Evolution of infant gut microbiome and potential links to personalized health

By

Umme Habiba

(Registration No: 00000401234)

Department of Sciences

School of Interdisciplinary Engineering & Sciences (SINES)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

Evolution of infant gut microbiome and potential links to personalized health

By

Umme Habiba

(Registration No: 00000401234)

A thesis submitted to the National University of Sciences & Technology, Islamabad,

in partial fulfillment of the requirements for the degree of

Master of Science in

Bioinformatics

Supervisor: Dr. Masood Ur Rehman Kayani

School of Interdisciplinary Engineering & Sciences (SINES)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

THESIS ACCEPTANCE CERTIFICATE

Certified that the final copy of MS Thesis written by Ms Umme Habiba (Registration No. 00000401234), of SINES (School/College/Institute) has been vetted by the undersigned, found complete in all respects as per NUST Statutes/ Regulations/ Masters Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for the award of Masters degree. It is further certified that necessary amendments as pointed out by GEC members and foreign/ local evaluators of the scholar have also been incorporated in the thesis.

Signature: 24/07/24

Name of Supervisor: Dr. Masood Ur Rehman Kayani

Date: 24/07/24

Signature (HOD):

Date: 26-7-2024

Dr. Fouzia Malik
HOD Sciences
Professor
SINES NUST
Islamabad

Dr. SYED IRTIZA ALI SHAH
Principal & Dean
SINES NUST, Sector H-12
Islamabad.

Signature (Dean/ Principal)

Date: 26/7/24

AUTHOR'S DECLARATION

I Umme Habiba hereby state that my MS thesis titled " Evolution of infant gut microbiome and potential links to personalized health " is my work and has not been submitted previously by me for taking any degree from the National University of Sciences and Technology, Islamabad, or anywhere else in the country/ world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Name of Student: Umme Habiba

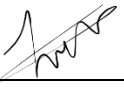
Date: June 24, 2024

PLAGIARISM UNDERTAKING

I solemnly declare that the research work presented in the thesis titled "**Evolution of infant gut microbiome and potential links to personalized health**" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and the National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above-titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred to/cited.

I undertake that if I am found guilty of any formal plagiarism in the above-titled thesis even after the award of MS degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and NUST, Islamabad have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature: 

Name: Umme Habiba

DEDICATION

I dedicate this thesis to my beloved parents and siblings, whose prayers, affection, and encouragement made my success possible, and to my teachers, whose support and guidance have been invaluable throughout my academic journey.

ACKNOWLEDGEMENTS

In the name of **Allah**, the Most Gracious, the Most Merciful. All praise is due to Him, the Lord of the Worlds, whose blessings have made this work possible. I am profoundly grateful for the guidance, patience, and strength that He has bestowed upon me throughout this journey. My deepest respect to the **Holy Prophet Muhammad (PBUH)** for his inspiring teachings and wisdom.

I am deeply grateful to my supervisor, **Dr. Masood Ur Rehman Kayani**, for his invaluable guidance, support, and encouragement. His insightful advice and unwavering patience have been crucial to the completion of this thesis. His mentorship has profoundly influenced my academic and personal growth.

I would also like to extend my sincere appreciation to my GEC members, **Prof. Ishrat Jabeen** and **Dr. Uzma Habib**, for their valuable feedback, helpful suggestions, and constant support. Their input has greatly enriched my work.

I am grateful to my **Microbiome Research Group** for their camaraderie and collaborative spirit. Special thanks to **Muhammad Adnan Tariq, Faheem Raziq**, **Haseeb Manzoor**, and **Muhammad Shahzaib** for their assistance, and support, and for creating an inspiring and productive research environment. I am also grateful to **Abdul Hanan, Muhammad Inam Rafique**, **Kashif Saleem,** and **Shanza Saeed**, whose unwavering support and encouragement have helped me navigate the challenges of my research.

TABLE OF CONTENTS

LIST OF TABLES

Page No.

LIST OF FIGURES

Page No.

ABSTRACT

The human gut microbiome plays a critical role in health and disease. This complex community of microorganisms is established early in life and undergoes significant changes during infancy, a period crucial for long-term health. In infants, the gut microbiome is particularly dynamic, with the mode of delivery being one of the initial factors influencing its composition. Despite ongoing research, there are gaps in understanding the extent of maternal microbial transmission, the selection process of these microbes, and the factors influencing this selection. Specifically, it remains unclear how the gut microbiome, once acquired, differs in terms of evolutionary patterns in infants born via vaginal delivery or C-section. The current body of research lacks thorough investigation into specific microbial taxa present in the infant gut, that drive the evolution of gut microbes, particularly in association with delivery mode, neonatal health, and disease outcomes. One way to explore these microbial dynamics is through the study of HGT events, which can reveal how genetic material is exchanged between different microbial species in the gut. Using the WAAFLE tool, this study identified major drivers of HGT in infants born through either of the delivery modes. In CSD, there were 447 known HGT drivers, VD, there were 360. These include transfer of genes, like *Salmonella enterica* transfers the *YagA* gene to *Klebsiella pneumoniae* in CSD infants. Post-translational modifications were observed with a positive log2 fold change of 6 in this HGT driver. This gene is involved in PFAM integrase catalytic activity and regulates genes related to metabolism, with its transcription being affected by stress and conditions leading to biofilm formation. In VD infants, *Butyricicoccus pullicaecorum* transfers the *lepB* gene to

Bifidobacterium longum, which is involved in signal peptidase activity, a valuable target for antimicrobial drug development and this transfer exhibited a positive log2 fold change of 5 in intracellular trafficking, secretion, and transport. Additionally, *Coprococcus catus* transfers the *alfA* gene to *Ruminococcus sp. 5_1_39BFAA* in VD infants, a gene associated with the degradation and metabolism of HMOs. *Ruminococcus sp_5_1_39BFAA* is negatively associated with lactose, showing a coefficient value of -1.84e+00 and an FDR of 9.052e-05. *Lachnospiraceae bacterium_2_1_46FAA* and *Ruminococcus gnavus* HGT drivers are found in VD infants and are positively associated with NICU-admitted infants, with coefficient values of 1.15e+00 and 1.12e+00, respectively, and FDR values of 2.281e-03 and 5.402e-03, respectively. The gene involved is xylB, which has antimicrobial properties and is involved in defense mechanisms. Notably, *Coprococcus catus* and *Lachnospiraceae bacterium_2_1_46FAA* were identified as a potential probiotic tailored to personalized health interventions.

Keywords: Infant gut microbiome, Vaginal delivery, C-section delivery, Horizontal gene transfer, Personalized health.

CHAPTER 1: INTRODUCTION

1.1 Human Gut Microbiome in health and disease

The human microbiome, which encompasses all the microorganisms residing in and on the human body, is linked to both the health and disease of the host. According to an estimate, the number of microorganisms that constitute the microbiome outnumber the total human cells by ten to one hundred [1]. Microorganisms, until recently considered to be pathogenic, have been regarded as acquired symbionts within the host [2]. The microbial community residing in the gut, collectively referred to as the human gut microbiome, has established a mutualistic association with the host and has been studied in the context of human health over the past few decades [3]. Among the key roles that the gut microbiome plays in digestion is the contribution to the digestive process through the production of enzymes responsible for carbohydrate degradation, such as carbohydrate-active enzymes (CAZymes) [4]); phosphate acetyltransferase for degrading dietary fiber; and malate L-lactate dehydrogenase, which is important in butanoate metabolism. [5]. The gut microbiome is also well known for producing short-chain fatty acids (SCFAs) through its members of *Bacteroides, Bifidobacterium*, and *Ruminococcus.* These SCFAs do much more, including metabolic processes, signaling, and growth promotion of certain bacteria. [6]. In addition, the microbiome is associated with energy harvest and the synthesis of useful B vitamins, influencing metabolic regulations, particularly through bile acid metabolism, which affects metabolic regulations and energy expenditure. [7]. The interaction of the gut microbiome with the immune system of the host is bidirectional and impacts the development and regulatory control of the immune response. The gut microbiome takes an active role in the protection of the host from potential pathogenic

invasions [8]. The exact molecular mechanisms are still unclear; however, recognition of commensal microorganisms by the innate immune system is thought to be central for immune system development [9]. Microbiome signals, which include SCFAs, affect myeloid-cell differentiation, innate lymphoid cell maturation, and the mucosal epigenome, influencing disease exposure and treatment responses [10].

Changes in the composition of the gut microbiome, its disrupted balance, or imbalances are associated with various diseases. In some—but not all—cases, the microbiome is believed to actively promote the etiology of a disease, mostly through inflammatory pathways include colorectal cancer [12] Crohn's disease [11], and autism [13]. Some bacterial species, usually part of the commensal microbiome, convert into opportunistic pathogens under defined circumstances. For example, *Helicobacter pylori* in the stomach, a commensal for the most part of the host's life, becomes a major risk factor of gastric adenocarcinoma under certain circumstances [14]. The gut microbiome has also been linked to obesity and metabolic syndromes. Excessive adiposity may alter the microbial community, creating a feedback loop that reinforces obesity [15]. Notably, Lachnospiraceae have been observed to contribute to the development of hyperglycemia [16]. The inflammatory pathways likely contribute to this relationship, given that obesity is characterized by low-grade inflammation [17]. A comprehensive metagenome-wide association study identified specific microbiome characteristics associated with type-2 diabetes, including reduced butyrate production, increased abundance of pathogenic bacteria, and enrichment in sulfate reduction and oxidative stress resistance [18]. The gut microbiome has long been linked to chronic inflammatory conditions such as Inflammatory Bowel Disease (IBD). Studies reveal distinctions between individuals with IBD and those

who are healthy, such as the reduced abundance of *F. prausnitzii* in Crohn's Disease and IBD, and increased β-lactamase-producing [17] and Enterobacteriaceae in severe ulcerative colitis (UC) [19].

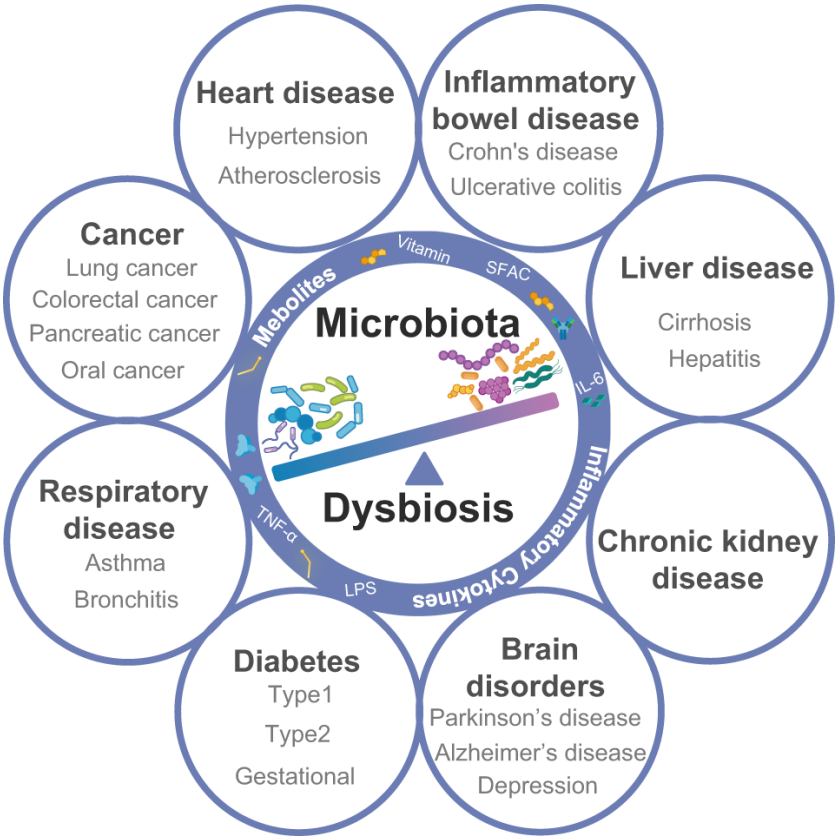


Figure 1.1 Dysbiosis in the human microbiome and its role in disease development.

1.2 Importance of the Infant Gut Microbiome

The development of both the immune system and gut microbiome in infants commences during the initial 1000 days following conception, encompassing the periods of pregnancy and the initial two years of life [20]. This "window of opportunity" is a crucial period for ensuring the healthy development of infants. At birth, both beneficial and harmful microbes can colonize the infant's gut and can confound the infant's health [21].

1.2.1 Composition, Function, and Importance of infant gut microbiome

The acquisition of the gut microbiome by infants typically begins at birth. [22]. The microbes encountered by infants at birth, particularly those from the maternal source, may significantly impact the infant's health and disease, or lead to long-lasting consequences [23]. Important members of the core composition of the infant gut microbiome include *Bifidobacteriales, Lactobacillales, Clostridiales, Prevotella*, *Bacteroidales*, and *Bacteroides Fragilis*, among others, although these may differ based on what the dominant populations and compositions are. The development and maturation of immunity acquired and innate in infancy are significantly aided by the presence of commensal or beneficial bacteria; for example, *Lactobacillus* and *Bifidobacterium* [24]. The acquisition of Clostridia species prevents infant colonization by bacterial pathogens. *Clostridiales* raise the colonization resistance of the gut and are considered a defensive mechanism against the attacks of certain pathogens on the gut of the infant [25]. Digestion of Human Milk Oligosaccharides (HMOs) takes place in the infant gut as a result of high abundance of *Bacteroides* and *Fragilis* [26]. Moreover, *Bacteroides thetaiotaomicron* facilitates the growth of short-chain fatty acid (SCFA)-producing bacteria when carbohydrates and human milk oligosaccharides (HMOs) are present. [27]. *Prevotella* plays a crucial role in glucose metabolism, as it can degrade pyruvate to acetate and formate [28]. Additionally, it is significant in various pathways related to drug, carbohydrate, and vitamin metabolism.

1.2.2 Factors influencing the development of infant gut microbiome

Factors influencing infant gut colonization include mode of delivery, vaginal delivery (VD) and C-section delivery (CSD), breastfeeding, gestational age, maternal and infant diets,

antibiotic use, environmental exposures, and maternal stress levels, all of which contribute to the establishment of the infant gut microbiome. (Figure 2.1).

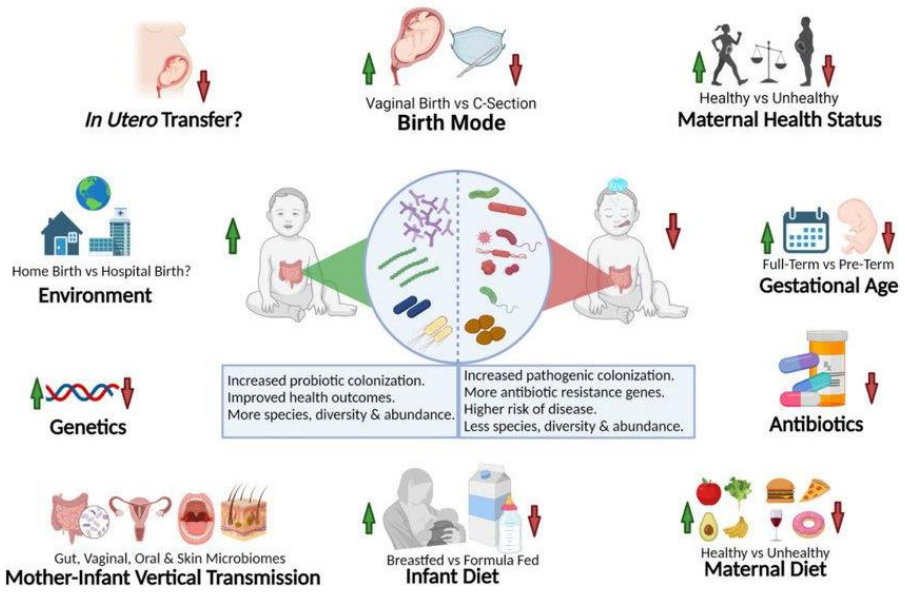


Figure 1.2 Factors shaping the infant gut microbiome affect microbial diversity, species abundance, and health outcomes.

Most importantly, the mode of delivery significantly influences the infant's gut microbiome. Infants delivered via VD "inherit" their microbiome from their mother's birth canal, acquiring bacteria such as *Lactobacillus* and *Prevotella* species. [29]. Conversely, infants delivered by CSD are colonized by different bacteria, such as *Corynebacterium, Staphylococcus*, *Propionibacterium,* and *Clostridium*, and exhibit lower levels of anaerobic bacteria like *Bacteroides* and *Bifidobacterium*.[30].

Breast milk fosters the growth of the gut microbiome by supplying probiotics and prebiotics, which help protect against pathogens. Breast milk predominantly contains microorganisms such as *Lactobacillus, Streptococcus, Bacteroides, Staphylococcus, Enterococcus, Clostridium,* and *Bifidobacterium*. [31]. Breast milk carries a unique set of microorganisms along with HMOs that are passed on to infants. These HMOs, which act

as prebiotics, encourage the growth of beneficial gut bacteria, prevent harmful pathogens from establishing themselves in the infant's gut, and promote overall health benefits. The composition of an infant's gut microbiome is influenced by exposure to various external environments during early development outside the uterus. Having siblings is associated witan increased Bifidobacterium and reduced abundance of *Peptostreptococcus* bacteria in infants [32]. The infant gut microbiome also shows diversity depending on geographical location, dietary patterns, and lifestyle. Studies indicate variances in the microbiome of rural versus urban infants. Furthermore, a study involving 605 infants from five European countries with diverse lifestyles and feeding practices found that infants from Northern European countries had a higher prevalence of *Bifidobacteria*.. In contrast, infants from Southern European countries had increased levels of *Bacteroides* and *Lactobacilli* [33]. When preterm infants are admitted to the neonatal intensive care unit (NICU), infants encounter microorganisms present in the hospital environment, with restricted exposure to microorganisms specific to their mother and family but increased contact with NICU staff microorganisms. These microorganisms include *Staphylococcus aureus, Enterococcus, Klebsiella, Pseudomonas aeruginosa, Acinetobacter,* and other *Enterobacteriaceae*. They are frequently found on NICU surfaces and are among the most common sources of nosocomial infections. CSD infants are more prone to receiving antibiotics, thereby elevating the risk of future health issues such as asthma [34], [35], obesity[35], and inflammatory bowel disease [36]. Antibiotics administered to infants have been associated with increased levels of *Enterobacteria,* whereas those given to mothers during pregnancy or breastfeeding led to a reduction in *Bacteroides* in their infants.

1.3 Mode of delivery and the infant gut microbiome

The mode of delivery is a significant factor influencing the composition of infant microbiome, with CSD being a crucial obstetric intervention for the safety of both mother and child. Despite its life-saving nature, there has been a recent trend of overuse of CSD. The proportion of infants delivered through CSD has been steadily increasing over the years. The rates of mothers undergoing a CSD were approximately 3.2% in 1990–91, 7.8% in 2006–07, 13.6% in 2012–13, and 19.6% in 2017–18[37]. These variations in delivery mode have been associated with differences in the gut microbiome of infants (Figure 3.1).

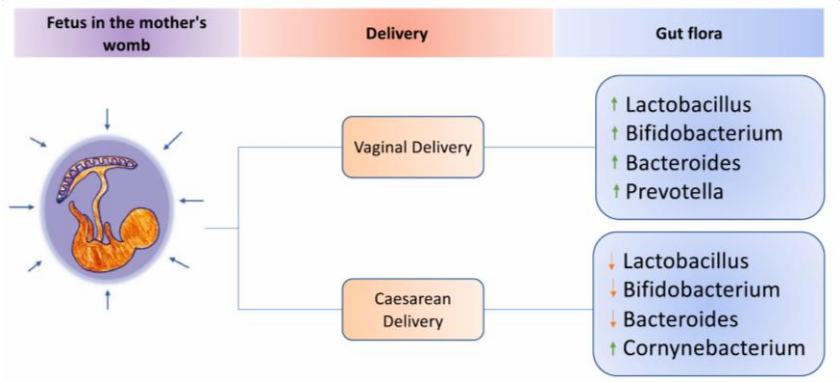


Figure 1.3 The Influence of Delivery Mode on the Formation of the Infant Gut Microbiome

1.3.1 Vaginal Delivery (VD) and Infant gut microbiome

Infants delivered through VD tend to have more diverse microbiome [38]. Greater richness and diversity of gut bacteria are often indicative of an individual's ability to defend against pathogenic invaders. Notably, Numerous studies have reported a notable increase in bacterial richness and diversity in both VD and CSD infants following the introduction of solid food. [39]. These findings suggest that while mode of delivery is not the sole a key factor influencing microbial composition, it remains one of the main determinants until the introduction of solid food. Hill et al. [40] found that the mode of delivery creates notable

differences in the infant microbiome, but these differences become less pronounced after the first 24 weeks of life. In contrast, additional studies have noted taxonomic disparities linked to the mode of delivery that continue up to seven years of age [41]. The duration of these microbial differences remains a matter of debate; nonetheless, it is clear that the mode of delivery does impact microbial composition. Furthermore, study by Dominguez-Bello et al. [42] identified evidence supporting the the direct transmission of bacteria from mother to infant. They discovered that in 75% of vaginal deliveries, the mother's vaginal bacteria were more similar to her infant's microbiome compared to those of other VD infants. These findings indicate vertical transmission of the unique vaginal microbiome. Another study found similar results, concluding that vaginal *Lactobacillus* species are passed from mother to infant during vaginal delivery. [43]. Dominguez-Bello et al. [44] also found that infants born through vaginal delivery had microbial communities with a high abundance of *Lactobacillus, Prevotella, Atopobium,* and *Sneathia* species. In their study, Dzidic et al. [45] similarly observed that the gut of vaginally delivered infants is predominantly dominated by vaginal bacteria, including species such as *Bifidobacterium, Klebsiella, Lactobacillus, Escherichia,* and *Prevotella*. According to study, the abundance of *Escherichia* was only associated with VD and age [46].

1.3.2 CSD and Infant gut microbiome

The difference in the infant gut microbiome in the VD and CSD categories is often considered to play a crucial role in the development of various pathologies related to metabolism, as well as the immune systems [47]. The difference is likely due to disrupted normal colonization, occurring when infants delivered by CSD, do not get exposed to maternal vaginal microbes, and encounter skin and environmental microbes [48]. Another

study has identified higher levels of *Haemophilus* and *Clostridium* species in adults delivered by CSD [49]. A study by Dominguez-Bello et al. [42] found that the microbiome of infants delivered by CSD was found to have higher proportions of *Staphylococcus* species. Dzidic et al. [45] also noted that the gut microbiome of infants born via CSD is mainly dominated by *Propionibacterium*, *Corynebacterium*, and *Staphylococcus*, which are commonly found on the skin. Other studies have reported an increased presence of *Veillonella* in infants delivered by CSD, in addition to the previously noted *Staphylococcus*, *Streptococcus*, *Corynebacterium*, and *Propionibacterium*. Shi et al. also observed that infants in this category frequently exhibited abundant colonization by *Bacillus licheniformis*, which is suggested to lead to microbial dysbiosis. Another study found that CSD-born infants had an increased abundance of *Clostridium perfringens*, which is toxigenic bacterium and associated with gastrointestinal illnesses [50]. Grönlund et al. [51] and Adlerberth et al. [52] revealed that although infants delivered by CSD eventually acquired *Lactobacillus*, *Bacteroides*, and *Bifidobacterium* in their gut, this colonization was delayed compared to VD infants. Similarly, Makino et al. [53] observed a delay in the colonization of *Bifidobacterium* in CSD infants, with slower bacterial counts that did not reach the levels observed in VD infants.

1.3.3 Association with health and diseases

Several research studies have highlighted the potential influence of the gut microbiome on infants' health, either through immune regulation or the microbiome-gut-brain axis [54]. For example, indole-3-lactic acid (ILA) produced by Bifidobacterium has been demonstrated to boost the expression of immunoregulatory galectin-1 in Th2 and Th17 cells during polarization, resulting in positive effects on infant regulatory immunity. [55].

Research on cytokine concentration observed higher blood serum levels of sIL-2R, sIL-4R, INF- γ, IL-1β, IL-6, and TNF- α in VD infants compared to CSD [56]. These cytokines receptors are associated with inflammation: sIL-2R acts as a receptor for IL-2, which regulates white blood cell activity; sIL-4R serves as a receptor for IL-4 and exhibits anti-inflammatory properties; INF-γ aids in fighting viral pathogens and regulating immune responses; IL-1β is believed to initiate labor; IL-6 may regulate placental and fetal growth; and TNF-α exerts pro-inflammatory effects [56]. *Bacteroides fragilis* significantly contributes to the development and maintenance of immune function, primarily due to its production of polysaccharide A. [57]. Polysaccharide A is vital for triggering T cell responses, ensuring T cell balance within the body, and preventing inflammation [58]. These variations in the development and function of the immune system are believed to have consequences beyond infancy and early childhood. Studies suggest that disruptions in the typical establishment of the immune system could lead to abnormalities in normal physiology throughout life. Therefore, it is crucial to explore and acknowledge the potential adverse health consequences associated with CSD.

Mild upper respiratory tract infections (URTIs) can be compared to potentially fatal illnesses like pneumonia or lower respiratory tract infections (LRTIs). Despite an overall decline in global mortality, it still ranks first among the causes of death for children under the age of five, accounting for five million of these deaths [59]. In a study conducted in New Hampshire, USA, on infants from the general population, early microbiome patterns were found to be associated with wheezing, diarrhea, and infant respiratory infections. During the initial year of life, a higher abundance of respiratory infections and related symptoms were linked to increased diversity in the gut microbiome of early infants [60].

The delivery of mode has been identified as a major influence on the development of the microbiome. [61]. In the context of respiratory health, it is hypothesized that lactic acid-producing bacteria, such *as Lactobacillus spp,* play a role in preventing infections [62]. VD infants have a higher colonization of *Lactobacillus* species compared to CSD. This variance in colonization is believed to potentially contribute to the increased prevalence of certain respiratory conditions and diseases in infants delivered by CSD [62]. One of the study also found a positive association between upper respiratory infections and *Veillonella,* more especially *Veillonella parvula* in CSD infants. According to a research study, CSD infants had a higher risk of upper respiratory tract infections when there was a higher relative abundance of *Corynebacterium* species [63].

In a meta-analysis of 26 epidemiological studies, Bager, Wohlfahrt, and Westergaard found that CSD infants had a moderately elevated risk of developing allergic rhinitis and asthma [64]. A study found that *Clostridium difficile* produces two unique exotoxins that are known to compromise the integrity of the intestinal epithelial cell barrier. These exotoxins initiate and then escalate the inflammatory response, which is thought to play a role in food hypersensitivity. [65]. Lundgren et al. revealed that infants delivered by CSD are more prone to developing dairy allergies compared to VD. The study hypothesized that this difference may arise from the reduced abundance of milk-digesting *Lactobacillus* in CSD infants. Similarly, Wampach et al. [66] observed a correlation between allergic diseases and lower levels of *Bacteroides* and *Bifidobacterium* species. Because such bacteria are not allowed to colonize in the case of CSD-born infants, these bacteria foster the development of allergic disease in CSD infants. Melli et al. [67] analyzed 21 different studies related to the microbial composition of the gut microbiome in allergic condition

cases. The study found that the microbial diversity level remains quite low in allergic infants compared to non-allergic infants. This was indicated by a higher number of *Bacteroidaceae* and an excess of *Firmicutes*. Chen et al. [68] studied 23 infants with food allergies and 22 healthy infants in a case-control study that was conducted in 2016. The finding indicated that alterations in the gut microbiome were linked to food allergies. Infants with food allergies have a lower diversity of the total microbiome. In infants with food allergies, there is a clear low level of *Bacteroidetes* bacteria and a high level of *Proteobacteria, Actinobacteria,* and *Firmicutes* at the phylum level when compared to healthy infants. Furthermore, variations have been reported among infants with food allergies in terms of the genera, such as the decreased numbers of *Veillonella* and *Bacteroides* and the increased numbers of *Subdoligranulum* and *Clostridium IV*. In a recent study, they assessed four clinically distinct allergic diseases identified in the large, comprehensively characterized CHILD cohort study at the age of five. They found that in infants who went on to develop allergic diseases, the pattern of changes in maturation is characterized by an increase in *C. innocuum, E. lenta, T. nexilis, E. faecalis,* and *E. coli*, as well as reduction in the bacterial species *F. saccharivorans, B. Wexlerae, E. hallii,* and *A. hadrus* [69]. It has been demonstrated recently that infants with food allergies show increased abundances of *Faecalibacterium prausnitzii* and *Ruminococcus gnavus*, along with reduction of several species of *Bacteroides, Bifidobacterium,* and other taxa with the ability to degrade fiber [70]. These results suggest that an elevated risk of allergic conditions is linked to imbalances in the gut microbiome.

Asthma, the most common chronic condition in infants, is a significant noncommunicable disease (NCD) impacting both adults and infants. About 52.9% of children with asthma

under five reported having experienced an attack [71]. The Centres for Disease Control and Prevention (CDC) report that between 2001 and 2020, there was a decrease in childhood asthma attacks [72]. Even though asthma is manageable, 50% of children with the condition are thought to have uncontrolled asthma [73]. McCauley et al. [74] characterized four distinct *Lactobacillus*-dominated vaginal microbiome clusters with unique compositions and functions. This was done by analyzing paired samples from mothers' vaginal microbiomes and infants' stool samples. These clusters exhibited distinct correlations with prenatal maternal exposures, including diet, stress, and farm exposure, as well as with infant IgE levels at one year of age. Mothers showed vaginal clusters predominantly characterized by *G. vaginalis* or *L. fornicalis*, according to McCauley et al. These clusters were linked to higher allergic sensitizations in their infants. Bacterial vaginosis is caused by the pathogen *G. vaginalis*, which also causes microbiome dysbiosis by inducing microbial production of the inflammatory endotoxin lipopolysaccharide (LPS) [75]. Comparing the gut microbiome of 319 infants, Arrieta et al. [76] discovered that within the first 100 days of life, infants at risk for asthma exhibit a temporary imbalance in their gut microbiome. Infants with a susceptibility to asthma display significantly reduced relative abundances of the bacterial genera *Rothia, Faecalibacterium, Veillonella*, and *Lachnospira*. Researchers found from a cohort of 319 human participants in the Canadian Healthy Infant Longitudinal Development (CHILD) Cohort that supports a connection between an elevated likelihood of asthma development and gut microbial imbalance in the initial 100 days of life, which is defined by reductions in four bacterial genera: *Rothia, Veillonella, Lachnospira,* and *Faecalibacterium* [77]. According to a study by Montoya-Williams et al. [78], the anti-inflammatory properties of *Lactobacillus* are proposed to

potentially inhibit the onset of asthma. In contrast, *Staphylococcus* and *Clostridium* species have been associated with conditions such as atopic dermatitis and asthma [79]. These results suggest that an elevated risk of allergic conditions is linked to imbalances in the gut microbiome.

The most prevalent illness following delivery, neonatal jaundice typically appears within the initial week of life. Jaundice develops within the initial week of life in around 60% of term infants and 80% of preterm infants. Jaundice in newborns is identified by elevated levels of total serum bilirubin [80]. Research has been conducted to explore the potential correlation between elevated levels of direct bilirubin and microbes, like *Bifidobacterium*[81]. According to a recent study, the gut microbiome dysbiosis linked to jaundice was characterized by a proliferation of potentially harmful bacteria, the *Enterobacteriaceae* family, a low-level biodiversity, and a decline in the microbiome with a beneficial potential, the genera *Bifidobacterium* and *Faecalibacterium*[82].

According to the study, researchers discovered that two strains of *Rhodopseudomonas palustris*, two strains of *Veillonella parvula*, three species of *Mycobacterium*, and five species of *Streptococcus* were upregulated in jaundiced infants[83]. Although the gut microbiome and neonatal jaundice are thought to be related, still the gut microbial characteristics of this illness are poorly understood.

Diarrheal illnesses continue to be the primary cause of mortality for children under five, contributing to 1.8 million child fatalities globally[84]. Even though diarrhea-related deaths have decreased over time, it is still a common reason for pediatric urgent care visits, particularly in certain economically challenged nations in Asia and Africa [85]. The main cause of infectious diarrhea is enteric bacterial pathogens. Bacterial pathogen-induced

diarrhea is a global health concern, especially in developing countries. Currently, the most common pathogens associated with diarrhea are thought to be *Clostridium difficile, Salmonella*, *Aeromonas, Campylobacter,* and *Shigella*[86], [87]. Recent research examined the gut microbiome characteristics of infants with diarrhea and those without it in several ways. The most prevalent bacteria in the group that causes diarrhea were *Proteobacteria*, which are also the microbial marker of dysbiosis in the gut microbiome [88]. Upon comparing the outcomes of the two cases at the genus level, Qingjie Fanthey et al. discovered that diarrhea was linked to a rise in *Klebsiella* and *Enterobacter* and a decrease in *Lactobacillus* [89].

1.4 Research gap and problem statement

The gut microbiome changes constantly throughout life, still, the vertical transfer from mother to infant is unquestionably a crucial developmental step. However, the extent of maternal transmission, the selection process of these microbes, and the factors influencing this selection remain unknown. Although the mode of delivery is the initial factor influencing an infant's gut microbiome, research suggests that infants born through either CSD or VD may exhibit similar microbes after a few years. A study has found that levels of *Lactobacillus* and *Bifidobacterium* species are significantly or insignificantly lower in infants delivered by CSD compared to VD. However, these differences tend to diminish by the time an infant reaches three years of age [90]. The observed trend of similar gut microbiome composition in infants could be influenced by factors such as the introduction of solid foods and weaning from exclusive milk diets occurring around the same time for both cases. However, despite this similarity, infants born via CSD may have an altered immune system and are at higher risk of diseases such as autoimmune diseases, allergies,

and asthma compared to VD. This change could potentially have enduring effects on human health. Understanding the long-term implications for human health requires the examination of how the relationships between the human host and its microbial communities have evolved and how these interactions may have shaped the development of the immune system's overall infant health, providing valuable insights into evolutionary patterns. However, the evolutionary patterns within the gut microbiome of infants delivered via either method remain uncertain, presenting a significant opportunity to identify personalized health biomarkers. In addition, by tracking the evolution of disease-causing microbes in the gut microbiome, it can become possible to prevent major infant diseases and reduce mortality rates in the future. The existing research lacks adequate exploration of specific microbial taxa present in the infant's gut, particularly with the mode of delivery, infant health, and diseases.

1.5 Objectives

- To characterize the differences in patterns of gut microbiome evolution at early-life stages in infants delivered by VD or CSD.
- To identify key microbial species driving the evolution of gut microbiome in both groups
- To establish a connection between microbiome evolution, infant health, and potential diseases.

CHAPTER 2: MATERIALS AND METHODS

2.1 Data Acquisition

The raw metagenome data was obtained from NCBI Sequence Read Archive (SRA) database using accession number PRJNA473126. The original study involved a total of 402 gut metagenome samples collected from infants delivered via either of the delivery mode (VD: n = 181, and CSD: n = 221) across different timepoints(M) starting from M0till M8 [91]. We were interested in understanding how the gut microbiome of infants evolves after acquisition from mothers at birth, thus, this dataset allowed us to answer our research question. Further details, including the metagenome sequencing along with relevant metadata is provided in Table 2.1 and Table 2.2.

Table 2.1 Major characteristics of infant data

	Mode of delivery	
	CSD	**VD**
Diet	Predominat: Mostly Formula-Fed, Cow's Milk Formula	Predominat: Breastfed, Cow's Milk Formula
Antimicrobial in last 7days	No: 202 Yes: 15	No: 169 Yes: 7
Maternal Diabetes	No: 196 Yes: 21	No: 176 Yes: 0
Neonatal Intensive Care Unit	No: 195 Yes: 22	No: 148 Yes: 28

Table 1.2 Infant Sample Counts at different timepoints

	Mode of delivery	
	CSD	**VD**
Timepoints (M)	**Number of samples**	
0	13	11
1	21	18
2	29	19
3	22	14
4	23	21
5	28	25
6	22	23
7	31	24
8	26	21

The SRA Toolkit is a collection of libraries and tools created by NCBI for communication with the SRA database [92]. By utilizing the *prefetch* command, the dataset mentioned above was retrieved from NCBI SRA. Then, the *fastq-dump* command was used to convert the metagenome samples from SRA to FASTQ format. For storage convenience, *gzip* command was used to compress these FASTQ paired-end files.

2.2 Preprocessing of sequence reads

Raw metagenome data can typically have compromised sequence quality and must be preprocessed before downstream analysis. The steps taken during preprocessing are described in the subsequent sections.

2.2.1 Quality Control

To identify potential quality issues in the data, FastQC [93], a quality control visualization tool, was used. These generated visualizations of the data quality which were subsequently curated for identification of major issues and selecting appropriate parameters for preprocessing. Some of the identified issues included adapter contamination, read duplication and short reads. Next, the preprocessing encompassing these requirements was implemented using fastp (v0.20.1) [94].

2.2.2 Host read removal

Another source of contamination in the sequencing data, is the potential presence of host reads. This can result in false-positive results. Hence, the data must be processed for the removal of host reads. To accomplish this, BBDuk script (from BBMap suit) was used [95]. The BBDuk script (https://github.com/Habiba8956/HGT-insight-hub employed the human genome (GRCh38) as a reference dataset to align metagenomic reads with the host genome, subsequently removing them.

2.2 Metagenome Assembly

For metagenome *de novo* assembly, we used MEGAHIT (v.2.4.3) [96]. The metagenome assembly was conducted using the KBase online server [98]. The MEGAHIT assembly was conducted with the following *k*-list 31, 59, 87, 115, 127 and a minimum contig length of 500.

2.3 Identification of HGTs using WAAFLE

WAAFLE (Workflow to Annotate Assemblies and Find LGT Events) (v.2.0.0) [101] is computational approach based on phylogenetics designed to detect novel HGTs and profile them from metagenome assemblies. Metagenome assemblies with WAFFLE to identify

potential HGT events was used. In the first step, assemblies were compared against the WAFFLE database using the *waafle_search* command. The resulting BLAST hits were then utilized for identifying open reading frames (ORFs), using the *waafle_genecaller* command. This yielded results in GFF (General Feature Format) which contains gene coordinates information. Finally, the metagenome assemblies, blast output and GFF files were used for identification of HGT events as well as for taxonomic classifications of the contigs involved in HGT events. WAFFLE results include both "*directed*" and "*cumulative set*" HGT events which refer to situations where donors are known and unknown, respectively.

2.3.1 Extraction of genes involved in HGT events

Next, FASTA sequences of genes involved in HGT events were identified and extracted from the genes predicted from the metagenome assemblies. For this, the coordinates of genes were extracted from the GFF files produced by WAAFLE. Then, the FASTA of the genes was obtained from the gene catalogue using the coordinates through a custom Python script.

2.3.2 Estimation of counts of extracted genes

Following gene extraction, *count* method from CoverM (8.6.1) was used to determine the number of reads mapping to each gene sequence. The resulting count tables for each sample were processed to create a matrix using a custom Python script. This script worked in two steps: (1) it summed up the counts of genes and (2) arranged the multisample counts such that each row corresponded to a unique query and columns represented counts in independent samples.

2.3.3 Estimation of Rate and Frequency of HGT events

In the next step, the rates and frequencies of HGT events were calculated, within samples as well as between clade pairs. These rates of HGT events identified by WAAFLE were normalized by the assembly size. For this, the number of HGT events was obtained for each sample from the outputs and was divided by the corresponding sample's metagenome assembly size in bp and finally normalized to rates per million. The rationale behind this step was to account for the inherent variability in assembly sizes, as larger assemblies naturally have a higher likelihood of capturing more events. Thus, this normalization enabled us to remove any potential bias that could have resulted due to a larger assembly size obtained due to deeper sequencing of certain samples.

This was estimated by calculating HGT events frequency between a pair of clades: A is acceptor, B is donor; the number of AD events was observed between A and B in all samples. Then the figure was normalized to the total number of genes assembled for both acceptor and donor across all samples and normalized to 1000 genes per assembly. Likewise, the frequency of directed HGT events from a donor to an acceptor was calculated by measuring the number of transfers from a donor to an acceptor across samples and then normalized to the total number of acceptor genes assembled across all samples and then scaled to 1000 genes. The normalization of this enabled an accurate average density of transfers from the donor to the acceptor, with particular importance on the recipient clade acceptor in HGT events.

2.4 Contig Abundance Estimation

CoverM (v0.6.1) [100] was applied to estimate the relative abundances of contigs in the metagenomes. For this, the contig identifiers obtained from the WAAFLE outputs were

utilized and their corresponding FASTA sequences were extracted from their respective metagenome assemblies. The specific parameters for estimating the abundances included: --min-read-percent-identity 95, and --min-read-aligned-percent 50. The 'Relative Abundance' method from the CoverM was used to calculate the percentage of total metagenomic reads that mapped to each contig. To concatenate multiple CoverM output files into a single matrix format, a customized Python script was prepared. This script performed two major functions: (1) it summed up the abundances of contigs that belonged to the same microbial taxa and (2) arranged the data such that abundances of each sample were represented in a single column. The first row of listed the name of the taxa that was idenfitied from WAAFLE.

2.5 Gene Annotation

HGT is conventionally linked to various molecular functions and to explore potential functional enrichments among genes involved in HGT events, eggNOG-mapper v.2.11. [104] is used for the identification of functions of genes. The extracted gene sequence files served as input. eggNOG-mapper was executed with default parameters to annotate the extracted gene sequences. This process involved mapping each gene to known orthologous groups and assigning putative functions. The output from eggNOG-mapper provided annotation files for both CSD and VD. These annotation files included information on the predicted functions of genes within each dataset, categorized according to Clusters of Orthologous Groups (COG). For this analysis, we processed the WAFFLE output and obtained the contigs that were assigned to the genes involved in HGT events. These steps collectively contributed to a comprehensive and functionally enriched representation of the gene sequences.

2.6 Statistical Analysis

2.6.1 Median comparison using Wilcoxon's test

In the microbiome study conducted by Falony et al., the comparison of median differences in alpha-diversity measures, the proportion of core genera, and the abundance of specific genera was accomplished using the Wilcoxon test for categorical variables. These non-parametric tests were employed to assess statistical significance, providing a robust method to evaluate differences within and between cases without relying on strong distributional assumptions. For the comparison of contig length, N50 and largest contig distribution in each case CSD and VD, the Wilcoxon test was applied using the QUAST report. This facilitated an examination of potential differences in contig sizes between cases.

In the analysis of HGT events, the Wilcoxon test was utilized in multiple scenarios. The HGT rate, and transferred gene frequency were compared between CSD and VD cases. Differences in HGTs between timepoints in each individual case were tested, as well as differences in HGTs between cases at an individual time point. This enabled a powerful statistical comparison of HGT dynamics with consideration for different experimental conditions and factors related to time.

2.6.2 Comparative Analysis of Alpha Diversity in CSD and VD

Alpha diversity describes the variety and distribution of species within one sample. It comprises evenness, which shows the relative abundance of each of those species in a sample, and richness, which is the number of distinct species. [103]. Because the Shannon method is well-established and frequently used, it was employed to compute the alpha diversity of the two examples, CSD and VD. The calculation of the Shannon index was done at each time point from 0 to 8 months for both cases, and the respective results were

plotted in a line plot to compare the trends in the alpha diversity between cases across time. We also performed Wilcoxon tests on the cases at each time point to produce p-values of the differences in alpha diversity between the cases to analyze the observed differences statistically.

2.6.2 Fold enrichment analysis using eggNOG-mapper

To identify meaningful changes in gene function between datasets, log2 fold change calculations were applied to COG categories from the annotation files. This statistical approach quantified differences in gene expression patterns over various timepoints. Heatmaps were then generated using R to visually display these differential expression patterns across the COG categories.

CHAPTER 3: RESULTS AND DISCUSSION

3.1 Preprocessing of the metagenome data

The mean raw read counts for CSD and VD were 4.73±1.87 M and 4.79±2.73 M respectively, as shown in Figure 3.1. These results indicate that the raw read counts for both cases are similar, thus the datasets were generated with similar sequencing depths. The mean counts for clean reads for the CSD and VD were 3.01±2.1 M and 3.09±2.1 M, respectively. Similarly, the read count after the removal of host-associated reads (HG trimmed) in CSD and VD was 3±1.39 M and 3.09±2.1 M, respectively (Figure 3.1).

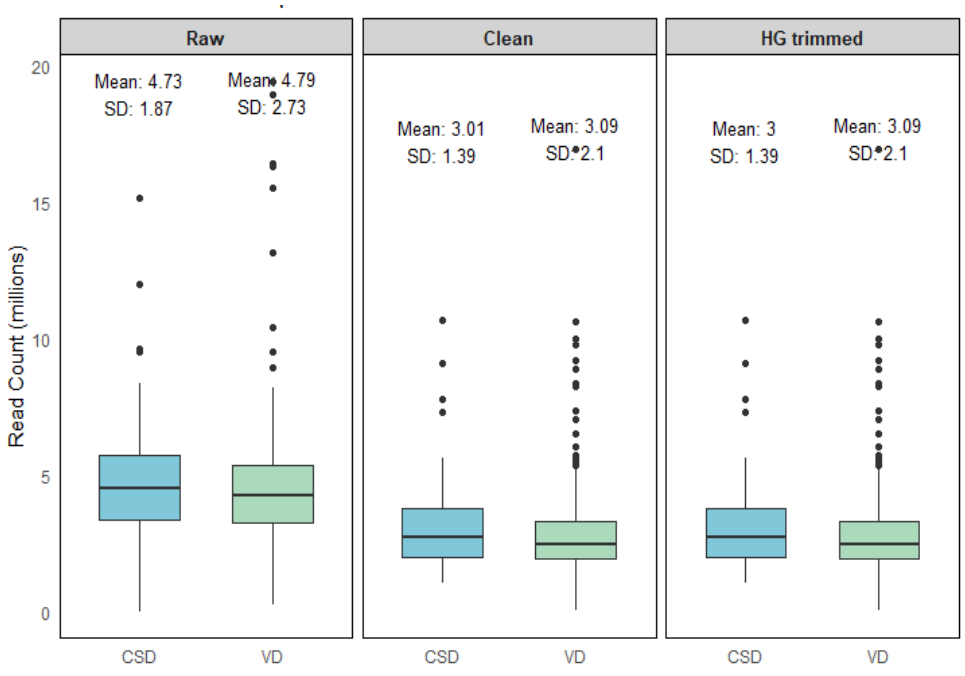


Figure 3.1: Comparison of read distribution in both cases, highlighting variations in data processing stages

Visualization of raw, clean, and HG-trimmed read qualities through FastQC indicated improved data quality after preprocessing. Furthermore, no significant difference was observed in the read counts in both cases.

3.2 Metagenome assembly and quality assessment

Metagenomes of both cases were assembled using MEGAHIT *de novo* metagenome assembler and their quality was assessed through QUAST. The main parameters evaluated for determining assembly quality include the total number of contigs, total assembly length, largest contig, and N50 length. These are discussed in detail below:

3.2.1 Total number of contigs in metagenome assemblies

The mean and std for the CSD are 2.17±1.38 kilobase pairs (Kbp), respectively. For the VD, the mean and std are 1.97±1.59 Kbp, respectively. This indicates that the number of contigs in both cases was relatively consistent, with the CSD having a relatively higher mean and std compared to the VD (Figure 3.2). However, no significant differences were observed between the cases, with p-value of 0.57.

3.2.2 Total length of metagenome assemblies

The total length of the metagenome assemblies for the CSD was 76.6±96.2 Mbp, whereas for the VD, it was 67.56±86.09 Mbp. This suggests that the total length of the metagenome assembly was comparatively higher for the CSD compared to the VD. However, no significant differences were observed between the cases, with p-value of 1. (Figure 3.2).

3.2.3 Largest contig size

The largest contig size of metagenome assemblies for the CSD was 307.15± 154 Kbp, whereas for the VD, it was 306.17± 146.39 Kbp. However, no significant differences were observed between the cases, with p-value of 0.91. (Figure 3.2).

3.2.4 N50 length

The N50 length for the CSD was 31±54 Kbp, whereas for the VD, it was 33.94±39.03 Kbp. This suggests that the N50 is slightly higher for the VD compared to the CSD (Figure 3.2).

However, no significant differences were observed between the cases, with p-value of 0.71. The results indicate that the metagenome assemblies for the CSD and VD were of similar quality, with no significant differences in the assembly metrics. This suggests that both cases were adequately represented in the assemblies. The assembled metagenomes of infants also enabled the investigation of HGTs involved in early life evolution and their impact on infant health and disease.

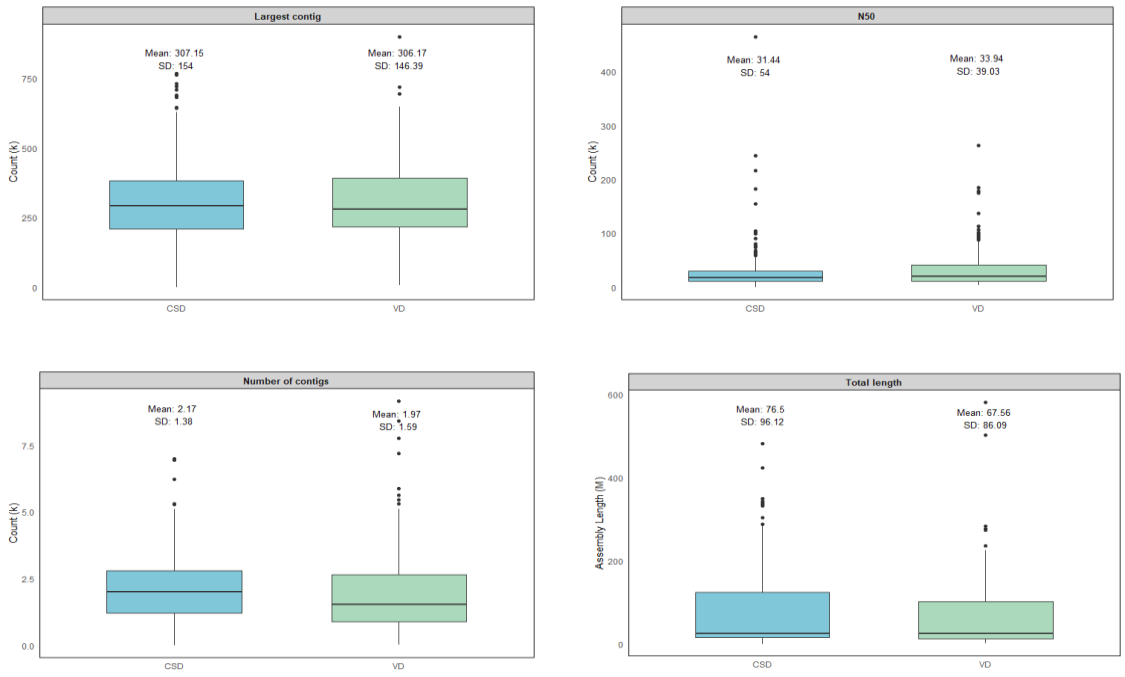


Figure 3.2: Visualization of quality assessment of metagenome assembly

3.3 Identification and distribution HGTs in infant gut microbiome

The observed pattern of HGT events in the infant gut microbiome, as depicted in Figure 3.3, 3.4 shows a notable difference between VD and CSD in the early stages of infancy. Specifically, VD infants exhibit a higher number of HGT events at M0, M1, M2, and M3 compared to CSD infants. However, by M5, M6, M7, and M8, the number of HGT events in both cases becomes more evenly distributed, with VD infants still showing a slight

advantage. Notably, no significant difference between the cases was observed other than at M4 (cumulative set of HGTs), as determined by the Wilcoxon rank-sum test (p-value = 0.028). This pattern of HGT events can be linked to the existing literature on the influence of delivery mode on the infant gut microbiome. This association of VD with a more diverse and mature infant gut microbiome has been validated by several studies. The increase in diversity and maturation might be the reason for the high number of HGT events seen in these infants in early infancy [105]. Increased exposure to the maternal microbiome during VD may facilitate the gene transfer between species, thus promoting faster development of the infant gut microbiome [105]. In contrast, less diversity and immaturity of the gut microbiome are associated with CSD, and events of HGT could therefore potentially be fewer in the very early stages of infancy [106]. This reduced exposure of the infant to maternal microbiota in CSD could limit gene transfer between species, thus delaying the development of gut microbiome. By months M3, M4, M5, and M6, the number of HGT events in both CSD and VD infants tends to equalize, likely due to the overall maturation of the infant gut microbiome [107].

Specific families of bacteria may dominate the infant gut microbiome and thus relate to the observed pattern of HGT events. For example, in M1, CSD tend to be dominated by Enterobacteriaceae, while VD will more likely be dominated by Bifidobacteriales. By M2, CSD tend toward Lachnospiraceae, whereas VD tend toward Ruminococcaceae. In M3, Enterobacteriaceae again tends to be highly abundant in CSD but Lachnospiraceae in VD. This pattern again changes in the subsequent months, when Eubacteriaceae is the dominant family at M4 in CSD, while Ruminococcaceae remain dominant in VD. At M5, Enterobacteriaceae again rises in CSD and Lachnospiraceae joins to be the dominant

family in VD. From M5 onwards, Lachnospiraceae is the most abundant bacterial family in both modes of delivery and remains so as the predominant group from M6 to M8.

In brief, these results suggested that VD is associated with an increased number of HGT events during this early period of life that would contribute to the establishment of a more rich and mature gut microbiome. HGT events modification may be dependent on the dominance of certain bacterial families in the infant gut. As such, Bifidobacteriales in VD could facilitate gene transfer between species, leading to an increase in the number of HGT events [106]. The fact that Lachnospiraceae is a dominant family in both delivery modes during this period might help explain the more equal distribution of HGT events. Nevertheless, it is close to being equally distributed between both delivery modes from M5 to M8; hence, other factors such as diet and environmental influences may turn out to be more relevant in shaping the infant gut microbiome over time.

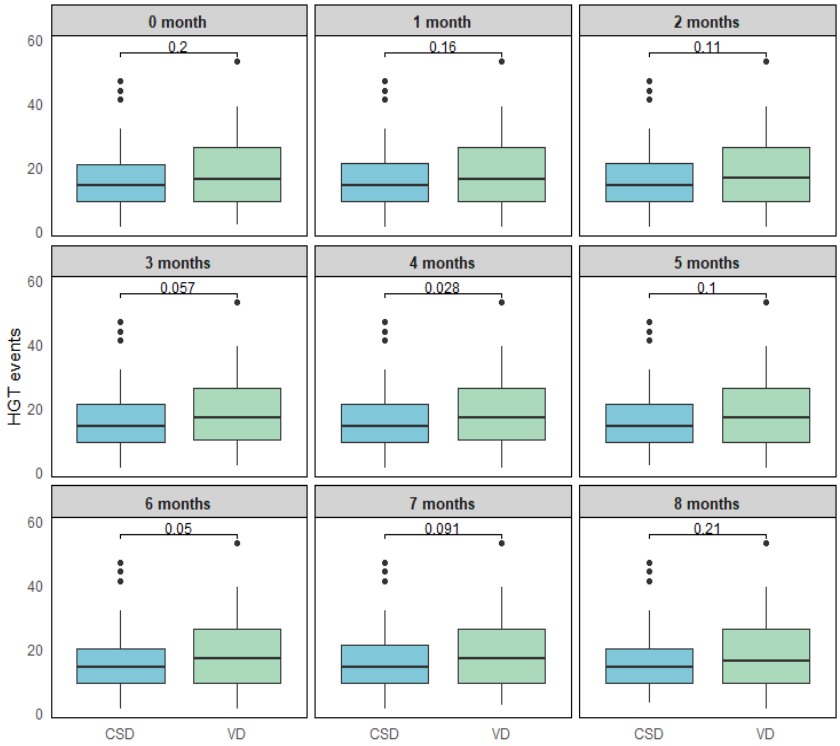


Figure 3.3: Median HGT events across different timepoints in CSD and VD (directed only)

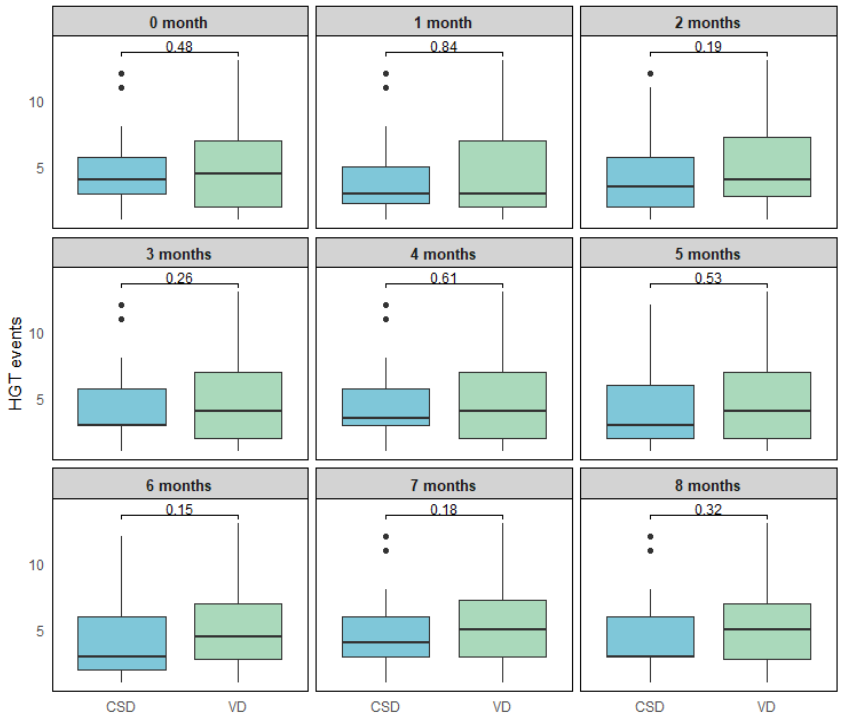


Figure 3.4: Median HGT events across different timepoints in CSD and VD (cumulative set of HGTs)

3.4 Rate of identified HGTs

The rate of HGT events in the infant gut microbiome varies notably between CSD and VD with a significant difference at M4 (directed only) (p-value = 0.04) in Figure 3.4.1 and Figure 3.4.2 shows rate of HGT events (cumulative set). For more appropriate comparison, the rate of HGTs was calculated for the two sample groups by dividing the HGT events with the assembly length (in Mbp). This normalization ensured handling the differences in assembly size between samples.

At M0, the CSD has a higher rate of HGT events (0.22 per Mbp of assembly) compared to the VD (0.18 per Mbp of assembly). This likely happens due to changes in the gut microbiota and reduced exposure to the maternal microbiome with CSD, which might influence how an infant's gut is initially colonized [105].

In contrast to the M1, the rate of HGT events for both delivery methods decreases, but VD displays a fairly constant rate. This might be a consequence of the continued influence of the maternal microbiome on the infant's gut through continued breastfeeding or other environmental factors [107]. From M1 to M2 and from M2 to M3, the rate of HGT events is constant in both modes of delivery. This time period is associated with the establishment of an infant gut microbiome, and stability in HGT events may be influenced by processes linked to maturation of the immune system or development of gut protective barriers [107]. While this HGT rate is maintained from M2-M3 to M3-M4 in VD, it starts to decrease in CSD. Again, this could be the result of continuous exposure in VD to the maternal microbiome, keeping a gut microbiome more diverse and resilient. From M4 to M5, HGT events remain constant in CSD and reduced in VD.

Now, going further from M5-M6, there is an increase in HGT events in CSD and a further drop in VD. This could result from increased exposure to environmental factors that should hence modify the gut microbiome composition [107]. Though in both cases there is a decline from M5-M6 to M6-M7 of HGT events, in CSD it still remains slightly higher. In the M7-M8 both cases the rate of HGT events levels off, with a slight uptick in CSD. This could be a result of the continuing development of the infant's immune system and stabilization of a gut microbiome [108].

In summary, the variation in the HGT rate over time would depend on how the gut microbiome matures from birth into adulthood, develops the immune system, and the effect of the environment.

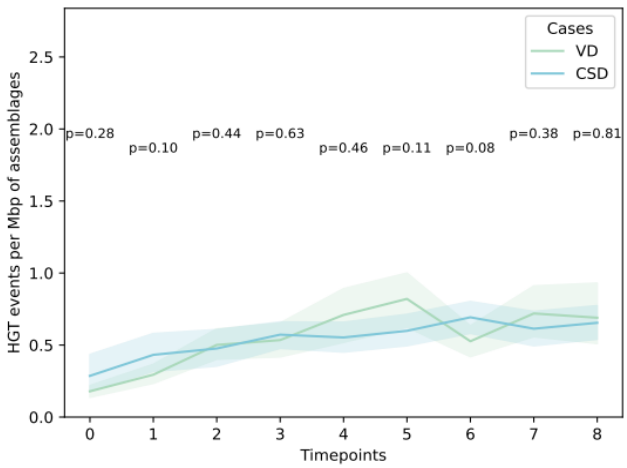


Figure 3.5: Rate of HGT events across different timepoints in CSD and VD (directed only)

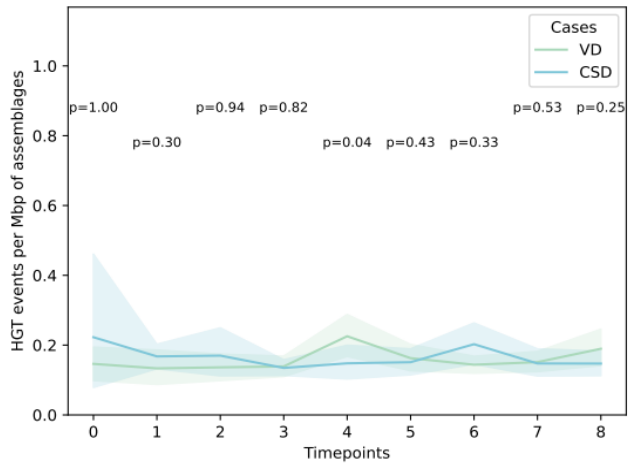


Figure 3.6: Rate of HGT events across different timepoints in CSD and VD (cumulative set of HGTs)

3.5 Frequency of identified HGTs

The frequency of identified HGT events in the infant gut microbiome varies notably between CSD and VD. Frequency of HGTs was calculated for the two sample groups by dividing the HGT events by the genes. This normalization allows us to count the number of transferred genes in a given set of genes, providing a more accurate representation of the HGT frequency.The results indicate that there is no significant difference in the frequency of HGT events between CSD and VD across all timepoints (directed, cumulative

set of HGTs) showing p values > 0.05 Figure 3.7, Figure 3.8. At M0, CSD had a slightly higher frequency of HGT events (0.3) compared to VD (0.2). This difference was maintained until M1-2, after which the frequency decreased and became similar for both cases. In the M2-3 period, there was no change in frequency.

However, in the M3-4 period, both CSD and VD showed an increase in frequency, with VD having a slightly higher value. First, the frequency of HGT events decreased during M4-M5, and then it continued falling to almost the same level by M5-M6. From M6-M8, it started rising again, though showing a minor high value for VD.

The first frequency difference of the CSD and VD can be explained through differences in the composition of maternal gut microbiomes and exposure to different environments that infants encounter [106]. The decrease in the frequency during M1-M2 could be because of the infant's gut microbiome adjusting itself to the new environment, losing some of the transient species [109]. Increased frequency during the period M3-M4 might indicate increased stability in gut microbiome, which allows HGT events to be more efficient. The drop in frequency during the M4-M5 might be attributed to gut microbiome reaching maturation and fewer needs for HGT events [107].

This steady increase during the period from M6 to M8 could be driven by further development of gut microbiome [107]. Specifically, these results highlight the dynamics of HGT events within infant gut microbiomes, and significant roles for temporal and environmental factors in these processes.

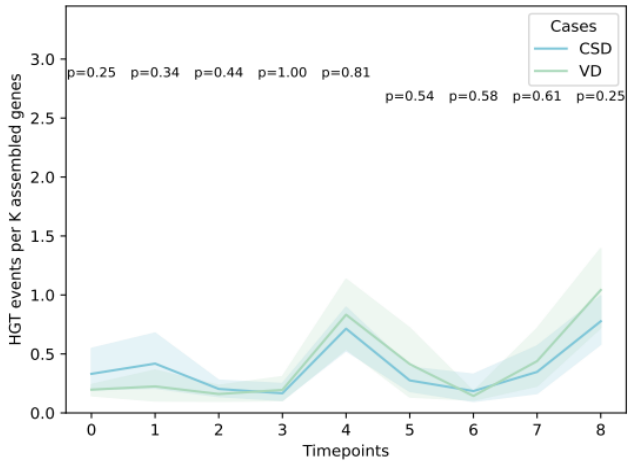


Figure 3.7: Frequency of HGT events across different timepoints in CSD and VD (directed only)

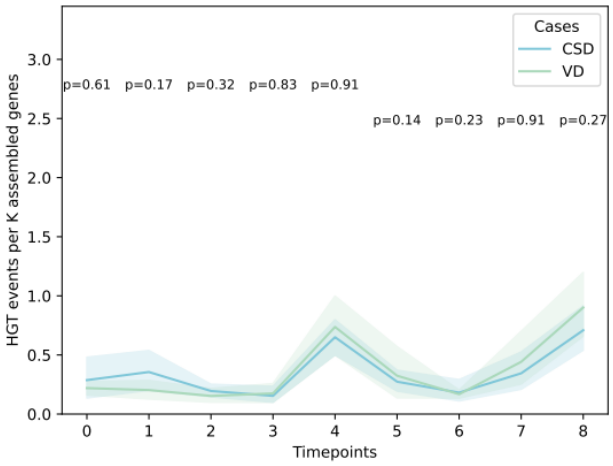


Figure 3.8: Frequency of HGT events across different timepoints in CSD and VD (cumulative set of HGTs)

3.6 Alpha diversity of identified HGTs

The results of α-diversity analysis for donor and acceptor species in both CSD and VD across different timepoints reveal distinct patterns and intersections. At M0, CSD has higher α-diversity than VD. This difference is maintained until M1, with a significant difference observed at p-value=0.03 shown in Figure 3.9 and 3.10 Thereafter, the α-diversity in CSD starts to decrease, while it continues to increase in VD until M3, at which point both cases intersect. In the CSD, α-diversity then keeps on going down until 5M

while VD maintains a higher level. There is no significant difference between two delivery modes with p-value under 0.05. In both cases, α-diversity again plateaus between M6 and M7. According to VD, this α-diversity is greater at the end of M8. These observations are consistent with reports in existing literature that describe the effect of mode of delivery on the infant gut microbiome. For example, Korpela et al. [110] recently reported that the gut microbiome was different in infants delivered by CSD when compared to VD. The authors postulated that this could be attributed to exposure to the maternal vaginal microbiome at birth during VD, in combination with other factors such as the introduction of solid foods and maturation of the infant immune system. They also indicated that the application of antibiotics, which is more frequent in the case of CSD infants, can perturb the gut microbiota and decrease α-diversity [111]. Another study by Yassour et al. [112] demonstrated lower diversity and different composition of gut microbiome in infants born by CSD versus VD. The investigators hypothesized that this disparity could be a result of reduced exposure to the maternal vaginal microbiome during CSD. Higher α-diversity of donor species at an earlier time point in CSD may indicate that a wider range of taxa is implicated in mediating HGT.

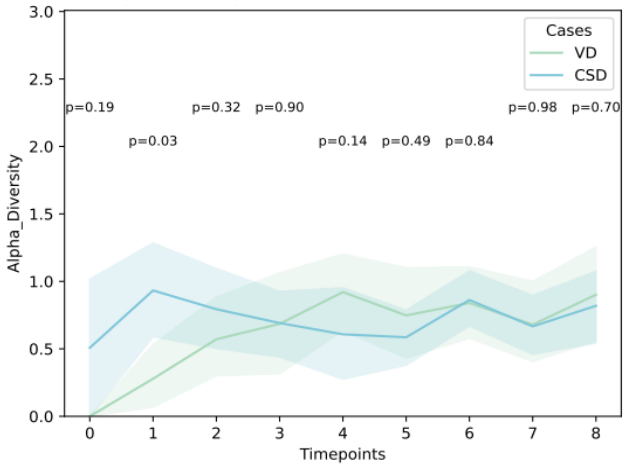


Figure 3.9: Alpha diversity across different timepoints in CSD and VD (directed only)

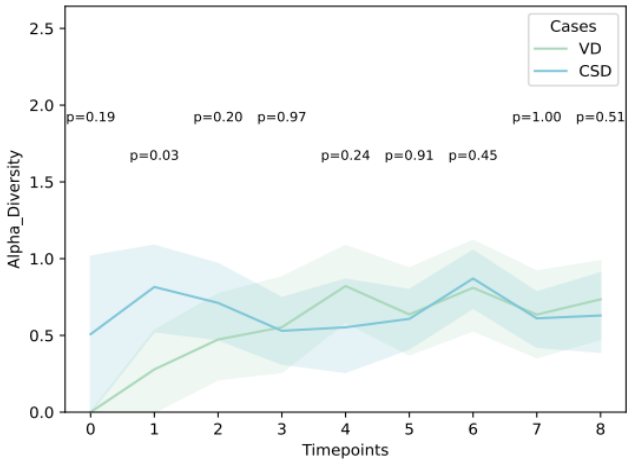


Figure 3.10: Alpha diversity across different timepoints in CSD and VD (cumulative set of HGTs)

3.7 Identifying unique driver of HGTs in CSD and VD

The results of α-diversity analysis reveal significant differences at the M1 timepoint. This difference is observed in both donor and acceptor species, with a p-value of 0.03 in both cases. The upset plot further supports this finding by showing a greater number of unique donor (Figure 3.11) and acceptor (Figure 3.12) species in CSD compared to VD at M1. The unique donor species in CSD at M1 include *Veillonella sp. oral taxon 158, Clostridiales genomosp. BVAB3, Anaerostipes hadrus, Streptococcus ferus, Oribacterium sinus, Escherichia coli, Dorea longicatena, Bifidobacterium animalis, and Clostridium clostridioforme*. In contrast, the unique donor species in VD at M1 are Neisseriaceae, Treponema, *Bacteroides vulgatus*, and *Streptococcus equinus*. The unique acceptor species in CSD at M1 are *Citrobacter freundii, Veillonella atypica, Megasphaera micronuciformis, Enterococcus malodoratus, Streptococcus thermophilus, Anaerostipes hadrus, Eubacterium ventriosum, and Roseburia hominis*. In VD, the unique acceptor species are *Parabacteroides merdae* and *Enterobacteriaceae*.

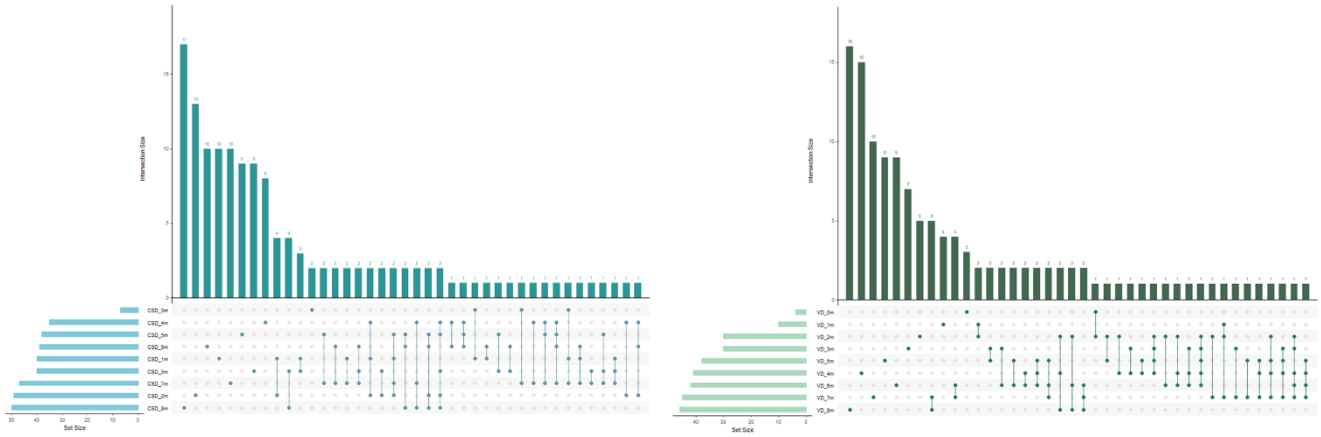


Figure 3.11: Visualization of unique and common donor species across different timepoints in CSD and VD

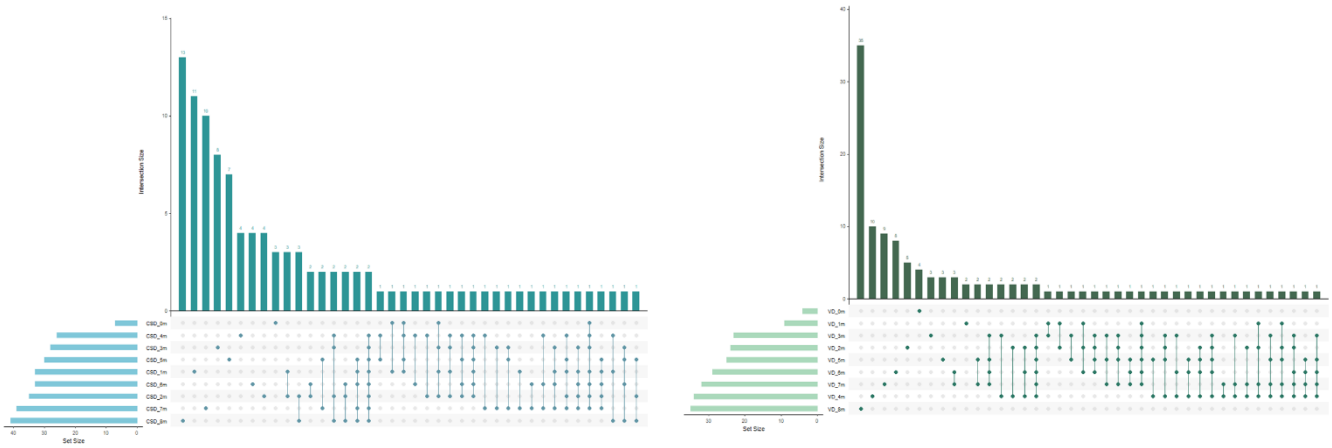


Figure 3.12: Visualization of unique and common acceptor species across different timepoints in CSD and VD

3.8 Comparative analysis of HGT events in CSD and VD

The donor-acceptor pair in CSD, involved in most HGT events (11 events), was *Klebsiella sp_4_1_44FAA-Collinsella aerofaciens*, followed by, *Streptococcus sp oral taxon 071-Streptococcus suis* (8 HGT events) shown in Figure 3.13. In contrast, the donor-acceptor pair in VD, involved in most HGT events (16 events), was *Coprococcus catus-Ruminococcus sp_5_1_39BFAA*, followed by, *Lachnospiracea bacterium 2 1 4 6 FAA-*

Ruminococcus gnavus(11 HGT events) shown in Figure 3.14. Furthermore, the donor-acceptor pairs in CSD and VD belong to different families. For example, *Klebsiella sp_4_1_44FAA* and *Collinsella aerofaciens* belong to different families (Enterobacteriaceae and Erysipelotrichaceae, respectively). Similarly, *Streptococcus sp oral taxon 071* and *Streptococcus suis* belong to the same family Streptococcaceae. In contrast, *Coprococcus catus* and *Ruminococcus sp_5_1_39BFAA* belong to different families (Lachnospiraceae and, Ruminococcaceae respectively), *while Lachnospiracea bacterium 2 1 4 6 FAA* and *Ruminococcus gnavus* belong to the same family Lachnospiraceae but different genera. *Lachnospiracea bacterium 2 1 4 6 FAA*, and *Ruminococcus gnavus* belongs to the genus *Ruminococcus.* Such differences in the donor-acceptor pairs and their families and genera may impact on the interaction of bacteria with each other during processes such as CSD and VD. For example, the Ruminococcaceae family is more abundant during the early gut microbiome of VD infants. This family has been found to be more prevalent in the gut microbiota of VD-born infants and to play an important role in the degradation of dietary intake, particularly during early life [106], [113]. In contrast, CSD infants often have a gut microbiome resembling their mothers' skin bacteria, dominated by *Corynebacterium* and *Staphylococcus spp.* [113]. Distinctions within the gut microbiome are thought to derive from a lack of exposure to maternal vaginal microbiome following VD.

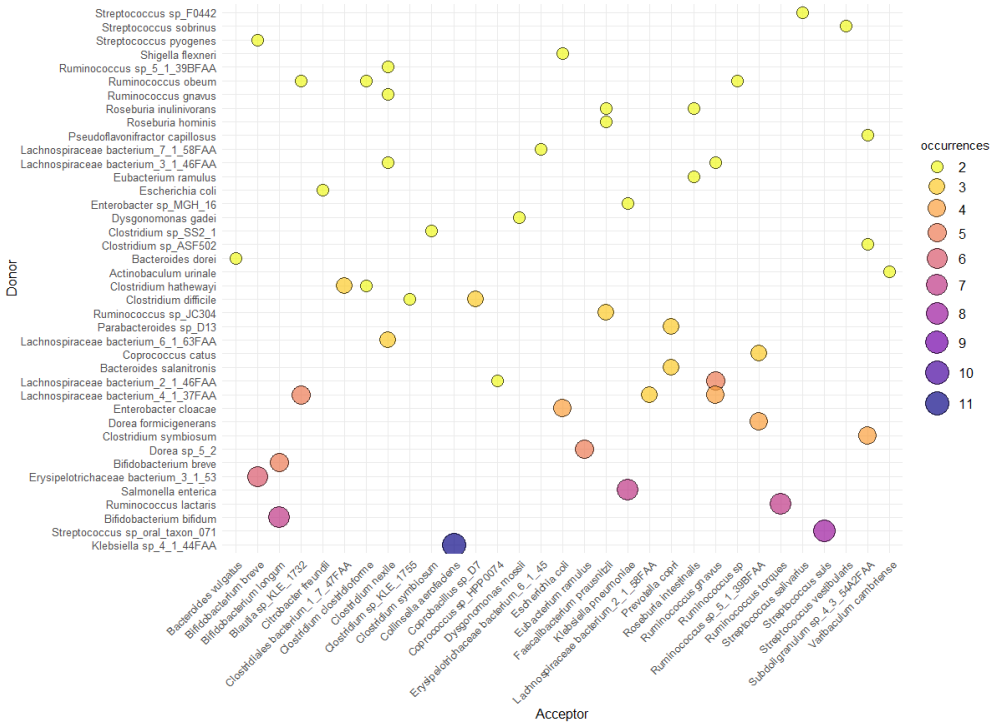


Figure 3.13: Visualization of HGT donor-acceptor pair in CSD, representing the number of occurrences

Figure 3.14: Visualization of HGT donor-acceptor pair in VD, representing the number of occurrences

3.9 Temporal Patterns of HGTs in Early-Life Gut Microbiomes

Understanding gene transfer between bacteria requires more than simply counting raw numbers of gene transfers. The reason is that raw counts can sometimes be misleading if certain bacteria are very abundant or very rare in the community. Instead, these counts have been normalised to show relative abundances, which will be helpful in understanding the dynamics of gene transfer. In CSD infants, the bacterium donor is mostly represented by *Clostridium nexile*. Over time, this bacterium starts transferring genes to other bacteria. For example, at M0, it transfers genes to the *Eubacterium ramulus*. At M2, it transfers to the *Coprobacillus sp. 3_3_56FAA* and M4 to *Roseburia intestinalis*. Another important donor is *Salmonella enterica*, which donates genes to *Klebsiella pneumoniae* at M5. The variability in the gene donor and the months where exchanges occur underscores the notion that the gut microbial community is still dynamic. This means that each event of transfer, more specifically from *C. nexile*, is changing the gut microbiome over time in ways that can have large consequences for health [114]. Compared with CSD, VD had different dominant donor species of gene transfer. The most abundant donor in VD was *Coprococcus catus*, which had gene transfer to *Ruminococcus sp_5_1_39BFAA* at M5. The second highest donor contribution is from *Butyricicoccus pullicaecorum*, which transfers genes to *Bifidobacterium longum* at multiple time points: M2, M3, and M6. These differences underline how some of the dominant donor species in VD, such as *C. catus* and *B. pullicaecorum*, belong to families—Ruminococcaceae and Butyricicoccaceae—common in the vaginal microbiome [111]. These findings suggest that the vaginal microbiome could be more important for shaping the gut microbiome of VD infants than that of CSD. The differences could exert very fundamental, long-lasting effects on the

health of infants. Figures 3.15 and 3.16 illustrate the gene flow over time between the donor and acceptor bacteria in CSD and VD infants, respectively.

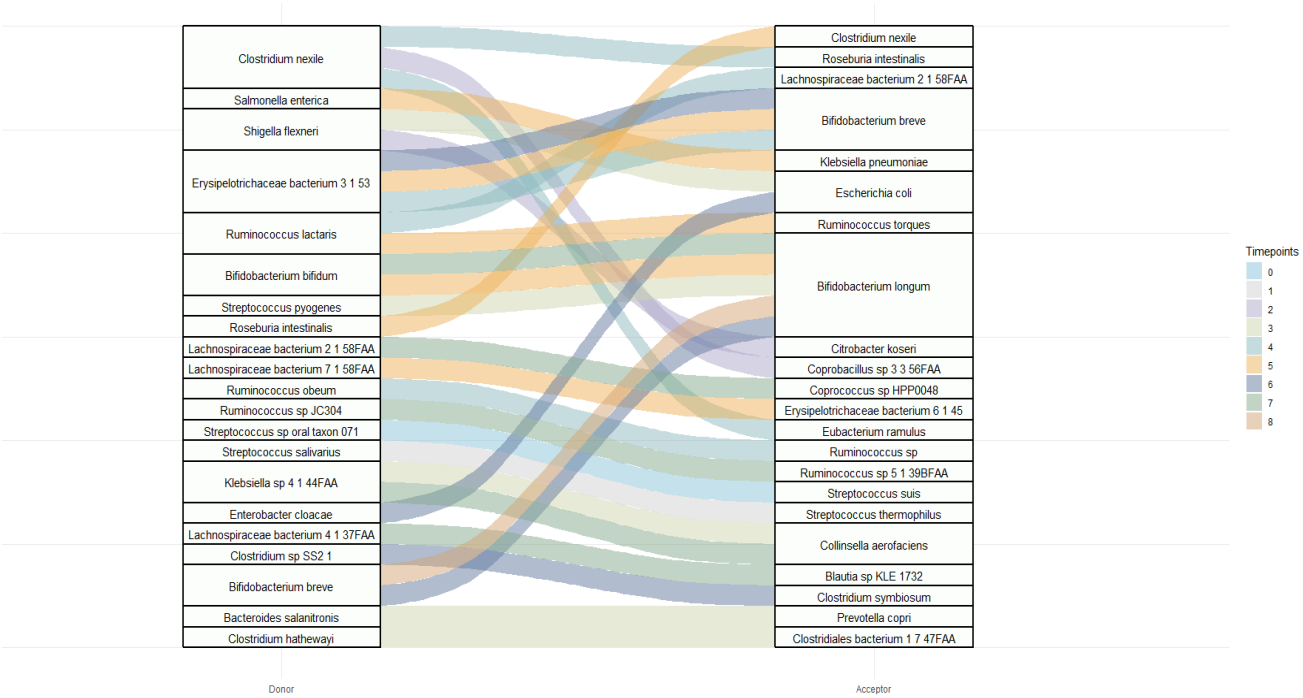


Figure 3.15: Visualization of temporal HGT patterns from donor to acceptor in CSD

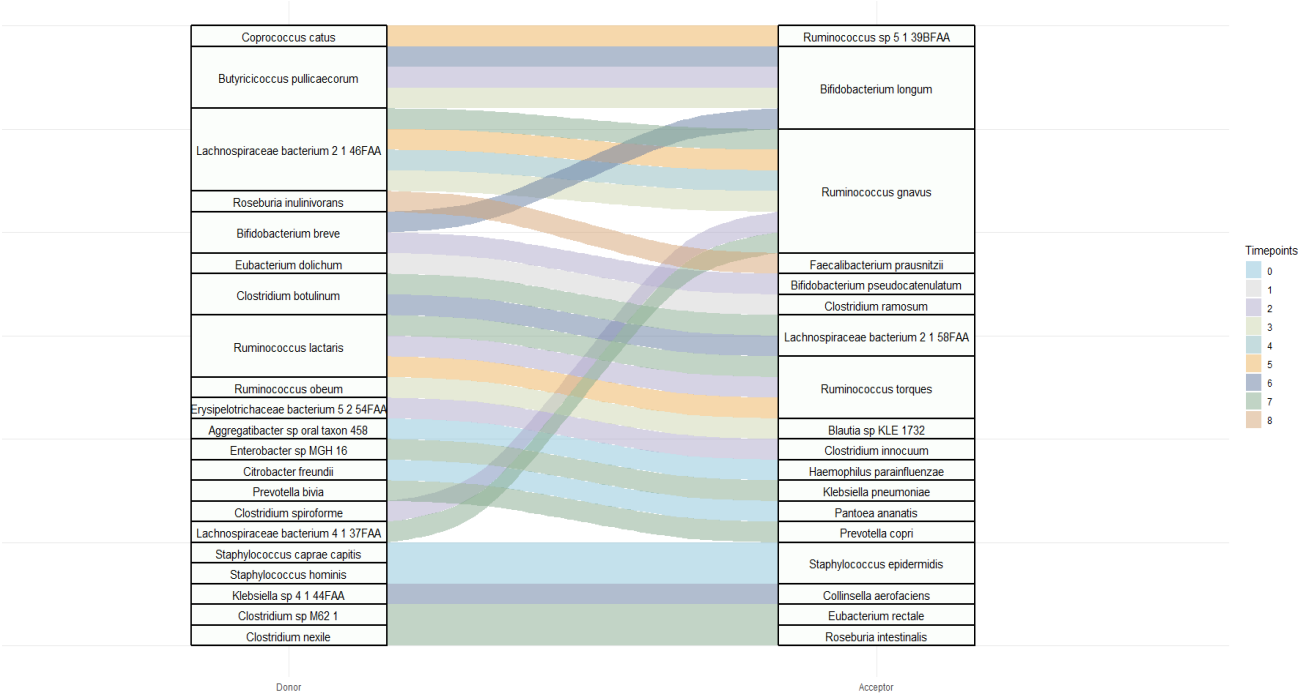


Figure 3.16: Visualization of temporal HGT patterns from donor to acceptor in VD

3.10 Gene Expression Profiles of Horizontal Gene Transfer in CSD and VD

We used DESeq2 to analyze how genes involved in HGT are expressed in CSD infants than those born via VD. To make sense of these genes, we extracted and annotated them with eggNOGmapper. We then identified the biological pathways they're involved in using COG categories and took a closer look at what these genes do. By mapping the gene reads, we were able to quantify gene expression and categorize the functions of HGT genes in both delivery methods. The summary of these findings is outlined below.

3.10.1 Differential Expression Analysis and Functional Annotation in CSD

In CSD, we identified and assigned functions to 3,753 genes, while 1,304 genes remained without a clear function. Of the annotated genes, 771 were involved in information storage and processing pathways, 599 were involved in cellular processing and signaling pathways, and 1079 were involved in metabolic pathways.

The function categories were divided into three: immunity, digestion, and cognition. These categories are crucial for infant health and development, especially in the context of personalized health care. Genes related to immunity play a very important role, for they will determine how prone an infant may be to infections, allergies, or autoimmune disorders [116], genes related to digestion affect how well nutrients are used and impact the diversity of gut microbes, overall metabolic health [117]. Moreover, gene-transfer-evolved cognitive-function genes can also have effects on neurological development, behavior, and mental health [118]. Such gene transfer events from these categories may give insights into how the gut microbiome influences overall health. Resulting understanding will, therefore, lead to tailored interventions addressing the unique health needs of each infant and maintaining good outcome in health throughout life.

It used count data of genes for this DESeq2 analysis. Log2 fold change was applied to understand the difference in gene expression. In this case, the M0 time point was used as a reference to compare the rest of the time points, starting from M1 up to M8. Figure 3.17 Heatmap: The heatmap depicted three clusters—a positive LFC, a negative LFC, and a near-zero LFC.

Changes in gene expression magnitude between time-points are quantified using the LFC scale. LFC 1 corresponds to a two-fold increase in gene expression and LFC –1 corresponds to two-fold decrease in gene expression. The more positive LFC is, the more increased the gene expression is relative to a reference time point.

Among the top two donor-acceptor pairs, *Klebsiella pneumoniae* is the acceptor receiving *Salmonella enterica*, transferring the YagA gene in CSD infants. This gene has PFAM integrase catalytic activity and regulates genes related to metabolism. Its transcription is influenced by stress and conditions leading to biofilm formation. It was found to be involved in the posttranslational modification function. This gene plays a crucial role in immunity due to its ability to integrate foreign DNA into the host genome. This can have a huge influence on the immune system of the infant.

In this respect, the expression of this gene was high, with a positive log2 fold change of LFC > 5 at timepoints M2, M5, M6, and M8, relative to other times. This could be attributed to several factors. Firstly, during these periods, the infant's immune system is still developing and hence more prone to infections. In addition to this, the integration of foreign DNA into the genome of the host could have long-term effects on the immune system, which could manifest itself more at such times. *Salmonella enterica* is a pathogenic bacterium that might cause severe infections and could transfer its genetic material to other

pathogenic bacteria, like *Klebsiella pneumoniae*. Gene transfer in this way could give rise to new, more harmful strains of bacteria that may further weaken an infant's immune system. Thus, the presence of these bacteria and their genetic material in the microbiome of an infant can have long-lasting effects on health and immunity.

Carbohydrate metabolism has changed across different time points; some remain almost with no change in LFC, while others show a slight decrease. One of the leading donor-acceptor pairs was the *Clostridium nexile* and *Eubacterium ramulus*, which were beneficial bacteria associated with a key function on the 4Fe-4S single-cluster domain by the gene *fdxA*. It is an important gene in the degradation of polysaccharides. This gene plays a significant role in the digestion of babies, given that it hydrolyzes complex carbohydrates to yield simple sugars that are absorbed as energy. In the process, it helps the gut microbiome derive energy from the dietary carbohydrate to further the growth and development of the infant. The expression of the gene encoding the 4Fe-4S single cluster domain was slightly higher at points M1, M4, and M8 compared to others. This might be the case because of the major changes in the diet of an infant across these time points. At M1, for example, when infants are fed by breast milk or formula, gut microbiome has to adapt the degradation of complicated carbohydrates. At M4, infants are on a weaning diet, and by M8, they are on much more varied diets that include solid foods. It clearly means that this will be required of the gut microbiome to adjust to break down more complex carbohydrates. The slightly higher expression of the 4Fe-4S single cluster domain at these times might be the microbiome’s way of adapting to these dietary changes, helping it efficiently process polysaccharides and provide energy for growth. Even though overall

gene expression showed a decrease (LFC < -2), the increased expression at these key times indicates that the gene remains active and important for digestion.

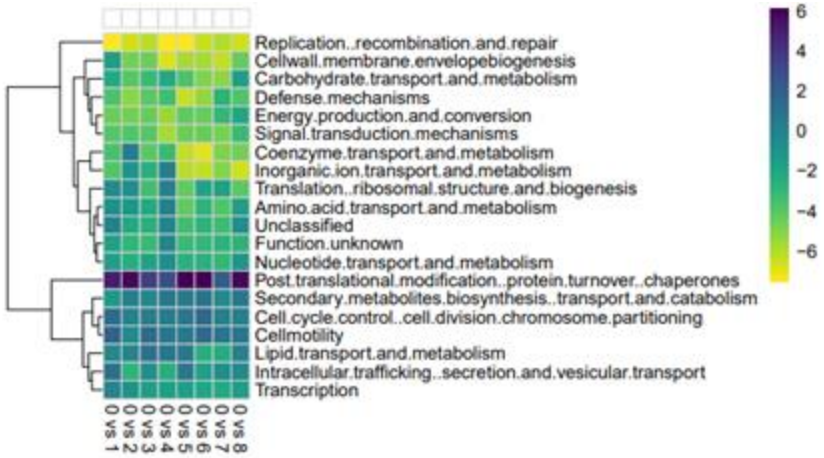


Figure 3.17: Heatmap depicting functional enrichment patterns across different timepoints in CSD

3.10.2 Differential Expression Analysis and Functional Annotation in VD

Similarly, for the VD, a total of 3622 genes were annotated with a function, while 1375 genes were of unknown function. Among the annotated genes, 771 were involved in information storage and processing pathways, 465 were involved in cellular processing and signaling pathways, and 1011 were involved in metabolic pathways.

Butyricicoccus pullicaecorum transfers the *lepB* gene to *Bifidobacterium longum*, which is involved in signal peptidase activity, a valuable target for antimicrobial drug development and this pair is significantly involved in the intracellular trafficking, secretion, and transport function, as indicated by a positive LFC in Figure 3.18. Specifically, the gene *lepB* was found to be the key player in this process. *lepB* plays a crucial role in the regulation of inflammation, stress response, and the production of neurotransmitters, all of which are essential for cognitive development [118]. At M1, M6, M7, and M8, the expression of SPP was notably higher, showing a significant increase compared to other times. That means that the changes in the SPP levels were accompanied by a change in gut

microbiota and the introduction of solid food into the baby's diet. Weaning processes and other dietary changes at those times probably affect both the gut microbiome and the SPP expression.

Association analysis showed that *Butyricicoccus pullicaecorum* and *Bifidobacterium longum* were significantly associated with the signal transduction function with an LFC above 3. This was particularly noted for the vanS gene, which seems to be involved in regulating immune responses. The gene is, therefore, an immune cell activator and the cytokines required in the course of controlling how a reaction from the immune system comes about [116]. In this way, increased expression levels of this gene in VD infants may enhance their immune response and reduce the risk of infections. High expression levels at M3 and M7 might also be related to the rapid development of the gut microbiome and immune system during both of these critical windows of growth. Dietary changes, especially coming off a weaning diet and onto solid foods, may also contribute to such variation in gene expression.

It transfer the *alfA* gene, related to carbohydrate metabolism, in *Coprococcus catus* to *Ruminococcus sp_5_1_39BFAA.* This gene pair represents the alpha-L-fucosidase gene, one of the major players in HMO degradation and metabolism. The overall expression was downregulated, showing an LFC of less than -5 for alpha-L-fucosidase, indicating that VD infants express this gene at lower levels. This may indicate that VD infants have reduced digestive and absorptive capability of HMOs, a significant source of nutrition for an infant. The trend in LFC values across time points is more nuanced in comparison. Although alpha-L-fucosidase, in general, is less expressed in VD infants compared to the other groups, it has increased a little at M1, M6, and M7. This might indicate that the gut

microbiome of VD infants is in a process of adaptation and development, which leads to a higher expression of genes in carbohydrate metabolism during these periods.

The output from Maaslin2 is shown in Figure 3.19, where there is a strong association of the *Ruminococcus sp_5_1_39BFAA* with carbohydrates like Lactose and GOS_FOS_Lactose, hence further confirming that this bacterium might be involved in breaking down carbohydrates. Thus, the negative association suggests *that Ruminococcus sp_5_1_39BFAA* promotes the degradation of these complex carbohydrates to form simpler sugars useful to the infant. This means that at certain times, expression of the gene increases, which may indicate gut microbiome adaptation and maturation in VD infants. This increase in alpha-L-fucosidase expression at specific points in time could be related to the weaning of the diet from mother's milk to solid foods for the infant.

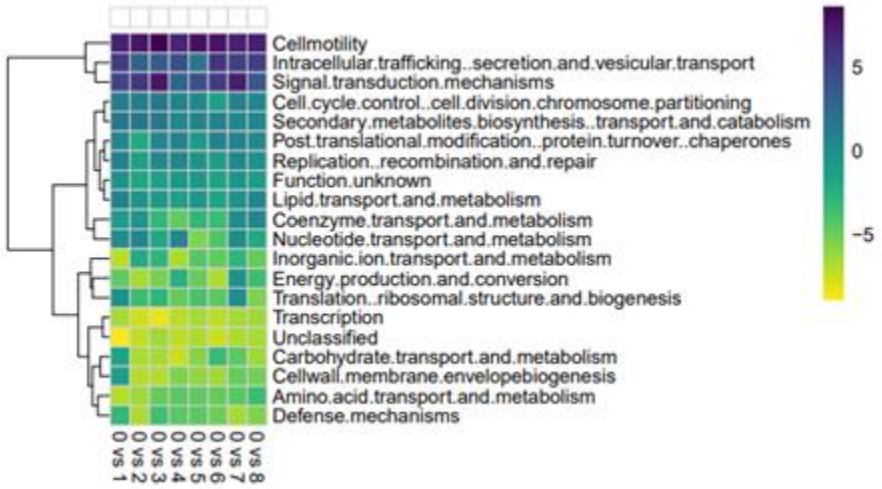


Figure 3.18: Heatmap depicting functional enrichment patterns across different timepoints in VD

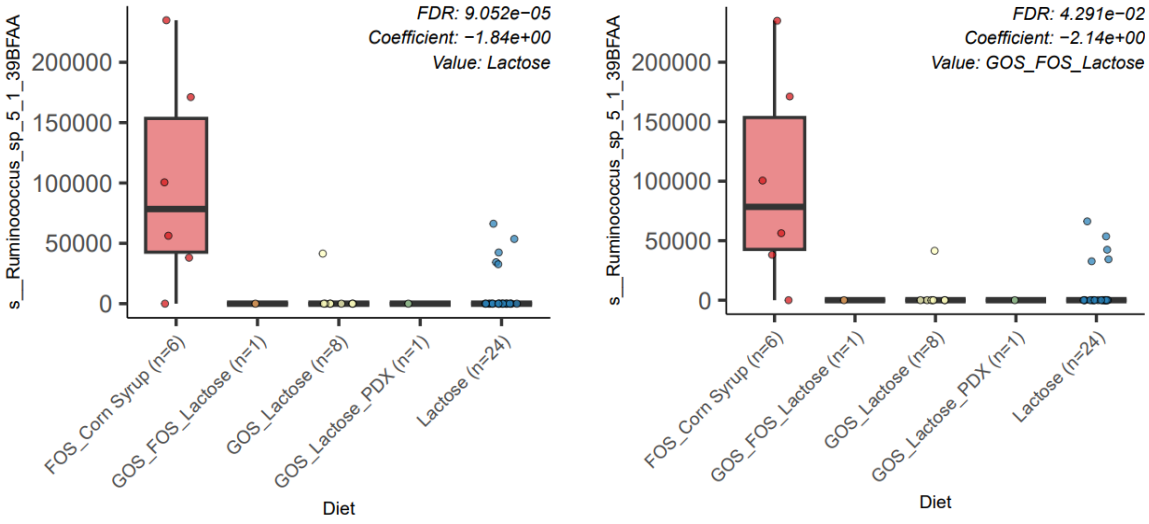


Figure 3.19: Impact of weaning diet on HGT driver in VD

CHAPTER 4: CONCLUSIONS AND FUTURE RECOMMENDATION

In conclusion, the findings highlight the dynamics of early life gut microbiome evolution and its potential implications for infant health and disease. Throughout the breastfeeding, weaning, and solid-food diet, distinct microbial compositions were observed between CSD and VD infants, shedding light on the intricate interplay between birth mode, dietary factors, and microbial colonization patterns.

In the early stages, CSD infants had a higher presence of *Enterobacteriaceae* and *Lachnospiraceae* families, while VD infants had more *Bifidobacteriales* and *Ruminococcaceae* families. Through HGT, this microbial environment was shaped by exchanging genetic material between specific bacteria. In the CSD infants, there were 447 known HGT drivers, whereas in the VD infants, there were 360. For example, in the case of CSD infants, Clostridium nexile was shown to transfer genes to Eubacterium ramulus, helping in breaking down polysaccharides with a log2 fold change of -2. In contrast, in VD infants, Butyricicoccus pullicaecorum transferred genes for Bifidobacterium longum that modulate inflammation and impact cognitive development, at a log2 fold change of +5.

Whereas during later stages, CSD infants drifted toward more Eubacteriaceae and Lachnospiraceae families, VD infants continued to have a robust representation of both Ruminococcaceae and Lachnospiraceae. In CSD infants, HGT transfer was noted, such as from *Salmonella enterica* to *Klebsiella pneumoniae*, and from *Butyricicoccus pullicaecorum* to *Bifidobacterium longum* in VD infants. These gene transfers thus had effects on processes such as posttranslational modification and the transduction of signals with a log2 fold change of +4. This association of *Coprococcus catus* and *Ruminococcus sp_5_1_39BFAA* has shown their crucial roles in carbohydrate metabolism, in particular

the degradation and fermentation of HMOs, in VD infants. *Ruminococcus sp_5_1_39BFAA* is negatively associated with lactose, showing a coefficient value of -1.84e+00 and an FDR of 9.052e-05. *Lachnospiraceae bacterium_2_1_46FAA* and *Ruminococcus gnavus* HGT drivers are found in VD infants and are positively associated with NICU-admitted infants, with coefficient values of 1.15e+00 and 1.12e+00, respectively, and FDR values of 2.281e-03 and 5.402e-03, respectively. The gene involved is xylB, which has antimicrobial properties and is involved in defense mechanisms.

In light of these findings, future recommendations must encompass a comprehensive approach to leverage the potential of the early-life gut microbiome for infant health optimization. Firstly, continued exploration of probiotic interventions tailored to specific microbial compositions, including species such as *Coprococcus catus* and *Lachnospiraceae bacterium_2_1_46FAA* , could offer promising avenues for therapeutic intervention. Secondly, longitudinal studies examining the intricate interplay between diet, microbial colonization, and health outcomes are imperative for informing targeted nutritional strategies. Furthermore, the development and refinement of microbiome-based therapies, such as fecal microbiota transplantation (FMT), hold promise for restoring microbial diversity and function in infants at risk of dysbiosis or associated health conditions.

REFERENCES

[1] A. Mullard and A. Mandavilli, ‘The inside story: the human body teems with microbes. In the first of two features, Asher Mullard looks at the global efforts to catalogue this vast “microbiome”. In the second, Apoorva Mandavilli meets the surgeons who have a rare opportunity to watch an ecosystem being established as they transplant guts from one person to another.’, *Nature*, vol. 453, no. 7195, pp. 578–581, May 2008, Accessed: Mar. 09, 2024. [Online].

[2] N. Cerf-Bensussan and V. Gaboriau-Routhiau, ‘The immune system and the gut microbiota: friends or foes?’, *Nature Reviews Immunology 2010 10:10*, vol. 10, no. 10, pp. 735–744, Sep. 2010, doi: 10.1038/nri2850.

[3] K. Hou *et al.*, ‘Microbiota in health and diseases’, *Signal Transduction and Targeted Therapy 2022 7:1*, vol. 7, no. 1, pp. 1–28, Apr. 2022, doi: 10.1038/s41392-022-00974-4.

[4] J. F. Wardman *et al.*, ‘Carbohydrate-active enzymes (CAZymes) in the gut microbiome’, *Nature Reviews Microbiology 2022 20:9*, vol. 20, no. 9, pp. 542–556, Mar. 2022, doi: 10.1038/s41579-022-00712-1.

[5] M. De Angelis *et al.*, ‘Diet influences the functions of the human intestinal microbiome’, *Scientific Reports 2020 10:1*, vol. 10, no. 1, pp. 1–15, Mar. 2020, doi: 10.1038/s41598-020-61192-y.

[6] P. Portincasa *et al.*, ‘Gut Microbiota and Short Chain Fatty Acids: Implications in Glucose Homeostasis’, *Int J Mol Sci*, vol. 23, no. 3, Feb. 2022, doi: 10.3390/IJMS23031105.

[7] J. Montenegro *et al.*, ‘Exploring the Influence of Gut Microbiome on Energy Metabolism in Humans’, *Advances in Nutrition*, vol. 14, no. 4, p. 840, Jul. 2023, doi: 10.1016/J.ADVNUT.2023.03.015.

[8] L. Y. Zhao *et al.*, ‘Role of the gut microbiota in anticancer therapy: from molecular mechanisms to clinical applications’, *Signal Transduction and Targeted Therapy 2023 8:1*, vol. 8, no. 1, pp. 1–27, May 2023, doi: 10.1038/s41392-023-01406-7.

[9] D. Zheng *et al.*, ‘Interaction between microbiota and immunity in health and disease’, *Cell Research 2020 30:6*, vol. 30, no. 6, pp. 492–506, May 2020, doi: 10.1038/s41422-020-0332-7.

[10] S. P. Wiertsema *et al.*, ‘The Interplay between the Gut Microbiome and the Immune System in the Context of Infectious Diseases throughout Life and the Role of Nutrition in Optimizing Treatment Strategies’, *Nutrients*, vol. 13, no. 3, pp. 1–14, Mar. 2021, doi: 10.3390/NU13030886.

[11] M. A. Núñez-Sánchez *et al.*, ‘Crohn’s Disease, Host–Microbiota Interactions, and Immunonutrition: Dietary Strategies Targeting Gut Microbiome as Novel Therapeutic Approaches’, *Int J Mol Sci*, vol. 23, no. 15, Aug. 2022, doi: 10.3390/IJMS23158361.

[12] R. Gao *et al.*, ‘Gut microbiota and colorectal cancer’, *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 36, no. 5, p. 757, May 2017, doi: 10.1007/S10096-016-2881-8.

[13] A. Fattorusso *et al.*, ‘Autism Spectrum Disorders and the Gut Microbiota’, *Nutrients*, vol. 11, no. 3, Mar. 2019, doi: 10.3390/NU11030521.

[14] V. I. Reshetnyak *et al.*, ‘Helicobacter pylori: Commensal, symbiont or pathogen?’, *World J Gastroenterol*, vol. 27, no. 7, p. 545, Feb. 2021, doi: 10.3748/WJG.V27.I7.545.

[15] C. D. Davis, ‘The Gut Microbiome and Its Role in Obesity’, *Nutr Today*, vol. 51, no. 4, p. 167, 2016, doi: 10.1097/NT.0000000000000167.

[16] M. Vacca *et al.*, ‘The Controversial Role of Human Gut Lachnospiraceae’, *Microorganisms*, vol. 8, no. 4, Apr. 2020, doi: 10.3390/MICROORGANISMS8040573.

[17] D. Khanna *et al.*, ‘Obesity: A Chronic Low-Grade Inflammation and Its Markers’, *Cureus*, vol. 14, no. 2, Feb. 2022, doi: 10.7759/CUREUS.22711.

[18] J. Wang *et al.*, ‘A metagenome-wide association study of gut microbiota in type 2 diabetes’, *Nature*, vol. 490, no. 7418, pp. 55–60, Oct. 2012, doi: 10.1038/NATURE11450.

[19] B. Khorsand *et al.*, ‘Overrepresentation of Enterobacteriaceae and Escherichia coli is the major gut microbiome signature in Crohn’s disease and ulcerative colitis; a comprehensive metagenomic analysis of IBDMDB datasets’, *Front Cell Infect Microbiol*, vol. 12, p. 1015890, Oct. 2022, doi: 10.3389/FCIMB.2022.1015890.

[20] R. C. Robertson *et al.*, ‘The Human Microbiome and Child Growth – First 1000 Days and Beyond’, *Trends Microbiol*, vol. 27, no. 2, pp. 131–147, Feb. 2019, doi: 10.1016/J.TIM.2018.09.008.

[21] M. Mathias, ‘Autointoxication and historical precursors of the microbiome–gut–brain axis’, *Microb Ecol Health Dis*, vol. 29, no. 2, p. 1548249, Nov. 2018, doi: 10.1080/16512235.2018.1548249.

[22] M. G. Dominguez-Bello *et al.*, ‘Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns’, *Proc Natl Acad Sci U S A*, vol. 107, no. 26, pp. 11971–11975, Jun. 2010, doi: 10.1073/PNAS.1002601107.

[23] Y. Yao *et al.*, ‘The Role of Microbiota in Infant Health: From Early Life to Adulthood’, *Front Immunol*, vol. 12, Oct. 2021, doi: 10.3389/FIMMU.2021.708472.

[24] Y. Yao *et al.*, ‘The Role of Microbiota in Infant Health: From Early Life to Adulthood’, *Front Immunol*, vol. 12, Oct. 2021, doi: 10.3389/FIMMU.2021.708472.

[25] Y. G. Kim *et al.*, 'Neonatal acquisition of Clostridia species protects against colonization by bacterial pathogens', *Science*, vol. 356, no. 6335, pp. 315–319, Apr. 2017, doi: 10.1126/SCIENCE.AAG2029.

[26] A. Marcobal *et al.*, 'Bacteroides in the Infant Gut Consume Milk Oligosaccharides via Mucus-Utilization Pathways', *Cell Host Microbe*, vol. 10, no. 5, p. 507, Nov. 2011, doi: 10.1016/J.CHOM.2011.10.007.

[27] L. W. Chia *et al.*, 'Bacteroides thetaiotaomicron Fosters the Growth of Butyrate-Producing Anaerostipes caccae in the Presence of Lactose and Total Human Milk Carbohydrates', *Microorganisms*, vol. 8, no. 10, pp. 1–13, Oct. 2020, doi: 10.3390/MICROORGANISMS8101513.

[28] P. Kovatcheva-Datchary *et al.*, 'Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella', *Cell Metab*, vol. 22, no. 6, pp. 971–982, 2015, doi: 10.1016/J.CMET.2015.10.001.

[29] N. T. Mueller *et al.*, 'The infant microbiome development: mom matters', *Trends Mol Med*, vol. 21, no. 2, p. 109, Feb. 2015, doi: 10.1016/J.MOLMED.2014.12.002.

[30] N. Shaterian *et al.*, 'Role of cesarean section in the development of neonatal gut microbiota: A systematic review', *Open Medicine*, vol. 16, no. 1, p. 624, Jan. 2021, doi: 10.1515/MED-2021-0270.

[31] K. Łubiech and M. Twarużek, 'Lactobacillus Bacteria in Breast Milk', *Nutrients*, vol. 12, no. 12, pp. 1–13, Dec. 2020, doi: 10.3390/NU12123783.

[32] M. F. Laursen *et al.*, 'Having older siblings is associated with gut microbiota development during early childhood', *BMC Microbiol*, vol. 15, no. 1, Aug. 2015, doi: 10.1186/S12866-015-0477-6.

[33] M. Fallani *et al.*, ‘Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres’, *Microbiology (N Y)*, vol. 157, no. 5, pp. 1385–1392, May 2011, doi: 10.1099/MIC.0.042143-0/CITE/REFWORKS.

[34] Z. Zhang *et al.*, ‘Association of infant antibiotic exposure and risk of childhood asthma: A meta-analysis’, *World Allergy Organ J*, vol. 14, no. 11, p. 100607, Nov. 2021, doi: 10.1016/J.WAOJOU.2021.100607.

[35] S. Y. Huh *et al.*, ‘Delivery by caesarean section and risk of obesity in preschool age children: a prospective cohort study’, *Arch Dis Child*, vol. 97, no. 7, pp. 610–616, Jul. 2012, doi: 10.1136/ARCHDISCHILD-2011-301141.

[36] M. Agrawal *et al.*, ‘Early life exposures and the risk of inflammatory bowel disease: Systematic review and meta-analyses’, *EClinicalMedicine*, vol. 36, Jun. 2021, doi: 10.1016/J.ECLINM.2021.100884.

[37] A. Amjad *et al.*, ‘Trends of caesarean section deliveries in Pakistan: secondary data analysis from Demographic and Health Surveys, 1990–2018’, *BMC Pregnancy Childbirth*, vol. 20, no. 1, pp. 1–13, Dec. 2020, doi: 10.1186/S12884-020-03457-Y/TABLES/4.

[38] O. C. Thompson-Chagoyán *et al.*, ‘Colonization and impact of disease and other factors on intestinal microbiota’, *Dig Dis Sci*, vol. 52, no. 9, pp. 2069–2077, Sep. 2007, doi: 10.1007/S10620-006-9285-Z/METRICS.

[39] L. Wampach *et al.*, ‘Colonization and succession within the human gut microbiome by archaea, bacteria, and microeukaryotes during the first year of life’, *Front Microbiol*, vol. 8, no. MAY, p. 738, May 2017, doi: 10.3389/FMICB.2017.00738/FULL.

[40] C. J. Hill *et al.*, ‘Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort’, *Microbiome*, vol. 5, no. 1, pp. 1–18, Jan. 2017, doi: 10.1186/S40168-016-0213-Y/FIGURES/8.

[41] E. Rinninella *et al.*, ‘What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases’, *Microorganisms*, vol. 7, no. 1, p. 14, Jan. 2019, doi: 10.3390/MICROORGANISMS7010014.

[42] M. G. Dominguez-Bello *et al.*, ‘Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns’, *Proc Natl Acad Sci U S A*, vol. 107, no. 26, pp. 11971–11975, Jun. 2010, doi: 10.1073/PNAS.1002601107.

[43] R. Nagpal and Y. Yamashiro, ‘Gut Microbiota Composition in Healthy Japanese Infants and Young Adults Born by C-Section’, *Ann Nutr Metab*, vol. 73 Suppl 3, pp. 4–11, Jul. 2018, doi: 10.1159/000490841.

[44] M. G. Dominguez-Bello *et al.*, ‘Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns’, *Proc Natl Acad Sci U S A*, vol. 107, no. 26, pp. 11971–11975, Jun. 2010, doi: 10.1073/PNAS.1002601107/SUPPL_FILE/PNAS.201002601SI.PDF.

[45] M. Dzidic *et al.*, ‘Gut Microbiota and Mucosal Immunity in the Neonate’, *Med Sci (Basel)*, vol. 6, no. 3, Jul. 2018, doi: 10.3390/MEDSCI6030056.

[46] M. Reyman *et al.*, ‘Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life’, *Nature Communications 2019 10:1*, vol. 10, no. 1, pp. 1–12, Nov. 2019, doi: 10.1038/s41467-019-13014-7.

[47] M. Barone *et al.*, ‘Gut barrier-microbiota imbalances in early life lead to higher sensitivity to inflammation in a murine model of C-section delivery’, *Microbiome*, vol. 11, no. 1, Dec. 2023, doi: 10.1186/S40168-023-01584-0.

[48] L. Wampach *et al.*, ‘Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential’, *Nature Communications 2018 9:1*, vol. 9, no. 1, pp. 1–14, Nov. 2018, doi: 10.1038/s41467-018-07631-x.

[49] F. Bäckhed *et al.*, ‘Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life’, *Cell Host Microbe*, vol. 17, no. 5, pp. 690–703, May 2015, doi: 10.1016/J.CHOM.2015.04.004.

[50] A. G. Shaw *et al.*, ‘Dynamics of toxigenic Clostridium perfringens colonisation in a cohort of prematurely born neonatal infants’, *BMC Pediatr*, vol. 20, no. 1, Feb. 2020, doi: 10.1186/S12887-020-1976-7.

[51] ‘Fecal Microflora in Healthy Infants Born by Different Method... : Journal of Pediatric Gastroenterology and Nutrition’. Accessed: Mar. 31, 2024. [Online]. Available: https://journals.lww.com/jpgn/fulltext/1999/01000/fecal_microflora_in_healthy_infants_born_by.7.aspx

[52] I. Adlerberth *et al.*, ‘Reduced Enterobacterial and Increased Staphylococcal Colonization of the Infantile Bowel: An Effect of Hygienic Lifestyle?’, *Pediatric Research 2006 59:1*, vol. 59, no. 1, pp. 96–101, Jan. 2006, doi: 10.1203/01.pdr.0000191137.12774.b2.

[53] H. Makino *et al.*, ‘Mother-to-Infant Transmission of Intestinal Bifidobacterial Strains Has an Impact on the Early Development of Vaginally Delivered Infant’s Microbiota’, *PLoS One*, vol. 8, no. 11, p. e78331, Nov. 2013, doi: 10.1371/JOURNAL.PONE.0078331.

[54] K. Vaher *et al.*, ‘Microbiome-gut-brain axis in brain development, cognition and behavior during infancy and early childhood’, *Developmental Review*, vol. 66, p. 101038, Dec. 2022, doi: 10.1016/J.DR.2022.101038.

[55] B. M. Henrick *et al.*, ‘Bifidobacteria-mediated immune system imprinting early in life’, *Cell*, vol. 184, no. 15, pp. 3884-3898.e11, Jul. 2021, doi: 10.1016/J.CELL.2021.05.030.

[56] A. Malamitsi-Puchner *et al.*, ‘The influence of the mode of delivery on circulating cytokine concentrations in the perinatal period’, *Early Hum Dev*, vol. 81, no. 4, pp. 387–392, 2005, doi: 10.1016/J.EARLHUMDEV.2004.10.017.

[57] E. B. Troy and D. L. Kasper, ‘Beneficial effects of Bacteroides fragilis polysaccharides on the immune system’, *Front Biosci (Landmark Ed)*, vol. 15, no. 1, pp. 25–34, Jan. 2010, doi: 10.2741/3603.

[58] M. Tanaka and J. Nakayama, ‘Development of the gut microbiota in infancy and its impact on health in later life’, *Allergol Int*, vol. 66, no. 4, pp. 515–522, Oct. 2017, doi: 10.1016/J.ALIT.2017.07.010.

[59] ‘Levels and trends in child mortality - UNICEF DATA’. Accessed: Mar. 13, 2024. [Online]. Available: https://data.unicef.org/resources/levels-and-trends-in-child-mortality-2024/

[60] Y. Moroishi *et al.*, ‘The relationship between the gut microbiome and the risk of respiratory infections among newborns’, *Communications Medicine 2022 2:1*, vol. 2, no. 1, pp. 1–8, Jul. 2022, doi: 10.1038/s43856-022-00152-1.

[61] J. C. Madan *et al.*, ‘Association of Cesarean Delivery and Formula Supplementation With the Intestinal Microbiome of 6-Week-Old Infants’, *JAMA Pediatr*, vol. 170, no. 3, pp. 212–219, Mar. 2016, doi: 10.1001/JAMAPEDIATRICS.2015.3732.

[62] T. Du *et al.*, ‘The Beneficial Role of Probiotic Lactobacillus in Respiratory Diseases’, *Front Immunol*, vol. 13, May 2022, doi: 10.3389/FIMMU.2022.908010.

[63] Y. Moroishi *et al.*, ‘The relationship between the gut microbiome and the risk of respiratory infections among newborns’, *Communications Medicine 2022 2:1*, vol. 2, no. 1, pp. 1–8, Jul. 2022, doi: 10.1038/s43856-022-00152-1.

[64] P. Bager *et al.*, ‘Caesarean delivery and risk of atopy and allergic disease: meta-analyses’, *Clin Exp Allergy*, vol. 38, no. 4, pp. 634–642, Apr. 2008, doi: 10.1111/J.1365-2222.2008.02939.X.

[65] C. Pothoulakis, ‘Effects of Clostridium difficile toxins on epithelial cell barrier’, *Ann N Y Acad Sci*, vol. 915, pp. 347–356, 2000, doi: 10.1111/J.1749-6632.2000.TB05263.X.

[66] L. Wampach *et al.*, ‘Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential’, *Nature Communications 2018 9:1*, vol. 9, no. 1, pp. 1–14, Nov. 2018, doi: 10.1038/s41467-018-07631-x.

[67] L. C. F. L. Melli *et al.*, ‘Intestinal microbiota and allergic diseases: A systematic review’, *Allergol Immunopathol (Madr)*, vol. 44, no. 2, pp. 177–188, Mar. 2016, doi: 10.1016/J.ALLER.2015.01.013.

[68] C.-C. Chen *et al.*, ‘Alterations in the gut microbiotas of children with food sensitization in early life’, *Pediatric Allergy and Immunology*, vol. 27, no. 3, pp. 254–262, May 2016, doi: 10.1111/PAI.12522.

[69] C. Hoskinson *et al.*, ‘Delayed gut microbiota maturation in the first year of life is a hallmark of pediatric allergic disease’, *Nature Communications 2023 14:1*, vol. 14, no. 1, pp. 1–14, Aug. 2023, doi: 10.1038/s41467-023-40336-4.

[70] F. De Filippis *et al.*, 'Specific gut microbiome signatures and the associated pro-inflamatory functions are linked to pediatric allergy and acquisition of immune tolerance', *Nature Communications 2021 12:1*, vol. 12, no. 1, pp. 1–11, Oct. 2021, doi: 10.1038/s41467-021-26266-z.

[71] '2020 National Health Interview Survey (NHIS) Data | CDC'. Accessed: Nov. 19, 2023. [Online]. Available: https://www.cdc.gov/asthma/nhis/2020/data.htm

[72] 'Asthma Data Visualizations | CDC'. Accessed: Nov. 19, 2023. [Online]. Available: https://www.cdc.gov/asthma/data-visualizations/default.htm

[73] 'AsthmaStats: Uncontrolled Asthma among Children, 2012–2014 | CDC'. Accessed: Nov. 19, 2023. [Online]. Available: https://www.cdc.gov/asthma/asthma_stats/uncontrolled-asthma-children.htm

[74] I. Tarrant and B. B. Finlay, 'Like mother, like child: The maternal microbiome impacts offspring asthma', *Cell Rep Med*, vol. 3, no. 8, p. 100722, Aug. 2022, doi: 10.1016/J.XCRM.2022.100722.

[75] S. Morrill *et al.*, 'Gardnerella vaginalis as a Cause of Bacterial Vaginosis: Appraisal of the Evidence From in vivo Models', *Front Cell Infect Microbiol*, vol. 10, p. 521471, Apr. 2020, doi: 10.3389/FCIMB.2020.00168/BIBTEX.

[76] M. C. Arrieta *et al.*, 'Early infancy microbial and metabolic alterations affect risk of childhood asthma', *Sci Transl Med*, vol. 7, no. 307, Sep. 2015, doi: 10.1126/SCITRANSLMED.AAB2271/SUPPL_FILE/7-307RA152_SM.PDF.

[77] M. C. Arrieta *et al.*, 'Early infancy microbial and metabolic alterations affect risk of childhood asthma', *Sci Transl Med*, vol. 7, no. 307, Sep. 2015, doi: 10.1126/SCITRANSLMED.AAB2271/SUPPL_FILE/7-307RA152_SM.PDF.

[78] L. M. Williams *et al.*, ‘Soluble fibre as a treatment for inflammation in asthma’, *J Nutr Intermed Metab*, vol. 18, p. 100108, Dec. 2019, doi: 10.1016/J.JNIM.2019.100108.

[79] L. Blicharz *et al.*, ‘Staphylococcus aureus: an underestimated factor in the pathogenesis of atopic dermatitis?’, *Advances in Dermatology and Allergology/Postępy Dermatologii i Alergologii*, vol. 36, no. 1, p. 11, 2019, doi: 10.5114/ADA.2019.82821.

[80] P. A. Dennery *et al.*, ‘Neonatal Hyperbilirubinemia’, *https://doi.org/10.1056/NEJM200102223440807*, vol. 344, no. 8, pp. 581–590, Feb. 2001, doi: 10.1056/NEJM200102223440807.

[81] S. Zhou *et al.*, ‘Association of serum bilirubin in newborns affected by jaundice with gut microbiota dysbiosis’, *J Nutr Biochem*, vol. 63, pp. 54–61, Jan. 2019, doi: 10.1016/J.JNUTBIO.2018.09.016.

[82] Y. Wang *et al.*, ‘Gut Microbiota Dysbiosis Is Associated with Altered Bile Acid Metabolism in Infantile Cholestasis’, *mSystems*, vol. 4, no. 6, Dec. 2019, doi: 10.1128/MSYSTEMS.00463-19/SUPPL_FILE/MSYSTEMS.00463-19-SF001.TIF.

[83] S. Zhou *et al.*, ‘Association of serum bilirubin in newborns affected by jaundice with gut microbiota dysbiosis’, *J Nutr Biochem*, vol. 63, pp. 54–61, Jan. 2019, doi: 10.1016/J.JNUTBIO.2018.09.016.

[84] ‘Incidence of diarrhoea among children aged <5 years (%)’. Accessed: Nov. 19, 2023. [Online]. Available: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/3148

[85] J. P. Nataro, ‘Diarrhea among children in developing countries’, *Adv Exp Med Biol*, vol. 764, pp. 73–80, 2013, doi: 10.1007/978-1-4614-4726-9_5/COVER.

[86] N. O. Kaakoush *et al.*, 'Global Epidemiology of Campylobacter Infection', *Clin Microbiol Rev*, vol. 28, no. 3, pp. 687–720, 2015, doi: 10.1128/CMR.00006-15.

[87] M. M. Levine *et al.*, 'Diarrhoeal disease and subsequent risk of death in infants and children residing in low-income and middle-income countries: analysis of the GEMS case-control study and 12-month GEMS-1A follow-on study', *Lancet Glob Health*, vol. 8, no. 2, pp. e204–e214, Feb. 2020, doi: 10.1016/S2214-109X(19)30541-8.

[88] M. Boolchandani *et al.*, 'Impact of international travel and diarrhea on gut microbiome and resistome dynamics', *Nature Communications 2022 13:1*, vol. 13, no. 1, pp. 1–19, Dec. 2022, doi: 10.1038/s41467-022-34862-w.

[89] Q. Fan *et al.*, 'The Impact of Age and Pathogens Type on the Gut Microbiota in Infants with Diarrhea in Dalian, China', *Can J Infect Dis Med Microbiol*, vol. 2020, 2020, doi: 10.1155/2020/8837156.

[90] R. Nagpal *et al.*, 'Sensitive quantitative analysis of the meconium bacterial microbiota in healthy term infants born vaginally or by cesarean section', *Front Microbiol*, vol. 7, no. DEC, p. 235273, Dec. 2016, doi: 10.3389/FMICB.2016.01997/BIBTEX.

[91] A. M. Baumann-Dudenhoeffer *et al.*, 'Infant diet and maternal gestational weight gain predict early metabolic maturation of gut microbiomes', *Nat Med*, vol. 24, no. 12, pp. 1822–1829, Dec. 2018, doi: 10.1038/S41591-018-0216-2.

[92] D. A. Benson *et al.*, 'GenBank', *Nucleic Acids Res*, vol. 38, no. Database issue, Nov. 2010, doi: 10.1093/NAR/GKP1024.

[93] 'Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data'. Accessed: Mar. 31, 2024. [Online]. Available: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[94] S. Chen *et al.*, ‘fastp: an ultra-fast all-in-one FASTQ preprocessor’, *Bioinformatics*, vol. 34, no. 17, pp. i884–i890, Sep. 2018, doi: 10.1093/BIOINFORMATICS/BTY560.

[95] ‘BBMap download | SourceForge.net’. Accessed: Mar. 31, 2024. [Online]. Available: https://sourceforge.net/projects/bbmap/

[96] D. Li *et al.*, ‘MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices’, *Methods*, vol. 102, pp. 3–11, Jun. 2016, doi: 10.1016/J.YMETH.2016.02.020.

[97] S. Nurk *et al.*, ‘MetaSPAdes: A new versatile metagenomic assembler’, *Genome Res*, vol. 27, no. 5, pp. 824–834, May 2017, doi: 10.1101/GR.213959.116/-/DC1.

[98] ‘Welcome to KBase Predictive Biology | KBase’. Accessed: Mar. 31, 2024. [Online]. Available: https://www.kbase.us/

[99] M. Wu and A. J. Scott, ‘Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2’, *Bioinformatics*, vol. 28, no. 7, pp. 1033–1034, Apr. 2012, doi: 10.1093/BIOINFORMATICS/BTS079.

[100] ‘GitHub - wwood/CoverM: Read coverage calculator for metagenomics’. Accessed: Apr. 08, 2024. [Online]. Available: https://github.com/wwood/CoverM

[101] ‘WAAFLE – The Huttenhower Lab’. Accessed: Apr. 08, 2024. [Online]. Available: https://huttenhower.sph.harvard.edu/waafle

[102] ‘humann2 – The Huttenhower Lab’. Accessed: Apr. 08, 2024. [Online]. Available: https://huttenhower.sph.harvard.edu/humann2

[103] A. Nadkarni *et al.*, ‘Feasibility, acceptability and cost-effectiveness of a brief, lay counsellor-delivered psychological treatment for men with alcohol dependence in primary

care: an exploratory randomized controlled trial', *Addiction (Abingdon, England)*, vol. 114, no. 7, p. 1192, Jul. 2019, doi: 10.1111/ADD.14630.

[104] 'DESEQ2 R Tutorial'. Accessed: Apr. 08, 2024. [Online]. Available: https://lashlock.github.io/compbio/R_presentation.html

[105] T. Vatanen *et al.*, 'Mobile genetic elements from the maternal microbiome shape infant gut microbial assembly and metabolism', *Cell*, vol. 185, pp. 4921-4936.e15, 2022, doi: 10.1016/j.cell.2022.11.023.

[106] T. Vatanen *et al.*, 'Mobile genetic elements from the maternal microbiome shape infant gut microbial assembly and metabolism', *Cell*, vol. 185, no. 26, p. 4921, Dec. 2022, doi: 10.1016/J.CELL.2022.11.023.

[107] C. Li *et al.*, 'Understanding Horizontal Gene Transfer network in human gut microbiota', *Gut Pathog*, vol. 12, no. 1, pp. 1–20, Jul. 2020, doi: 10.1186/S13099-020-00370-9/TABLES/1.

[108] K. Hou *et al.*, 'Microbiota in health and diseases', *Signal Transduct Target Ther*, vol. 7, no. 1, Dec. 2022, doi: 10.1038/S41392-022-00974-4.

[109] A. H. Behling *et al.*, 'Horizontal gene transfer after faecal microbiota transplantation in adolescents with obesity', *Microbiome*, vol. 12, no. 1, Dec. 2024, doi: 10.1186/S40168-024-01748-6.

[110] C. J. Stewart *et al.*, 'Temporal development of the gut microbiome in early childhood from the TEDDY study', *Nature*, vol. 562, no. 7728, pp. 583–588, Oct. 2018, doi: 10.1038/S41586-018-0617-X.

[111] F. Bäckhed *et al.*, ‘Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life’, *Cell Host Microbe*, vol. 17, no. 5, pp. 690–703, May 2015, doi: 10.1016/J.CHOM.2015.04.004.

[112] E. J. De Muinck and P. Trosvik, ‘Individuality and convergence of the infant gut microbiota during the first year of life’, *Nature Communications 2018 9:1*, vol. 9, no. 1, pp. 1–8, Jun. 2018, doi: 10.1038/s41467-018-04641-7.

[113] C. Zhang *et al.*, ‘The Effects of Delivery Mode on the Gut Microbiota and Health: State of Art’, *Front Microbiol*, vol. 12, p. 724449, Dec. 2021, doi: 10.3389/FMICB.2021.724449.

[114] A. Lerner *et al.*, ‘Potential Effects of Horizontal Gene Exchange in the Human Gut’, *Front Immunol*, vol. 8, no. NOV, p. 27, Nov. 2017, doi: 10.3389/FIMMU.2017.01630.

[115] C. R. Hsu *et al.*, ‘Klebsiella pneumoniae Translocates across the Intestinal Epithelium via Rho GTPase- and Phosphatidylinositol 3-Kinase/Akt-Dependent Cell Invasion’, *Infect Immun*, vol. 83, no. 2, p. 769, 2015, doi: 10.1128/IAI.02345-14.

[116] H. M. Natama *et al.*, ‘Genetic variation in the immune system and malaria susceptibility in infants: a nested case–control study in Nanoro, Burkina Faso’, *Malar J*, vol. 20, no. 1, pp. 1–14, Dec. 2021, doi: 10.1186/S12936-021-03628-Y/FIGURES/2.

[117] C. Xiao *et al.*, ‘Associations of dietary diversity with the gut microbiome, fecal metabolites, and host metabolism: results from 2 prospective Chinese cohorts’, *Am J Clin Nutr*, vol. 116, no. 4, p. 1049, Oct. 2022, doi: 10.1093/AJCN/NQAC178.

[118] A. Lerner *et al.*, ‘Potential effects of horizontal gene exchange in the human gut’, *Front Immunol*, vol. 8, no. NOV, p. 298585, Nov. 2017, doi: 10.3389/FIMMU.2017.01630/BIBTEX.