Volume 10:1-12

UJEMI+

The Undergraduate Journal of Experimental Microbiology & Immunology (+Peer Reviewed)

Bacterial Composition and Metabolic Pathways Differ in Six-Month-Old Infants from Peru and California, USA

Stella Lin, Kevin Le, Risa Fox, Michael Qiu, Dennis Xie

Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada

SUMMARY The infant gut microbiome rapidly develops with age and is influenced by various factors. The potential influence of an infant's geographical origin on these adaptations should be considered. We compared six-month-old infants from two distinct locations — Iquitos, Peru, and San Diego, California, USA —to identify geography-related differences in infant gut microbiomes. This comparison was conducted by analyzing microbial diversity, composition, and functional phenotype in fecal samples. Beta diversity analysis suggested statistically significant differences in microbial communities between the two infant cohorts. Functional analysis using the PICRUSt2 software revealed an overrepresentation of different bacterial metabolic pathways associated with infant development in both infants from Peru and California. However, taxonomic results showed that the bacteria commonly associated with these metabolic pathways, namely, *Faecalibacterium*, *Lactobacillus*, and *Bifidobacterium,* were more abundant in the infants from Peru than from California. These findings varied in similar patterns between the two cohorts, suggesting a potential difference in the microbiome of the infants from these geographical locations. This highlights the need for further studies to directly uncover and characterize microbial variations between infants from different regions through direct location comparisons between multiple groups as well as data collection in understudied geographical locations and demographics. Ultimately, this may improve the quality of results and encourage further studies on diverse geographical locations and demographics to expand our knowledge on the infant intestinal microbiota and different functional activity, worldwide.

INTRODUCTION

The establishment of a healthy gut microbiome in early life is critical for long-term growth and development (1). Shaped by factors such as host genetics, prenatal environment, and delivery mode (2), the infant microbiome undergoes rapid development (3). Further modulated by postnatal factors such as antibiotic treatment, diet, and environmental exposure (2), the microbiome reaches a stable state with diversity and composition resembling that of adults by the age of three-years (3). Once established, the composition of the gut microbiota remains relatively stable throughout adult life and is usually only altered by bacterial infections, antibiotic treatments, surgery, and long-term changes in lifestyle and diet (4). In particular, the composition and diversity of the gut microbiota is important as they perform various functions such as protecting against enteropathogens, aiding in food digestion, and contributing to immune function (5). Therefore, establishing a microbiome that is both balanced and diverse is critical for an infant's immediate and long-term health. Another reason why the diversity and composition of the infant microbiome are critical is that its bacterial pathways may regulate bodily functions. For example, brain activity and cognitive function have been suggested to be

Published Online: September 2024

Citation: Lin, Le, Fox, Qiu, Xie. 2024. Bacterial composition and metabolic pathways differ in six-month-old infants from Peru and California, USA. UJEMI+ 10:1-12

Editor: Shruti Sandilya, University of British Columbia

Address correspondence to: https://jemi.microbiology.ubc.ca/

modulated by the microbiota through the gut-brain axis, which regulates the production, transportation, and functioning of neurotransmitters (6).

Though there is robust research on the development of the infant gut microbiome and its importance in early life outcomes, studies comparing the microbial differences between infants from differing geographical locations are limited. Some studies, such as those conducted by Collado et al. (7) and Kemppainen et al. (8), have shown that geographic location impacts the overall structure of the infant microbiome. Collado et al. compared two populations of mothers and infants from China and Spain and found that geographical location had a greater impact on the infant microbiota than perinatal factors such as delivery mode and maternal body mass index (7). Kemppainen et al. compared infant microbiomes from the USA, Germany, Sweden, and Finland, and found that bacterial community diversity differed significantly by geographical location, and further found that the risk for type I diabetes was positively correlated with location (8).

Although some studies comparing geographical locations have already been conducted, further studies conducted with different permutations of locations, including locations that vary ethnically and geographically, will help add to a more holistic understanding of location-based impacts. Our study compares Iquitos, Peru and San Diego, California, USA, which have different ethnogeographic populations that are known to have distinct genetic backgrounds, regional diets, and cultural practices (9). Also, mid- and high- income nations may not have the same level of healthcare and sanitation practices (10).

Our research aims to investigate the microbial difference between cohorts of six-month-old infants from these two geographical locations. Specifically, we will consider differences in microbial diversity, composition, and functional phenotype of the gut microbiota in the infants from these regions. A key focus of this study will be the examination of differences in the representation of metabolic pathways and their potential implications for infant development and health. Our data will be sourced by the integration of two datasets: fecal microbiome data collected from healthy and anemic infants collected by McClorry et al. (11), and fecal microbiome data collected from infant-mother dyads by Rhee et al. (12). The results of this study will contribute to the growing body of knowledge regarding geographic variations in microbial composition and biochemical pathways. Research on underrepresented locations, such as Peru, is important for enhancing the generalizability and validity of findings, and potentially uncovering and understanding health disparities and dispelling biases towards these populations. This may encourage further exploration of diverse locations to increase the understanding of different infant gut microbiomes worldwide.

METHODS AND MATERIALS

Dataset Description. Two different datasets were used in this study. The first dataset was derived from the Infant Feeding Practices Study II, conducted in California, USA (12). This study included stool samples from infant-mother pairs collected at different time points of life from two-weeks to twelve-months of age, totaling 325 samples. The metadata provided information on the infant's age, sex, weight, and eating behaviors as well as dietary, perinatal, and medical history. The second dataset was from a study by Shannon McClorry et al. (11), which investigated the association between anemia and microbiome in infants. A total of 193 infants at either six-months or twelve-months of age were recruited to the Moronacocha Health Center in Iquitos, Peru for serum and stool sample collection. The metadata describes each subject's age, sex, weight, diet, and blood attributes, such as blood iron content, and anemia status. Stool samples from both studies were subjected to sequencing using the Illumina MiSeq platform. Both studies focused on the V4 region of the 16S rRNA gene and were amplified using 515fbc and 806r primers.

Sample filtering. As our study focused on infants, we removed parental data from the Infant Feeding Practices Study II (12). Additionally, infants with anemia or a missing anemia status were excluded from the study by Shannon McClorry et al. (11). Both datasets included samples at two common time points: six-month and twelve-month-old infants. Due to an insufficient number of samples from the twelve-month time point, we focused solely on the six-month-old infant data. To account for the significant impact of breastfeeding on the gut

microbiome development and microbial diversity (13), we filtered the infant data to include those who were either exclusively breastfed or fed a combination of breast milk and formula. Infants with combination diets were excluded from the study if the quantity of breast milk versus formula was not measured. After filtering, 15 samples from the infant feeding (12) and 46 samples from the anemia dataset (11) were included in the downstream analysis.

Data processing using the Quantitative Insights into Microbial Ecology version 2 (QIIME2). The 16S RNA sequences from each dataset were imported and demultiplexed using QIIME. The Divisive Amplicon Denoising Algorithm 2 (DADA2) pipeline was used for quality control and denoising for the generation of amplicon sequence variants (ASVs) (14). The truncation length used in both datasets was 150 to ensure a median Phred score of at least 30, while maintaining consistent read length between the two datasets.

Taxonomic analysis. A new classifier was trained with the SILVA 16S database using 515F and 806R V4 region primer sequences (15, 16). Taxonomy information was assigned to the processed sequences using the trained classifier. Additional filtering steps removed mitochondria and chloroplast associated sequences.

Diversity analysis and differential Abundance analysis. Diversity analysis was performed in R (version 4.3.2) (17) and RStudio (version 2023.12.1+402) (18) using the R packages: phyloseq (version 1.44.0) (19), tidyverse (version 2.0.0) (20), ape (version 5.7-1) (21), and vegan (version 2.6-4) (22). Both datasets were rarefied to a sequencing depth of 1600. This depth reached a plateau of the rarefaction curve in both datasets, providing a good estimate of the species richness while retaining all but one sample from the Peru dataset. Alpha diversity between the two datasets was assessed using the Shannon Diversity Index, which was selected to account for both the richness and evenness of species in each dataset. Statistical significance was determined using the Wilcoxon rank sum test. Beta diversity was compared using Bray-Curtis dissimilarity, which was chosen for its effectiveness in handling differences in sample sizes and species abundance. Statistical significant for beta diversity was assessed using permutation multivariate analysis of variance (PERMANOVA). Differential abundance analysis was carried out using the package DESeq2 (version 1.42.0) in RStudio (23). For all statistical testing, a p or p-adj value <0.05 was considered significant.

Microbiome functional analysis. The PICRUSt2 (version 2.5.2) pipeline was used to predict the functional profile of the gut microbiome community in each sample (24). Annotated files were exported for downstream analysis in RStudio using the ggpicrust2 package (version 1.7.1) (25). Differential abundance analysis (DAA) was conducted using the DESeq2 method integrated in ggpicrust2 to assess the MetaCyc pathway dataset.

RESULTS

Diversity analysis revealed a significant difference in microbial community composition between infants from Peru and California. Alpha-diversity, as assessed by Shannon's index, found no significant differences between the two cohorts (Figure S1). However, beta-diversity analysis using Bray-Curtis dissimilarity indicated distinct clustering between the two cohorts, as visualized as a PCoA plot showing distinct clustering along the y-axis (Figure 1). This observation was supported by the PERMANOVA test, which confirmed a significant difference in gut microbiome composition ($p <0.001$). The unique clustering observed in the PCoA plot suggests that geographical location may be a potential factor influencing gut microbiome community composition.

Divergent phylum-level communities in the gut microbiome of infants from different geographic locations. Taxonomic analysis provided an initial overview of the variance in gut microbiome composition between infants from the two locations. The taxonomic bar plot revealed distinct differences in the gut microbiome at the phylum level between the two cohorts (Figure 2). Upon closer examination, the contribution of Actinobacteria and Firmicutes to the total gut microbial population was higher in the Peru cohort, while the contribution of Proteobacteria was higher in the California cohort.

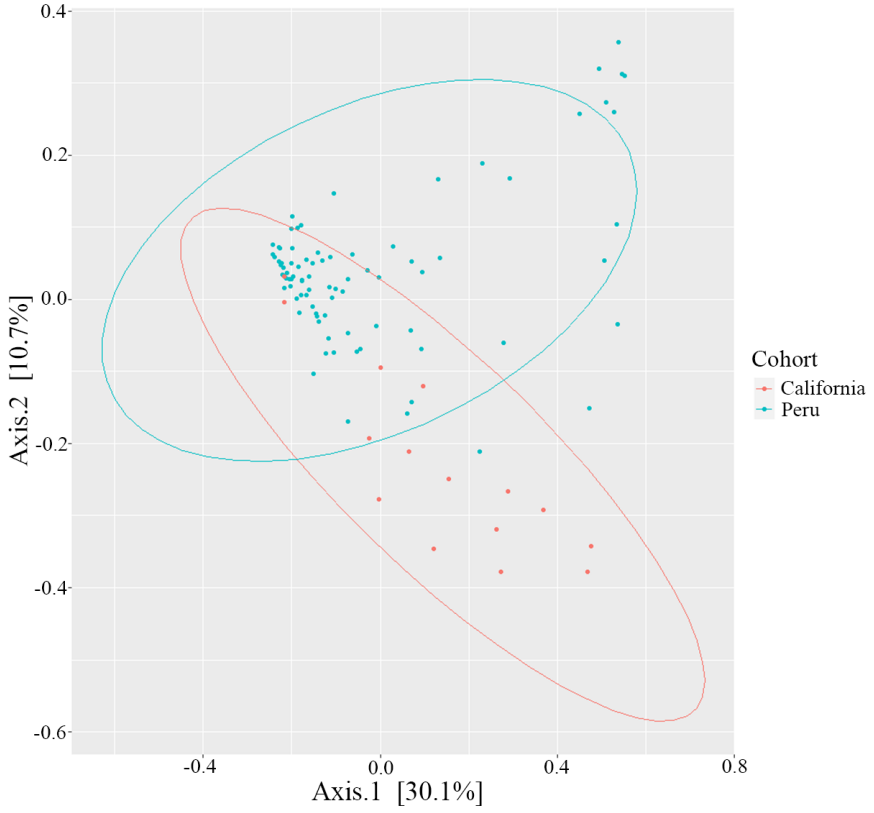

FIG. 1 Figure 1. Beta diversity shows significant differences between cohorts from different geographical locations. Principal Coordinates analysis (PCoA) plot of beta diversity by Bray Curtis dissimilarity between Peru (blue) cohort and California (red) cohort. PERMANOVA test was used to demonstrate a statistically significant difference between the two groups ($p < 0.001$).

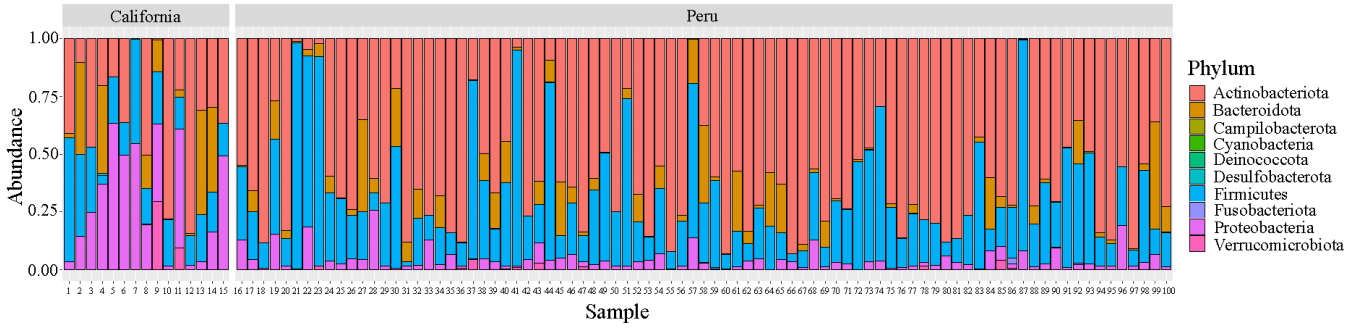

FIG. 2 Taxonomic bar plot showing distinct gut microbiome makeup in different geographical locations. The taxonomic bar plot showed the relative abundance of gut bacteria taxonomic composition at the phylum level between the two cohorts California (left; samples 1-15), and Peru (right; samples 16-100). The samples, originally processed from FASTQ files, are represented by a numerical identifier. Ten distinct bacterial phyla are represented, revealing distinct compositional differences.

More differentially abundant genera are overrepresented in infants from Peru than infants from California. To compare the relative abundance of different taxonomic groups between infants from California and Peru, differential abundance analysis was conducted using DESeq2 (Figure 3), with Peruvian infants as the control group. The analysis identified 65 genera that were significantly more abundant in Peruvian infants compared to Californian infants, and 11 genera that were more abundant in Californian infants compared to Peruvian infants. Notably, *Lactobacillus, Bifidobacterium, Clostridium*, and *Faecalibacterium* were significantly more abundant in Peruvian infants, with a log2 fold change ranging from 5 to 10. *Lactobacillus, Clostridium*, and *Faecalibacterium* are all members of the Firmicutes phylum.

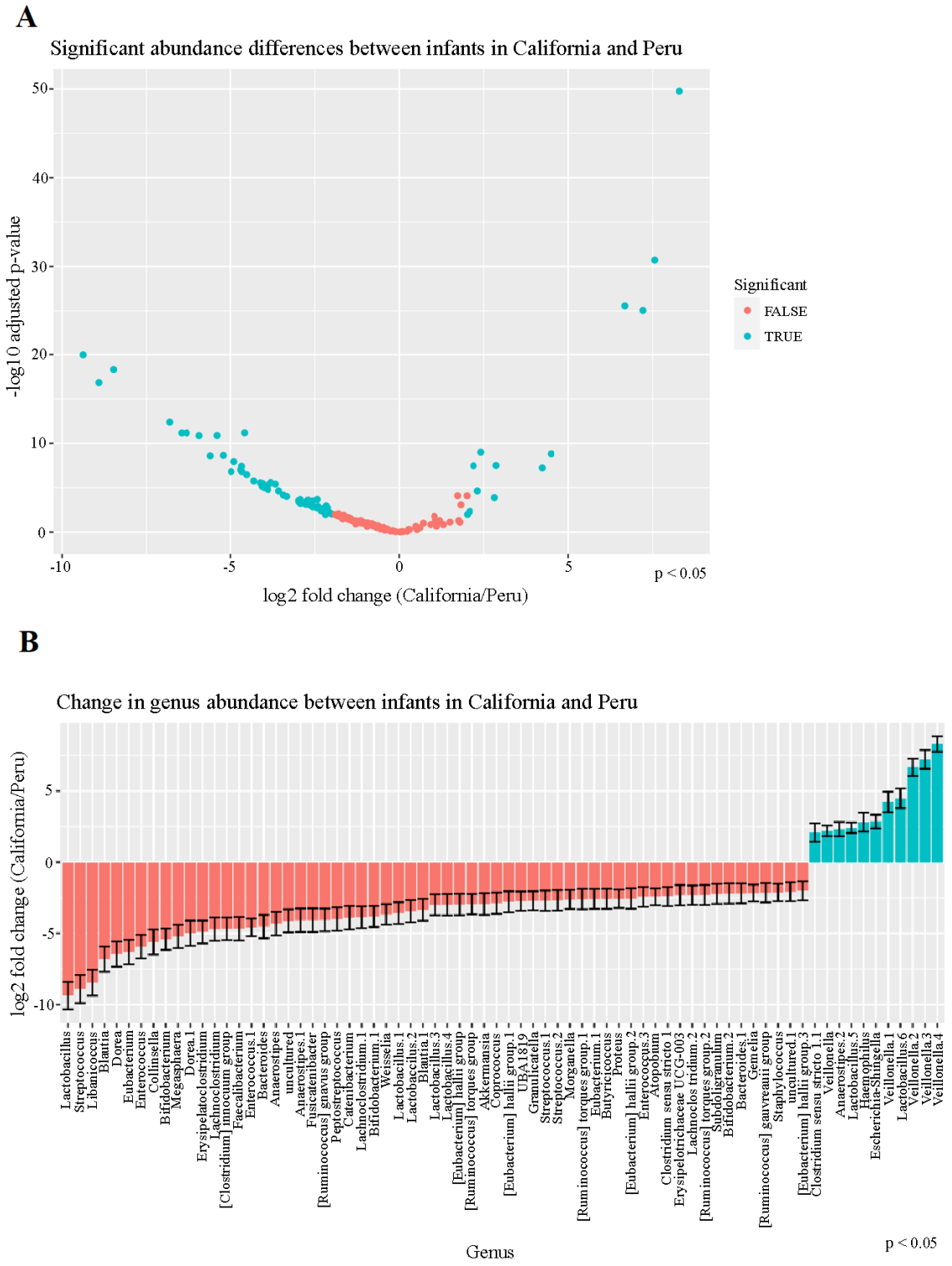

FIG. 3 Many genera have decreased abundance and few have increased abundance in Californian infants when compared to Peruvian infants. Differential expression analysis using DESeq2 showing a bar plot (B) and volcano plot (A) showing the log2 fold change between different genera of the gut microbiome of infants from two different locations, California and Peru (control). A positive log2 fold change shown on the right-hand side of the graph (A and B) shows increased abundance of genera in Californian infants compared to Peruvian infants. A negative log2 fold change shown on the left-hand side of the graph (A and B) show increased abundance of genera in Peruvian infants compared to Californian infants. The error bars on plot B represent the calculated standard deviation (SD) of the log2 fold change. Data points are shown for n = 61.

More metabolic pathways critical for infant growth and development are overrepresented in California relative to Peru. The PICRUSt2 analysis pipeline was used to predict gut bacterial functional pathways using the MetaCyc dataset, aiming to assess whether geography is a potential driving factor behind the functional changes in the microbiome. Differential abundance analysis using DESeq2 identified 20 differentially abundant metabolic pathways in Californian infants, where 13 were overrepresented and 7 were underrepresented relative to Peruvian infants (Figure 4). In Californian infants, pathways related to neurological development were notably overrepresented. These include 4-hydroxyphenylacetate degradation, phenylacetate degradation I (aerobic), superpathway of phenylethylamine degradation, and the urea cycle. Additionally, pathways related to key vitamin synthesis were also overrepresented in the California infants. These include: 1,4-dihydroxy-2-naphthoate biosynthesis I, superpathway of phylloquinol biosynthesis, and biotin biosynthesis II. Conversely, in Peruvian infants, certain metabolic pathways with implications for infant health and development were overrepresented, relative to California. Notably, pyrimidine deoxyribonucleotides de novo biosynthesis IV and pyrimidine deoxyribonucleotides biosynthesis from CTP were overrepresented, both of which are associated with neurological development.

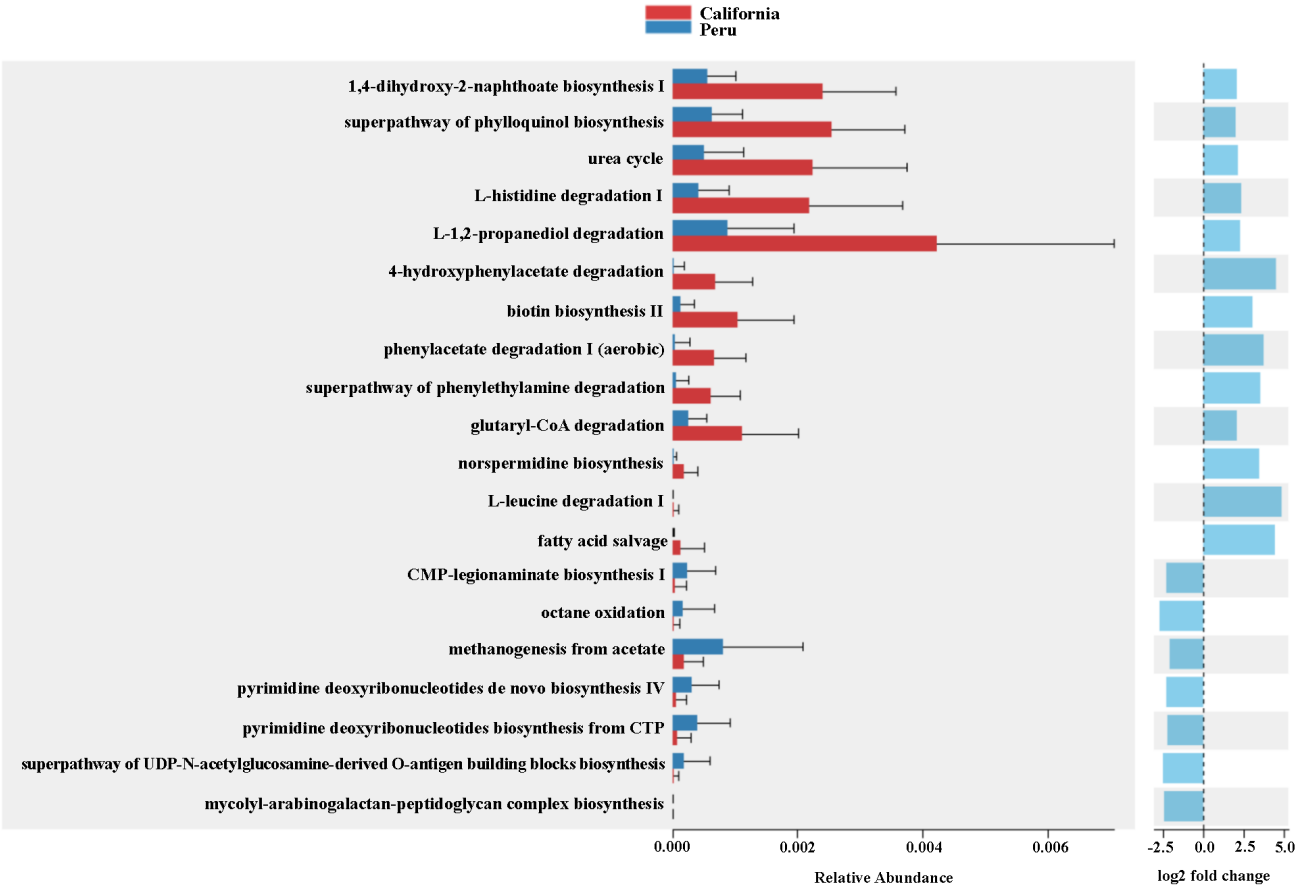

FIG. 4 Predicted differential abundance of metabolic pathways. Differential abundance analysis using DESeq2 conducted with PICRUSt2 visualized as a bar plot depicts relative abundance of metabolic pathways (shown on the left-hand side of the graph) found in infants of different geographical locations, California (in red) and Peru (in blue; control). Log2 fold change was calculated (shown on the right-hand side of the graph), where a positive log2 fold change represents an upregulated pathway, while a negative log2 fold change represents an underrepresented pathway.

DISCUSSION

We used previously published metadata by McClorry et al. and Rhee et al. to investigate differences in the intestinal microbiome of healthy six-month-old infants from two distinct geographical locations: San Diego, California, USA and Iquitos, Peru. Our primary objective was to understand how geographical variations might influence microbial composition and infant health and development. Our initial analyses using alpha diversity metrics did not reveal significant differences, suggesting that the microbiomes within Peru and California display comparable richness and evenness. However, analyses with beta diversity using Bray-Curtis Dissimilarity highlighted significant differences, demonstrating pronounced dissimilarity in the microbial community composition between the two locations.

Taxonomic analysis identified Actinobacteria, Firmicutes, and Proteobacteria as the key phylum exhibiting differential distribution in the Peru and California cohorts (Figure 2). Further examination using differential abundance analysis revealed compositional differences within the Firmicutes phylum present between Peru and California. Specifically, the Peru cohort showed greater abundance in *Lactobacillus*, *Faecalibacterium*, *Clostridium*, and *Eubacterium*, while *Veillonella* was more prevalent in California (Figure 3). *Bifidobacteria,* a genus within the Actinobacteria phylum, was also found to be more prevalent in Peru. However, as there was an imbalance in sample sizes between the cohorts, with California being six times smaller, genera with fewer appearances may be a result of rarity of certain ASVs rather than geographical differences. To explore the functional implications of the taxonomic composition differences, we conducted a PICRUSt2 analysis to assess how variations in microbial composition might impact metabolic processes.

Results from the bacterial functional abundance using PICRUSt2 revealed that infants from San Diego, California, USA had a higher number of overrepresented bacterial metabolic pathways compared to infants from Iquitos, Peru. Specifically, pathways associated with phenylethylamine (PEA) degradation, specifically, 4-hydroxyphenylacetate degradation, phenylacetate degradation I (aerobic), and the superpathway of PEA degradation, were notably overrepresented in the California cohort (Figure 4). PEA is a trace amine metabolite of phenylalanine (26), a nutrient found in breast milk. Phenylalanine is metabolized into various neurotransmitter precursors and hormones, such as tyrosine (27). PEA typically is short-lived due to its rapid degradation by monoamine oxydase-B (MAO-B) (26). However, a case study from Ghozlan et al. (2004) reported that PEA might be more toxic than phenylalanine and other metabolites when MAO-B activity is low, potentially leading to irreversible brain injury (26). Moreover, the improper degradation of PEA is associated with an increased risk of infants developing phenylketonuria (PKU), a condition linked to cognitive and behavioral dysfunctions (26). In our study, the various pathways involved in PEA degradation were overrepresented in the California cohort, suggesting that infants from California may have a higher level of activity for PEA degradation compared to Peru. Contrary to our results, a recent study conducted by Hillert et al. (2020) examined the epidemiology of PKU in 64 countries, in which PKU was observed to be more prevalent in the USA (1:25,000 of its total population) compared to in Peru (1:46,970 of its total population) (28). This discrepancy highlights the need for further investigation to associate these findings with a broader set of epidemiological data.

PEA degradation also produces phenylacetate (PAA), which is oxidized into phenylacetyl-CoA. Phenylacetyl-CoA provides microbes with energy for degrading aromatic compounds, such as amines (29), which are then excreted through the urea cycle. Efficient recycling of amines is important to reduce the infant's risk of developing hyperammonemia, characterized by elevated levels of ammonia in the blood (30). Hyperammonemia is associated with neurological deficits, such as seizure and encephalopathy (31). Our results showcase a two-fold overrepresentation of the urea cycle in Californian infants compared to Peruvian infants. Although minimal studies have been conducted to correlate the prevalence of hyperammonemia to different geographical regions, our findings suggest that infants in California may exhibit higher ammonia recycling activity, potentially reducing their risk for developing hyperammonemia (28). Furthermore, excessive accumulation of PAA in the blood can lead to neuronal degeneration and an elevated risk of developing PKU, particularly following its reaction with glucose to form phenylacetylglutamine (PAG) (32, 33). In addition to its role in decreasing hyperammonemia, the urea cycle is capable of excreting

phenylacetylglutamine through urine (29), emphasizing the importance of the urea cycle in its role to excrete neurotoxic molecules.

Conversely, our investigation revealed an overrepresentation of metabolic pathways involved in the biosynthesis of pyrimidine deoxyribonucleotides, which play a role in neurological development. Specifically, the pathways for pyrimidine deoxyribonucleotides de novo biosynthesis IV and pyrimidine deoxyribonucleotides biosynthesis from CTP were overrepresented in the Peru data compared to the California data (Figure 4). Pyrimidine deoxyribonucleotides is key for nucleic acid synthesis, carbohydrate metabolism, and cell proliferation (34). Weichsel and Clark (1977) demonstrated the involvement of pyrimidine biosynthesis enzymes with cerebellum cell division, in which enhanced enzyme activity leading to the biosynthesis of pyrimidine deoxyribonucleotides was positively correlated with improved brain growth in neonatal undernourished rats (35). Proper development of the cerebellum is essential for an infant's motor coordination, cognition, behaviour, language, memory, and learning (36). Thus, our findings suggest that infants from Peru potentially have higher metabolic activity that supports early neurological development.

An overrepresentation of bacterial metabolic pathways crucial for synthesizing adequate vitamin concentration for growth and development was observed in infants from California, compared to infants from Peru. Results revealed several overrepresented metabolic pathways related to the synthesis of vitamin K. These include the 1,4-dihydroxy-2-naphthoate biosynthesis, and the superpathway of phylloquinone biosynthesis (Figure 4). These pathways are essential for the production of menaquinone (vitamin K2) and phylloquinone (vitamin K1), respectively (37, 38). Vitamin K is essential for blood coagulation and bone metabolism (39). Insufficient levels of vitamin K have been linked to an increased susceptibility of developing life-threatening bleeding disorders and bone disorders such as osteoporosis (40). Our functional analysis revealed that both vitamin K-related pathways were overrepresented in infants from California, suggesting a lower risk of them developing vitamin K-related disorders such as Vitamin K deficiency bleeding (VKDB), characterized by impaired blood clotting (40). These findings align with a previous study that identified a significantly lower incidence of VKDB in North America compared to other regions (41). Although the study does not explicitly disclose the specific regions involved in the comparison, this underscores the need for further studies to explore regional variations and underlying influences on VKDB prevalence in infants (41).

The overrepresentation of metabolic pathways in Californian infants involved in synthesizing vitamin B was also observed (Figure 4); specifically, the biotin (vitamin B7) biosynthesis pathway. Vitamin B7 is an essential cofactor for enzymes involved in the metabolism of fatty acids, glucose, and amino acids (42). Vitamin B7 deficiency in infants is known to cause a delay in brain development (42). Our functional analysis revealed a 3-fold greater representation of biotin synthesis pathways in California compared to Peru, potentially increasing the risk of infants from Peru to B7 deficiency. This finding aligns with literature where individuals of Hispanic origin, including those in the USA, are more prone to biotin deficiency (42). Biotin-deficiency can also increase the risk of infants developing encephalopathy (43) and experiencing neurodevelopmental difficulties, ataxia, mental disabilities, and seizures due to neurotoxicity (44). Further studies examining biotin levels across different regions may be beneficial for supporting optimal neurological development in mid-income nations.

Contrary to the observed overrepresentation of vitamin K2 synthesis pathways in California, our analysis of bacterial genera revealed higher levels of bacteria which may be responsible for vitamin K2 synthesis in Peru. Menaquinone isoform synthesizers from the genera *Eubacteria* and *Veillonella* were more abundant in Peru than in California respectively (45). This suggests that *Veillonella* may have a more significant role in vitamin K2 synthesis compared to *Eubacteria*. Similarly, *Lactobacillus*, which is associated with the increased production of neurotransmitters (46), was more abundant in Peru. This highlights the complexity in associating microorganisms to implications of differentially represented metabolic pathways.

Limitations Several limitations affect our study, primarily due to the use of two separate and unrelated datasets. Methodological variations between the datasets may introduce non-

geographical biases. For example, factors such as the exact timing and storage conditions of the fecal sample collection from McClorry et al. (11) were not standardized, potentially contributing to the variations found in our analyses unrelated to geographical location. Although we aimed to filter for infants with similar feeding methods, variations in feeding time and frequency could still introduce inconsistencies between our datasets. Additionally, the Peru dataset focused on six-month-old infants, whereas the California dataset included infants of various ages, resulting in a smaller sample size (n=15 for California verses n=88 for Peru). This imbalance in sample size potentially reduces statistical power, which may have obscured our taxonomic composition results. With such limitations in inherent experimental variation and low sample size, we acknowledge that perceived microbial differences between the two cohorts may not be strictly due to geographical variables. In addition, since only two groups were compared, the specific conditions of the study do not allow us to make general conclusions about the geographical effect on the infant gut microbiome. Future studies that standardize methods across multiple cohorts would prevent these limitations and enable stronger conclusions.

Conclusions Comparisons of microbial characteristics between six-month-old infants from Iquitos, Peru and San Diego, California, USA revealed notable differences in intestinal microbiome composition and metabolic pathways. We observed overrepresentation of metabolic pathways related to neurological health and essential vitamin synthesis in both cohorts. Specifically, the Peru data showed greater representation of pyrimidine deoxyribonucleotide synthesis, suggesting a potential benefit in cerebellum development. Conversely, Californian infants exhibited overrepresentation in pathways associated with PEA degradation, vitamin B synthesis, vitamin K synthesis, and the urea cycle which may suggest a decreased risk for conditions such as PKU, vitamin deficiencies, and hyperammonemia. Interestingly, microorganisms linked to these pathways were more prevalent in the Peru data. These findings underscore the complexity of associating microbial profiles to health outcomes and highlight significant variability between the two cohorts. However, numerous confounding variables can impact results, making it challenging to make direct geographical correlations. These observations emphasize the need for further research to standardize data collection sources and explore the implications of geographical and microbial variations on infant health.

Future Directions To address the limitations of the current study, future research should focus on standardizing data collection methods and equalizing sample size across cohorts. Additionally, expanding sample diversity and conducting more comprehensive investigations into various neurological pathways could provide more robust insights on the gut-brain axis.

While the study revealed differences associated with geographical location, future studies should aim to further explore regional variations by including sample sources from a broader range of locations. For instance, samples could be collected from different states in the USA, expanding to more South American countries, and other continents. This approach would allow more specific attribution of factors such as climate, economic status, and genetics on bacterial abundance and its effects on developmental pathways.

Moreover, further research could utilize neurological tests, such as the plantar reflex, to investigate how specific genera impacts an infant's neurological development. A previous study leveraged the point and gaze test and identified associations between higher abundances of Actinobacteria and lower abundances of Firmicutes with improved neurological development (47). Exploring these relationships would help to further comprehend the complex interactions between microbial composition and developmental outcomes, and aid to understand the opposing results found between the functional and compositional analysis.

ACKNOWLEDGEMENTS

We are grateful to the Department of Microbiology & Immunology, Faculty of Science, The University of British Columbia, Vancouver, B.C., Canada for the funding, resources, and facilities that made this research possible. Specifically, we would like to express gratitude to Dr. Evelyn Sun and the MICB475 teaching team for providing us with the instruction and

material that supported our project. Of note, we thank Avril Metcalfe-Roach for the invaluable guidance and feedback throughout the entire research process. Additionally, we would like to extend our appreciation to McClorry et al. and Rhee et al. for providing the research dataset and metadata information that was used in this study. Finally, we are thankful to the anonymous reviewer for providing constructive feedback to ensure the integrity of our research.

CONTRIBUTIONS

Risa was responsible for writing the introduction and the results regarding differential expression analysis using DESeq.
Kevin contributed to the writing of the abstract and final sections of the discussion.
Stella contributed to writing the results for PICRUSt2 and the metabolic pathway components of the discussion.
Michael contributed to writing portions of the discussion including compositional analysis and future directions.
Dennis contributed to the writing of the methods and the result and discussion on diversity and taxonomy analysis.
All co-authors contributed to the editing of the manuscript and were involved in the progression of the research aims throughout the study.

REFERENCES

1. **Grech A, Collins CE, Holmes A, Lal R, Duncanson K, Taylor R, Gordon A.** 2021. Maternal exposures and the infant gut microbiome: A systematic review with meta-analysis. *Gut Microbes* **13**.
2. **Tamburini S, Shen N, Wu HC, Clemente JC.** 2016. The microbiome in early life: Implications for health outcomes. *Nature Medicine* **22**:713–722.
3. **Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M , Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI.** 2012. Human gut microbiome viewed across age and geography. *Nature* **486**:222–227.
4. **Rodríguez JM, Murphy K, Stanton C, Ross RP, Kober OI, Juge N, Avershina E, Rudi K, Narbad A, Jenmalm MC, Marchesi JR, Collado MC.** 2015. The composition of the gut microbiota throughout life, with an emphasis on early life. *Microbial Ecology in Health & Disease* **26**.
5. **Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R.** 2012. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**:220–230.
6. **Chen Y, Xu J, Chen Y.** 2021. Regulation of neurotransmitters by the gut microbiota and effects on cognition in neurological disorders. *Nutrients* **13**:2099.
7. **Cheng Y, Selma-Royo M, Cao X, Calatayud M, Qi Q, Zhou J, Zeng L, Garcia-Mantrana I, Collado MC, Han B.** 2022. Influence of geographical location on maternal-infant microbiota: Study in two populations from Asia and Europe. *Frontiers in Cellular and Infection Microbiology* **11**.
8. **Arrieta M-C, Stiemsma LT, Amenyogbe N, Brown EM, Finlay B.** 2014. The intestinal microbiome in early life: Health and disease. *Frontiers in Immunology* **5**.
9. 2014. Country classification, p. 143–150. *In World economic situation and prospects*. essay, United Nations, New York.
10. **Kemppainen KM, Ardissone AN, Davis-Richardson AG, Fagen JR, Gano KA, León-Novelo LG, Vehik K, Casella G, Simell O, Ziegler AG, Rewers MJ, Lernmark Å, Hagopian W, She J-X, Krischer JP, Akolkar B, Schatz DA, Atkinson MA, Triplett EW.** 2014. Early childhood gut microbiomes show strong geographic differences among subjects at high risk for type 1 diabetes. *Diabetes Care* **38**:329–332.
11. **McClorry S, Zavaleta N, Llanos A, Casapía M, Lönnerdal B, Slupsky CM.** 2018. Anemia in infancy is associated with alterations in systemic metabolism and microbial structure and function in a sex-specific manner: An observational study. *The American Journal of Clinical Nutrition* **108**:1238–1248.
12. **Rhee KE.** 2020. Eating behavior development in infants. Project: PRJEB39437. Retrieved from: https://www.ebi.ac.uk/ena/browser/view/PRJEB39437
13. **Cong X**, **Judge M**, **Xu W**, **Diallo A**, **Janton S**, **Brownell EA**, **Maas K**, **Graf J**. 2017. Influence of feeding type on gut microbiome development in hospitalized preterm infants. *Nursing Research* **66**:123–133.
14. **Callahan BJ**, **McMurdie PJ**, **Rosen MJ**, **Han AW**, **Johnson AJ**, **Holmes SP**. 2016. Dada2: High-resolution sample inference from Illumina Amplicon Data. *Nature Methods* **13**:581–583.
15. **Quast C**, **Pruesse E**, **Yilmaz P**, **Gerken J**, **Schweer T**, **Yarza P**, **Peplies J**, **Glöckner FO**. 2012. The Silva ribosomal RNA gene database project: Improved Data Processing and web-based tools. *Nucleic Acids Research* **41**.
16. **Yilmaz P**, **Parfrey LW**, **Yarza P**, **Gerken J**, **Pruesse E**, **Quast C**, **Schweer T**, **Peplies J**, **Ludwig W**, **Glöckner FO**. 2013. The Silva and "all-species living tree project (LTP)" taxonomic frameworks. *Nucleic Acids Research* **42**.

17. **R Core Team**. 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
18. **RStudio Team.** 2021. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA.
19. **McMurdie PJ**, **Holmes S**. 2013. Phyloseq: An R package for reproducible interactive analysis and graphics of Microbiome Census Data. *PLoS ONE* **8**.
20. **Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H**. 2019. "Welcome to the tidyverse.". *Journal of Open Source Software* **4**: 1686.
21. **Paradis E**, **Schliep K**. 2018. APE 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**:526–528.
22. **Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J**. 2023. vegan: Community Ecology Package. R package version 2.6-5. https://github.com/vegandevs/vegan.
23. **Love MI**, **Huber W**, **Anders S**. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with deseq2. *Genome Biology* **15**.
24. **Douglas GM**, **Maffei VJ**, **Zaneveld JR**, **Yurgel SN**, **Brown JR**, **Taylor CM**, **Huttenhower C**, **Langille MG**. 2020. PICRUST2 for prediction of metagenome functions. *Nature Biotechnology* **38**:685–688.
25. **Yang C**, **Mai J**, **Cao X**, **Burberry A**, **Cominelli F**, **Zhang L**. 2023. GGPICRUST2: An R package for picrust2 predicted functional profile analysis and visualization. *Bioinformatics* **39**.
26. **Ghozlan A, Varoquaux O, Abadie V.** 2004. Is monoamine oxydase-B a modifying gene and phenylethylamine a harmful compound in phenylketonuria? Molecular Genetics and Metabolism 83(4):337-340.
27. **Banta-Wright SA, Shelton KC, Lowe ND, Knafl KA, Houck GM.** 2012. Breast-feeding Success Among Infants with Phenylketonuria. *Journal of Pediatric Nursing* **27:**319-327.
28. **Hillert A, Anikster Y, Belanger-Quintana A, Burlina A, Burton BK, Carducci C, Chiesa AE, Christodoulou J, Đorđević M, Desviat LR, Eliyahu A, Evers RAF, Fajkusova L, Feillet F, Bonfirm-Freitas PE, Gižewska M, Gundorova P, Karall D, Kneller K, Kutsev SI, Leuzzi V, Levy HL, Lichter-Konecki U, Muntau AC, Namour F, Oltarzewski M, Paras A, Perez B, Polak E, Polyakov AV, Porta F, Rohrbach M, Scholl-Bürgi S, Spécola N, Stoppioni V, Thöny B, Trefz FK, Vockley J, Yu Y, Zschocke J, Hoffmann GF, Garbade SF, Blau N.** 2020. The Genetic Landscape and Epidemiology of Phenylketonuria. *American Journal of Human Genetics* **107**(2):234-250.
29. **Jiao M, He W, Ouyang Z, Shi Q, Wen Y.** 2022. Progress in structural and functional study of the bacterial phenylacetic acid catabolic pathway, its role in pathogenicity and antibiotic resistance. *Frontiers in Microbiology* **13**.
30. **Savy N, Brossier D, Brunel-Guitton C, Ducharme-Creview L, Pont-Thibodeau, GD, Jouvet P.** 2022. Acute pediatric hyperammonemia: current diagnosis and management strategies. *Dove Press* **10**:105-115.
31. **Ali R, Nagalli S.** 2023. Hyperammonemia. StatPearls [Internet].
32. **Mokhtarani M, Diaz GA, Rhead W, Berry SA, Lichter-Konecki U, Feigenbaum A, Schulze A, Longo N, Bartley J, Berquist W, Gallagher R, Smith W, McCandless SE, Harding C, Rocky DC, Vierling JM, Mantry P, Ghabril M, Brown RS, Dickinson K, Scharschmidt BF**. 2013. Elevated phenylacetic acid levels do not correlate with adverse events in patients with urea cycle disorders or hepatic encephalopathy and can be predicted based on the plasma PAA to PAGN ratio. *Elsevier* **110**:4.
33. **Wang X, Tseng J, Mak C, Poola N, Vilchez RA.** 2021. Exposures of Phenylacetic Acid and Phenylacetylglutamine Across Different Subpopulations and Correlation with Adverse Events. *Clinical Pharmacokinetics* **60**:1557-1567.
34. **Pesini A, Iglesias E, Bayona-Bafaluy MP, Garrido-Pérez N, Meade P, Gaudó P, Jiménez-Salvador I, Andrés-Benito P, Montoya J, Ferrer I, Pesini P, Ruiz-Pesini E.** 2019. Brain pyrimidine nucleotide synthesis and Alzheimer disease. *Aging* **11**(19):8433-8462.
35. **Weichsel ME, Clark BR.** 1977. Pyrimidine Metabolism during Restorative Brain Growth after Neonatal Undernutrition in the Rat. *Pediat. Res.* **11**:293-297.
36. **Spoto G, Amore G, Vetri L, Quatrosi G, Cafeo A, Gitto E, Nicotera AG, Rosa GD.** 2021. Cerebellum and Prematurity: A Complex Interplay Between Disruptive and Dysmaturational Events. *Frontiers in Systems Neuroscience* **15**:655164.
37. **Scaramuzza AE.** 2022. The Impact of Vitamin K2 (Menaquinones) in Children's Health and Diseases: A Review of the Literature. *Children (Basel)* **9**(1):78.
38. **Clarke P, Mitchell SJ, Shearer MJ.** 2015. Total and Differential Phylloquinone (Vitamin K1) Intakes of Preterm Infants from All Sources during the Neonatal Period. *Nutrients* **7**(10):8308-20.
39. **Lippi G, Franchini M.** 2011. Vitamin K in neonates: facts and myths. Blood *Transfusion* **9**(1):4-9.
40. **Eden RE, Daley SF, Coviello JM.** 2023. Vitamin K Deficiency. *StatPearls [Internet]*.
41. **Neal M, Crowther M, Douketis J, Verhovsek M, Stidley C, Garcia D.** 2008. Geographic variation in the reversal of vitamin K antagonist-associated coagulopathy. *Critical Care* **12**.

42. **Saleem F, Soos MP. Biotin Deficiency.** [Updated 2023 Feb 20]. In: StatPearls [Internet]. Treasure Island (FL): *StatPearls Publishing*; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK547751/
43. **Saleem H, Simpson B**. 2023. Biotinidase Deficiency. *StatPearls* [Internet].
44. **Zempleni J, Hassan YI, Wijeratne SSK.** 2008. Biotin and biotinidase deficiency. *Expert Rev Endocrinol Metab* **3**(6):715-724.
45. **Yan H, Chen Y, Zhu H, Huang WH, Cai XH, Li D, Lv YJ, Zhao S, Zhou HH, Luo FY, Zhang W, Li X.** 2022, The Relationship Among Intestinal Bacteria, Vitamin K and Response of Vitamin K Antagonist: A Review of Evidence and Potential Mechanism. *Front. Med.* **9**.
46. **Dicks LMT, Hurn D, Hermanus D.** 2021. Gut Bacteria and Neuropsychiatric Disorders. *Microorganisms*, **9**(12):2583.
47. **Hunter S, Flatten E, Petersen C, Gervain J, Werker JF, Trainor LJ, Finlay BB.** 2023. Babies, bugs and brains: How the early microbiome associates with infant brain and behavior development. *PLoS ONE* **18**(8).