Transactions on Materials, Biotechnology and Life Sciences
ISSN: 2960-1932, eISSN: 2960-2319 | Volume 8
MMTCS 2025

Power of Poop: Identification and Analysis of Gut Microbiome Biomarkers in Diarrhea

Chenxi Hu

Tabor Academy, 66 Spring Street, Marion, MA, USA, 02738

Chu26@taboracademy.org

Abstract. The human gut microbiome, a vast collection of billions of bacteria in the gastrointestinal tract, plays a crucial role in maintaining host health through homeostasis. Diarrhea, which affects approximately 179 million people annually, significantly alters the gut microbiome. While infectious diarrhea is characterized by an increase in pathogens such as *Escherichia coli*, the changes in gut microbiome following non-infectious and chronic diarrhea remain unclear. Recent advances in metagenomic sequencing have enabled us to investigate the changes in gut microbiome. To identify biomarkers in diarrhea and characterize the gut microbiome, we integrated datasets from 12 published studies into a comprehensive dataset, comprising 1015 healthy samples and 2302 diarrhea-related disease samples, among which approximately 90% are non-infectious. Using linear regression model and machine learning, we identified microbial taxa enriched in healthy samples, including *Bacteroides uniformis*, *Bacteroides vulgatus*, *Akkermansia muciniphila*, *Firmicutes bacterium*, and *Alistipes*. In contrast, *Flavonifractor plautii*, *E. coli*, and *Prevotella* were more abundant and prevalent in diarrhea samples. For microbial function analysis, we found that pathways such as phytol degradation and aerobic respiration I (cytochrome C) were enriched in healthy samples, whereas other pathways including glycerol degradation, methanogenesis, isoprene biosynthesis, and taxadiene biosynthesis, were more commonly detected in diarrhea samples. We employed graph theory and inferred a denser but less stable network from diarrhea samples compared to the network of healthy controls. These results indicate that healthy and diarrhea gut microbiomes have distinct features characterizing diarrhea alterations. We anticipate that these biomarkers can be used for personalized dietary interventions and targeted, cost-effective diagnosis and treatments of diarrhea that consider complex gut microbiome interactions.

Keywords: Gut Microbiome, Diarrhea, Taxonomic Biomarkers, Functional Pathway, Machine Learning.

1. Introduction

Diarrhea is a common digestive condition causing over 1.3 million deaths and affecting 179 million people annually [1]. It can arise from infections or chronic diseases, such as inflammatory bowel disease (IBD) [2]. Diarrhea is classified by its duration—acute (a few days), persistent (2-4 weeks), or chronic (over 4 weeks)—and by cause, such as bacterial pathogens (e.g., *Clostridium difficile* and Shiga toxin-producing *Escherichia coli*) or viruses (e.g., Rotavirus). Stool type also differentiates types of diarrhea: secretory diarrhea produces watery stool, while inflammatory diarrhea often results in bloody stool. IBD, a noninfectious form of chronic inflammatory bowel disease, affects over 1.6 million people in the U.S. [3] and is characterized by persistent gut inflammation and a gut microbiome imbalance.

Traditional studies on diarrhea have primarily focused on pathogens, often overlooking the role of the gut microbiota and its complex interactions with possible drug targets and disease pathogens. Traditional diarrhea medicine often overlooks how pathogens interact with the gut microbiome and whether the microbiota influences diarrhea prevention [4]. Recent advances in sequencing technologies and bioinformatics tools have enabled deeper insights through methods like 16S rRNA gene and metagenomics sequencing, allowing detailed taxonomic and functional profiling of the gut microbiomes [5]. Studies have observed shifts in microbiota composition during acute, noninflammatory diarrhea caused by *Vibrio cholerae* and enterotoxigenic *E. coli*, with the microbiota

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

transitioning from *Bacteroides*-dominant to pathogen-dominant states [6]. While it is established that diarrhea alters the microbiota, the effects of non-acute and non-infectious forms—particularly IBD—remain poorly understood. A comprehensive analysis across diverse types of diarrhea, including both acute and chronic cases, is therefore crucial.

This study investigates the diversity and dynamic changes of the gut microbiome in diarrhea patients. It also aims to identify novel biomarkers that could be used to monitor diarrhea progression and inform the development of more effective, personalized, and targeted treatments.

2. Materials and Methods

2.1. Gut microbiome data: access and processing

Relative abundance and pathway abundance data are retrieved from the "CuratedMetagenomics" package (v3.8.0) [7]. Data from samples with diarrhea-related diseases are extracted and filtered, with a sample size of $n > 20$ for each condition. In addition, healthy samples from the corresponding 13 studies in the disease list are included for control. The body site of sample collection is limited to stool. In total, we obtained 1015 healthy control samples and 2302 diarrhea-related samples, among which 90% are non-infectious diarrhea-related. Additional filtering ensures samples across all age groups (child, school age, adult, and senior) have sample sizes $n>300$ and country groups with $n > 40$. The TKI-dependent diarrhea group is filtered out for DESeq2 analysis, which requires normalized count data.

2.2. Gut microbial composition and diversity analysis

For microbial composition profiling, we used bar plots to compare the taxonomic composition of different disease types. Each bar represents a different sample, with different colors representing different phyla. Further ecological analysis of alpha and beta diversity is done using the 'vegan' package (v 2.6-8) [8]. We calculate the Shannon alpha diversity index for both samples. Pairwise t-test, pairwise Wilcoxon-test, and ANOVA are used for the significance test when comparing the samples' alpha diversity index difference. Additionally, beta diversity is plotted using PCoA. Bray-Curtis dissimilarity index is calculated. Prevalence of each microbe is calculated by dividing the number of samples in which it appeared by the total number of samples. They were plotted to show the relationship between prevalence and average relative abundance in both healthy samples and non-healthy (diarrhea) samples separately.

2.3. Core gut microbiome/function identification and network inference

Before further analysis, the assay matrix is filtered to remove rare taxa based on the following criteria: prevalence of 10% and a relative abundance of 0.1%. Following the definition of the core microbiome by Lahti and Shetty et al. [9], the core microbiome for both healthy and unhealthy categories are identified. In the core microbiome of each disease category, the filtering criteria are set such that the prevalence of bacterial species exceeds 80%.

The correlation between microbes in the assay matrix is calculated using Spearman's rank correlation coefficient. After filtering, only species with more than 20% prevalence and a relative abundance greater than 0.001 are retained. The generated adjacency matrix is then filtered based on correlation value ($|r| > 0.3$) and significance value ($P < 0.01$). A heat map is created to illustrate the correlation between taxa. As with the retrieval of relative abundance data, the pathway abundance data comprises $n = 50{,}126$ pathways. Next, pathways with a prevalence greater than 5% and 0.001% pathway abundance are retained, followed by plotting a heat map. Network results are visualized in Cytoscape (v3.10.2), with non-healthy samples formatted according to Markov clustering. Only the top ten nodes with the highest relative abundance are included in the network for both healthy and diarrhea graphs.

2.4. Differential abundance of species and function analysis between conditions

DESeq2 is used to identify differentially abundant species among diarrhea-related samples with count data of species as well as functional pathways [10]. Samples are sorted based on disease categories: either healthy or non-healthy. A volcano plot will show the log2FoldChange and adjusted p-value. Species with log2FoldChange < -1 will be considered enriched in healthy samples; conversely, species with log2FoldChange > 1 will be deemed enriched in non-healthy samples. Species with an adjusted p-value < 0.01 will be regarded as significant. Aggregated relative abundance (aRA) is calculated for both healthy and non-healthy enriched species, and the top 5 differentially abundant species in each case are identified. Additionally, multiple factor DESeq2 is performed with disease categories (including IBD, CDI, acute diarrhea, and healthy), country, and age category. Again, only microbes with a p-value < 0.01 are considered significant and are highlighted by color in the volcano plots.

2.5. Random Forest

Random forest is a supervised machine learning method for decision tree classification. The method we used is modified from chapter 7 of Exploring microbiome analysis using R by N. Fauziyyahis [11]. It is employed here to confirm the DESeq2 results. In this analysis, we utilized machine learning to classify healthy and non-healthy samples based on the host gut microbiome's relative abundance, DESeq2 relative abundance results, and DESeq2 functional pathway results. Each dataset is randomly divided into training (80%) and testing (20%) groups. Then, the top six species/pathways that have the most significant influence on the healthy/non-healthy groupings are identified and visualized.

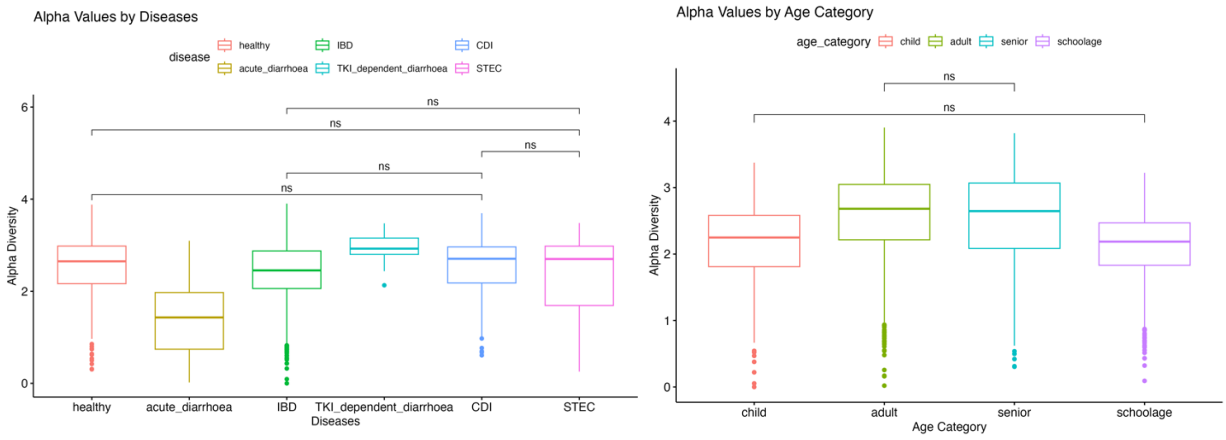
2.6. Data Availability

All the analyses will be performed in open-source R, and all the codes used in this project will be uploaded to a publicly available GitHub repository (https://github.com/tutuhu125/Diarrhea-Study).

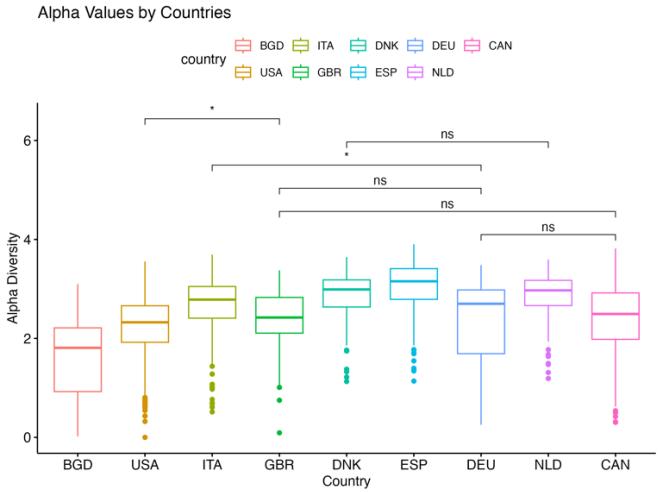
3. Results

3.1. Microbial diversity analysis

All t-tests generated results in the same significance range as Wilcoxon tests.



(a) (b)



(c)

Figure 1. In the interaction pairings, ns indicates p-value > 0.05, and * indicates p-value ≤ 0.05.

All non-indicated pairings have p-value ≤ 0.01. (a) Alpha diversity as a function of diseases. In the alpha diversity analysis, comparing samples from different conditions, the pairwise t-test revealed that acute diarrhea has significantly lower alpha diversity compared to CDI, healthy, IBD, STEC, and TKI-dependent diarrhea samples (all p-values < 0.01). Additionally, TKI-dependent diarrhea's alpha diversity is significantly higher than that of CDI, IBD, STEC, and healthy samples (p-value < 0.01). Healthy samples exhibit significantly higher alpha diversity than IBD samples (p-value < 0.01). The pairwise Wilcoxon test confirmed these results. (b) Alpha diversity as a function of age category. For age categories, adults have significantly higher alpha diversity than children and the school-age group (p-value < 0.01), while seniors have significantly higher alpha diversity than children and the school-age group (p-value < 0.01). Again, the Wilcoxon test corroborated these findings (all p-values < 0.01). (c) Alpha diversity as a function of country. In terms of countries, Bangladesh has significantly lower alpha diversity compared to Canada, Germany, Denmark, Spain, Great Britain, Italy, the Netherlands, and the USA (all p-values < 0.01). Additionally, USA samples have higher alpha diversity than those from Bangladesh but lower alpha diversity than those from Denmark, Spain, Italy, and the Netherlands (all p-values < 0.01). From ANOVA, country has the biggest influence on alpha diversity with F-value of 134.165, followed by disease with F-value 54.304. Any pair of countries, diseases, or age categories had a limited influence on alpha diversity as a paired influence.

In our beta-diversity PCoA analysis, no clear clusters are evident in the plots distinguishing between non-healthy and healthy samples (Figure 2). Similar results are evident in the beta-diversity PERMANOVA test. Consequently, the lack of specificity in the beta-diversity results prompted us to conduct a DESeq2 analysis and a machine learning (random forest) approach to identify significant biomarkers in both healthy and non-healthy samples.

When comparing the beta diversity between samples using the PCoA plot, no clear clusters were observed. The PERMANOVA test indicated that 84.42% of the variance cannot be explained by the influence of disease, age categories, and country (Figure 2).

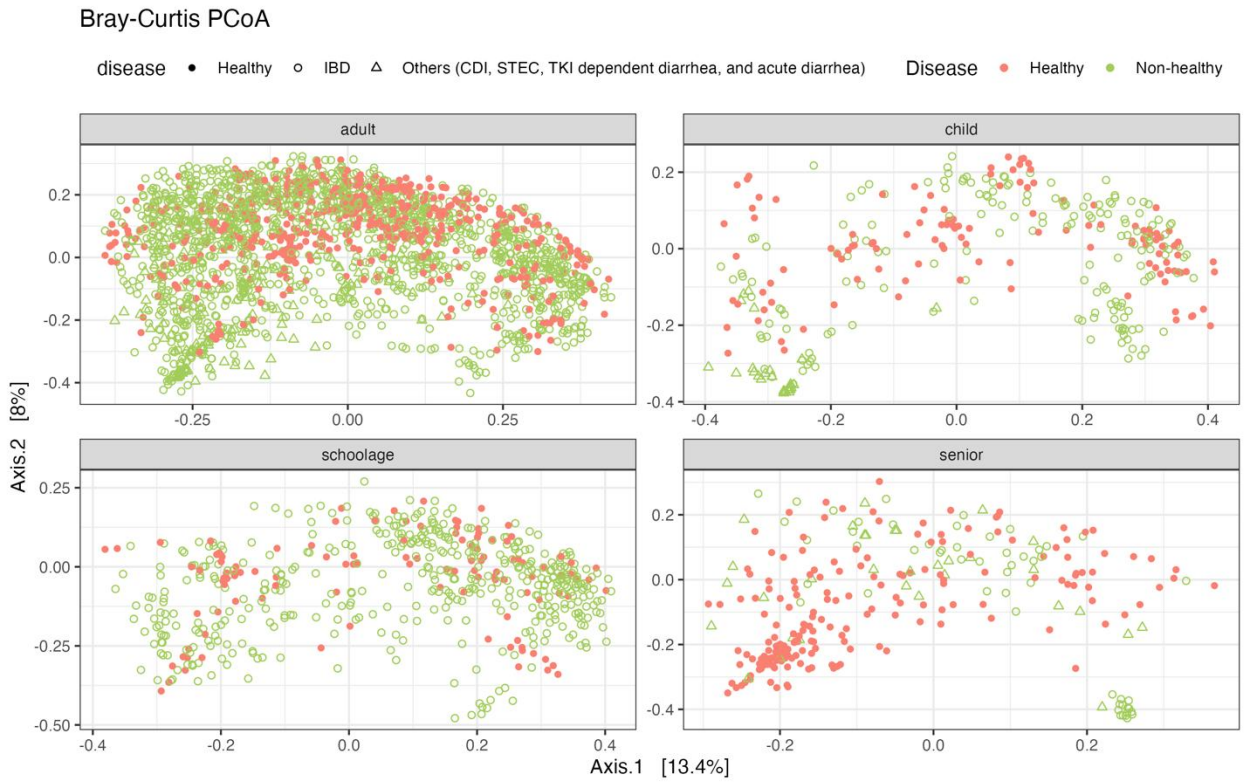


Figure 2. PCoA plots based on age categories. Each dot represents one sample, noted by its color as either healthy (red) or non-healthy (green).

3.2. Microbial community composition

For the community profiling at the phylum level, healthy and IBD samples appeared to align closely, with Bacteroidetes, Firmicutes, and Proteobacteria as the main phyla. IBD shows a slightly higher abundance of Proteobacteria. In acute diarrhea samples, there is a significant portion of Proteobacteria compared to other disease conditions (Figure 3). In prevalence plots, nonhealthy samples indicate that E. coli has a prevalence of > 0.75 and a relative abundance of > 0.15. For healthy samples, both B. uniformis and B. vulgatus exhibit a prevalence of greater than 0.75 and a relative abundance of greater than 0.05 (Figure 4).

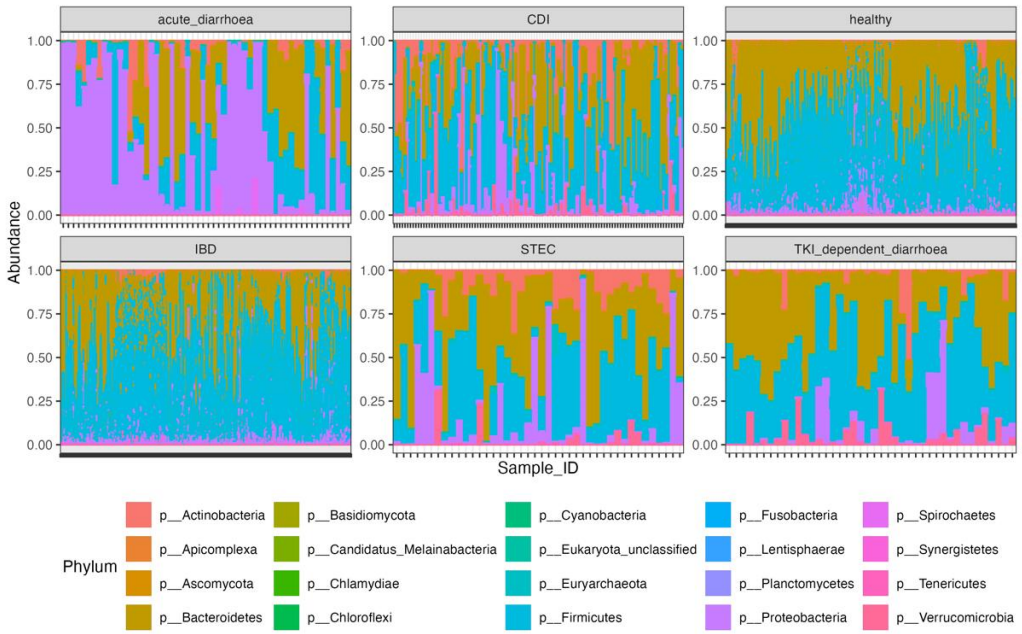


Figure 3. Community profiling on Phylum Level.

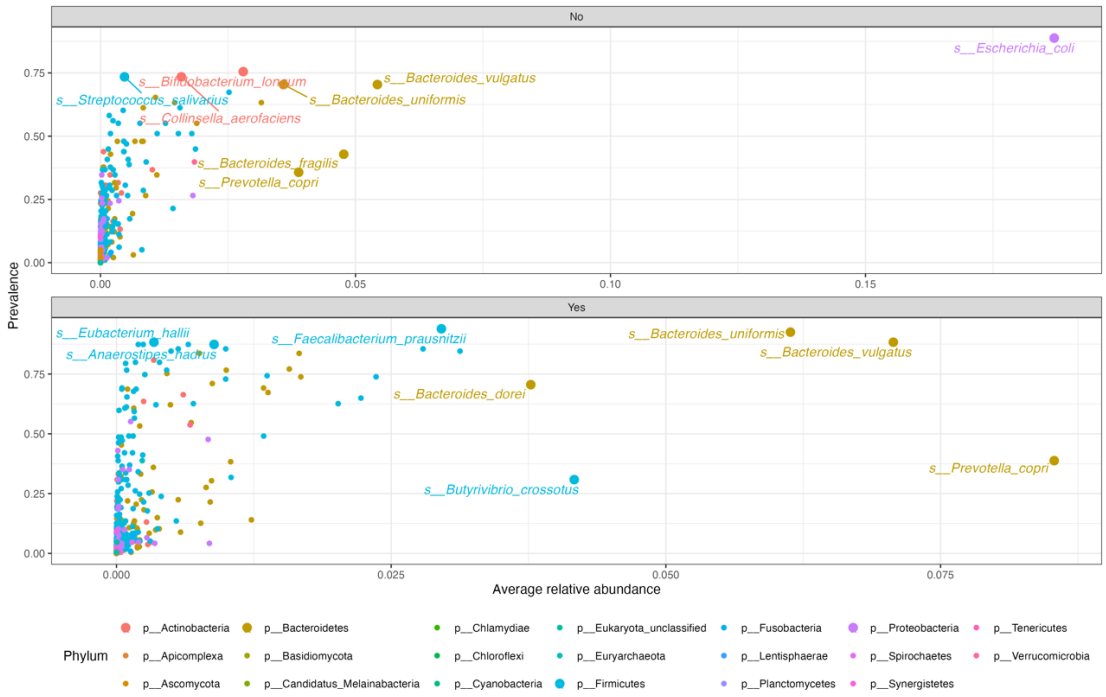


Figure 4. Prevalence Plots for Healthy and Nonhealthy Samples. “No” represents non-healthy gut microbiome prevalence, while “yes” represents healthy gut microbiome prevalence.

3.3. Differential abundant species of healthy and non-healthy gut microbiomes

From the non-healthy and healthy categories, 85 species are identified as more abundant in the healthy samples, while 121 species are found to be more abundant in the non-healthy samples (Figure 5). The top five differentially abundant healthy species, which exhibit the highest baseMeans, include Firmicutes bacterium, Brachyspira, Alistipes, Ruminococcus, and Anaerococcus tetradi. Conversely, the top five non-healthy abundant species consist of Bacteroides fluxus, Fusobacterium ulcerans, Prevotella, Prevotella buccae, and Eubacterium.

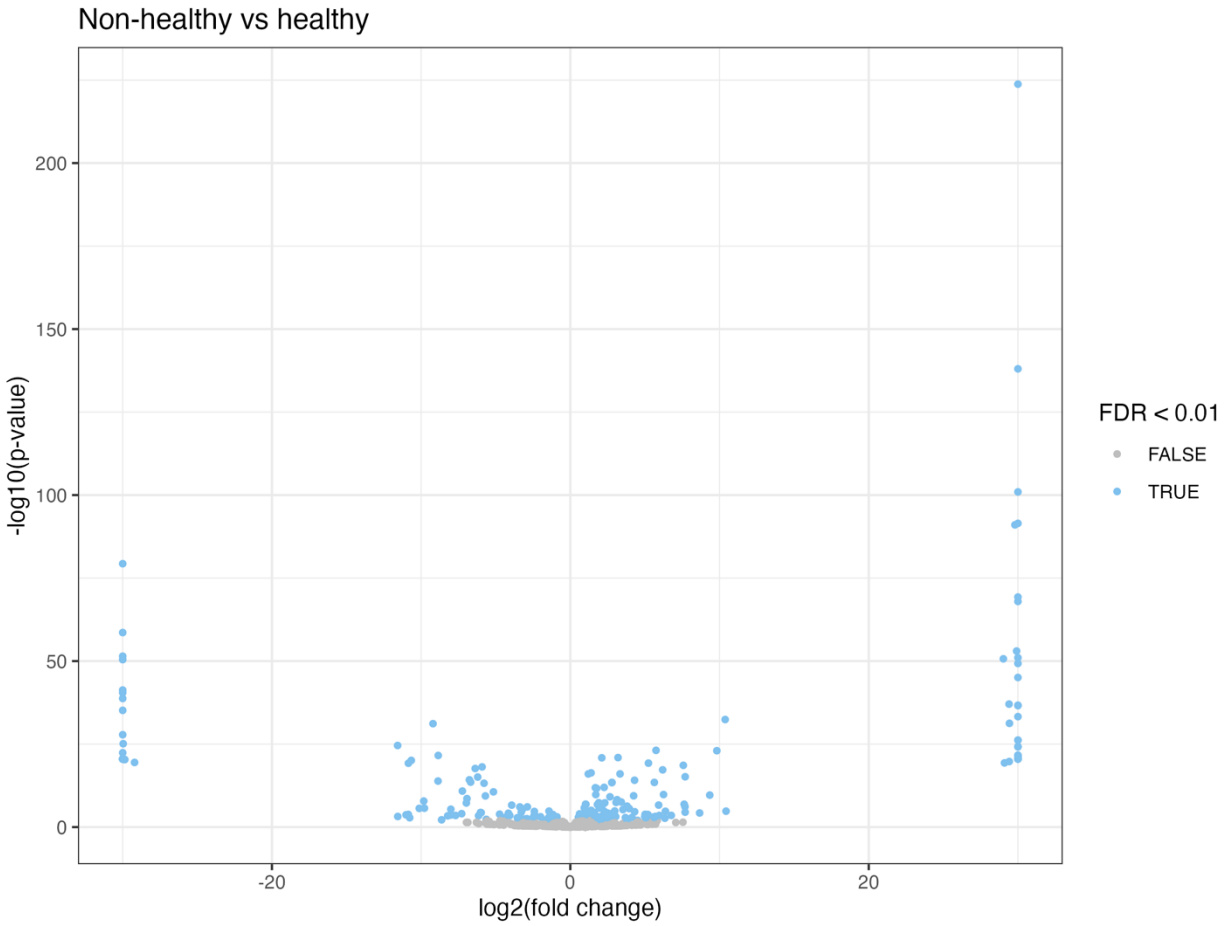


Figure 5. Non-healthy vs. healthy volcano plot of differentially abundant species.

For functional pathway DESeq2 analysis, 4 pathways are differentially more abundant in healthy samples and 4 in non-healthy samples (Figure 6). The top 4 differentially abundant healthy pathways include aerobic respiration I (cytochrome C), aerobic respiration II (cytochrome C), phytol degradation, and isoleucine biosynthesis II. The top 4 differentially abundant non-healthy pathways include isoprene biosynthesis I, super pathway of glycerol degradation to 1,3-propanediol, taxadiene biosynthesis, and methanogenesis from acetate.

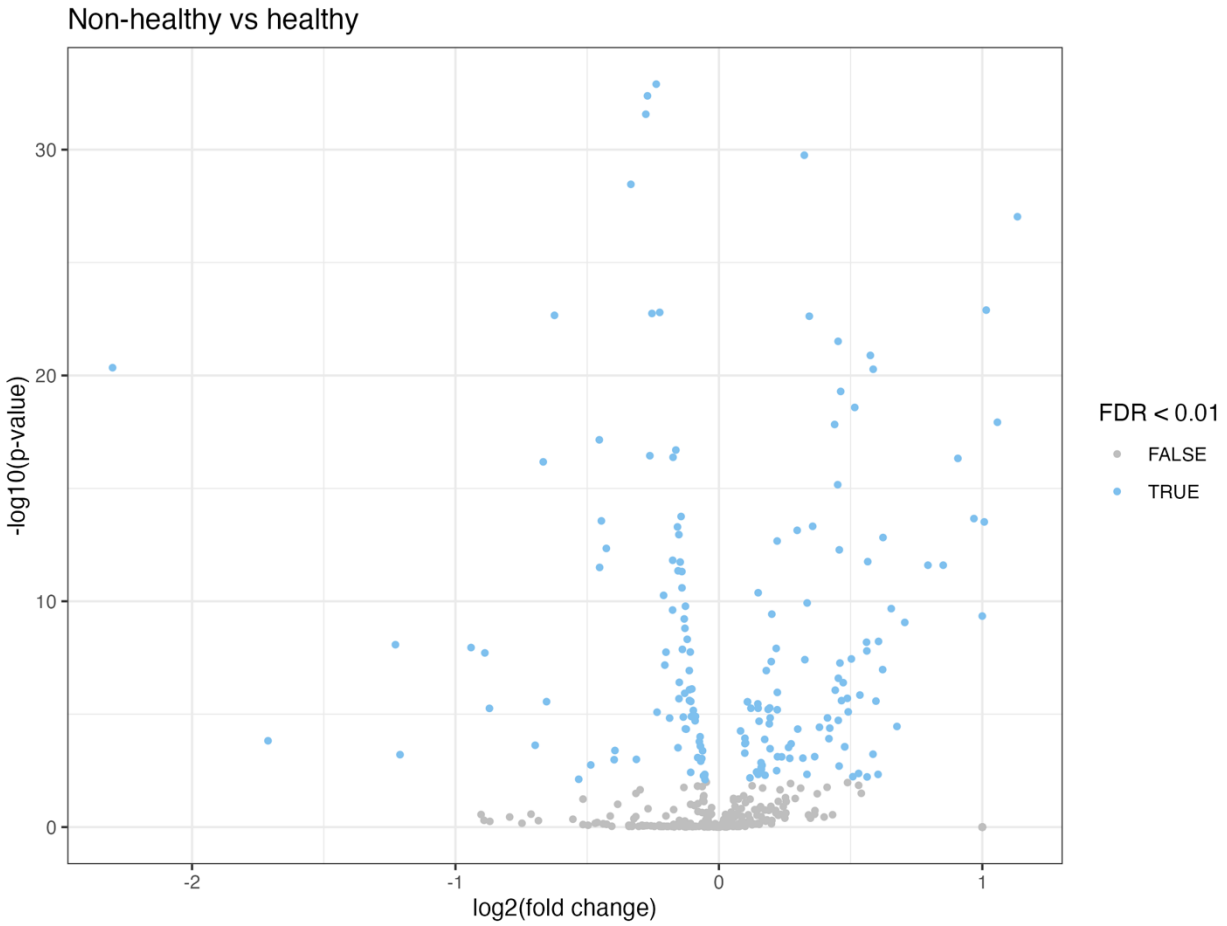


Figure 6. Non-healthy vs. healthy volcano plot of differentially abundant pathways.

3.4. Core microbiome and function comparison between conditions

The healthy samples' core microbiome consists of 3 species: *Bacteroides uniformis, Bacteroides vulgatus,* and *Faecalibacterium prausnitzii*. The non-healthy samples core microbiome includes 2 species: *F. prausnitzii* and *Flavonifractor plautii*. *F. prausnitzii* exists in both healthy and non-healthy core microbiomes.

3.5. Network analysis of diarrhea-enriched species and pathways

For the network graphs, the size of the nodes represents the relative abundance of the given species. Different colors denote different phyla: green for *Actinobacteria*, purple for *Bacteroidetes*, orange for *Euryarchaeota*, yellow for *Firmicutes*, blue for *Proteobacteria*, and pink for *Verrucomicrobia*. The thickness of edges denotes the correlation value. The red edge indicates a positive correlation, while the blue edge indicates a negative correlation. Identified biomarkers are presented in larger fonts.

In the correlation network for healthy samples, 9 main clusters are identified. The healthy network has a network density of 0.12932692. Both positive and negative correlations were found (Figure 7). Negative correlations were found between clusters 1, 3, 9, each connecting to cluster 2 via negative correlation. Prominent members from the healthy network include *Eubacterium eligens* from cluster 1, *Ruminococcus bromii* from cluster 3, *Bacteroides uniformis* from cluster 7, *Escherichia coli* from cluster 8, *Roseburia faecis, Eubacterium rectale, Faecalibacterium prausnitzil, Barnesiella intestinihominis, Parabacteroides distasonis,* and *Lachnospira pectinoschiza* from cluster 9.

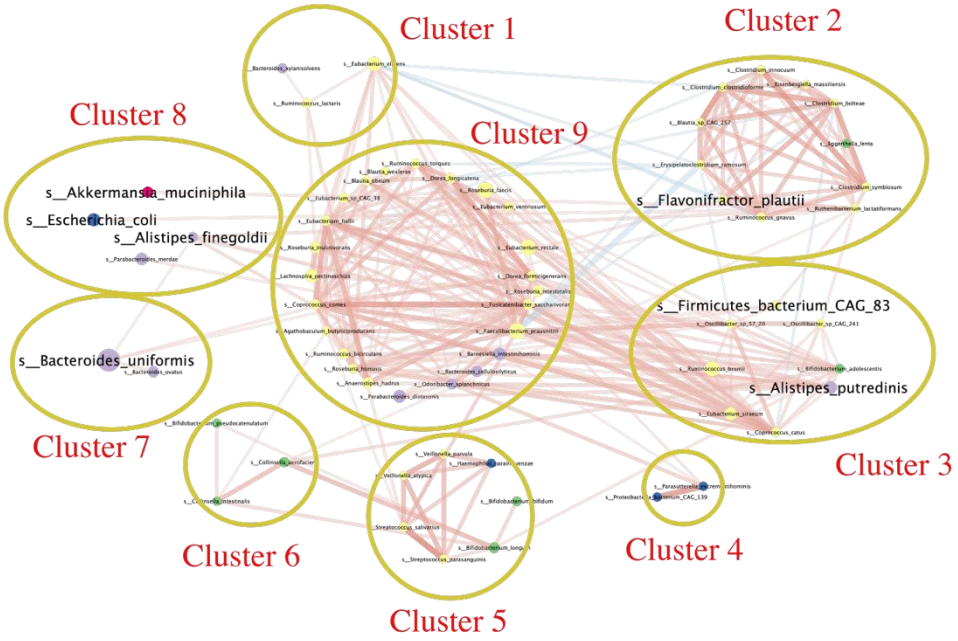


Figure 7. Healthy samples correlation network.

The non-healthy network consists of 4 clusters. The non-healthy network has a network density of 0.3816216. Both positive and negative correlations were observed in the network. Negative correlations were observed between clusters 2 and 3, which were each negatively correlated to cluster 4. Prominent members of the non-healthy include *Prevotella copri* in cluster 4; *Bacteroides dorel*, *Ruminococcus bromii*, *Anaerostipes hardus*, *Alitipes putredinis*, *Bacteroides uniformis*, *Bacteroides vulgatus*, *Bifidobacterium longum*, *Faecalibacterium prausnitzii*, and *Eubacterium rectale* in cluster 2.

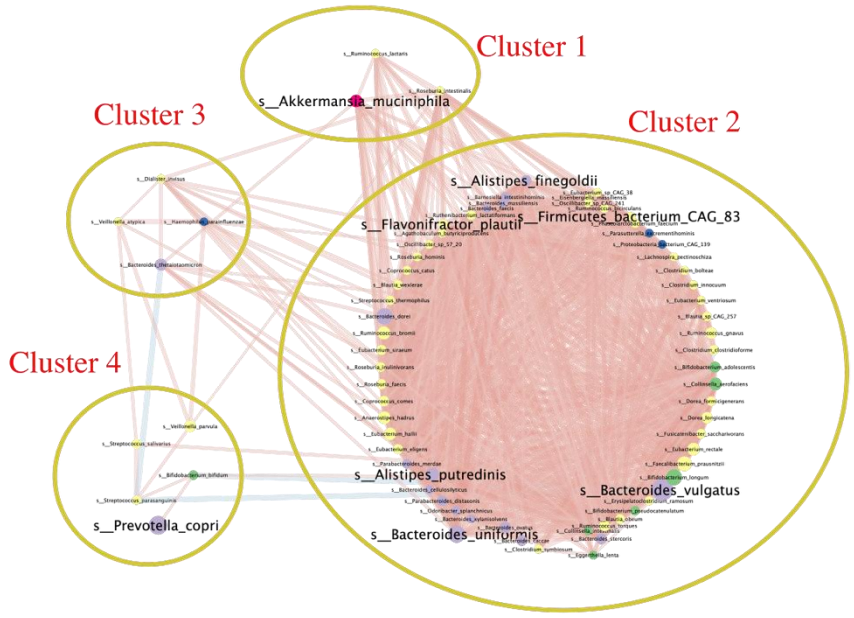


Figure 8. Non-healthy samples correlation network.

3.6. Random Forest identifies discriminative species

From the diarrhea counts dataset, the top 6 discriminative species that the random forest (RF) identified as either healthy or non-healthy are listed. *Akkermansia muciniphila*, *Alistipes putredinis*,

Coprococcus eutactus, *Firmicutes bacterium*, *Gemmiger formicilis*, and *Ruminococcus torques* are all more abundant in healthy than non-healthy samples (Figure 9).

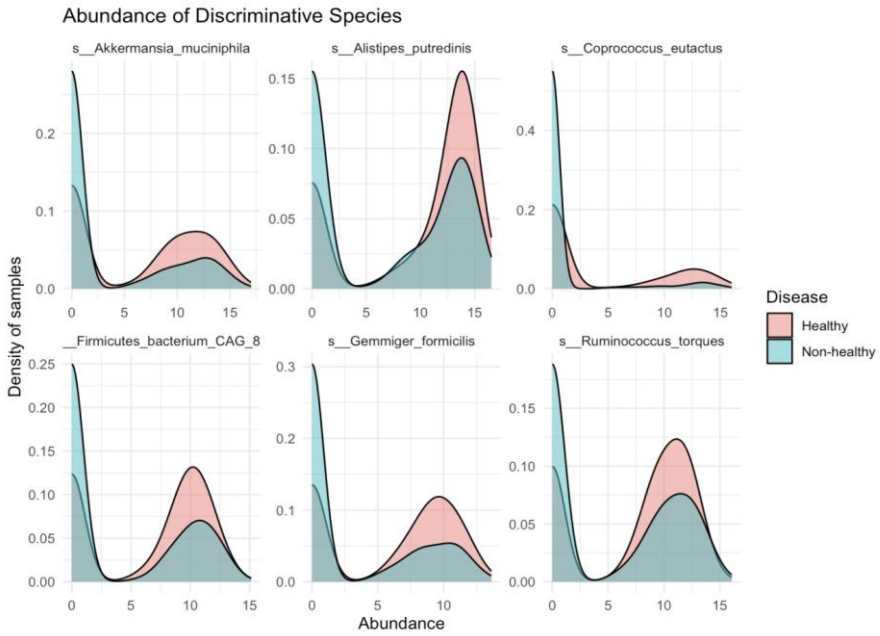


Figure 9. Top abundance of discriminative species in non-healthy and healthy species.

From the DESeq2-filtered diarrhea counts dataset, the top 6 discriminative species that random forest recognized in healthy and nonhealthy species are identified. 5 of them, *Akkermansia muciniphila*, *Collinsella aerofaciens*, *Oscilibacter*, *Prevotella buccalis*, and *Ruminococcus bromii* are all more abundant in healthy than non-healthy samples (Figure 10). *Flavonifractor plautii* is more abundant in non-healthy samples than in healthy ones.

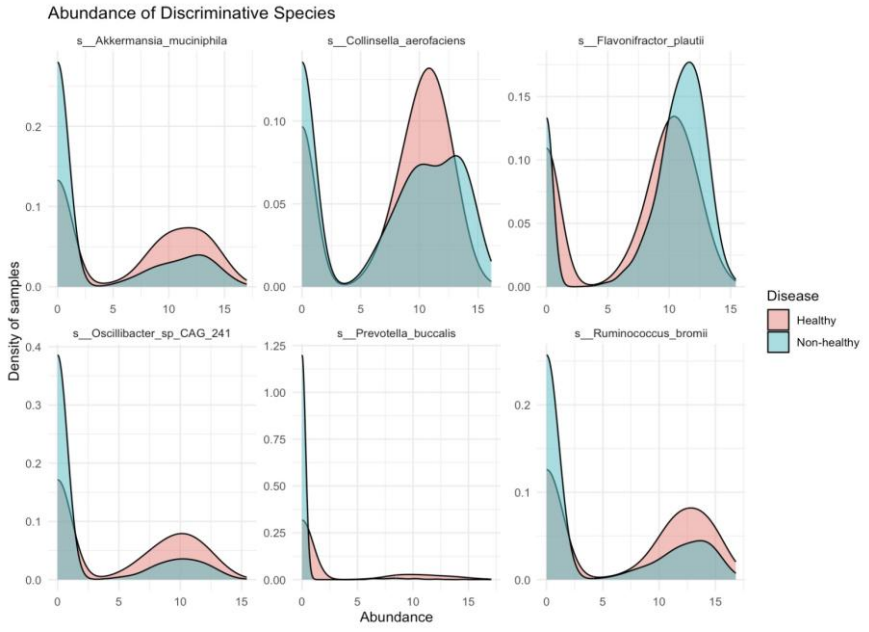


Figure 10. Top abundance of discriminative species in non-healthy and healthy species in DESeq2 filtered samples.

From the DESeq2-filtered diarrhea functional pathway dataset, the top 6 discriminative gene pathways that RF recognized in healthy and non-healthy pathways are identified. 4 pathways, including super pathway of glycerol degradation to 1,3-propanediol, methanogenesis from acetate, isoprene biosynthesis I, taxadiene biosynthesis engineered, are all more abundant in non-healthy than

healthy samples (Figure 11). On the other hand, 2 pathways, including phytol degradation and aerobic respiration I (cytochrome c) are more abundant in healthy than non-healthy samples.

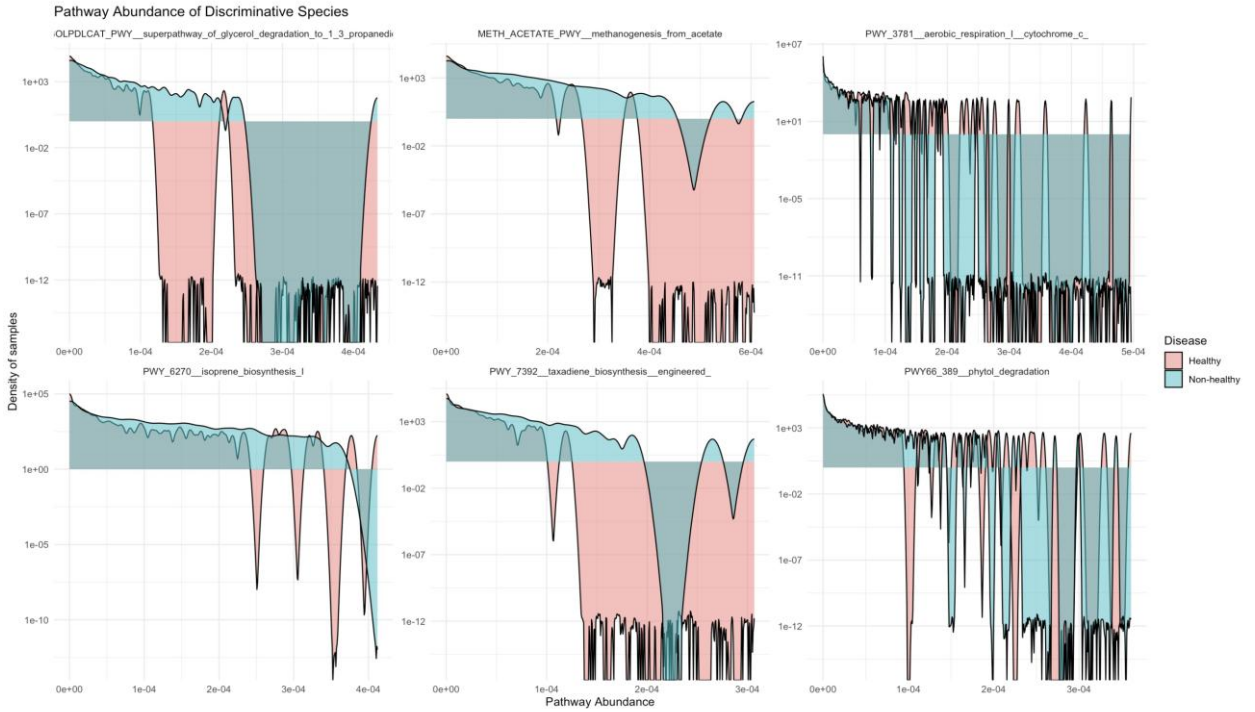


Figure 11. Top abundance of discriminative functional pathway in non-healthy and healthy pathways in DESeq2 filtered samples. The graph is log10 transformed.

4. Discussion

We found that different countries have a profound impact on a person's gut microbiome composition. The TKI-dependent diarrhea group shows significantly higher alpha diversity compared to the rest of the disease groups (Figure 1a). While it may seem counterintuitive that diarrhea samples have higher alpha diversity than healthy samples, the TKI-dependent diarrhea cohort mainly consists of stools collected from patients in Italy, France, and the Netherlands [12]. As shown by our result (Figure 3), Canada, Germany, Denmark, Spain, Great Britain, Italy, Netherland, and USA all have significantly higher alpha diversity than Bangladesh. This is due to most of the Bangladesh samples being diagnosed with acute diarrhea, which ranks the lowest alpha diversity among all disease categories. Thus, the fact that TKI-dependent diarrhea has the highest alpha diversity among all disease categories is due to its sample's focus in specific countries, which renders a higher alpha diversity overall. Indeed, our ANOVA result again confirms that country has the biggest influence over determining the level of alpha diversity, showcased by the country's highest F-value of 134.165, followed by disease with an F-value of 54.304.

Indeed, gut microbiome composition reflects a person's diet, which varies among different countries. Study by Nayama et al. that investigates gut microbiome difference of children in Asia found that the *Prevotella* group is more abundant in gut microbiome of people from Thailand than in other Asian countries [13]. This difference reflects their diet of high intake of resistant starch heavy on rice. On the other hand, Japanese children has less diversified gut microbiome with more *Bifidobacterium* abundance, reflecting their highly hygienic lifestyle and unique diets.

Overall, we identify biomarkers of healthy samples as the following. *Bacteorides uniformis* and *Bacteroides vulgatus*, as suggested by the core microbiome and prevalence plot (Figure 4) result. *Akkermansia muciniphila*, as suggested by RF relative abundance analysis (Figure 9) and RF DESeq2 analysis (Figure 10); *Firmicutes bacterium*, as suggested by the DESeq2 relative abundance and RF

relative abundance result (Figure 10). *Alistipes*, as suggested by DESeq2 relative abundance analysis and RF analysis (Figure 9).

Biomarkers that characterize non-healthy samples include *Flavonifractor plautii*, as suggested by core microbiome result and RF DESeq2 result (Figure 10). *Escherichia coli*, as suggested by prevalence plot (Figure 4) and aligned with the result of the community profiling plot (Figure 3), as *E. coli* is under the *Proteobacteria* phylum and *Proteobacteria* phylum is dominant in the acute diarrhea category. *Prevotella*, as suggested by the DESeq2 relative abundance result and its high abundance in the non-healthy network analysis (Figure 8).

Our results indeed aligned with published literatures. For healthy biomarkers, studies have suggested *B. uniformis* is one of the most prevalent and abundant bacterial species in mammalian species. It directly competes with pathogens in gut microbiome over consuming host-derived amino acids and monosaccharides. It also increases the abundance of symbiotic bacteria, making it potential therapy for intestinal barrier dysfunction [14]. *B. vulgatus* is part of the genera *Bacteroides*, often an indication of healthy gut microbiome [15]. It also can protect against *Escherichia coli*-induced colitis [16]. Low level of *Akkermansia muciniphila* has been found in association with several diseases, such as obesity, diabetes, and inflammation. *A. mucinphila* also helps build and maintain a healthy gut barrier, protecting against inflammation from the start [17]. *Firmicutes* represented over 90% of the gut microbiome, alongside with *Bacteroides*. *Firmicutes* help digest dietary fiber [18]. Another species, namely the *Alistipes*, is a fairly new genus of bacteria that has protective effects on colitis [19].

For the non-healthy biomarkers, studies have suggested *Flavonifractor plautii* is responsible for the degradation of beneficial anticarcinogenic flavonoids, causing CRC in India [20]. For *E. coli*, oftentimes, following infectious diarrhea, *Enterobacteriaceae* bloom happens in the gut [21]. Studies found that high abundance of *Prevotella* tend to induce inflammation [22].

We also find the non-healthy samples' microbiomes are much less stable than that of healthy samples. In the correlation analysis, we found the non-healthy network has a network density of 0.3816216, higher than healthy network, with network density of 0.12932692. This difference indicates that the non-healthy network is more interconnected than the healthy network. Thus, if there is a disruption in one of the species in non-healthy network, it will have a bigger repercussion than a disruption in healthy network. In addition, we overserved more negative correlations in healthy than non-healthy network, indicating more stable connection. We also observed that the healthy network has more clusters than the non-healthy network, showing that will be less prone to external effects.

In general, we also identify biomarker pathways that characterizes healthy and non-healthy samples. In healthy samples, we identify 2 significant pathways, including phytol degradation and aerobic respiration I (cytochrome c), both indicated by the result of DESeq2 and RF (Figure 11). For non-healthy samples, we identify 4 significant pathways, including glycerol degradation to 1,3-propanediol, methanogenesis from acetate, isoprene biosynthesis I, and taxadiene biosynthesis (engineered) (Figure 11), suggested by both DESeq2 and RF result.

According to literatures, phytol degradation can be employed to digest and absorb phytol after phytol-rich diet [23]. Aerobic respiration I (cytochrome c) accounts for the electron transport chain in cellular respiration, a common feature for aerobic microbes. Glycerol degradation to 1,3-propanediol is often facilitated by Clostridia as well as Enterobacteriaceae, suggesting the enrichment of these two species [24]. These species also often suggest a diarrhea/non-healthy gut microbiome. Methanogenesis from acetate suggested the presence of methanogen [25]. Methanogens are often archaea, which were associated with various human diseases [26]. Isoprene biosynthesis served as an intermediate for bacteria cell wall synthesis [27]. *E. coli* frequently produces taxadiene via the taxadiene biosynthesis pathway [28], suggesting enrichment of *E coli* where taxadiene biosynthesis pathway is enriched.

By highlighting these biomarkers and pathways, we provide insights in what can be the potential species and functional pathway biomarkers for both diarrhea and healthy samples. This can thus be

used for further indication when diagnosing diarrhea, or targeted treatment of diarrhea specific to the biomarkers. It marks a potential shift from treating diarrhea with medication to adjusting diets, thus eradicating diarrhea from its root cause, enabling personalized medicine. Additionally, knowing the biomarkers would be beneficial for future diagnosis. To that end, instead of sequencing the entire gur microbiome, only more efficient sequencing technique (16srRNA amplicon sequencing or PCR) is needed to characterize healthy versus diarrhea gut microbiome.

5. Conclusion

This study provides precise insights into biomarkers highlighting gut microbiome alterations following diarrhea. Through comprehensive analysis based on a large dataset, we observed healthy samples have species *Bacteroides uniformis, Bacteroides vulgatus, Akkermansia muciniphila, Firmicutes bacterium,* and *Alistipes* enriched, as well as pathways Phytol degradation and Aerobic respiration I (cytochrome C) upregulated. For diarrhea samples, we observed species *Flavonifractor plautii*, *Escherichia coli,* and *Prevotella* enriched as well as pathways including Glycerol degradation, Methanogenesis, Isoprene biosynthesis, and Taxadiene biosynthesis upregulated. These biomarkers marked significant shift in gut microbiome after diarrhea, and they could be targeted for observing dysbiosis in diarrhea and therefore effectively treating it. Potential ways of mitigation and treatment include targeted microbe transplant, balanced nutrient supplement, or medications considering impacts on gut microbiome interactions. Overall, this study underscores the intricate relationship between diarrhea and gut microbiome changes. While this study paves a way for biomarkers in diarrhea, future investigations, especially through clinical studies, are needed to further validate the biomarkers beyond mere computational modeling.

Acknowledgement

I wish to convey my heartfelt gratitude to Dr. Rui Guan for her continuous support, encouragement, and for spending time debugging code with me. Her expertise and background inspired my research.

I would also like to thank myself. Without my passion for science, persistence in work, and absolutely crazy ideas, I would not have completed this project successfully.

References

[1] GBD Diarrhoeal Diseases Collaborators (2017). Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet. Infectious diseases, 17(9), 909–948. https://doi.org/10.1016/S1473-3099(17)30276-1

[2] WHO. (n.d.). Diarrhoeal disease. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease

[3] Johns Hopkins Medicine. (n.d.). Inflammatory bowel disease. Retrieved December 17, 2024, from https://www.hopkinsmedicine.org/health/conditions-and-diseases/inflammatory-bowel-disease

[4] Li, Y., Xia, S., Jiang, X., Feng, C., Gong, S., Ma, J., Fang, Z., Yin, J., & Yin, Y. (2021). Gut Microbiota and Diarrhea: An Updated Review. Frontiers in cellular and infection microbiology, 11, 625210. https://doi.org/10.3389/fcimb.2021.625210

[5] Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., ... & Wong, G. K. S. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. Frontiers in microbiology, 7, 459.

[6] David, L. A., Weil, A., Ryan, E. T., Calderwood, S. B., Harris, J. B., Chowdhury, F., Begum, Y., Qadri, F., LaRocque, R. C., & Turnbaugh, P. J. (2015). Gut microbial succession follows acute secretory diarrhea in humans. mBio, 6(3), e00381-15. https://doi.org/10.1128/mBio.00381-15

[7] Pasolli, E., Schiffer, L., Manghi, P., Renson, A., Obenchain, V., Truong, D. T., Beghini, F., Malik, F., Ramos, M., Dowd, J. B., Huttenhower, C., Morgan, M., Segata, N., & Waldron, L. (2017). Accessible, curated metagenomic data through ExperimentHub. Nature methods, 14(11), 1023–1024. https://doi.org/10.1038/nmeth.4468

[8] Dixon, P. (2003), VEGAN, a package of R functions for community ecology. Journal of Vegetation Science, 14: 927-930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

[9] Lahti, L., & Shetty, S. (2017). microbiome R package. Bioconductor. https://doi.org/10.18129/B9.bioc.microbiome

[10] Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014). https://doi.org/10.1186/s13059-014-0550-8

[11] Fauziyyah, Nabiilah Ardini (2020). Classification using Microbiome. In Exploring microbiome analysis using R. https://microbiome.netlify.app/classification-using-microbiome

[12] Ianiro, G., Punčochář, M., Karcher, N., Porcari, S., Armanini, F., Asnicar, F., Beghini, F., Blanco-Míguez, A., Cumbo, F., Manghi, P., Pinto, F., Masucci, L., Quaranta, G., De Giorgi, S., Sciumè, G. D., Bibbò, S., Del Chierico, F., Putignani, L., Sanguinetti, M., Gasbarrini, A., … Segata, N. (2022). Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases. Nature medicine, 28(9), 1913–1923. https://doi.org/10.1038/s41591-022-01964-3

[13] Nakayama, J., Watanabe, K., Jiang, J. et al. Diversity in gut bacterial community of school-age children in Asia. Sci Rep 5, 8397 (2015). https://doi.org/10.1038/srep08397

[14] Yan, Y., Lei, Y., Qu, Y. et al. Bacteroides uniformis-induced perturbations in colonic microbiota and bile acid levels inhibit TH17 differentiation and ameliorate colitis developments. npj Biofilms Microbiomes 9, 56 (2023). https://doi.org/10.1038/s41522-023-00420-5

[15] Liu Liyun , Xu Mingchao , Lan Ruiting , Hu Dalong , Li Xianping , Qiao Lei , Zhang Suping , Lin Xiaoying , Yang Jing , Ren Zhihong , Xu Jianguo, Bacteroides vulgatus attenuates experimental mice colitis through modulating gut microbiota and immune responses, Frontiers in Immunology, vol. 13, 2022, https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2022.1036196, 10.3389/fimmu.2022.1036196

[16] Waidmann, M., Bechtold, O., Frick, J. S., Lehr, H. A., Schubert, S., Dobrindt, U., Loeffler, J., Bohn, E., & Autenrieth, I. B. (2003). Bacteroides vulgatus protects against Escherichia coli-induced colitis in gnotobiotic interleukin-2-deficient mice. Gastroenterology, 125(1), 162–177. https://doi.org/10.1016/s0016-5085(03)00672-3

[17] Cani, P.D., Depommier, C., Derrien, M. et al. Akkermansia muciniphila: paradigm for next-generation beneficial microorganisms. Nat Rev Gastroenterol Hepatol 19, 625–637 (2022). https://doi.org/10.1038/s41575-022-00631-9

[18] Parker, B. J., Wearsch, P. A., Veloo, A. C. M., & Rodriguez-Palacios, A. (2020). The genus Alistipes: Gut bacteria with emerging implications to inflammation, cancer, and mental health. Frontiers in Immunology, 11, 906. https://doi.org/10.3389/fimmu.2020.00906

[19] Parker, B. J., Wearsch, P. A., Veloo, A. C. M., & Rodriguez-Palacios, A. (2020). The Genus Alistipes: Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health. Frontiers in immunology, 11, 906. https://doi.org/10.3389/fimmu.2020.00906

[20] Gupta, A., Dhakan, D. B., Maji, A., Saxena, R., P K, V. P., Mahajan, S., Pulikkan, J., Kurian, J., Gomez, A. M., Scaria, J., Amato, K. R., Sharma, A. K., & Sharma, V. K. (2019). Association of Flavonifractor plautii, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India. mSystems, 4(6), e00438-19. https://doi.org/10.1128/mSystems.00438-19

[21] Hao Chung The, Son-Nam H Le, Dynamic of the human gut microbiome under infectious diarrhea, Current Opinion in Microbiology, Volume 66, 2022, Pages 79-85, ISSN 1369-5274, https://doi.org/10.1016/j.mib.2022.01.006.

[22] Larsen J. M. (2017). The immune response to Prevotella bacteria in chronic inflammatory disease. Immunology, 151(4), 363–374. https://doi.org/10.1111/imm.12760

[23] J. Gloerich, D.M. van den Brink, J. P.N. Ruiter, N. van Vlies, F.M. Vaz, R. J.A. Wanders, S. Ferdinandusse,

[24] Metabolism of phytol to phytanic acid in the mouse, and the role of PPARα in its regulation, Journal of Lipid Research, Volume 48, Issue 1, 2007, Pages 77-85, ISSN 0022-2275, https://doi.org/10.1194/jlr.M600050-JLR200.

[25] Biebl, H., Menzel, K., Zeng, A. P., & Deckwer, W. D. (1999). Microbial production of 1,3-propanediol. Applied microbiology and biotechnology, 52(3), 289–297. https://doi.org/10.1007/s002530051523

[26] Das, S., & Dash, H. R. (2020). Microbial and Natural Macromolecules: Synthesis and Applications. Academic Press.

[27] Duller, S., & Moissl-Eichinger, C. (2024). Archaea in the Human Microbiome and Potential Effects on Human Infectious Disease. Emerging Infectious Diseases, 30(8), 1505-1513. https://doi.org/10.3201/eid3008.240181.

[28] Kuzuyama, T., & Seto, H. (2012). Two distinct pathways for essential metabolic precursors for isoprenoid biosynthesis. Proceedings of the Japan Academy. Series B, Physical and biological sciences, 88(3), 41–52. https://doi.org/10.2183/pjab.88.41

[29] Qiulong Huang, Charles A Roessner, Rodney Croteau, A.Ian Scott, Engineering Escherichia coli for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol, Bioorganic & Medicinal Chemistry, Volume 9, Issue 9, 2001, Pages 2237-2242, ISSN 0968-0896, https://doi.org/10.1016/S0968-0896(01)00072-4.