Aga Khan University, Washington University in St. Louis, University of Virginia, and Cincinnati Children's Hospital Medical Center, with funding from the Bill and Melinda Gates Foundation (114). Our subset comprised three groups: 30 malnourished infants with paired samples at 3-6 and 9 months, 30 malnourished infants with paired samples at 10-14 months, and 30 healthy controls sampled after 12

months of age. Caregivers of malnourished infants received complementary feeding education at enrollment (3-6 months), with anthropometric measurements recorded monthly (114). Infants maintaining WHZ < -2 at 9 months received a two-month intervention of daily *Acha Mum*, a locally-produced ready-to-use supplementary food (RUSF; ingredients: chickpea, edible oil, dried milk, sugar, vitamins [A, B1, B2, B3, B5, B6, B9, B12, C, D, E, K, H], zinc, folate, iodine, calcium etc. emulsifier) (114). Study staff monitored compliance through weekly home visits, with additional visits as needed (114). A previous Phase 1 study demonstrated 90.5% intervention adherence (114). The RUSF intervention was administered for 57.8 ± 6.3 days beginning at 9.89 ± 0.99 months (122). Non-responders (WHZ < -2 post-intervention) without evidence of celiac disease or other identifiable growth failure etiologies underwent esophagogastroduodenoscopy (EGD) for investigation of underlying pathophysiology (114).

Fecal samples were collected by CHWs (approximately 1 g per participant per timepoint) within 10 minutes of production and transferred to liquid nitrogen pre-charged dry shippers without additives or preservatives (Taylor Wharton, CX-100) for transfer to and storage at -80°C at Aga Khan University (AKU) research facilities prior to shipping on dry ice to Washington University in St. Louis (WUSTL), USA. Sample collection occurred at 3-6, 9, 10, and 14 months for malnourished infants and at ≥ 12 months for healthy controls. At WUSTL, frozen fecal samples underwent pulverization in liquid nitrogen using a sterile mortar and pestle. A 50 mg aliquot of pulverized material was transferred to a 2 mL screw-cap vial containing. A solution consisting of 500 μ L of phenol: chloroform: isoamyl alcohol (25:24:1), 210 μ L of 20% SDS and 500 μ L of buffer (200 mM NaCl, 200 mM Trizma base, 20 mM EDTA) containing 500 μ L of 0.1 mm diameter zirconia/silica beads was added to each sample, followed by bead beating (Mini-Beadbeater-8; Biospec) to extract DNA. DNA was purified (Qiaquick columns, Qiagen), eluted in 70 μ L of Tris-EDTA (TE) buffer, and quantified (Quant-iT dsDNA broad range kit; Invitrogen).

Residual dietary plant DNA and animal DNA in stool samples was assessed using FoodSeq, as previously described (46,118). Briefly, FoodSeq libraries were generated using a two-step PCR

protocol. Primary amplification was conducted using locus-specific primers (*trnL g-b* for plants and 12SV5F/12SV5R76 for animals) with Illumina overhang adapter sequences, a human blocking primer (DeBarba14 HomoB77 for 12SV5), and SYBR Green for qPCR. PCR reactions were performed in a total volume of 10 μ l using either AccuStart II PCR SuperMix (12SV5) (118) or previously established conditions (*trnL*) (46). Cycling conditions included an initial denaturation at 94°C for 3 minutes, followed by 35 cycles of denaturation, annealing, and extension. Each PCR batch included positive and negative controls, and batches were repeated if controls failed (118). Secondary PCR was conducted in a 50 μ l reaction volume using KAPA HiFi polymerase to add Illumina adapters and 8 bp dual indices for sample multiplexing (118). The final libraries were cleaned, pooled separately for *trnL* and 12SV5, and sequenced on independent Illumina runs (118).

Curated reference databases were constructed by compiling a list of edible plant and animal taxa from global food surveys and reference sources (118). Sequences containing *trnL* or 12SV5 regions were retrieved from GenBank and RefSeq, filtered for primer binding sites with a $\leq 20\%$ mismatch tolerance, and trimmed to the target amplicon regions (118). Identical sequences from different accessions were de-duplicated, while intra-taxon variability and conserved sequences across taxa were retained (118). Demultiplexed reads were processed using BBDuk (adapter trimming), Cutadapt (primer filtering and additional trimming), and DADA2 (denoising, merging, and amplicon sequence variant (ASV) inference) (118). Reads were quality-filtered using an expected error threshold of 2 and truncated at the first base with a quality score ≤ 2 (118). For *trnL*, ASVs were assigned using DADA2's `assignSpecies` function, with exact sequence matching to the custom *trnL* reference database (118). In cases of ambiguous matches, taxa were assigned to the last common ancestor (e.g., a sequence matching both wheat and rye was classified at the Poaceae family level) (118). For 12SV5, ASVs were assigned using DADA2's `assignTaxonomy`, due to reduced performance of exact sequence matching caused by PCR polymerase-induced mismatches (118). Suspected contaminants were identified using `decontam` and removed based on DNA quantitation data (118). ASV count tables, taxonomic assignments, and metadata were structured using `phyloseq` (version 1.38.0) (118).

For statistical analyses, taxonomic read counts were converted to relative abundances and normalized through center-log-ratio transformation to preserve overall data structure followed by removal of taxa with fewer than 5 sequence reads, removal of taxa with no assignment at the superkingdom level in trnL and removal of human ASVs from the 12SV5 data. Plant dietary compositionality was analyzed using Principal Component Analysis (PCA, '*prcomp*' function, parameters: centered, not scaled). Plant FoodSeq richness (pFR) was calculated as the number of unique taxa with at least one read count in each sample as previously described (46). The Dietary Diversity Score (DDS) was derived from 24-hr dietary recall data (123). The data collection process, as well as the complete methodology for standardization and homogenization has been previously described (123). The reported food items were categorized into 19 predefined food groups following a standardized food classification system (123). Food groups considered included: cereals and their products; pulses, legumes, and their products; nuts, seeds, and their products; leafy vegetables and their products; vegetables and their products; starchy roots and tubers; flesh meat and their products; fish, seafood, and their products; organ meats; fruits and their products; beverages; milk and their products; sweets and their products; spices, condiments, and seasoning foods; fats and oils; non-food items; supplements; and eggs and their products. Although standard DDS calculations incorporate up to ten food groups, for this study we classified food in greater detail for improved comparison with genomic dietary data resolution. Each participant's dietary intake was assessed to determine whether at least one item from each food group was consumed within the recall period. A binary scoring system was applied, where the presence of a food group was assigned a score of '1', while its absence was assigned a score of '0'. The DDS was calculated as the sum of all consumed food groups:

$$DDS = \sum_{i=1}^n \text{Score } i$$

where Score represents the presence (1) or absence (0) of the *i*th food group, and n is the total number of food groups considered (19 in this study). In terms of comparative testing, *Wilcoxon rank-sum test* was used for comparisons between unpaired samples and *wilcoxon signed-rank test* for paired

samples. Relationships spanning all timepoints were assessed using linear mixed effects models with participant ID as a random effect and gender and PCR batch as additional fixed effects.

3.4.1 Data availability

Deidentified data, data dictionaries, code notebooks, and clinical metadata for reproducing manuscript results from processed trnL-P6 and 12SV5 data can be provided upon request and will be shared in compliance with relevant study agreements, ethical standards, and regulations.

4. Conclusion

4.1 Summary

The goal of this thesis is to establish dietary genomics as a complementary tool in the malnutrition intervention toolbox for monitoring dietary intervention uptake and impact in diverse contexts. Traditional self-reported dietary assessments suffer from recall bias and limited accuracy, whereas dietary genomics provides high-resolution, objective dietary data with orthogonal sources of error. Combined with the universal nature of DNA which allows assessment from any sample that can yield DNA regardless of the demographic, geographic or cultural context, dietary genomics is particularly well suited for populations where self-reported dietary assessment has been challenging such as children and low- and middle-income settings. Despite its potential, dietary genomics has not yet been tested in for measuring adherence and dietary shifts in childhood malnutrition interventions. This work investigates the use of stool-based dietary genomics for assessing dietary quality and compliance with food provisioning interventions across diverse settings, from children with obesity in Durham, USA, to malnourished infants in rural Pakistan. We demonstrate that dietary genomics can detect broad shifts towards low-energy diets following food provisioning and provide assessments of adherence to food provisioning regimens. Additionally, our findings establish the feasibility of stool genomics for detecting RUSF/RUTF consumption in malnourished infants, offering a novel tool for compliance monitoring in settings where conventional tracking methods are challenging.

In Chapter 2, I address two pressing questions in dietary intervention research for pediatric obesity management: (1) whether food provisioning can serve as an effective tool for improving dietary quality in children with obesity, and (2) whether the impact of such an intervention can be robustly detected using objective biomarkers of diet. Through a randomized clinical trial, we demonstrated that food provisioning significantly improved dietary quality relative to usual care, a finding that held across both traditional dietary assessment methods and novel genomic dietary biomarker approaches. Furthermore, our results suggest the potential for sustained dietary improvements beyond the

intervention period, with tentative shifts in gut microbiome composition indicative of enhanced fiber degradation. While “Food-is-Medicine” interventions have demonstrated success in adults, pediatric applications remain limited (69). A recent review of 36 studies found only one pediatric-specific intervention and 3–4 household-based implementations (69). By providing direct evidence for dietary improvements in children, this work supports the expansion of “Food-is-Medicine” strategies in pediatric populations and presents a scalable model for implementation. Additionally, this study advances dietary assessment methodology by integrating a genomic dietary biomarker alongside traditional methods. While metabolomic biomarkers have been proposed for dietary assessment, their implementation at the dietary pattern level remains limited. By leveraging a genomic biomarker from stool samples, we introduce an alternative, scalable approach that captures comprehensive dietary information from a single analysis. This innovation lays a new roadmap for dietary assessment, particularly in nutritional epidemiology and intervention studies, where objective measures of intake are essential. A key advantage of genomic dietary biomarkers is their ability to enable cross-cohort comparability. Unlike traditional assessments, this method facilitates direct dietary comparisons across diverse populations and study designs. If stool samples are available, dietary data can even be retrospectively reconstructed, allowing for the development of reference datasets against which future interventions can be evaluated. As genomic dietary data continues to expand, datasets like ours will offer critical insights into US pediatric diets relative to adult dietary patterns and global populations. Finally, by integrating concepts from pediatric obesity, “Food-is-Medicine”, genomic markers of diet, and gut microbiome analysis, this work spans multiple fields with implications for both research and clinical practice. Moreover, because food provisioning directly addresses the needs of patient populations who may benefit from structured dietary interventions, these findings also carry broad public appeal.

In Chapter 3, I address two critical questions in the management of acute childhood malnutrition: (1) whether dietary genomics can serve as an objective biomarker of adherence to ready-to-use therapeutic and supplementary foods (RUTF/RUSF) in malnourished infants, and (2) whether

genomic dietary assessment can capture broader dietary patterns and socioeconomic factors influencing treatment efficacy. Through a retrospective study of infants with moderate acute malnutrition in rural *Matiari*, Pakistan, we demonstrated significant genomic detection of chickpea DNA, the primary ingredient of the locally produced RUSF *Acha Mum* administered to infants, at levels distinctly elevated compared to both healthy controls and post-intervention samples as well as captured widespread unexpected dietary behaviors such as the consumption of tea that could negatively impact treatment efficacy. Applying genomic dietary assessment in this population provides both a methodological advancement and the first high-resolution dietary profile of complementary feeding in malnourished Pakistani infants— an understudied yet high-burden setting (6,113). Our findings confirm prior reports of age-inappropriate early complementary feeding (113) and reveal widespread tea consumption, a potential barrier to nutritional recovery that may go undetected in self-reported assessments. These insights underscore the need for dietary guidelines that reflect actual consumption patterns to improve feasibility and adherence in public health programs and by enabling objective, population-level dietary tracking, dietary genomics can support the development of nutrition policies and national and regional dietary recommendations based on direct biological evidence. Our findings further introduce a new approach to monitoring therapeutic food interventions and establish dietary genomics as a potential biomarker of adherence to RUTF/RUSF therapy. This discovery enables the development of targeted, DNA-based assays for compliance monitoring. While broad DNA sequencing may not be practical for widespread use in low-resource settings, targeted PCR assays could be designed to detect specific food DNA in stool samples. Given that *trnL-P6* DNA can be amplified in human stool without cold-chain preservation (53) and that infant stool is may be more readily collected than adult samples (from diapers etc.), this method has the potential for a scalable, cost-effective solution for tracking dietary interventions in large-scale programs.

4.2 Ongoing and future research

The results presented in this thesis suggest several avenues for future research categorized as follows:

- Increased research transparency through return of genomic data: A fascinating possibility emerging from this work is the ability to bridge the gap between dietary research and participant engagement by providing personalized genomic dietary feedback. Parents lack objective tools to assess their children's dietary quality, particularly in an environment where dietary misinformation and the marketing of highly processed, nutrient-poor foods are widespread (39). Traditional measures such as weight tracking are insufficient, as weight changes occur gradually and may not provide immediate reinforcement for healthier dietary habits. At the same time, a growing movement in biomedical research calls for greater transparency and equitable participant-researcher relationships (124). Historically, the field has prioritized caution in returning study results, leading to concerns about "helicopter research," where investigators collect deeply personal data but disengage without follow-up once the study is over (124). However, research participants increasingly expect access to their personal data, particularly when it pertains to their own health or that of their families (124). Providing dietary feedback represents an opportunity to not only enhance participant satisfaction but also support informed decision-making around nutrition and health. To address this, we developed the Dietary Data Return Report (DDRR) (Appendix C: Fig. 16), an infographic designed to give participants accessible, personalized insights into their dietary patterns. The DDRR provides information on food diversity, intake across major food groups (e.g., grains, legumes, vegetables, fruits), and a comparison of individual dietary intake to other participants in the study (Appendix C: Fig. 16). Ongoing work in the CHOICES cohort (Chapter 2) assesses the impact of genomic dietary data return on participant study satisfaction and utility in motivating positive behavior change (Appendix C: Table 12). Preliminary findings from 12 out of a total 17 30-40 min interviews with parent-child pairs suggest parents

find genomic dietary feedback valuable for better understanding their children's diets and making informed dietary choices (Appendix C: Fig. 17, Table 13).

- Dietary genomics for large-scale epidemiological and environmental surveillance: This work further highlights the potential of dietary genomics for large-scale epidemiological studies that investigate the broader environmental and socioeconomic forces shaping nutrition. One such application is in wastewater-based epidemiology to provide a high-resolution, cost-effective approach to monitoring dietary patterns across entire communities at a fraction of the cost of conventional survey-based methods at this scale, allowing for the integration of dietary data with economic, geographic, and cultural factors to reveal how urbanization, resource availability, and infrastructure shape dietary choices. Another application involves understanding the environmental drivers of dietary resilience. Ongoing work applies FoodSeq to assess the impact of deforestation on dietary quality in 514 individuals across 19 villages from rural western Honduras (Appendix D: Fig. 18). Forests play a critical role in local diets by providing diverse, nutrient-rich foods— including wild fruits, vegetables, mushrooms, and bushmeat— that contribute to dietary resilience (125). However, deforestation threatens these resources, reducing biodiversity and limiting community adaptation to environmental change. Preliminary findings suggest associations between deforestation and elevation with both dietary diversity and composition changing significantly with elevation (Appendix D: Fig. 18, Fig. 19).
- Digestion and transit dynamics of dietary DNA: Another direction enabled by this work is the ability to move beyond dietary monitoring to investigate fundamental biological questions about dietary DNA digestion and metabolism. While lipid, carbohydrate, and protein digestion are well characterized, little is known about the digestion kinetics of food DNA, despite estimates suggesting daily intake ranges from one to twenty grams (126). Prior studies have demonstrated enzymatic degradation

of DNA in gastric juices and intestinal salvage pathways for nucleotides, but the rate of dietary DNA transit through the digestive tract remains unexplored (127). Understanding these processes would not only increase our understanding of the mechanisms of dietary DNA digestion but also facilitate the development of DNA-based spike-in markers and assays by defining the transit time and detection dynamics of dietary DNA fragments from stool. To address this gap, a future fixed-order within-subjects pilot study aims to quantify dietary DNA transit time and detection limits by providing study-specific single ingredient food doses to participants and collecting stool samples over five days to be analyzed via qPCR to track marker DNA degradation and clearance. Additionally, ongoing work in murine models is investigating the broader physiological fate of dietary DNA by quantifying DNA degradation efficiency across key host factors known to influence macronutrient digestion, including age, body composition, intestinal transit time, and the microbiome. By comparing the degradation of plant-derived DNA to total protein and caloric energy, this work will establish how dietary DNA is processed relative to other macronutrients.

Appendix A

Table 6: Demographic descriptive characteristics at baseline between subjects who were included in the analysis and those who were excluded due to lost to follow-up or discontinued.

	excluded	included	Total
	(N=17)	(N=33)	(N=50)
Child's race			
Black or African American	8 (47.1%)	12 (36.4%)	20 (40.0%)
White	2 (11.8%)	4 (12.1%)	6 (12.0%)
Other	3 (17.6%)	15 (45.5%)	18 (36.0%)
Two or more races	0 (0%)	2 (6.1%)	2 (4.0%)
Missing	4 (23.5%)	0 (0%)	4 (8.0%)
Child's ethnicity			
Hispanic or Latino	4 (23.5%)	18 (54.5%)	22 (44.0%)
Not Hispanic or Latino	7 (41.2%)	14 (42.4%)	21 (42.0%)
Prefer not to answer	3 (17.6%)	1 (3.0%)	4 (8.0%)
Missing	3 (17.6%)	0 (0%)	3 (6.0%)
Child's sex			
Female	10 (58.8%)	16 (48.5%)	26 (52.0%)
Male	6 (35.3%)	17 (51.5%)	23 (46.0%)
Missing	1 (5.8%)	0 (0%)	1 (2.0%)
Child's age (years)			
Mean (SD)	8.9 (\pm 1.3)	8.9 (\pm 1.5)	8.9 (\pm 1.446)
Median [Q1, Q3]	9.0 [8.0, 10.0]	9.0 [8.0, 10.0]	9.0 [8.0, 10.0]
[Min, Max]	[7.0, 11.0]	[6.0, 11.0]	[6.0, 11.0]
Child's BMI at baseline			
Mean (SD)	30.3 (\pm 5.1)	29.5 (\pm 4.8)	29.7 (\pm 4.9)
Median [Q1, Q3]	29.1 [26, 32.7]	28.8 [26, 34.4]	29.1 [26, 34.0]
[Min, Max]	[23.3, 40.8]	[21.7, 38.4]	[21.7, 40.8]
Child's weight at baseline (pounds)			
Mean (SD)	139.9 (\pm 16.9)	134.9 (\pm 37.7)	135.5 (\pm 35.8)
Median [Q1, Q3]	138.4 [132.0, 146.4]	121.2 [110.9, 157.8]	125.2 [111.8, 157.5]
[Min, Max]	[121, 161.8]	[73.4, 217.5]	[73.4, 217.5]
Missing	13 (76.5%)	1 (3.0%)	14 (28.0%)
Having other family member(s) with obesity			
Yes	2 (11.8%)	8 (24.2%)	10 (20.0%)
No	10 (58.8%)	15 (45.5%)	25 (50.0%)
Missing	5 (29.4%)	10 (30.3%)	15 (30.0%)

Parent's education			
Below college	7 (41.2%)	15 (45.5%)	22 (44.0%)
Some colleges or higher	8 (47.1%)	18 (54.5%)	26 (52.0%)
Prefer not to answer	1 (5.9%)	0 (0%)	1 (2.0%)
Missing	1 (5.9%)	0 (0%)	1 (2.0%)
Parent's race			
Black or African American	8 (47.1%)	11 (33.3%)	19 (38.0%)
White	3 (17.6%)	5 (15.2%)	8 (16.0%)
Other	3 (17.6%)	12 (36.4%)	15 (30.0%)
Two or more races	0 (0%)	1 (3.0%)	1 (2.0%)
Missing	3 (17.6%)	4 (12.1%)	7 (14.0%)
Parent's ethnicity			
Hispanic or Latino	4 (23.5%)	19 (57.6%)	23 (46.0%)
Not Hispanic or Latino	10 (58.8%)	11 (33.3%)	21 (42.0%)
Prefer not to answer	2 (11.8%)	0 (0%)	2 (4.0%)
Missing	1 (5.9%)	3 (9.1%)	4 (8.0%)

Table 7: Summary of study deviation notes for participants lost to follow-up or discontinued.

Reason for study non-completion	Usual Care	Intervention	Unassigned	Total
Participant not responding to communication	2	5	0	7
Non-compliance with study protocol ^a	2	3	0	5
Difficulty collecting stool sample ^b	3	2	0	5
Participant personal reason	2	2	0	4
Rescheduling request	1	2	0	3
None	0	0	2	2
Participant not happy with randomization	1	0	0	1
Problem with shipping service ^c	0	1	0	1
Multiple deviations were noted for six participants, leading to them being counted more than once in the same category or counted across multiple categories.				
^a Compliance issues related to not following study processes at designated times e.g. not collecting stool sample before grocery provision or missing multiple samples.				
^b Stool sample collection difficulties primarily consisted of participant or caregiver unwillingness to collect samples.				
^c Problem with shipping service related to failed delivery by shipping service of baseline samples.				

Table 8: Intervention and usual care participants experience with stool sampling. Assessed from participants who completed the study.

Experience with stool sampling	Intervention week 8 sample		Usual care week 8 sample		Total
	Provided	Not provided	Provided	Not provided	
Positive	2	2	9	2	15
Negative	0	4	1	1	6
Mixed: Paper problem ^a	2	2	2	0	6
Mixed: Shipping service problem ^b	0	3	1	1	5
Mixed: Technical problem ^c	0	1	0	0	1

^a“Paper problem” referred to the breaking of the provided stool catching paper during sample collection.
^b“Shipping service problem” referred to problems with FedEx picking up the samples.
^c“Technical problem” referred to difficulty in timing bowel movement properly for stool collection.

Table 9: Descriptive statistics of DSQ predicted intake of calcium (DSQcalc) (mg) in both treatment groups at baseline, week 4, and week 8.

	Usual care (17)	Intervention (16)
Baseline		
Mean (SD)	1146 (± 273)	1133 (± 303)
Median [Q1, Q3]	1078 [933, 1425]	1073 [846, 1368]
[Min, Max]	[764, 1593]	[768, 1720]
Unknown	0	1
Week 4		
Mean (SD)	1204 (± 342)	881 (± 188)
Median [Q1, Q3]	1035 [938, 1515]	846 [777, 891]
[Min, Max]	[771, 1858]	[718, 1508]
Week 8		
Mean (SD)	1153 (± 269)	936 (± 169)
Median [Q1, Q3]	1083 [947, 1328]	872 [829, 1029]
[Min, Max]	[797, 1632]	[774, 1331]
Unknown	1	2

American dietary guidelines recommend 1,000 mg calcium/day for children aged 6-8 years and 1,300 mg/day for children aged 9-11 years (66).



Figure 6: Total order price and item price distribution across all 64 grocery orders.

The intervention group received a total of 64 grocery orders (matching expected number given 16 participants receiving groceries for four weeks each). **(a)** Boxplot of total order price. The median total price was \$96.45 (mean= \$95.75) with a minimum order total of \$62.43 and a maximum order total of \$115.97. Any order totals below the maximum allowed for a given family was at the discretion of the parents. Order totals above \$100 were a result of difficulty in curating component items to fit within the \$100 threshold for that order. **(b)** Density graph of the item prices across all grocery orders. The median item price was \$3.28 (mean= \$4.20) with the maximum item price being \$39.70 (1 instance; Salmon fillets).

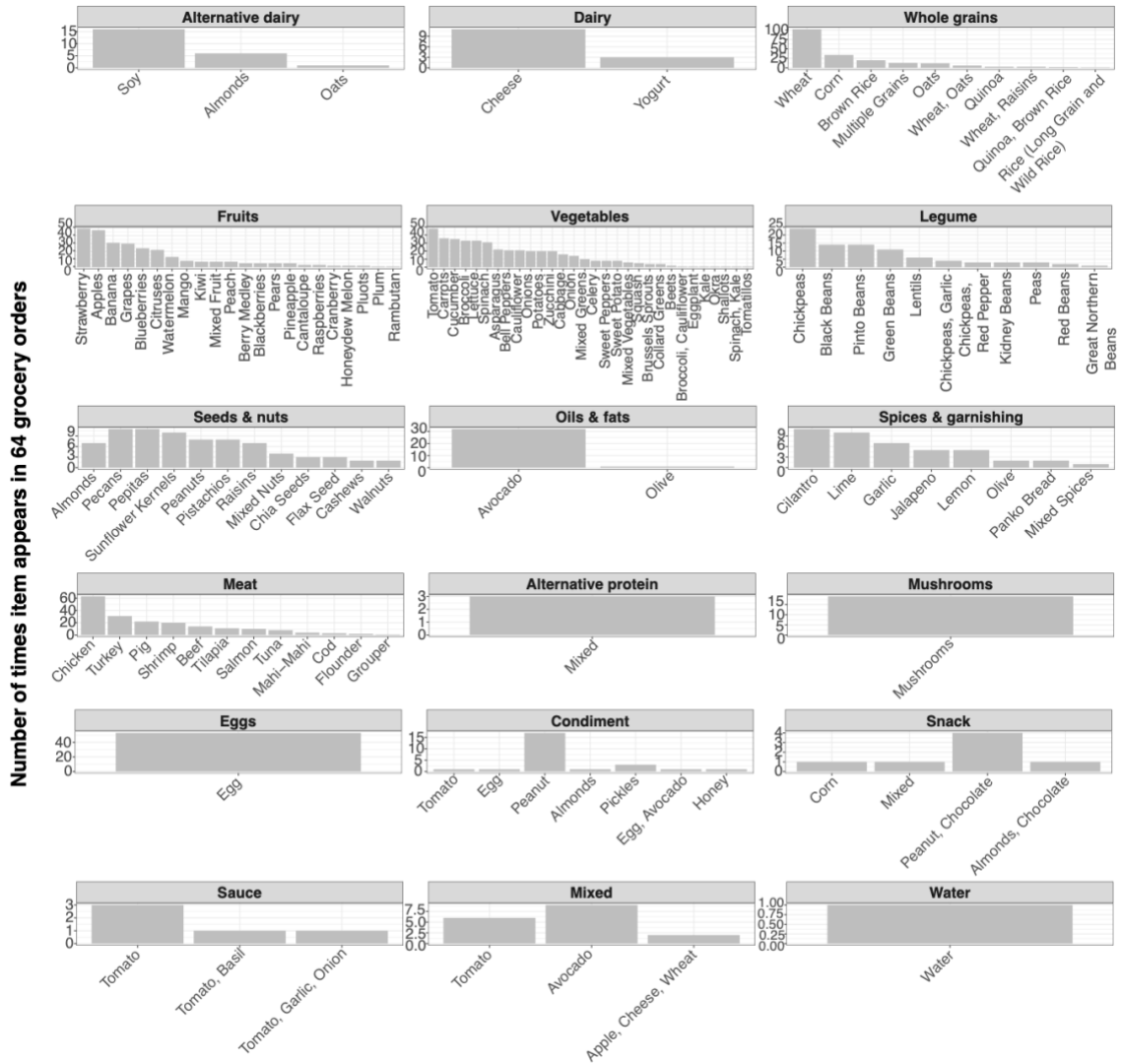


Figure 7: Counts of food item types from 15 different food categories across all grocery orders.

Across the 64 grocery orders, 1500 food items were ordered in total. These items were categorized into 15 overarching food categories. The mixed category was assigned to foods that had more than one primary ingredient (e.g. Salsa). Foods that were botanically fruits but commonly eaten as vegetables were assigned to the vegetables category (e.g. Tomato). The highest variety of foods were found in the fruits and vegetables categories.

Comparison of DSQ dietary intake scores to USDA dietary guidelines

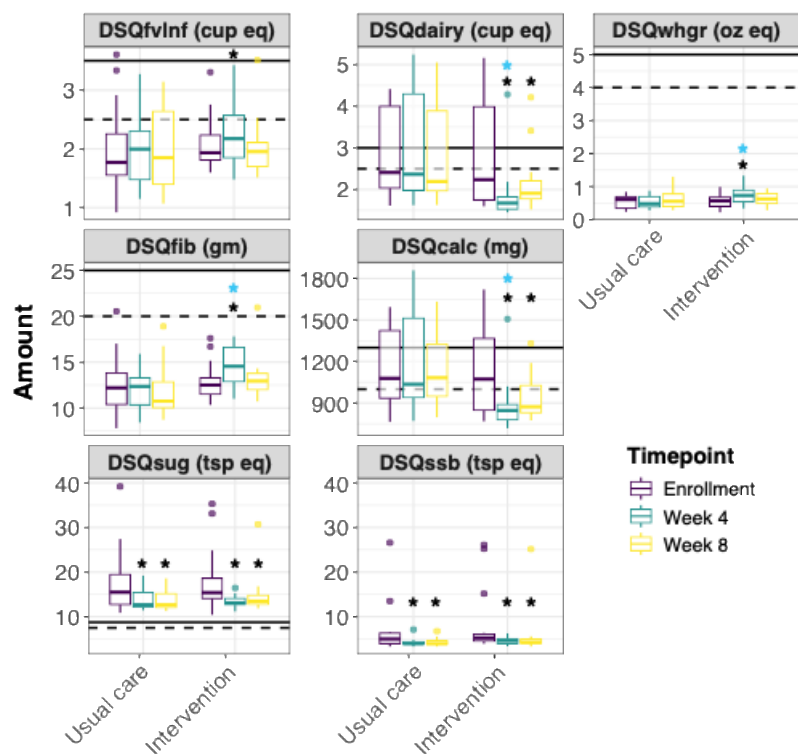


Figure 8: Dietary intake distribution of participants throughout the study, compared to USDA 2020-2025 dietary guidelines (66).

The box plots display the median, 25th and 75th percentiles, and outliers for DSQ intake categories. Category abbreviations: DSQfvlnf = DSQ predicted intake of fruits/ vegetables including legumes and excluding French fries, DSQdairy = DSQ predicted intake of dairy, DSQwhgr = DSQ predicted intake of whole grains, DSQfib = DSQ predicted intake of fiber, DSQcalc = DSQ predicted intake of calcium, DSQsug = DSQ predicted intake of total added sugars, DSQssb = DSQ predicted intake of added sugars from sugar-sweetened beverages. For all categories except DSQsug and DSQssb, the black dashed line represents the minimum guideline for children aged 2-8 years (based on a 1200 kcal/-day diet), while the black solid line represents the minimum guideline for male children aged 9-13 years (based on a 1600 kcal/day diet). The higher male threshold was chosen for stringency, as the guidelines do not provide a specific recommendation for an average 1500 kcal/day intake. For DSQsug, the black dashed line represents the maximum guideline for children aged 2-8 years (1200 kcal/day), and the black solid line represents the maximum for female children aged 9-13 years (1400 kcal/day). The female threshold was chosen for stringency due to the lack of a specific recommendation for a 1500 kcal/day intake. No separate guidelines are provided for DSQssb, as it is subsumed under DSQsug in the USDA guidelines. The USDA recommendations for fruit and vegetable intake were combined, as the DSQ provides a single score for both. **Black asterisks** denote significant within-group outcomes, while **blue asterisks** indicate significant between-group outcomes at each timepoint.

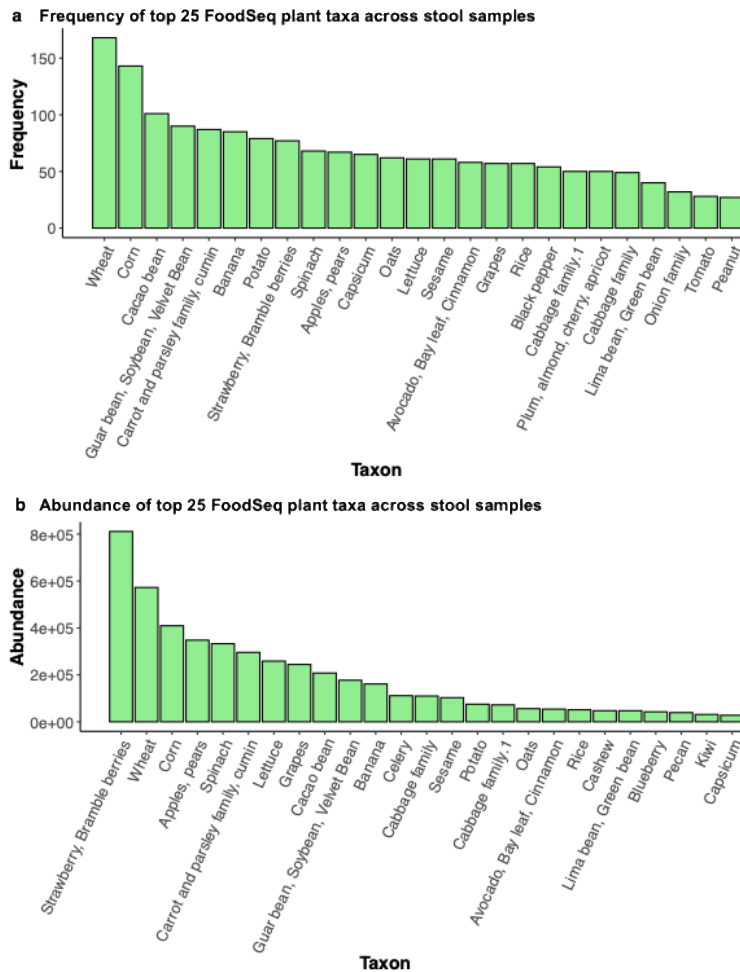


Figure 9: Frequency across samples and total abundance of top 25 plant taxa detected in enrolled participants using FoodSeq.

(a) Taxon frequency refers to the presence/absence of a given taxon across the 180 stool samples included in this study. Staple crops such as wheat and corn were the most frequently detected followed by cacao bean, the primary ingredient of chocolate. **(b)** Taxon abundance refers to the cumulative read counts of a given taxon across the 180 stool samples included in this study. Wheat and corn were again some of the most abundant taxa found in the samples. However, the topmost abundant taxon was now “Strawberries, Bramble berries”.

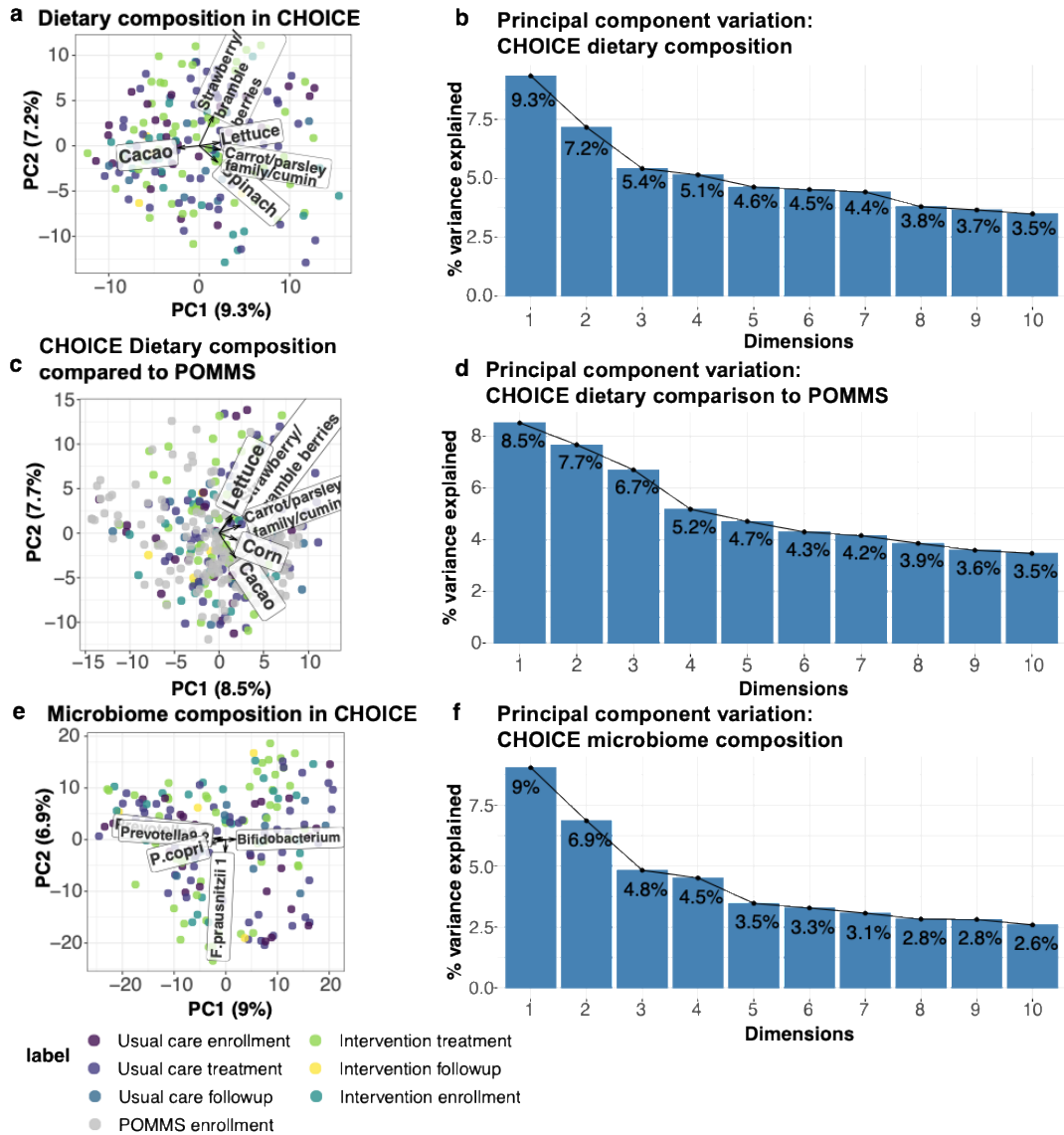


Figure 10: PCA and scree plots of FoodSeq and Microbiome PCA .

(a, c, e). Axes display the first PCs capturing the majority of the variation in the data. Each point on the plot represents the PC1 and PC2 coordinates of each stool sample analyzed. Top 5 factor loadings for each PCA overall are annotated on the PCA plots, length of the arrow represents magnitude of contribution to the PCs and angle of the arrow represents direction of contribution. The spread and clustering of the points is one visual representation of the compositional differences in the data. (b, d, f) PCA scree plots of FoodSeq and Microbiome data. Scree plots show PCs on the x-axis (Dimensions) and the percentage of variance explained by each PC on the y-axis. Number of PCs capturing a significant portion of the variance is identified by the “elbow” of the line.

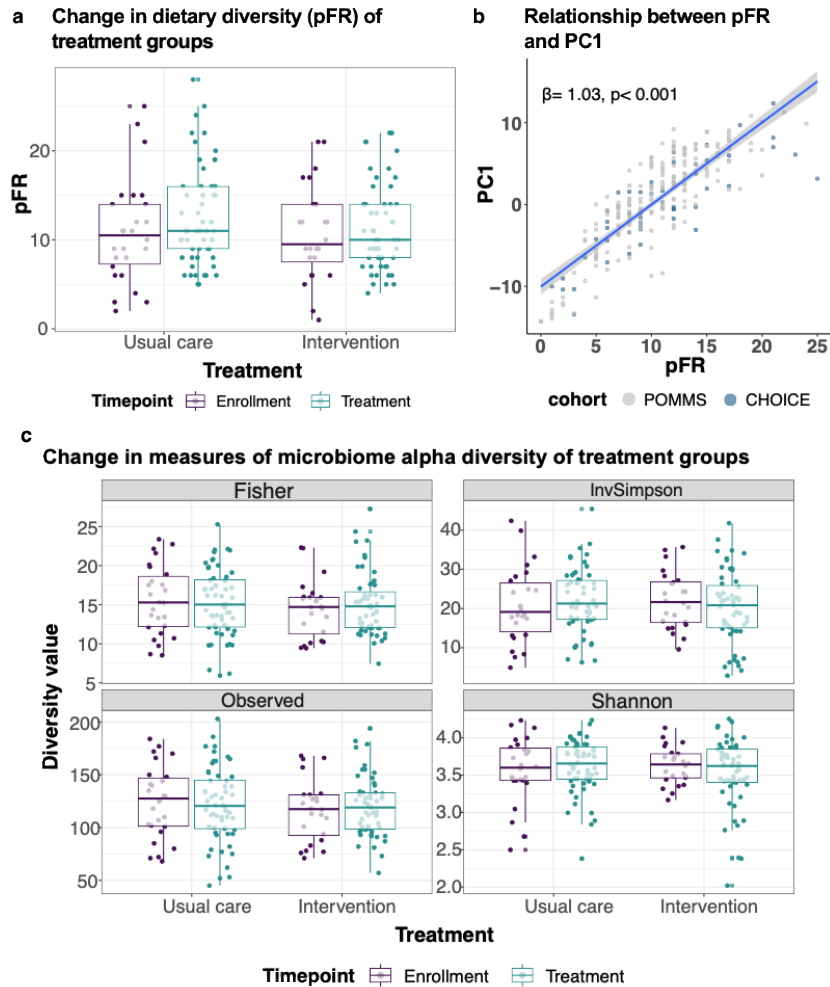


Figure 11: FoodSeq dietary diversity (pFR) and microbiome alpha diversity across treatment groups and relationship between pFR and PC1 from integrated cohort FoodSeq PCA.

(a) No Change in pFR (dietary diversity) from enrollment to treatment in usual care and intervention. (linear mixed effects model). **(b)** Strong association between pFR and PC1 from the integrated cohort analysis; only enrollment samples included from both cohorts (linear mixed effects model accounting for cohort and participant ID). **(c)** No change in four metrics of microbiome alpha diversity from enrollment to treatment in either usual care or intervention. (linear mixed effects model).

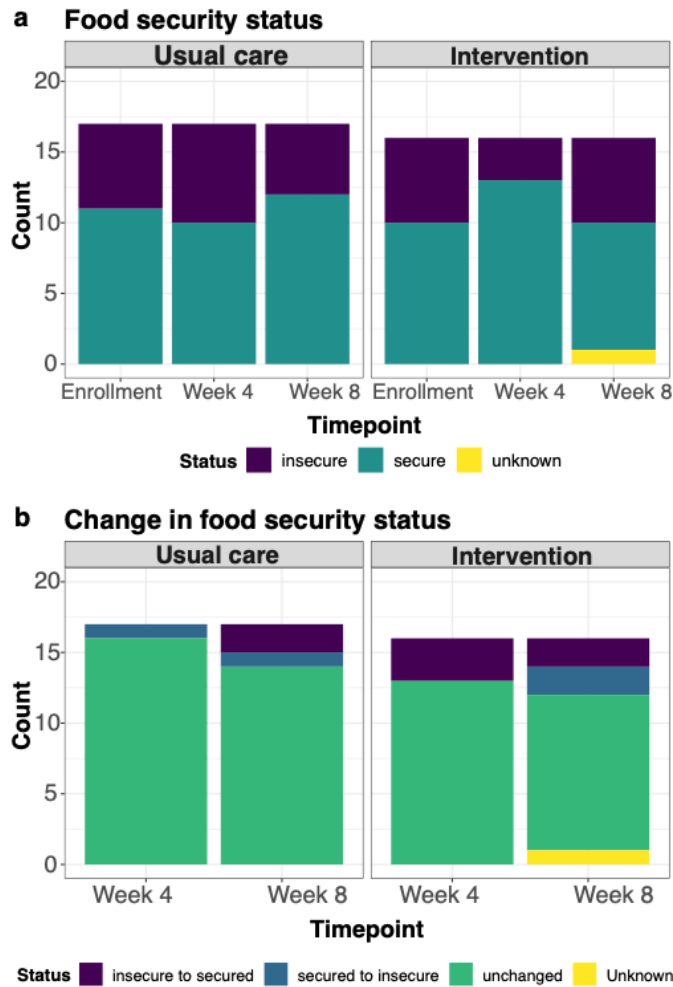


Figure 12: Food security status shifts through course of study.

(a) Food security status of intervention and usual care at enrollment, week 4, week 8. (b) Change in food security status of intervention and usual care at from baseline to week 4 and baseline to week 8. The relatively rarity of shifts in food security status across both groups indicated stability in food security status for the majority of the families studied.

Table 10: Top taxa identified as drivers of PC1 and PC2 in FoodSeq PCAs.

Amplicon Sequence Variant (ASV)	Label	Possible Species
AATCCCGTTTTATGAA AACAAACAAGGGTTTC AGAAAGCGAGAATAA ATAAAG	Strawberry, bramble berries	Fragaria chiloensis, Fragaria chiloensis subsp. lucida, Fragaria chiloensis subsp. pacifica, Fragaria daltoniana, Fragaria gracilis, Fragaria iinumae, Fragaria mandshurica, Fragaria moschata, Fragaria nilgerrensis, Fragaria nipponica, Fragaria nubicola, Fragaria orientalis, Fragaria

		<p>pentaphylla, <i>Fragaria</i> sp. 301, <i>Fragaria</i> sp. CFRA 538, <i>Fragaria vesca</i>, <i>Fragaria vesca</i> f. alba, <i>Fragaria vesca</i> subsp. americana, <i>Fragaria vesca</i> subsp. bracteata, <i>Fragaria vesca</i> subsp. vesca, <i>Fragaria virginiana</i>, <i>Fragaria virginiana</i> subsp. glauca, <i>Fragaria virginiana</i> subsp. platypetala, <i>Fragaria virginiana</i> subsp. virginiana, <i>Fragaria viridis</i>, <i>Fragaria</i> x <i>ananassa</i>, <i>Rosa rugosa</i>, <i>Rosa rugosa</i> f. alboplana, <i>Rubus alexeterius</i>, <i>Rubus amphidasys</i>, <i>Rubus arcticus</i>, <i>Rubus arcticus</i> subsp. acaulis, <i>Rubus assamensis</i>, <i>Rubus australis</i>, <i>Rubus bambusarum</i>, <i>Rubus caesius</i>, <i>Rubus calophyllus</i>, <i>Rubus calycinus</i>, <i>Rubus canescens</i>, <i>Rubus chamaemorus</i>, <i>Rubus chingii</i>, <i>Rubus clinocephalus</i>, <i>Rubus corchorifolius</i>, <i>Rubus coreanus</i>, <i>Rubus crataegifolius</i>, <i>Rubus cuneifolius</i>, <i>Rubus ellipticus</i>, <i>Rubus fockeanus</i>, <i>Rubus geoides</i>, <i>Rubus hawaiiensis</i>, <i>Rubus henryi</i>, <i>Rubus hirsutus</i>, <i>Rubus hoffmeisterianus</i>, <i>Rubus hybrid</i>, <i>Rubus ichangensis</i>, <i>Rubus illecebrosus</i>, <i>Rubus indotibetanus</i>, <i>Rubus innominatus</i>, <i>Rubus irenaeus</i>, <i>Rubus lambertianus</i>, <i>Rubus lasiococcus</i>, <i>Rubus lasiostylus</i>, <i>Rubus lineatus</i>, <i>Rubus macilentus</i>, <i>Rubus nepalensis</i>, <i>Rubus nishimuranus</i>, <i>Rubus nivalis</i>, <i>Rubus niveus</i>, <i>Rubus odoratus</i>, <i>Rubus palmatus</i>, <i>Rubus parvifolius</i>, <i>Rubus pectinarioides</i>, <i>Rubus pectinellus</i>, <i>Rubus pentagonus</i>, <i>Rubus phoenicolasius</i>, <i>Rubus plicatus</i>, <i>Rubus pungens</i>, <i>Rubus rosifolius</i>, <i>Rubus sanctus</i>, <i>Rubus saxatilis</i>, <i>Rubus sengorensis</i>, <i>Rubus</i> sp. EC-2017, <i>Rubus</i> sp. SC-2017, <i>Rubus spectabilis</i>, <i>Rubus stellatus</i>, <i>Rubus swinhoei</i>, <i>Rubus tephrodes</i>, <i>Rubus thomsonii</i>, <i>Rubus treutleri</i>, <i>Rubus tricolor</i>, <i>Rubus trifidus</i>, <i>Rubus ulmifolius</i>, <i>Rubus ursinus</i>, <i>Rubus urticifolius</i>/ <i>Rubus chingii</i>/<i>Rubus crataegifolius</i></p>
CTCCTTTTTC AAAAG CAAGAATAAAAAAAAA G	Spinach	<i>Spinacia oleracea</i> , <i>Spinacia tetrandra</i>
ATCCTATTTTCCAAAA ACAAACAAAGGCCCA GAAGGTGAAAAAAG	Carrot and parsley family, cumin	<i>Anethum foeniculum</i> , <i>Anethum graveolens</i> , <i>Angelica archangelica</i> , <i>Cuminum cyminum</i> , <i>Daucus carota</i> , <i>Pastinaca sativa</i> , <i>Petroselinum crispum</i>
ATCCTGTTTTATGAAA ATAAACAAGGGTTTCA	Apples, pears	<i>Aronia melanocarpa</i> , <i>Malus domestica</i> , <i>Malus domestica</i> subsp. chinensis, <i>Malus sylvestris</i> , <i>Pyrus communis</i> , <i>Pyrus communis</i>

TAAACCGAAAATAAAA AAG		subsp. caucasica, Pyrus pashia, Pyrus pyrifolia, Pyrus ussuriensis, Pyrus x bretschnideri
ATCACGTTTTCCGAAA ACAAACAACGGTTCAG AAAGCGAAAATCAAA AAG	Lettuce	Lactuca sativa
ATCCTATTATTTTATTA TTTTACGAAACTAAAC AAAGGTTTCAGCAAGC GAGAATAATAAAAA AG	Cacao	Theobroma cacao
ATCCTGTTTTCCGAAA ACAAAGAAAAGTTCAT AAAGTGATAATAAAAA AGG	Guar bean, soybean	Cyamopsis tetragonoloba, Glycine max
ATCCGTGTTTTGAGAG GGGGTTCICGAACT AGAATACAAAGGAAA AG	Oats	Avena sativa
ATCCTGGGTTACGCG AACAAAACAGAGTTTA GAAAGCGG	Cabbage family	Brassica oleracea, Brassica oleracea var. capitata, Brassica oleracea var. gemmifera, Brassica oleracea var. gongylodes, Brassica oleracea var. italica
ATCCATGTTTTGAGAA AACAAAGCGGTTCTCGA ACTAGAACCCAAAGG AAAAG	Rice	Oryza sativa
AATCCGTGTTTTGAGA AAACAAGGGTTCIC GAACTAGAATACAAA GAAAAG	Wheats	Secale cereale, Secale cereale subsp. ancestrale, Secale cereale subsp. dighoricum, Secale cereale subsp. rigidum, Secale cereale subsp. segetale, Secale cereale subsp. vavilovii, Triticum aestivum, Triticum carthlicum var. fuliginosum, Triticum dicoccoides, Triticum dicoccum, Triticum ispananicum, Triticum macha, Triticum monococcum, Triticum monococcum subsp. aegilopoides, Triticum spelta, Triticum spelta var. duhamelianum, Triticum timopheevii, Triticum timopheevii subsp. araraticum, Triticum turgidum, Triticum turgidum subsp. durum, Triticum urartu, Triticum urartu var. nigrum
AAATCCCTTTTTGAA AAACAAGTGGTTCICA AACTAGAACCCAAAG GAAAAG	Corn	Zea mays
AAATCCTGTTTTCCGA AAACAACCAAGGGTTC AGAAAACGATAATAAA AAAAG	Grapes	Vitis rotundifolia, Vitis vinifera, Vitis vinifera subsp. caucasica, Vitis vinifera subsp. sylvestris/Vitis vinifera

ATCCTGTTTTCTGAAA ACAAACAAAGGTTCA GAAAAAAG	Potato, goji berry	Lycium barbarum, Lycium barbarum var. auranticarpum, Lycium chinense, Physalis angulata, Physalis philadelphica, Solanum americanum, Solanum betaceum, Solanum nigrum, Solanum tuberosum, Solanum tuberosum subsp. andigenum
ATCCTTTTTTTGGCA AACAAATAAAGATTCC GAAAGAAAAAGG	Blueberry	Vaccinium pallidum, Vaccinium vitis-idaea
ATCCGTGTTTTGAGAA GGGATTCTCGAACTA GAATACAAAGGAAAA G	Barley	Hordeum vulgare, Hordeum vulgare subsp. spontaneum, Hordeum vulgare subsp. vulgare, Hordeum vulgare var. distichon
ATCCTGTTTTCAGAAA ACAAGGGTTCAGAAA GCGAGAACCAAAAAA AGGATAG	Avocado, Bay leaf, Cinnamon	Persea americana, Persea americana var. americana, Persea americana var. americana x 'guatemalensis', Persea americana var. costaricensis, Persea americana var. drymifolia, Persea americana var. guatemalensis, Cinnamomum aromaticum, Cinnamomum verum, Laurus nobilis, Persea americana, Persea americana var. costaricensis, Persea americana var. guatemalensis
ATCCTATTTTCTGAAA ACAAATAAAGGTTTCAG AAAAAAG	Capsicum	Capsicum annuum, Capsicum baccatum, Capsicum baccatum var. baccatum, Capsicum baccatum var. pendulum, Capsicum baccatum var. praetermissum, Capsicum chinense, Capsicum frutescens, Capsicum pubescens
ATCCTATTTTTCGGAA AACACGAGTTCAGAA AGCGAAAAAGG	Black pepper	Piper guineense, Piper longum, Piper nigrum, Piper sarmentosum
ATCCTTATTTTGAGAA AACAAAGGTTTATAAA ACTAGAATTTAAAAG	Banana	Musa acuminata, Musa acuminata subsp. malaccensis, Musa acuminata subsp. siamea, Musa acuminata subsp. truncata, Musa acuminata var. zebrina, Musa balbisiana, Musa balbisiana var. balbisiana
ATCCTGGGTACGCG AACAAAACAGAGTTTA GAAAGCGG	Cabbage family.1	Brassica oleracea, Brassica oleracea var. capitata, Brassica oleracea var. gemmifera, Brassica oleracea var. gongylodes, Brassica oleracea var. italica

Appendix B

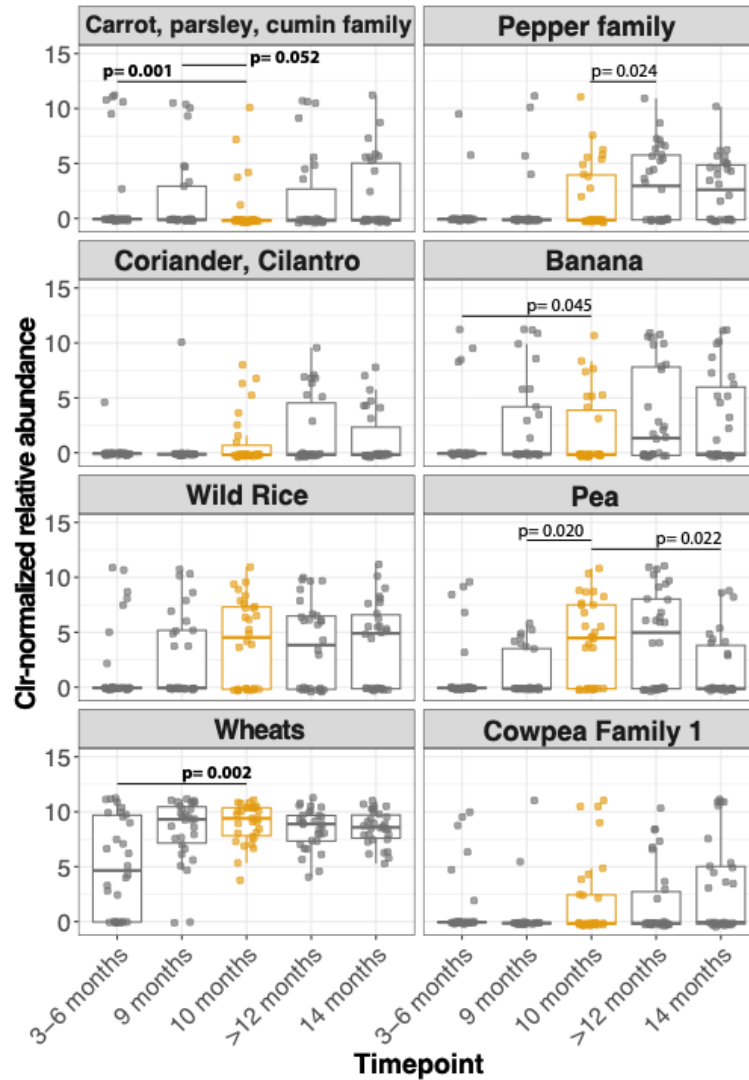


Figure 13: Major dietary taxa DNA detection patterns do not align with RUSF administration.

Clr-normalized DNA relative abundances of top 10 taxa driving dietary composition as identified. By PCA biplot compared across all timepoints. The 10-month timepoint when RUSF was administered is highlighted in orange. Box plots show median and interquartile ranges, while individual dots represent samples from each subject. *Wilcoxon signed-rank test* for 3-6 months vs 9 months and 10 months vs 14 months and *Wilcoxon rank-sum test* for all other timepoint comparisons were conducted. Significance results of statistical testing are marked. Significance values in bold remain significant after Benjamini-Hochberg multiple testing correction (including Chickpea and Tea timepoint comparison tests).

Table 11: Top taxa identified as drivers of PC1 and PC2 in dietary principal components analysis.

Label	Component Taxa	Amplicon Sequence Variant (ASV)
-------	----------------	---------------------------------

Wheats	Rye, bread wheat, emmer wheat, spelt, einkorn, wild einkorn, domesticated hulled wheat, rivet wheat, durum wheat	ATCCGTGTTTGTGAGAAAACAAGGGGT TCTCGAACTAGAATACAAAGGAAAAG
Tea	Common Camellia, Japanese Camellia	ATCCTGTTTTTCTAAAACAAACAAAGA TCCGAAAGCGAAAATAAAAAAAG
Banana	Banana, , Fehi Banana, Hueta, Tahitian Red Cooking Banana, Baby Pink Banana, Fuzzy Banana, Pink Banana, Pink Fruiting Banana, Pink Velvet Banana, Red Banana, Self-Peeling Banana, plantain	ATCCTTATTTTTGAGAAAACAAGGTT TATAAAACTAGAATTAAAAAG
Cowpea Family 1	Mung bean, cowpea	ATCCTGTTTTCTGAAAACAAGAAAA ATTAAGAAAGTTATAATAAAAAAGG
Chickpea	Chickpea	ATCCTGCTTTCGGAAAACAACAAAA AAAGTTCAGAAAGTTAAAAATCAAAAA AG
Carrot, parsley, cumin family	Fennel, dill, angelica, Arracacha (common), Peruvian Carrot, Peruvian Parsnip, White Carrot, cumin, wild carrot, parsnip, parsley	ATCCTATTTTTCCAAAAACAACAAAG GCCCAGAAGGTGAAAAAAG
Pea	Pea	ATCCTTCTTTCTGAAAACAATAAAAG TTCAGAAAGTGAAAATCAAAAAAG
Wild rice	Asian wild rice, African wild rice	ATCCATGTTTTGAGAAAACAAGCGGT TCTCGAACTAGAACCCAAAGGAAAAG
Pepper Family	Bell peppers, sweet/Italian peppers, serrano, cayenne, paprika, jalapeno, aji amarillo, aji limon, ciolla sella, NA, habanero, scotch bonnet, trinidad scorpions, bhut jolokia, carolina reaper, tabascos, indian chiles, rocoto and manzano peppers	ATCCTATTTTTCTGAAAACAATAAAAG GTTTCAGAAAAAAG
Coriander, Cilantro	Coriander, cilantro	ATCCTATTTTTCCAAAAACAACAAAG GCCCAGAAGGTGAAAAAAG

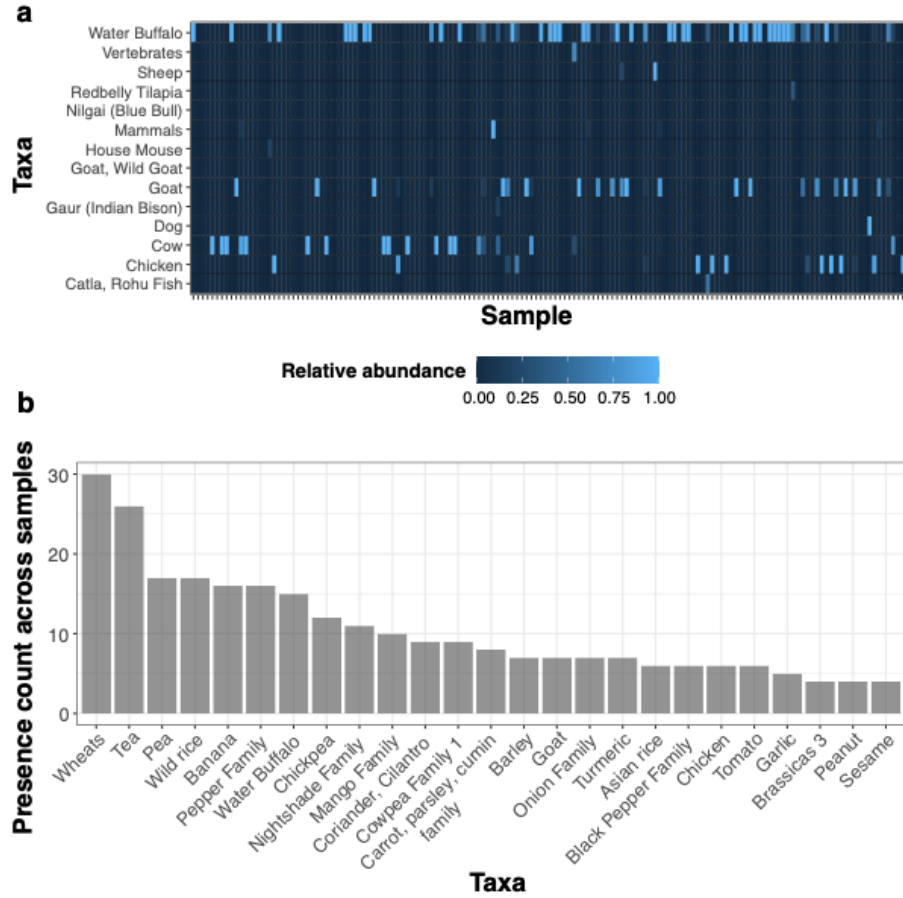


Figure 14: Dietary DNA Composition in Stool Samples: Heatmap of Vertebrate Taxa Abundances and Presence Counts of Common Food Taxa in Healthy Controls.

(a) Heatmap of relative abundances of vertebrate taxa DNA post-filtration of human DNA and taxa with less than 5 total reads. **(b)** Bar plot showing the presence counts of top 25 food (plant and vertebrate combined) taxa detected in participant stool samples in healthy controls (>12 months).

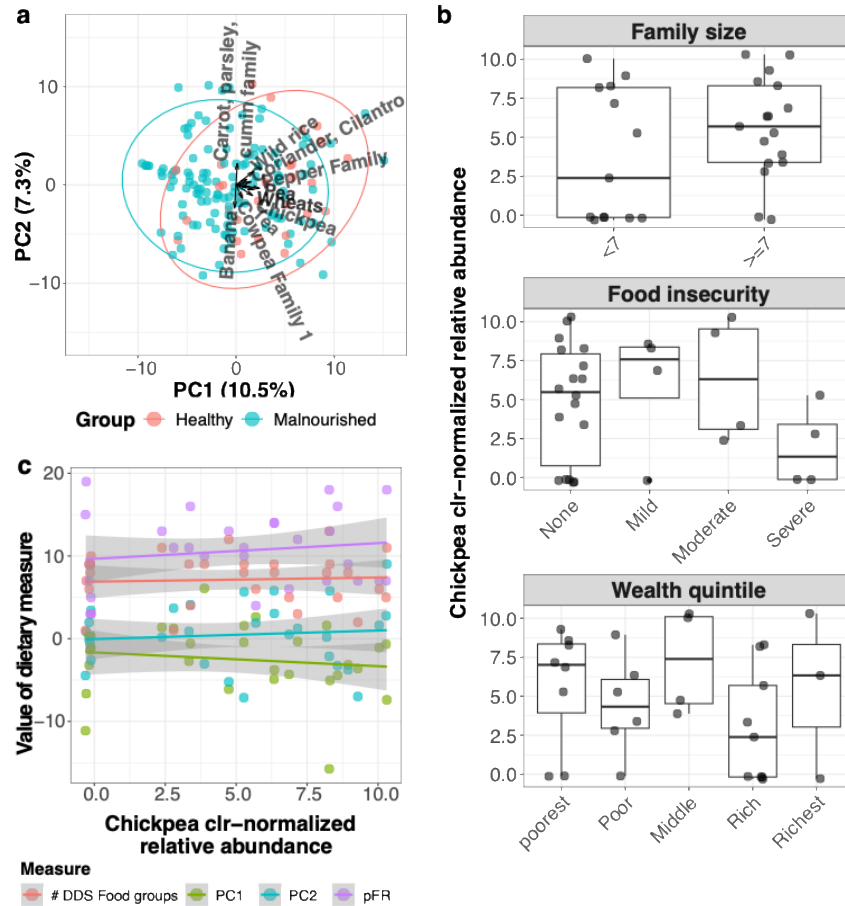


Figure 15: Health status does not impact dietary composition and chickpea DNA detection is not impacted by socioeconomic or other dietary factors.

(a) Principal Component Analysis (PCA) of plant dietary composition with biplot indicating relative contributions of top 10 dietary components colored by health status (healthy controls vs malnourished samples) and clusters denoted by ellipses. Health status compared along PC1 and PC2 values using linear mixed effect models with age, gender, and PCR batch as fixed effects and sample ID as random effects. No significant impact detected when accounting for age. **(b)** Chickpea Clr-normalized relative abundances at the 10-month timepoint compared across family size, food insecurity status, wealth quintile of surveyed households using *Wilcoxon rank-sum test*. Box plots show median and interquartile ranges, while individual dots represent samples from each subject. No significant difference detected. **c**, Chickpea Clr-normalized relative abundances compared at the 10-month timepoint against dietary factors such as traditional and genomic dietary diversity as well as the first two principal components from a plant dietary composition PCA excluding chickpea (each PC approximating a broader dietary pattern). Linear mixed effects model with participant ID as random effect and PCR batch, and gender as fixed effects. Shaded areas represent 95% confidence intervals. No significant relationship detected.

Appendix C

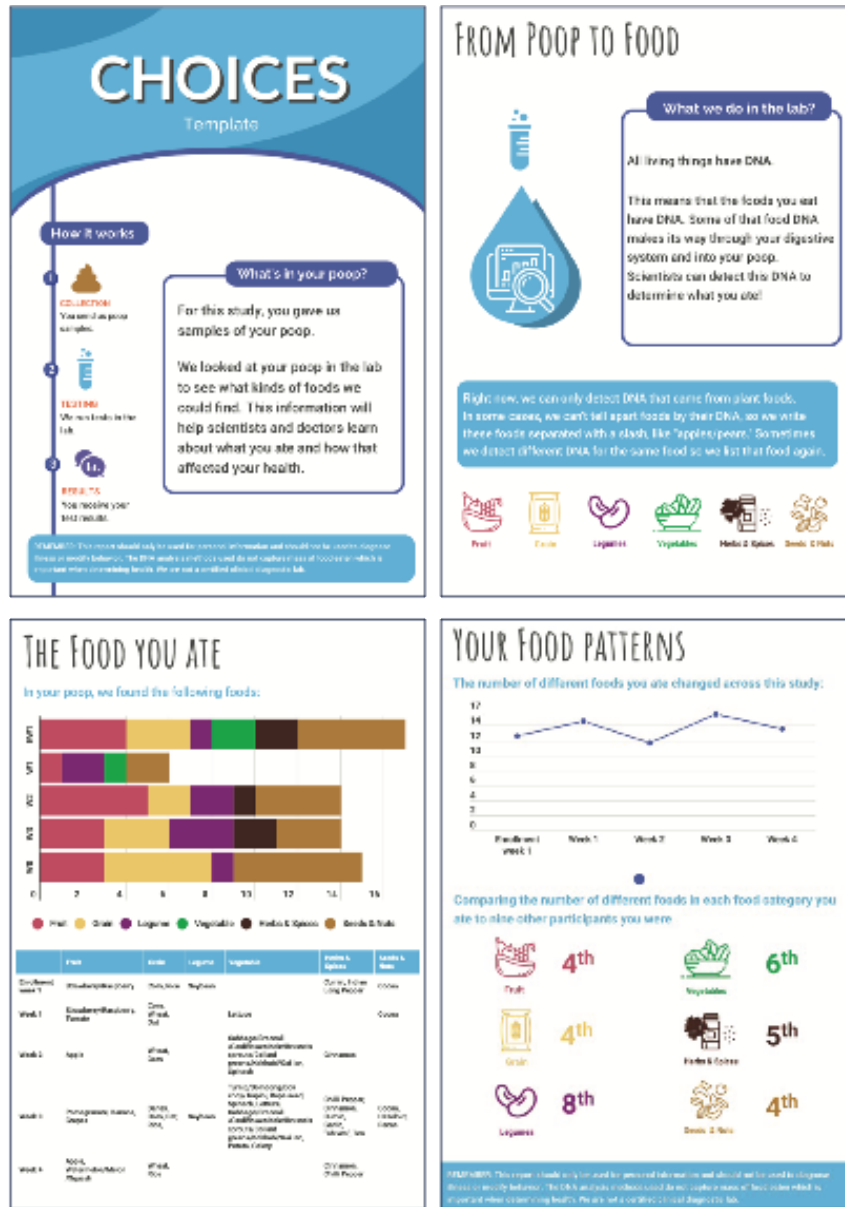


Figure 16: Example of a personalized genomic dietary data return report provided to 17 participants (parents and children) in the CHOICES study.

Each report was generated using genomic dietary data derived from participants' weekly stool samples collected during the study. Reports were individualized to reflect the plant-based foods detected in their samples and included visualizations of dietary patterns over time. Interviews were conducted one-on-one, with each participant receiving their customized report during the session. The interviewer guided the participant through the report and figures while asking questions to assess their reactions and feelings about the report's utility for improving health and their experience seeing their dietary data presented in this format. Interviews lasted 30–40 minutes and were conducted in English or Spanish based on the participant's language preference.

Table 12: Coding scheme used for analyzing participant interviews.

Category	Code	Description of content captured
Motivation	Health Concern	Language/sentiments indicating interest for perceived benefit to existing health conditions.
	Health Conscious	Language/sentiments indicating interest in a proactive health approach.
	Curiosity	Language/sentiments indicating general curiosity rather than specific health needs.
	Parental Surveillance	Language/sentiments indicating interest in greater knowledge and monitoring of children's dietary intake.
	Financial Compensation	Language/sentiments indicating interest in financial incentives or compensation.
Frequency-Timing	Report Frequency	How often participants prefer receiving reports (monthly, weekly, etc.).
	Turnaround	How quickly participants prefer results to be provided after sample submission. (days, weeks, months).
Value of Report	Verification	Language/sentiments commenting on potential of tracking diet, mentions of accuracy or objectivity, confirmation of knowledge.
	Vs. Financial Compensation	Compares whether participants value the dietary reports more or less than financial compensation provided. (Greater, Lesser, and Special case = would consider instead of financial compensation if certain conditions are met).
	Behavior-Regulation	Statements that imply personal responsibility, accountability, intent to change or improve behavior.
	Chart	Greatest preference for which one of the four charts presented (Food groups chart (1 st), Taxa table (2 nd), Dietary diversity chart (3 rd), Participant ranking (4 th))
	Discovery & Insight	Language/expressions of surprise, inquisitiveness, unexpected results without behavior change specified.
	Approachable	Language/sentiments capturing user-friendliness of report.
	Accomplishment	Language/sentiments capturing a sense of progress in health goals.
	Potential for Cross-Use	Language/sentiments indicating interest in broad quantitative comparisons with other people.
Grievance-Barrier	More Utility	Suggestions/expressions indicating desire for greater utility from the report.
	More Clarity	Suggestions/expressions indicating unclear or confusing information in report.
	More Visuals	Suggestions/expressions indicating desire for more graphic representations of data and/or explanations.
	Stool Sampling	Captures hesitations or challenges related to stool sample collection.
	Chart	Captures issues or concerns with a specific chart (Food groups chart (1 st), Taxa table (2 nd), Dietary diversity chart (3 rd), Participant ranking (4 th))

Motivation	Health concern	“Because one eats and doesn’t know how much damage it can do, to say it like that. [CHILD] is sick of the stomach, so how much damage it can do to them. So, yes it can tell me what they ate and I can change it to make it better.”
	Health conscious	“Well, I think the thing that interested me the most was that it could change [CHILD]’s eating and well [CHILD] was eager to try many things that they hadn’t tried.”
	Curiosity	“I just thought it would be interesting to see, you know, the food choices I make, what comes out of, you know, looking at my stool sample and recommendations for what I should eat versus what I shouldn't?”
	Parental surveillance	“Yes, I believe that it was for the same, to know their, about the foods that [CHILD] was eating.”
Report value	Verification	“Yes, I find it useful. It kind of tells you, you know what you're actually putting in your body, makes you more aware of what you're actually eating. What you're putting in your body. I like it”
	Behavior regulation	“So it does show me like, "Oh, wow, it really doesn't matter every day even because I'm like, I don't have time today, I don't have time to stop here because today I don't have time to do this today," and it's showing me that it doesn't matter, that you should attempt every day to have those different foods.”
	Chart	“I like the first chart. The color chart, yeah. So you can quickly just with the vivid colors, say OK, what am I eating and not eating enough of? What am I eating in the most? And you know the least of I like this one, this one.”
	Vs financial compensation	“Well yes, like I said, the compensation is not a requirement for me. [...] But, if there wasn’t any compensation but the report of what is consumed, for me it would be excellent, there would be no problem if I had the time to participate again.”
	Discovery & Insight	“It's just interesting. I think this is amazing. I think this is wonderful. [...]It would just make me be more inquisitive, or especially when I see a rise. Oh, what was going on that week where we had that a rise and what we're eating.”
	Approachable	“I do like the way it was presented. It made it look very clear and easy to read versus a bunch of words across the screen, and I get annoyed reading it, so this was very easy to understand the graphs.”
	Concerns/barriers	Chart
More utility		“Cause I'd rather know what I'm doing wrong to correct it vs just saying “oh, we did this, oh, good job, good job, good job”. I wanna see like “oh my god, I really have to change this” [...] If you include like a percentage this is

		the percent you have in vegetable but this is the percent you need in vegetable to be like in standard”
	More clarity	“Oh yeah, I had to ask you what did the chart with the percentage [ranking chart], what it mean. So I would think like some sort of description, like letting you know that once this 40% means that and what is recommended right?””
	More visuals	“New charts, visual. We like, I like talking to, you know, might to describe some people is long.”
Desired frequency	Report Frequency	“Maybe like once a month. So maybe she's not getting in enough vegetables. I know that. So the following month is OK, we're gonna push more vegetables.”
	Turnaround	“A month or within a month.”

Appendix D

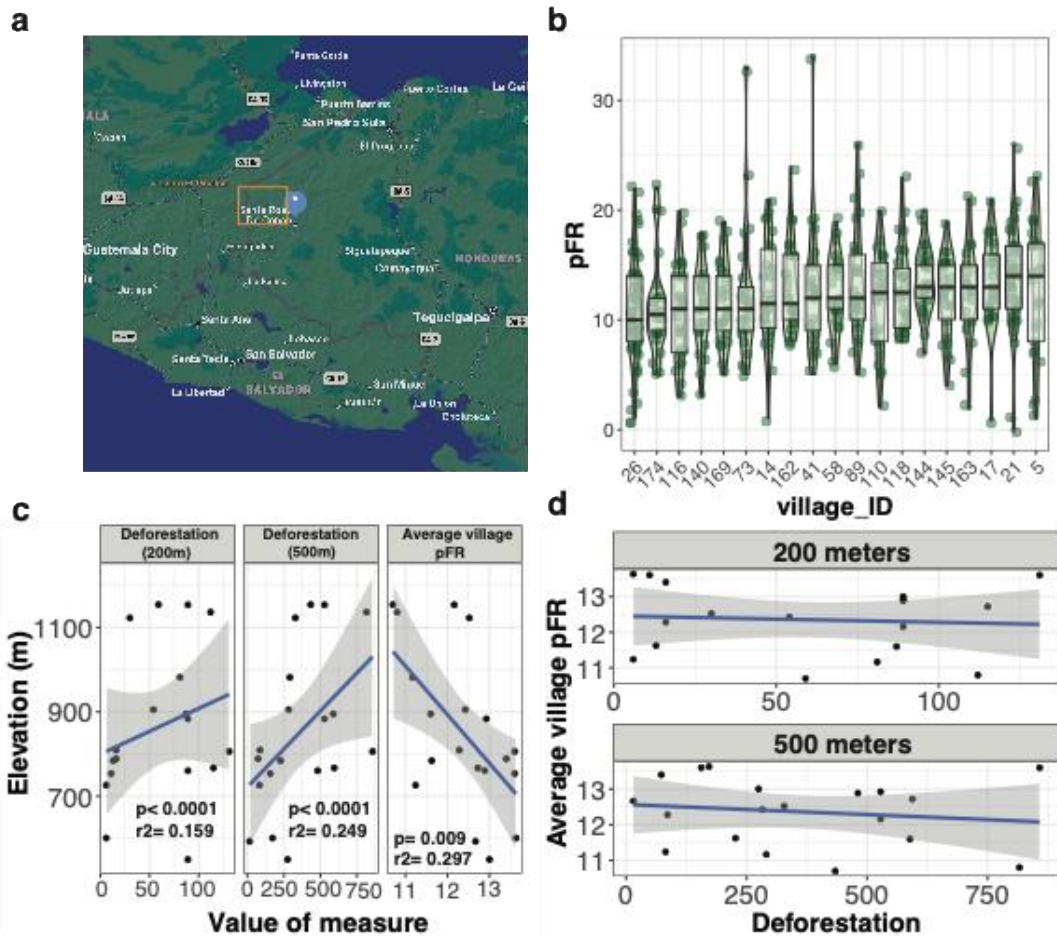


Figure 18: Genomic dietary diversity (pFR) and its relationship with deforestation and elevation across 19 villages in western Honduras.

(a) Map showing the study region (orange box) where stool samples were collected from 517 adults (approximately 27 per village) for FoodSeq analysis. **(b)** Distribution of genomic dietary diversity (pFR) across the 19 villages, ordered from lowest to highest median pFR. **(c)** Significant relationships observed between deforestation (at 200m and 500m), elevation, and average village pFR, with elevation positively associated with both deforestation and pFR (linear regression controlled for sequencing batch, $p < 0.05$). **(d)** Relationship between average village pFR and deforestation at 200m and 500m radii around village centroids, showing no significant associations. Deforestation data were derived from Landsat time-series analyses, defining ‘forest’ as pixels with at least 70% tree cover and calculating deforestation as the proportion of forest lost between 2001 and 2016 within 200m and 500m around each village centroid. Linear regressions accounted for sequencing batch effects.

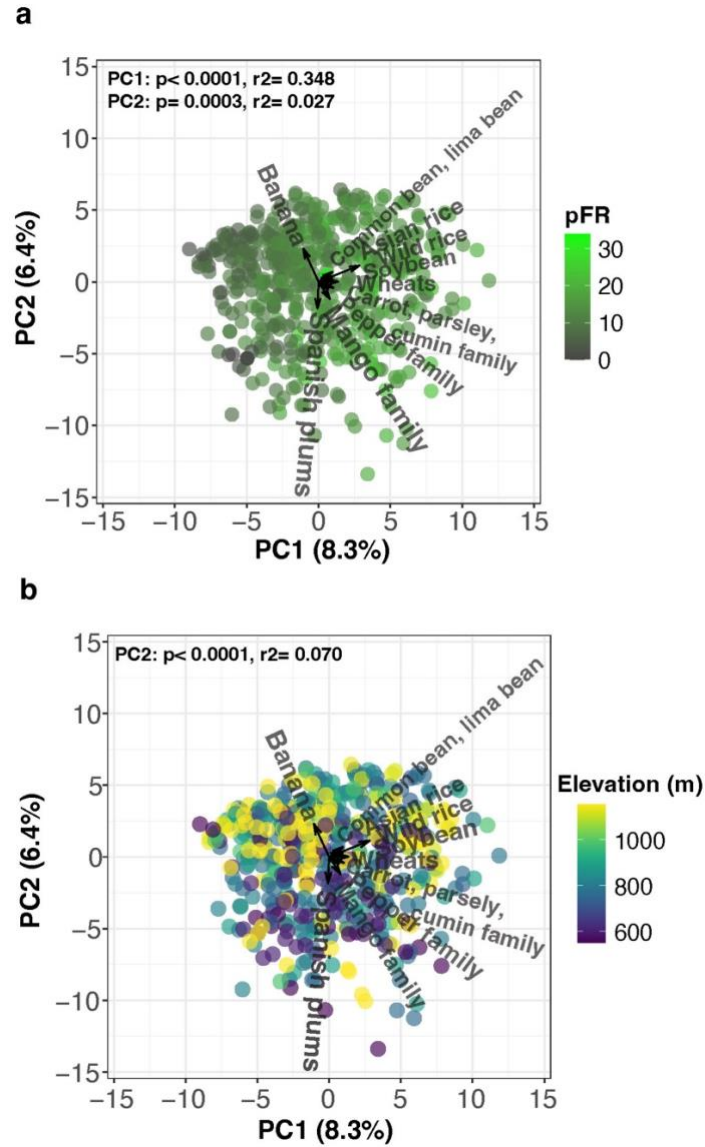


Figure 19: Principal Component Analysis (PCA) of plant dietary composition across samples, with biplots indicating the relative contributions of the top 10 dietary components.

(a) Points are colored by genomic dietary diversity (pFR) at each time point. Linear regression results (accounting for sequencing batch) indicate a significant association between pFR and both PC1 and PC2. **(b)** Points are colored by elevation level, with statistical testing by linear regression (accounting for sequencing batch) showing a weak but significant association between PC2 and elevation.

Bibliography

1. THE 17 GOALS | Sustainable Development [Internet]. [cited 2023 Oct 25]. Available from: <https://sdgs.un.org/goals>
2. UN. Goal 2 | Department of Economic and Social Affairs [Internet]. [cited 2025 Feb 11]. Available from: https://sdgs.un.org/goals/goal2#targets_and_indicators
3. WHO. Fact sheets - Malnutrition [Internet]. [cited 2022 Jul 13]. Available from: <https://www.who.int/news-room/fact-sheets/detail/malnutrition>
4. Popkin BM, Corvalan C, Grummer-Strawn LM. Dynamics of the double burden of malnutrition and the changing nutrition reality. *The Lancet*. 2020 Jan 4;395(10217):65–74.
5. The double burden of malnutrition: policy brief [Internet]. [cited 2022 Jul 19]. Available from: <https://www.who.int/publications-detail-redirect/WHO-NMH-NHD-17.3>
6. UNICEF. UNICEF Nutrition Strategy 2020-2030.pdf [Internet]. [cited 2025 Feb 11]. Available from: <https://www.unicef.org/media/92031/file/UNICEF%20Nutrition%20Strategy%202020-2030.pdf>
7. Cusick SE, Georgieff MK. The Role of Nutrition in Brain Development: The Golden Opportunity of the “First 1000 Days.” *J Pediatr*. 2016 Aug;175:16–21.
8. Morales F, Montserrat-de la Paz S, Leon MJ, Rivero-Pino F. Effects of Malnutrition on the Immune System and Infection and the Role of Nutritional Strategies Regarding Improvements in Children’s Health Status: A Literature Review. *Nutrients*. 2023 Dec 19;16(1):1.
9. Ebbeling CB, Pawlak DB, Ludwig DS. Childhood obesity: public-health crisis, common sense cure. *The Lancet*. 2002 Aug 10;360(9331):473–82.
10. World Bank Blogs [Internet]. [cited 2025 Feb 17]. World Bank country classifications by income level for 2024-2025. Available from: <https://blogs.worldbank.org/en/opendata/world-bank-country-classifications-by-income-level-for-2024-2025>
11. WHO. The Double Burden of Malnutrition Policy Brief [Internet]. [cited 2025 Feb 12]. Available from: <https://iris.who.int/bitstream/handle/10665/255413/WHO-NMH-NHD-17.3-eng.pdf?sequence=1>
12. Imdad A, Sadiq K, Bhutta ZA. Evidence-based prevention of childhood malnutrition. *Current Opinion in Clinical Nutrition & Metabolic Care*. 2011 May;14(3):276.
13. Hawkes C, Ruel MT, Salm L, Sinclair B, Branca F. Double-duty actions: seizing programme and policy opportunities to address malnutrition in all its forms. *Lancet*. 2020 Jan 11;395(10218):142–55.
14. Jebeile H, Kelly AS, O’Malley G, Baur LA. Obesity in children and adolescents: epidemiology, causes, assessment, and management. *The Lancet Diabetes & Endocrinology*. 2022 May 1;10(5):351–65.
15. Lenters L, Wazny K, Bhutta ZA. Management of Severe and Moderate Acute Malnutrition in Children. In: Black RE, Laxminarayan R, Temmerman M, Walker N, editors. *Reproductive,*

- Maternal, Newborn, and Child Health: Disease Control Priorities, Third Edition (Volume 2) [Internet]. Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2016 [cited 2025 Feb 21]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK361900/>
16. A Short History of SNAP | Food and Nutrition Service [Internet]. [cited 2025 Feb 13]. Available from: <https://www.fns.usda.gov/snap/history#1964>
 17. Supplemental Nutrition Assistance Program (SNAP) | Food and Nutrition Service [Internet]. [cited 2025 Feb 13]. Available from: <https://www.fns.usda.gov/snap/supplemental-nutrition-assistance-program>
 18. What Can SNAP Buy? | Food and Nutrition Service [Internet]. [cited 2025 Feb 13]. Available from: <https://www.fns.usda.gov/snap/eligible-food-items>
 19. WIC: USDA's Special Supplemental Nutrition Program for Women, Infants, and Children | Food and Nutrition Service [Internet]. [cited 2025 Feb 13]. Available from: <https://www.fns.usda.gov/wic>
 20. World Bank [Internet]. [cited 2025 Feb 13]. Social protection in Mexico: great opportunities and a new big challenge ahead. Available from: <https://www.worldbank.org/en/news/feature/2012/12/20/social-protection-in-mexico-great-opportunities-and-a-new-big-challenge-ahead>
 21. World Bank [Internet]. [cited 2025 Feb 13]. Takaful and Karama: A Social Safety Net project that Promotes Egyptian Women Empowerment and Human Capital. Available from: <https://www.worldbank.org/en/results/2020/11/17/takaful-and-karama-a-social-safety-net-project-that-promotes-egyptian-women-empowerment-and-human-capital>
 22. Valizadeh P, Ng SW. The New School Food Standards and Nutrition of School Children: Direct and Indirect Effect Analysis. *Econ Hum Biol.* 2020 Dec;39:100918.
 23. Bleich SN, Vercammen KA, Zatz LY, Frelter JM, Ebbeling CB, Peeters A. Interventions to prevent global childhood overweight and obesity: a systematic review. *The Lancet Diabetes & Endocrinology.* 2018 Apr 1;6(4):332–46.
 24. Hampl SE, Hassink SG, Skinner AC, Armstrong SC, Barlow SE, Bolling CF, et al. Clinical Practice Guideline for the Evaluation and Treatment of Children and Adolescents With Obesity. *Pediatrics.* 2023 Jan 9;151(2):e2022060640.
 25. Novo Nordisk [Internet]. [cited 2025 Feb 18]. News details. Available from: https://www.novonordisk-us.com/content/nncorp/us/en_us/media/news-archive/news-details.html
 26. Research C for DE and. FDA approves treatment for chronic weight management in pediatric patients aged 12 years and older. FDA [Internet]. 2022 Jun 27 [cited 2025 Feb 18]; Available from: <https://www.fda.gov/drugs/news-events-human-drugs/fda-approves-treatment-chronic-weight-management-pediatric-patients-aged-12-years-and-older>
 27. Research C for DE and. FDA approves first treatment for weight management for people with certain rare genetic conditions. FDA [Internet]. 2024 Aug 9 [cited 2025 Feb 18]; Available from:

<https://www.fda.gov/drugs/news-events-human-drugs/fda-approves-first-treatment-weight-management-people-certain-rare-genetic-conditions>

28. Windus JL, Stewart SJ, Adam MTP, Dodd CT, Burrows TL, Collins CE, et al. Exploring Dietary Assessment Methods Used to Measure Individual Dietary Intakes in Low- and Middle-Income Countries and Under-Served Populations in High-Income Countries. *Nutrients*. 2025 Jan;17(2):360.
29. Coates JC, Colaiezzi BA, Bell W, Charrondiere UR, Leclercq C. Overcoming Dietary Assessment Challenges in Low-Income Countries: Technological Solutions Proposed by the International Dietary Data Expansion (INDDEX) Project. *Nutrients*. 2017 Mar 16;9(3):289.
30. Raber M, Liao Y, Rara A, Schembre SM, Krause KJ, Strong L, et al. A systematic review of the use of dietary self-monitoring in behavioural weight loss interventions: delivery, intensity and effectiveness. *Public Health Nutrition*. 2021 Dec;24(17):5885–913.
31. Burke LE, Wang J, Sevick MA. Self-Monitoring in Weight Loss: A Systematic Review of the Literature. *Journal of the American Dietetic Association*. 2011 Jan 1;111(1):92–102.
32. Olukotun O, Seal N. A Systematic Review of Dietary Assessment Tools for Children Age 11 Years and Younger. *ICAN: Infant, Child, & Adolescent Nutrition*. 2015 Jun 1;7(3):139–47.
33. Rollo ME, Williams RL, Burrows T, Kirkpatrick SI, Bucher T, Collins CE. What Are They Really Eating? A Review on New Approaches to Dietary Intake Assessment and Validation. *Curr Nutr Rep*. 2016 Dec 1;5(4):307–14.
34. 24-hour Dietary Recall (24HR) At a Glance | Dietary Assessment Primer [Internet]. [cited 2021 Sep 26]. Available from: <https://dietassessmentprimer.cancer.gov/profiles/recall/>
35. ASA24® Dietary Assessment Tool | EGRP/DCCPS/NCI/NIH [Internet]. [cited 2025 Feb 17]. Available from: <https://epi.grants.cancer.gov/asa24/>
36. Walker JL, Ardouin S, Burrows T. The validity of dietary assessment methods to accurately measure energy intake in children and adolescents who are overweight or obese: a systematic review. *Eur J Clin Nutr*. 2018 Feb;72(2):185–97.
37. Livingstone MBE, Robson PJ, Wallace JMW. Issues in dietary intake assessment of children and adolescents. *British Journal of Nutrition*. 2004 Oct;92(S2):S213–22.
38. Pérez-Rodrigo C, Artiach Escauriaza B, Artiach Escauriaza J, Polanco Allúe I. Dietary assessment in children and adolescents: issues and recommendations. *Nutr Hosp*. 2015 Feb 26;31 Suppl 3:76–83.
39. Adamo KB, Brett KE. Parental Perceptions and Childhood Dietary Quality. *Matern Child Health J*. 2014 May 1;18(4):978–95.
40. Eliason J, Acciai F, DeWeese RS, Vega-Lopez S, Ohri-Vachaspati P. Children’s Consumption Patterns and Their Parent’s Perception of a Healthy Diet. *Nutrients*. 2020 Aug;12(8):2322.
41. El Kinany K, Garcia-Larsen V, Khalis M, Deoula MMS, Benslimane A, Ibrahim A, et al. Adaptation and validation of a food frequency questionnaire (FFQ) to assess dietary intake in Moroccan adults. *Nutrition Journal*. 2018 Jun 12;17(1):61.

42. Maruvada P, Lampe JW, Wishart DS, Barupal D, Chester DN, Dodd D, et al. Perspective: Dietary Biomarkers of Intake and Exposure— Exploration with Omics Approaches. *Advances in Nutrition*. 2020 Mar 1;11(2):200–15.
43. Diener C, Holscher HD, Filek K, Corbin KD, Moissl-Eichinger C, Gibbons SM. Metagenomic estimation of dietary intake from human stool. *Nat Metab*. 2025 Feb 18;1–14.
44. Reese AT, Kartzinel TR, Petrone BL, Turnbaugh PJ, Pringle RM, David LA. Using DNA Metabarcoding To Evaluate the Plant Component of Human Diets: a Proof of Concept. *mSystems* [Internet]. 2019 Oct 29 [cited 2021 Jan 21];4(5). Available from: <https://msystems.asm.org/content/4/5/e00458-19>
45. Schneider J, Mas-Carrió E, Jan C, Miquel C, Taberlet P, Michaud K, et al. Comprehensive coverage of human last meal components revealed by a forensic DNA metabarcoding approach. *Sci Rep*. 2021 Dec;11(1):8876.
46. Petrone BL, Aqeel A, Jiang S, Durand HK, Dallow EP, McCann JR, et al. Diversity of plant DNA in stool is linked to dietary quality, age, and household income. *Proceedings of the National Academy of Sciences*. 2023 Jul 4;120(27):e2304441120.
47. Staats M, Arulandhu AJ, Gravendeel B, Holst-Jensen A, Scholtens I, Peelen T, et al. Advances in DNA metabarcoding for food and wildlife forensic species identification. *Anal Bioanal Chem*. 2016 Jul 1;408(17):4615–30.
48. Pompanon F, Deagle BE, Symondson WOC, Brown DS, Jarman SN, Taberlet P. Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology*. 2012;21(8):1931–50.
49. Ando H, Mukai H, Komura T, Dewi T, Ando M, Isagi Y. Methodological trends and perspectives of animal dietary studies by noninvasive fecal DNA metabarcoding. *Environmental DNA*. 2020;2(4):391–406.
50. Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, et al. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res*. 2007 Feb;35(3):e14.
51. Kartzinel TR, Pringle RM. Multiple dimensions of dietary diversity in large mammalian herbivores. *Journal of Animal Ecology*. 2020;89(6):1482–96.
52. McCann JR, Bihlmeyer NA, Roche K, Catherine C, Jawahar J, Kwee LC, et al. The Pediatric Obesity Microbiome and Metabolism Study (POMMS): Methods, Baseline Data, and Early Insights. *Obesity*. 2021;29(3):569–78.
53. Aqeel A, Kay MC, Zeng J, Petrone BL, Yang C, Truong T, et al. Grocery intervention and DNA-based assessment to improve diet quality in pediatric obesity: a pilot randomized controlled study. *Obesity*. 2025;33(2):331–45.
54. St. Pierre C, Ver Ploeg M, Dietz WH, Pryor S, Jakazi CS, Layman E, et al. Food Insecurity and Childhood Obesity: A Systematic Review. *Pediatrics*. 2022 Jun 13;150(1):e2021055571.
55. Tester JM, Rosas LG, Leung CW. Food Insecurity and Pediatric Obesity: a Double Whammy in the Era of COVID-19. *Curr Obes Rep*. 2020;9(4):442–50.

56. Ortiz-Marrón H, Ortiz-Pinto MA, Urtasun Lanza M, Cabañas Pujadas G, Valero Del Pino V, Belmonte Cortés S, et al. Household food insecurity and its association with overweight and obesity in children aged 2 to 14 years. *BMC Public Health*. 2022 Oct 17;22(1):1930.
57. Coleman-Jensen A, Rabbitt MP, Gregory CA, Singh A. Household Food Security in the United States in 2021 [Internet]. [cited 2023 Jun 22]. Available from: <http://www.ers.usda.gov/publications/pub-details/?pubid=104655>
58. Eicher-Miller HA, Zhao Y. Evidence for the age-specific relationship of food insecurity and key dietary outcomes among US children and adolescents. *Nutrition Research Reviews*. 2018 Jun;31(1):98–113.
59. Tester JM, Xiao L, Tinajero-Deck L, Juarez L, Rosas LG. Food Insecurity Influences Weight Trajectory in Children with Obesity. *Childhood Obesity*. 2022 Oct;18(7):437–44.
60. Barlow SE, and the Expert Committee. Expert Committee Recommendations Regarding the Prevention, Assessment, and Treatment of Child and Adolescent Overweight and Obesity: Summary Report. *Pediatrics*. 2007 Dec 1;120(Supplement_4):S164–92.
61. Evans A, Banks K, Jennings R, Nehme E, Nemecek C, Sharma S, et al. Increasing access to healthful foods: a qualitative study with residents of low-income communities. *International Journal of Behavioral Nutrition and Physical Activity*. 2015 Jul 27;12(1):S5.
62. Palmer SM, Knoblauch ST, Winham DM, Hiller MB, Shelley MC. Putting Knowledge into Practice: Low-Income Women Talk about Food Choice Decisions. *International Journal of Environmental Research and Public Health*. 2020 Jan;17(14):5092.
63. Van Hul M, Cani PD. The gut microbiota in obesity and weight management: microbes as friends or foe? *Nat Rev Endocrinol*. 2023 May;19(5):258–71.
64. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014 Jan;505(7484):559–63.
65. Thompson FE, Midthune D, Kahle L, Dodd KW. Development and Evaluation of the National Cancer Institute’s Dietary Screener Questionnaire Scoring Algorithms. *The Journal of Nutrition*. 2017 Jun 1;147(6):1226–33.
66. Dietary Guidelines for Americans, 2020–2025.
67. Hu FB, Rimm EB, Stampfer MJ, Ascherio A, Spiegelman D, Willett WC. Prospective study of major dietary patterns and risk of coronary heart disease in men. *The American Journal of Clinical Nutrition*. 2000 Oct 1;72(4):912–21.
68. CDC. Centers for Disease Control and Prevention. 2023 [cited 2024 Apr 19]. New Adult Obesity Maps. Available from: <https://www.cdc.gov/obesity/data/prevalence-maps.html>
69. Food-is-Medicine_Peer-Reviewed-Research-in-the-U.S.1.pdf [Internet]. [cited 2023 Jul 27]. Available from: https://chlpi.org/wp-content/uploads/2013/12/Food-is-Medicine_Peer-Reviewed-Research-in-the-U.S.1.pdf
70. Rolls BJ. Dietary energy density: Applying behavioural science to weight management. *Nutrition Bulletin*. 2017;42(3):246–53.

71. Khoury CK, Bjorkman AD, Dempewolf H, Ramirez-Villegas J, Guarino L, Jarvis A, et al. Increasing homogeneity in global food supplies and the implications for food security. *Proceedings of the National Academy of Sciences*. 2014 Mar 18;111(11):4001–6.
72. Hill AM, Katcher HI, Flickinger BD, Kris-Etherton PM. 20 - Human Nutrition Value of Soybean Oil and Soy Protein. In: Johnson LA, White PJ, Galloway R, editors. *Soybeans* [Internet]. AOCS Press; 2008 [cited 2023 Nov 29]. p. 725–72. Available from: <https://www.sciencedirect.com/science/article/pii/B9781893997646500238>
73. Gupta S, Variyar PS. Chapter 12 - Guar Gum: A Versatile Polymer for the Food Industry. In: Grumezescu AM, Holban AM, editors. *Biopolymers for Food Design* [Internet]. Academic Press; 2018 [cited 2023 Nov 29]. p. 383–407. (Handbook of Food Bioengineering). Available from: <https://www.sciencedirect.com/science/article/pii/B9780128114490000128>
74. List GR. 1 - Soybean Lecithin: Food, Industrial Uses, and Other Applications. In: Ahmad MU, Xu X, editors. *Polar Lipids* [Internet]. Elsevier; 2015 [cited 2023 Nov 29]. p. 1–33. Available from: <https://www.sciencedirect.com/science/article/pii/B9781630670443500054>
75. Abrams SA. Bone Health in School Age Children: Effects of Nutritional Intake on Outcomes. *Front Nutr*. 2021 Nov 19;8:773425.
76. Patridge EF, Bardyn TP. Research Electronic Data Capture (REDCap). *Journal of the Medical Library Association*. 2018 Jan 12;106(1):142–4.
77. Hager ER, Quigg AM, Black MM, Coleman SM, Heeren T, Rose-Jacobs R, et al. Development and Validity of a 2-Item Screen to Identify Families at Risk for Food Insecurity. *Pediatrics*. 2010 Jul 1;126(1):e26–32.
78. Dolinsky DH, Armstrong SC, Walter EB, Kemper AR. The Effectiveness of a Primary Care–Based Pediatric Obesity Program. *Clin Pediatr (Phila)*. 2012 Apr 1;51(4):345–53.
79. National Heart, Lung and Blood Institute. Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents: Summary Report. In: *Pediatric Clinical Practice Guidelines & Policies* [Internet]. 14th ed. American Academy of Pediatrics; 2014 [cited 2023 Aug 28]. p. 1099–1099. Available from: <https://publications.aap.org/aapbooks/book/482/chapter/5802783/Expert-Panel-on-Integrated-Guidelines-for>
80. Soltani S, Arablou T, Jayedi A, Salehi-Abargouei A. Adherence to the dietary approaches to stop hypertension (DASH) diet in relation to all-cause and cause-specific mortality: a systematic review and dose-response meta-analysis of prospective cohort studies. *Nutrition Journal*. 2020 Apr 22;19(1):37.
81. Tan J, McKenzie C, Potamitis M, Thorburn AN, Mackay CR, Macia L. Chapter Three - The Role of Short-Chain Fatty Acids in Health and Disease. In: Alt FW, editor. *Advances in Immunology* [Internet]. Academic Press; 2014 [cited 2023 Jul 20]. p. 91–119. Available from: <https://www.sciencedirect.com/science/article/pii/B9780128001004000039>
82. Devkota S, Wang Y, Musch MW, Leone V, Fehlner-Peach H, Nadimpalli A, et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10^{-/-} mice. *Nature*. 2012 Jul 5;487(7405):104–8.

83. Aslam H, Ruusunen A, Berk M, Loughman A, Rivera L, Pasco JA, et al. Unravelled facets of milk derived opioid peptides: a focus on gut physiology, fractures and obesity. *International Journal of Food Sciences and Nutrition*. 2020 Jan 2;71(1):36–49.
84. The Microsetta Initiative [Internet]. [cited 2023 Oct 5]. The Microsetta Initiative – Researching Global Microbiomes. Available from: <https://microsetta.ucsd.edu/>
85. Dietary Screener Questionnaire in the NHANES 2009-10: Background | EGRP/DCCPS/NCI/NIH [Internet]. [cited 2024 May 17]. Available from: <https://epi.grants.cancer.gov/nhanes/dietscreen/>
86. Data Processing & Scoring Procedures Using Current Methods (Recommended) | EGRP/DCCPS/NCI/NIH [Internet]. [cited 2023 Aug 29]. Available from: <https://epi.grants.cancer.gov/nhanes/dietscreen/scoring/current/>
87. Hewawitharana SC, Thompson FE, Loria CM, Strauss W, Nagaraja J, Ritchie L, et al. Comparison of the NHANES dietary screener questionnaire to the Automated Self-Administered 24-Hour Recall for Children in the Healthy Communities Study. *Nutrition Journal*. 2018 Nov 27;17(1):111.
88. Food Security in the U.S. - Survey Tools | Economic Research Service [Internet]. [cited 2024 Feb 21]. Available from: <https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/survey-tools>
89. Baer TE, Scherer EA, Fleegler EW, Hassan A. Food Insecurity and the Burden of Health-Related Social Problems in an Urban Youth Population. *Journal of Adolescent Health*. 2015 Dec;57(6):601–7.
90. Holmes ZC, Silverman JD, Dressman HK, Wei Z, Dallow EP, Armstrong SC, et al. Short-Chain Fatty Acid Production by Gut Microbiota from Children with Obesity Differs According to Prebiotic Choice and Bacterial Community Composition. *mBio*. 2020 Aug 11;11(4):10.1128/mbio.00914-20.
91. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA*. 2011 Mar 15;108(supplement_1):4516–22.
92. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012 Aug;6(8):1621–4.
93. Release 138.1 [Internet]. [cited 2023 Aug 29]. Available from: <https://www.arb-silva.de/documentation/release-1381/>
94. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front Microbiol* [Internet]. 2017 [cited 2021 Jun 2];8. Available from: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02224/full>
95. Zhao J, Li Z, Gao Q, Zhao H, Chen S, Huang L, et al. A review of statistical methods for dietary pattern analysis. *Nutrition Journal*. 2021 Apr 19;20(1):37.

96. Schwedhelm C, Iqbal K, Knüppel S, Schwingshackl L, Boeing H. Contribution to the understanding of how principal component analysis–derived dietary patterns emerge from habitual data on food consumption. *The American Journal of Clinical Nutrition*. 2018 Feb;107(2):227–35.
97. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016 Apr 13;374(2065):20150202.
98. Bispo R, Marques F. Stability of principal components under normal and non-normal parent populations and different covariance structures scenarios. *Journal of Statistical Computation and Simulation*. 2023 May 3;93(7):1060–76.
99. Birhanu F, Yitbarek K, Bobo FT, Atlantis E, Woldie M. Undernutrition in children under five associated with wealth-related inequality in 24 low- and middle-income countries from 2017 to 2022. *Sci Rep*. 2024 Feb 9;14(1):3326.
100. Gallegos D. Effects of Food and Nutrition Insecurity on Global Health. *New England Journal of Medicine*. 2025 Feb 12;392(7):686–97.
101. Patlán-Hernández AR, Stobaugh HC, Cumming O, Angioletti A, Pantchova D, Lapègue J, et al. Water, sanitation and hygiene interventions and the prevention and treatment of childhood acute malnutrition: A systematic review. *Maternal & Child Nutrition*. 2022;18(1):e13257.
102. Subramanian S, Huq S, Yatsunenkov T, Haque R, Mahfuz M, Alam MA, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature*. 2014 Jun;510(7505):417–21.
103. Syed S, Ali A, Duggan C. Environmental Enteric Dysfunction in Children. *Journal of Pediatric Gastroenterology and Nutrition*. 2016;63(1):6–14.
104. Martins VJB, Toledo Florêncio TMM, Grillo LP, Franco M do CP, Martins PA, Clemente APG, et al. Long-Lasting Effects of Undernutrition. *Int J Environ Res Public Health*. 2011 Jun;8(6):1817–46.
105. Teshale G, Debie A, Dellie E, Gebremedhin T. Evaluation of the outpatient therapeutic program for severe acute malnourished children aged 6–59 months implementation in Dehana District, Northern Ethiopia: a mixed-methods evaluation. *BMC Pediatrics*. 2022 Jun 28;22(1):374.
106. Hussain I, Habib A, Ariff S, Khan GN, Rizvi A, Channar S, et al. Effectiveness of management of severe acute malnutrition (SAM) through community health workers as compared to a traditional facility-based model: a cluster randomized controlled trial. *Eur J Nutr*. 2021 Oct 1;60(7):3853–60.
107. Kangas ST, Salpèteur C, Nikièma V, Ritz C, Friis H, Briend A, et al. Predictors of time to recovery and non-response during outpatient treatment of severe acute malnutrition. *PLOS ONE*. 2022 May 31;17(5):e0267538.
108. Lenters LM, Wazny K, Webb P, Ahmed T, Bhutta ZA. Treatment of severe and moderate acute malnutrition in low- and middle-income settings: a systematic review, meta-analysis and Delphi process. *BMC Public Health*. 2013 Sep 17;13(3):S23.

109. A wonder 'food' for the world's children | UNICEF [Internet]. [cited 2024 Feb 28]. Available from: <https://www.unicef.org/nutrition/RUTF>
110. Purwestri RC, Scherbaum V, Inayati DA, Wirawan NN, Suryantan J, Bloem MA, et al. Impact of Daily versus Weekly Supply of Locally Produced Ready-to-Use Food on Growth of Moderately Wasted Children on Nias Island, Indonesia. *International Scholarly Research Notices*. 2013;2013(1):412145.
111. Beggs B, Bustos M, Brubacher LJ, Little M, Lau L, Dodd W. Facilitators and barriers to implementing complex community-based interventions for addressing acute malnutrition in low- and lower-middle income countries: A scoping review. *Nutr Health*. 2024 Jul 1;30(3):447–62.
112. Tadesse E, Berhane Y, Hjern A, Olsson P, Ekström EC. Perceptions of usage and unintended consequences of provision of ready-to-use therapeutic food for management of severe acute child malnutrition. A qualitative study in Southern Ethiopia. *Health Policy and Planning*. 2015 Dec 1;30(10):1334–41.
113. Manikam L, Sharmila A, Dharmaratnam A, Alexander EC, Kuah JY, Prasad A, et al. Systematic review of infant and young child complementary feeding practices in South Asian families: the Pakistan perspective. *Public Health Nutr*. 2018 Mar;21(4):655–68.
114. Iqbal NT, Syed S, Sadiq K, Khan MN, Iqbal J, Ma JZ, et al. Study of Environmental Enteropathy and Malnutrition (SEEM) in Pakistan: protocols for biopsy based biomarker discovery and validation. *BMC Pediatr*. 2019 Dec;19(1):1–17.
115. Sarfraz A, Ahmed S, Muhammad S, Rehman N, Soomro SI, Qureshi K, et al. Standard RUTF vs. locally-made RUSF for acutely malnourished children: A quasi-experimental comparison of the impact on growth and compliance in a rural community of Pakistan. *PLOS ONE*. 2023 Jul 12;18(7):e0287962.
116. Micha R, Coates J, Leclercq C, Charrondiere UR, Mozaffarian D. Global Dietary Surveillance: Data Gaps and Challenges. *Food Nutr Bull*. 2018 Jun;39(2):175–205.
117. Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*. 2011 Nov 1;39(21):e145–e145.
118. L. Petrone B, Bartlett A, Jiang S, Korenek A, Vintila S, Tenekjian C, et al. A pilot study of metaproteomics and DNA metabarcoding as tools to assess dietary intake in humans. *Food & Function*. 2025;16(1):282–96.
119. WHO Guideline for complementary feeding of infants and young children 6-23 months of age [Internet]. [cited 2025 Jan 22]. Available from: <https://www.who.int/publications/i/item/9789240081864>
120. Deb GK, Nahar TN, Duran PG, Presicce GA. Safe and Sustainable Traditional Production: The Water Buffalo in Asia. *Front Environ Sci* [Internet]. 2016 May 18 [cited 2024 Nov 18];4. Available from: <https://www.frontiersin.org/journals/environmental-science/articles/10.3389/fenvs.2016.00038/full>
121. Bento Lab [Internet]. [cited 2024 Dec 2]. Bento Lab | A Portable PCR Workstation. Available from: <https://bento.bio/>

122. Iqbal NT, Chen RY, Griffin NW, Hibberd MC, Khalid A, Sadiq K, et al. A shared group of bacterial taxa in the duodenal microbiota of undernourished Pakistani children with environmental enteric dysfunction. *mSphere*. 2024 May 14;9(6):e00196-24.
123. Soomro SI, Jamil Z, Memon N, Ahmed S, Umrani F, Choudhri IA, et al. Nutrient dataset development via FAO/INFOODS approach for infant nutritional survey in rural Matiari, Pakistan. *Journal of Food Composition and Analysis*. 2024 Sep 1;133:106471.
124. Returning Individual Research Results to Participants: Guidance for a New Research Paradigm [Internet]. [cited 2025 Feb 27]. Available from: <https://nap.nationalacademies.org/read/25094/chapter/1>
125. Hall CM, Rasmussen LV, Powell B, Dyngeland C, Jung S, Olesen RS. Deforestation reduces fruit and vegetable consumption in rural Tanzania. *Proceedings of the National Academy of Sciences*. 2022 Mar 8;119(10):e2112063119.
126. Jonas DA, Elmadfa I, Engel KH, Heller KJ, Kozianowski G, König A, et al. Safety considerations of DNA in food. *Ann Nutr Metab*. 2001;45(6):235–54.
127. Liu Y, Zhang Y, Dong P, An R, Xue C, Ge Y, et al. Digestion of Nucleic Acids Starts in the Stomach. *Sci Rep*. 2015 Jul 14;5:11936.

Biography

Ammara Aqeel grew up in Faisalabad, Pakistan. She earned her Bachelor of Science in Biology from Duke University in 2019 as a Karsh International Merit Scholar, graduating *summa cum laude* and as an early inductee to Phi Beta Kappa. Following graduation, she worked as an Associate in Research in Dr. Lingchong You's laboratory in the Department of Biomedical Engineering at Duke's Pratt School of Engineering. In 2020, she began her Ph.D. in Molecular Genetics and Microbiology at the Duke University School of Medicine and joined Dr. Lawrence David's lab in 2021 to pursue her doctoral research. During this time, Ammara also served in multiple leadership roles across campus: as Co-President of the Duke Advanced Degree Consulting Club (DACC), as a Senator representing the non-clinical School of Medicine in the Graduate and Professional Student Government, as a Ph.D. Fellow representing Ph.D. students at the Innovation and Entrepreneurship (I&E) Program at the Fuqua School of Business, and as a peer mentor at the Office of Biomedical Graduate Education, supporting first-year Ph.D. students. She also worked for two years as a Fellow at the Duke Office of Translation and Commercialization. Ammara will graduate from Duke with her Ph.D. in May 2025.