Available Online: 3 August 2026 DOI: 10.54254/3029-0821/2026.35786

Associations among modifiable lifestyle factors and dominant intestinal bacterial genus

Yilin Zhang

School of Health Economics and Management, Nanjing University of Chinese Medicine, Nanjing, China

012423154@njucm.edu.cn

Abstract. Modifiable lifestyle factors influence the structure of the gut microbiome; however, most studies have isolated individual exposures and focused on continuous diversity indices rather than categorical community phenotypes. As a result, the relative weights of various coexisting lifestyle habits in determining the composition of the dominant intestinal bacterial genus are still unknown. Genus-level relative abundance data and self-reported lifestyle metadata for 336 adults from the American Gut Project were analysed. Remove unclassified taxa and re-normalise abundances, then assign a dominant genus to each sample based on its most abundant taxon. Associations with the frequency of alcohol consumption, exercise frequency, sleep duration, dietary pattern, antibiotic use history, body mass index and demographic factors were analysed using univariate tests with False Discovery Rate (FDR) correction, multinomial logistic regression, PERMANOVA, redundancy analysis and differential abundance tests. Dataset-of-origin was explicitly modelled to rule out technical batch effects. After controlling for batch effects in a multivariate model, alcohol consumption frequency (importance = 0.0230), exercise frequency (importance = 0.0209) and sleep duration (importance = 0.0168) were the top two modifiable predictors of dominant genus classification. Only the frequency of alcohol consumption was significantly associated with the global community structure (PERMANOVA, F = 2.03, p = 0.01, FDR = 0.04), and Bacteroides showed a graded increase in relative abundance from 10% for never-drinkers to 39% for daily drinkers. Exercise frequency was positively correlated with increased Rothia abundance (FDR = 0.039). The dataset identifier was a relatively strong overall predictor of the dominant genus assignment (importance = 0.0462) and exceeded the combined contribution of all lifestyle variables. Alcohol consumption, exercise and sleep duration are all independently associated with the composition of dominant intestinal bacteria in adults after controlling for technical batch effects. The above results present new high-priority targets for microbiome modification and also indicate that explicit batch correction must be performed during aggregated microbiome analysis.

Keywords: gut microbiome, dominant genus, lifestyle factors, batch effects

1. Introduction

The human gut microbiome is a metabolic system in the gut that can alter energy balance and immune-regulatory and neuroendocrine signals of the host by producing Short-Chain Fatty Acids (SCFAs) and secondary bile acids [1-3]. Perturbations to this ecosystem are called dysbiosis, and they have been

mechanistically linked to obesity, type 2 diabetes, and chronic inflammatory diseases [2]. Therefore, the first aim will be to identify modifiable reasons for changes in gut microbiota in chronic diseases.

Food intake, exercise, sleep, alcohol consumption and antibiotic use are known to alter the environment of the adult gut microbiome [3, 4]. High-fibre and Mediterranean diets generally increase the abundance of SCFA-producing genera, such as *Faecalibacterium* and *Bifidobacterium*, and reduce alpha diversity [5]. A 2025 systematic review further verified that diets rich in fibre, phytonutrients and unsaturated fatty acids selectively promote the growth of *Akkermansia*, *Bacteroides*, *Faecalibacterium* and *Roseburia*, and thus have an important impact on the composition of the gut microbiome [6]. A plant-rich diet can promote the growth of butyrate-producing bacteria in the gut; *Faecalibacterium* is a representative such organism that has been frequently identified as an increase under these circumstances and may offer a path to diet-induced modifications of the gut microbiota that improve cardiometabolic health [7]. Exercise can also increase the diversity of gut bacteria, and it is relatively safe. A meta-analysis of obese and type 2 diabetic individuals showed a significant increase in Shannon index (SMD = 0.40 [0.15, 0.65] for obesity; SMD = 0.48 [0.08, 0.88] for type 2 diabetes) and enrichment of butyrate-producing taxa, such as *Roseburia* and *Akkermansia muciniphila* [8]. A separate 2024 meta-analysis of 25 studies including 1,044 adults confirmed the above results and found a significant increase in Shannon index (WMD = 0.05, 95% CI: 0.00-0.09) with different responses for men and women of different age groups [9]. Additionally, a 2025 systematic review and meta-analysis of randomised controlled trials in overweight and obese individuals found that exercise also had a positive effect on beta diversity (pooled Bray-Curtis dissimilarity = 4.56; 95% CI: 1.77–11.80; p = 0.002), thus providing more support for the idea that exercise improves the structure of the gut microbial community [10].

Conversely, circadian disruption, chronic alcohol consumption and antibiotic use have also been shown to cause dysbiosis, a reduction in microbial diversity and an increase in pro-inflammatory pathobionts [11-13]. A 2024 review in the field of sleep has shown that irregular sleep cycles and insomnia lead to stress responses that alter the composition of gut microbiota and disrupt circadian rhythms; thus, sleep, stress, and the gut microbiome are closely interconnected [14]. In 2025, some studies have also found that both circadian rhythm disruption and sleep deprivation can harm the structure and function of the gut microbiota, thus leading to metabolic disorders and inflammation in the host [15]. A systematic review of 50 clinical studies has shown that gut dysbiosis is a common feature of individuals with alcohol use disorder, with reduced microbial diversity and impaired gut barrier function [16]. In 2025, a comprehensive study was conducted that identified several links in the gut-brain axis among alcohol-induced changes in the gut microbiota and their resulting neuroinflammation and behaviour modification [17]. According to a 2024 review on antibiotics, treatment rapidly reduces the diversity of the microbiome; even partial recovery takes as long as six months, and at the same time, the gut resistome is enriched, leading to an increase in antimicrobial resistance genes [18]. In addition to the above reasons, chronic psychological stress has also been shown to cause dysbiosis and promote neuroinflammatory pathways through the microbiota-gut-brain axis [19]. Although psychological stress was not explicitly measured in the present study, it is well-established that stress affects gut microbiota, so lifestyle and the microbiome are also interconnected to some extent.

Many studies have been conducted on the above factors of life, but three deficiencies remain. First, most studies have examined the exposure in isolation and thus failed to assess the combined effects of multiple behaviours in free-living populations [20]. Second, most existing research focuses on continuous measures of alpha/beta diversity or differential abundance of individual taxa and rarely examines how lifestyle factors influence the classification of dominant genera; these genera can be considered a discrete phenotype that reflects the single most abundant taxon in a sample and may serve as a clinically relevant biomarker of gut

ecosystem status. Third, technical confounders, such as dataset batch effects and sample source heterogeneity, often mask biological signals in microbiome studies, and few studies have formally quantified the relative contribution of technical versus lifestyle factors [21].

The present study systematically explores the association among modifiable lifestyle factors and the genus classification of dominant intestinal bacteria to address the above deficiencies. A total of 336 human fecal samples have been collected, and their corresponding metadata, such as diet, exercise frequency, alcohol consumption, sleep time, antibiotic use history and body mass index, have been standardised. Our analytical framework uses both univariate association tests with False Discovery Rate (FDR) correction and multinomial logistic regression to examine the independent and joint effects of lifestyle factors, and explicitly controls for dataset-of-origin batch effects. Designate the dominant genus as an indivisible, readily interpretable phenotype and quantify the relative contributions of multiple lifestyle exposures in a single model; thus, a general view of how daily behaviour collectively shapes the gut microbial ecosystem is provided, along with high-priority lifestyle targets for follow-up intervention research.

2. Materials and methods

2.1. Data source and study population

Relative abundance tables at the phylum and genus levels, and participant metadata were obtained from the American Gut Project via the Qiita platform. The set of data included 336 fecal samples and comprehensive information on the living conditions, physical measurements and origins of the samples. All participants had given informed consent according to the original project protocols.

2.2. Definition of dominant genus

The taxon with the highest relative abundance in the sample was taken as the dominant genus. Low-prevalence genera were grouped as "*Other*" and a discrete outcome variable for association analysis was thus formed.

2.3. Statistical analysis

Univariate analysis of the association between each individual predictor and the main genus classification was performed using a chi-squared test (for categorical variables) or the Kruskal-Wallis rank-sum test (for continuous/ordinal variables). Benjamini-Hochberg False Discovery Rate (FDR) correction was used for all univariate comparisons, and an FDR < 0.05 was taken as statistically significant. Multinomial logistic regression was used to examine the independent effects of lifestyle factors on the dominant genus, controlling for technical covariates (dataset identifier) and demographic variables in a model. Variable importance was calculated as the sum of the absolute regression coefficients for all outcome categories. Bray-Curtis dissimilarities were computed for genus-level relative abundances. Permutational Multivariate Analysis of Variance (PERMANOVA) [22] was conducted with 999 permutations to investigate the effect of each categorical factor on the whole community structure, and FDR correction was applied to all factors. Principal Coordinates Analysis (PCoA) was used for ordination visualisation. Kruskal-Wallis tests with FDR correction were employed to identify genera exhibiting different relative abundances at various levels of the above lifestyle factors. Hellinger-transformed genus abundances were used as the dependent variables in Redundancy Analysis (RDA) to explore multivariate relationships with predictor variables. Pearson correlation coefficients between lifestyle factors and the relative abundances of genera were calculated,

hierarchically clustered, and an overall heatmap of lifestyle-microbiota associations was generated. To rule out batch effects in the results, the cohort was split based on the sample source, and the primary association tests were carried out within these divisions. All the above analyses were performed in Python with the standard scientific computing library.

3. Results

Raw 16S rRNA gene amplicon sequencing data and curated participant metadata were obtained from the American Gut Project via the Qiita platform. A total of 336 fecal samples were collected, and all the lifestyle, anthropometric and sample-of-origin data were properly recorded. The analysis workflow was as follows: Exclude taxa without classification, re-normalise the abundance of the remaining taxa to a total of 1 per sample, and then conduct multivariate modelling and differential abundance tests. All analyses were conducted on the pre-processed phylum- and genus-level relative abundance tables provided by the American Gut Project (Figure 1).

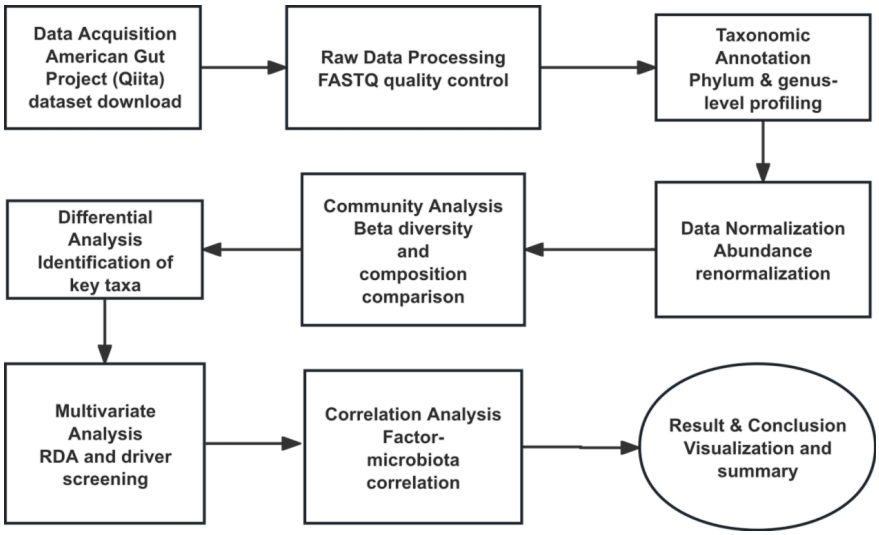

Figure 1. Schematic overview of the analytical workflow

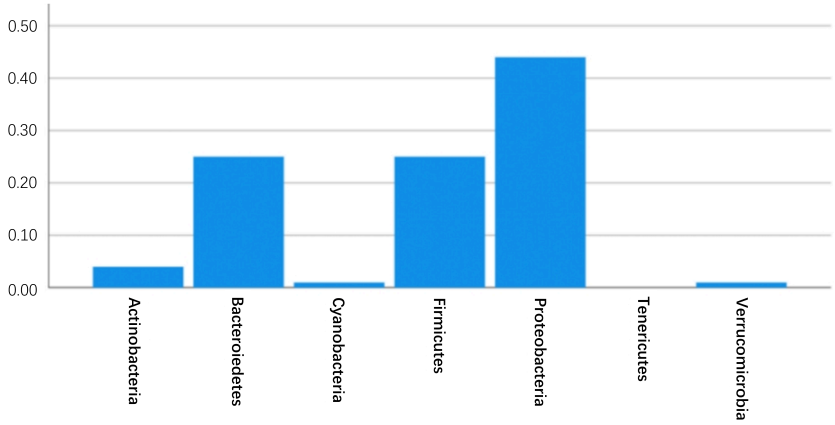

Figure 2. Phylum-level relative abundance of microbiota

Proteobacteria (mean relative abundance 45%), *Firmicutes* (27%) and *Bacteroidetes* (26%) were the main phyla (Figure 2) in the cohort, with small proportions of *Actinobacteria* (4%), *Cyanobacteria* (1%), *Tenericutes* (0.5%) and *Verrucomicrobia* (0.2%). At the genus level, the top taxa with the highest mean relative abundances in all samples (Figure 3) were *Bacteroides*, *Enterobacter*, *Faecalibacterium*, *Akkermansia*, *Bifidobacterium*, *Prevotella*, *Parabacteroides*, *Pseudomonas*, and *Coprococcus*.

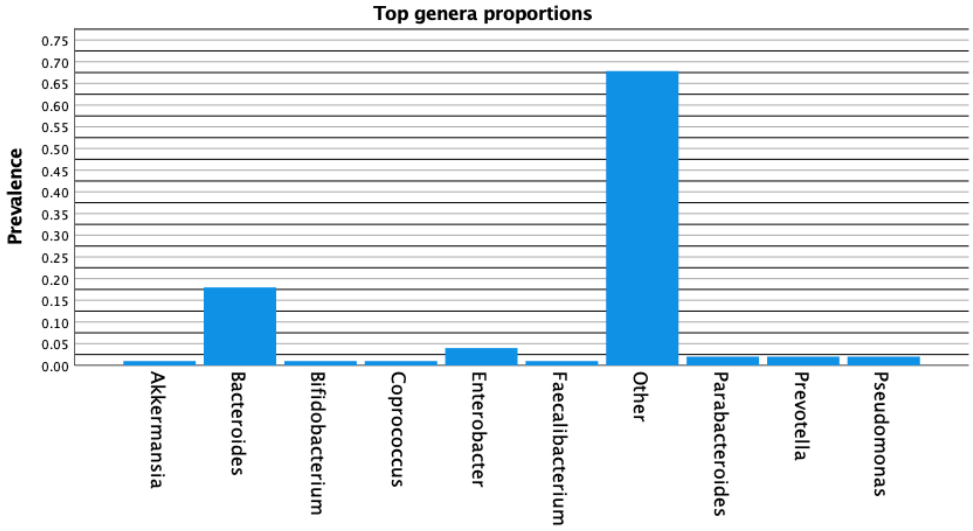

Figure 3. Genus-level relative abundance of microbiota

3.1. Definition of dominant genus and overall distribution

For each sample, the taxon with the highest relative abundance in that sample was designated as the dominant genus. The distribution of the dominant genera had a head-and-tail shape. *Corynebacterium* and *Bacteroides* were the most frequently occurring dominant genera, followed by a decreasing gradient of less abundant genera such as *Enterobacter*, *Pantoea*, *Streptococcus*, *Pseudomonas*, *Parabacteroides* and *Prevotella*. Therefore, due to the above skewness, statistical methods suitable for sparse categorical data were selected.

3.2. Univariate and Multivariate Associations with Dominant Genus

Univariate and Multivariate Associations with Dominant Genus Univariate chi-square tests, with False Discovery Rate (FDR) correction for all comparisons, show that the dataset source identifier (*dataset*) is the most significant single predictor of dominant genus classification (χ^2 = 149.8, FDR = 1.30×10^{-22}). Among the lifestyle-related variables, *diet_type*had the strongest association (χ^2 = 63.16, FDR = 0.0062); *antibiotic_history*and *alcohol_frequency* were moderately associated but not statistically significant (FDR = 0.249 for both).

3.3. multivariate modelling and global community structure

To separate the independent effects of multiple lifestyle exposures and control for technical confounders, a multinomial logistic regression model was used, and dominant genus served as the polytomous outcome. Variable importance is defined as the sum of the absolute regression coefficients for each outcome category, so *dataset* was still the most influential predictor (importance = 0.0462). Adjusting for dataset effects, *alcohol consumption frequency* (importance = 0.0230), *exercise frequency* (importance = 0.0209) and *sleep duration*

(importance = 0.0168) were found to be the top modifiable lifestyle factors associated with the dominant genus composition. Geographic variables (*latitude* and *longitude*) and *diet_type* were moderately significant; *bmi_cat*, *antibiotic_history*, and *sex* contributed to model discrimination to a lesser extent (Figure 4).

Permutational Multivariate Analysis of Variance (PERMANOVA) was performed on Bray-Curtis dissimilarities to determine how much of the total variation in the microbial community was explained by each factor. Consistent with the regression results, *sample_source* explained the largest proportion of the variation (F = 6.67, p = 0.01, FDR = 0.04) (Figure 5). Among the lifestyle variables, only the frequency of alcohol consumption (Figure 6) showed a behavioural effect at the community level of compositional change (F = 2.03, p = 0.01, FDR = 0.04). Principal Coordinates Analysis (PCoA) ordination visually confirmed this association by showing that the samples were distributed in different clusters according to the frequency of alcohol consumption along the first two principal coordinate axes (PCo1: 27.6%; PCo2: 8.8%), although there was significant overlap; that is to say, many factors influenced the changes in gut microbiota. On the other hand, none of the other lifestyles or anthropometric parameters reached statistical significance after FDR correction (all FDR > 0.32).

Redundancy Analysis (RDA) of Hellinger-transformed genus abundances also showed the same results. The first two constrained axes explained 41.1% and 12.0% of the variance. The main cause of the separation in ordination was a sample source, and *Streptococcus*, *Corynebacterium* and *Rothia* loaded heavily on RDA1, while *Bacteroides*, *Pseudomonas* and *Enterobacter* contributed significantly to RDA2. As shown in the ordination plot, although the sample source accounts for a large proportion of the global community structure, lifestyle factors, specifically the frequency of alcohol consumption, have a relatively small impact on genus-level composition (Figure 7).

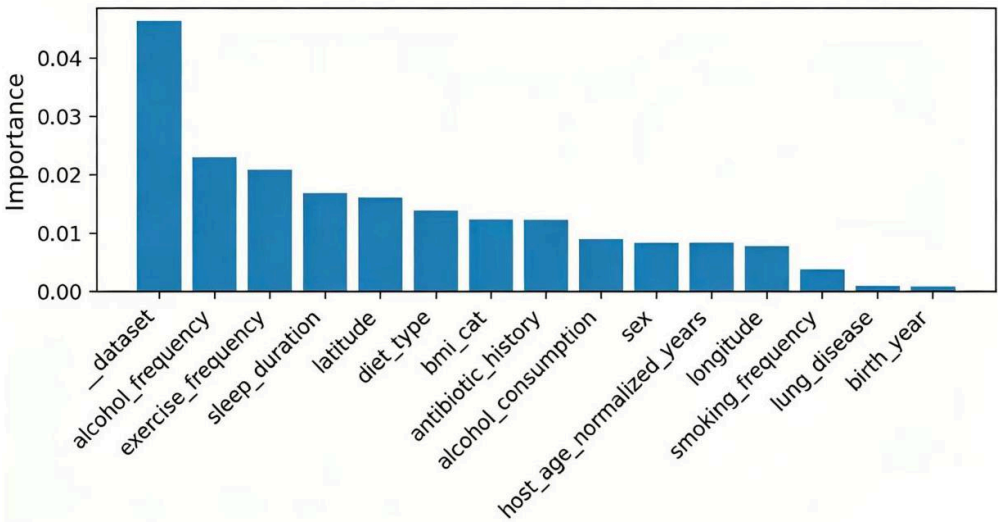

Figure 4. Variable importance from multinomial logistic regression

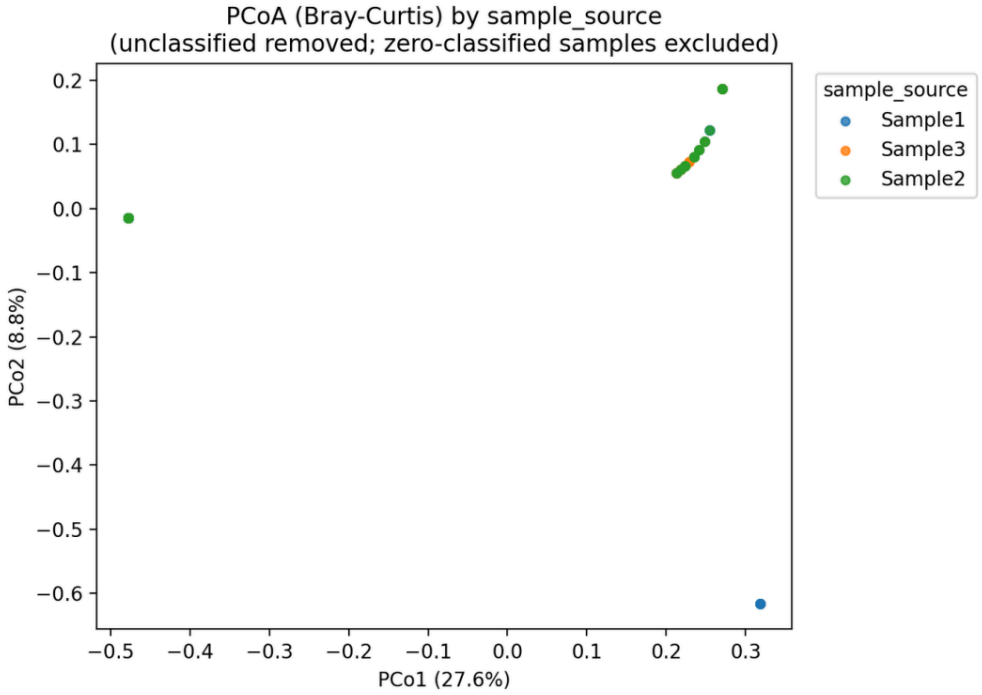

Figure 5. PCoA colored by sample source

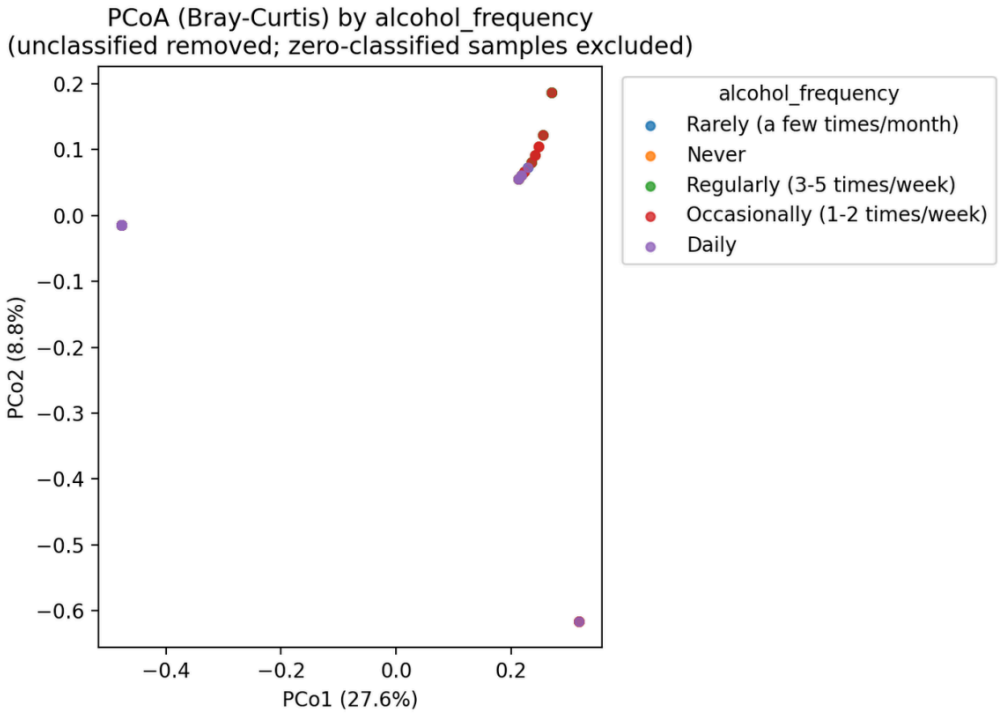

Figure 6. PCoA colored by alcohol consumption frequency

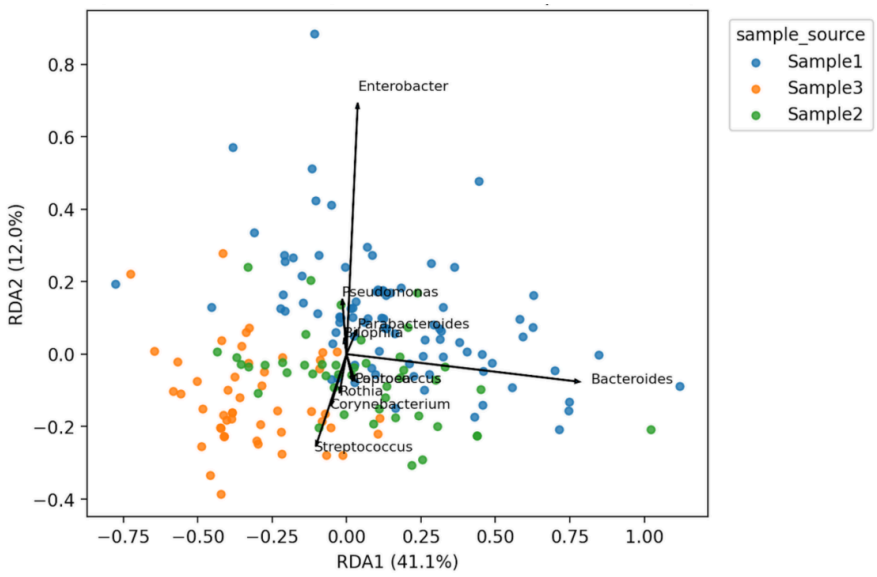

Figure 7. Redundancy analysis (RDA)

3.4. Compositional shifts at the phylum and genus levels in various lifestyle strata

To explore more in detail the microbial signatures of the most influential lifestyle and technical factors, compositional profiles were further divided based on *sample_source* and *alcohol_frequency.*

3.4.1. Variation among sample sources

There was a significant difference in phyla (Figure 8) and genera (Figure 9). *Firmicutes* was the most numerous in *Sample 1* (57%), and *Bacteroidetes* accounted for 46% in *Sample 3*. *Sample 1* had a relatively higher abundance of *Bacteroides* (mean = 0.26), and *Sample 3* had a relatively lower abundance (mean = 0.04) at the genus level. On the other hand, the "*Other*" category of low-abundance genera was relatively large in *Sample 3* and more aligned with a diverse, less skewed taxonomic structure.

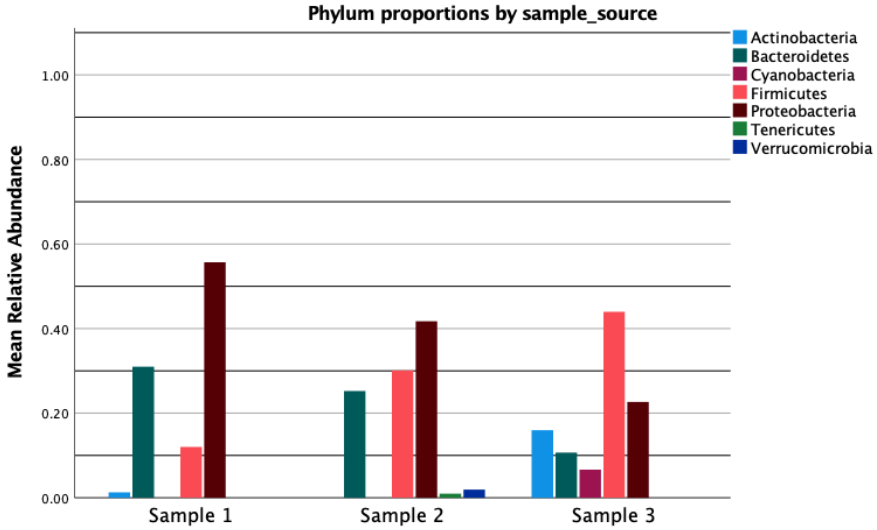

Figure 8. Phylum-level microbiota composition across sample sources

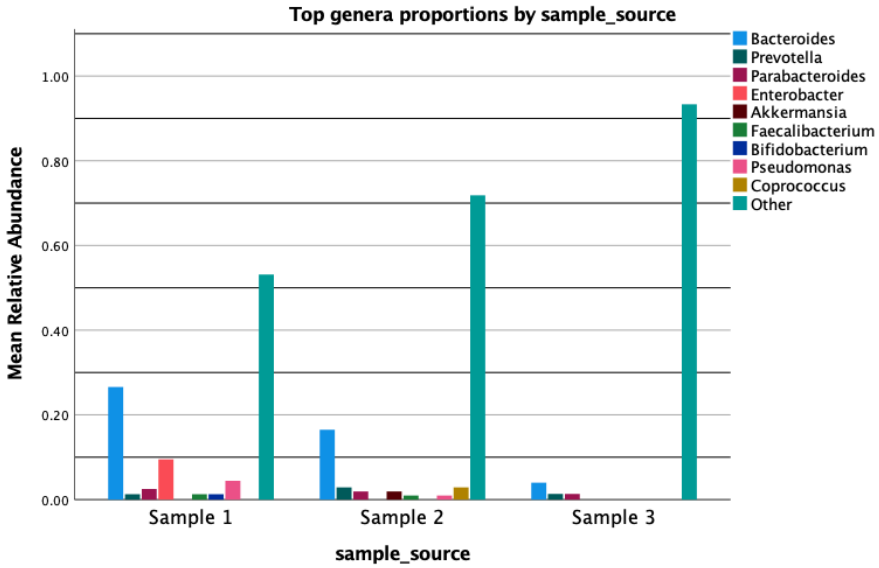

Figure 9. Genus-level microbiota composition across sample sources

3.4.2. Variation according to alcohol consumption frequency

Graded shifts in phyla-level composition across alcohol-drinking groups. *Bacteroidetes* relative abundance increased gradually with rising levels of alcohol consumption; that is, never-drinkers had 0.18 and daily drinkers had 0.38. *Firmicutes* abundance decreased at the same time (Figure 10).

Bacteroides increased in mean relative abundance with an increase in alcohol consumption at the genus level: 0.10 (*Never*), 0.17 (*Rarely*), 0.17 (*Occasionally*), 0.16 (*Regularly*), and 0.39 (*Daily*). *Enterobacter* had a non-linear relationship; it was highest among occasional drinkers (0.10) and lowest among daily drinkers (0.03). The above trends in composition are shown visually by stacked bar charts and heatmaps (Figure 11).

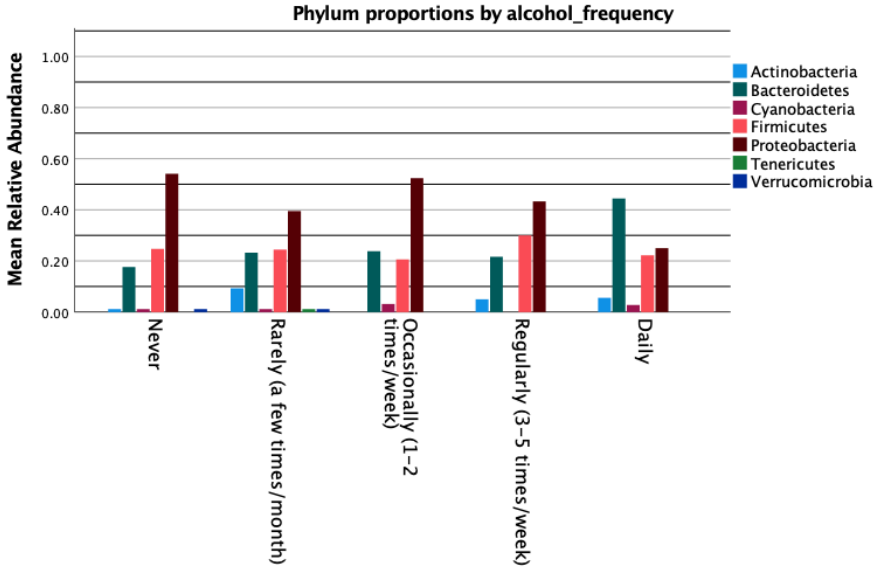

Figure 10. Phylum-level microbiota composition across alcohol drinking frequencies

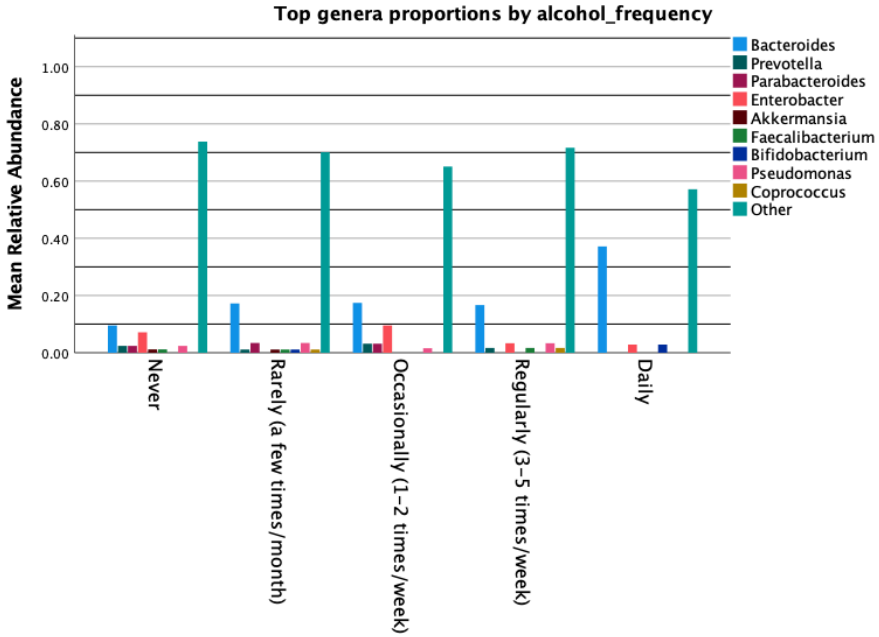

Figure 11. Genus-level microbiota composition across alcohol drinking frequencies

3.5. Differential abundance and correlation analysis

Kruskal-Wallis tests with FDR correction were used to identify genera with significantly different relative abundances at different levels of *sample_source* and *alcohol_frequency*. For *sample_source*, the most significantly different genera were *Streptococcus* (FDR = 3.99×10^{-5}), *Bacteroides* (FDR = 6.27×10^{-5}), *Enterobacter* (FDR = 0.002), *Corynebacterium* (FDR = 0.013), and *Rothia* (FDR = 0.040); therefore, both technical batch and geographic origin significantly affected the genus-level profiles (Figure 12). The top-ranked differentially abundant taxa for *alcohol_frequency* were *Bacteroides* (FDR = 0.091), *Enterobacter* (FDR = 0.381), *Erwinia*, *Corynebacterium*, *Ruminococcus* and *Dialister* (Figure 13). Although most of the FDR values were above the 0.05 threshold, the gradient observed for *Bacteroides* aligns with the multivariate importance ranking and PERMANOVA results, suggesting that alcohol consumption is a consistent modifier of the gut microbial community.

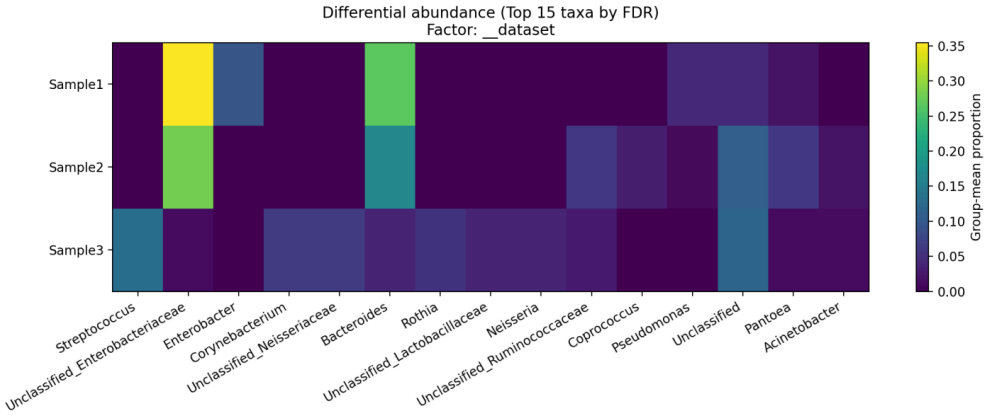

Figure 12. Heatmap of differentially abundant genera across sample sources

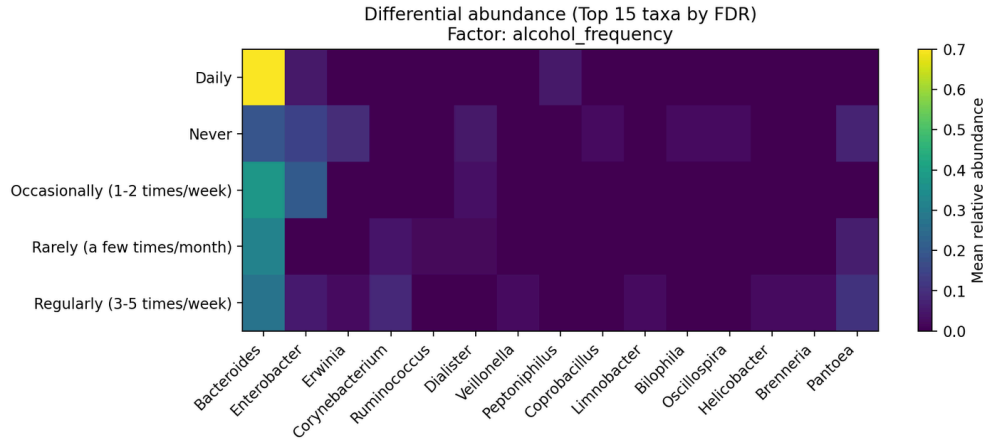

Figure 13. Heatmap of differentially abundant genera across alcohol drinking frequencies

Pearson correlation coefficients were calculated between each lifestyle factor and the relative abundance of individual bacterial genera to obtain an integrated, taxon-resolved view of the relationship among lifestyles and the microbiota; then, both these factors were hierarchically clustered. As shown in the resulting heatmap, the distribution of association strengths is relatively even. *Exercise_frequency* and *alcohol_frequency* showed the highest mean absolute correlations with genus-level abundances, followed by geographic coordinates (*longitude*, *latitude*) and *sleep_duration*. *antibiotic_history*, *sex* and *bmi_cat* showed relatively weak correlations with bacterial taxonomic profiles. Among the bacterial genera, some groups that are frequently found in the environment, such as the *Enterobacteriaceae* family and *Pseudomonas*, showed significant positive or negative correlations with specific lifestyle exposures; thus, taxon-specific ecological sensitivities were observed (Figure 14).

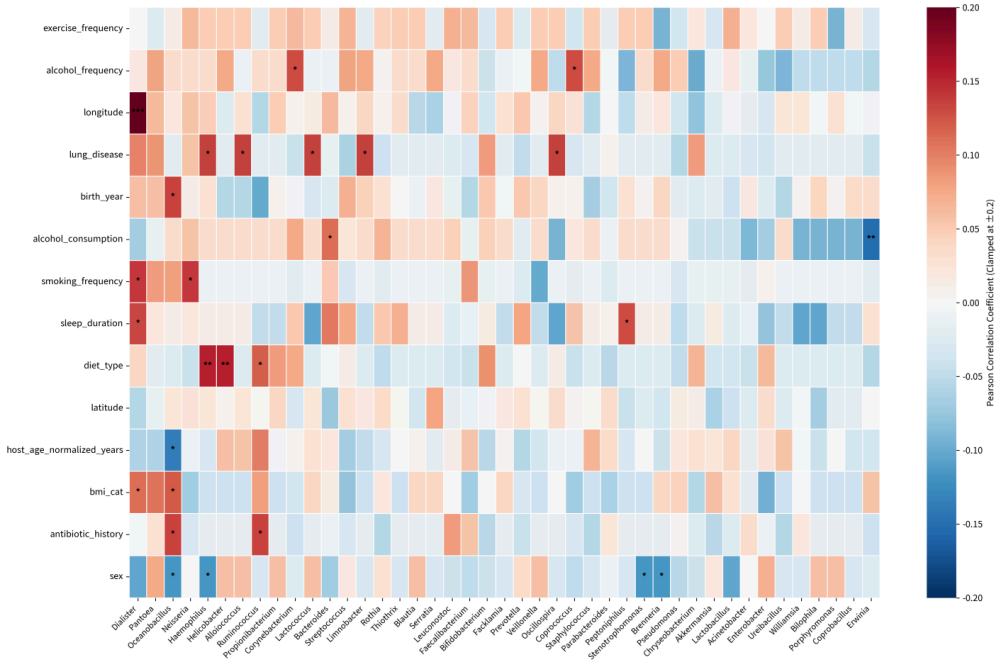

Figure 14. Pearson correlations between lifestyle factors and bacterial genera

3.6. Sensitivity analysis

Batch effects and robustness of lifestyle associations given the overwhelming influence of *dataset* across all analytical frameworks, stratified sensitivity analysis by *sample_source* was Conducted. Within each source stratum, the directionality of the association between *alcohol consumption frequency* and *exercise_frequency* with dominant genus classification remained the same as in the pooled analysis, but the statistical power was reduced due to a smaller sample size. Based on the above observation, although technical batch effects are the most significant source of variation in the aggregated microbiome data, reproducible biological signals related to modifiable lifestyle factors can also be detected; therefore, strict batch correction and standardised analytical protocols are required for multi-cohort studies.

4. Discussion

This paper provides a general assessment of the relationships between modifiable lifestyle factors and the abundance of the main genus of intestinal bacteria in 336 adults from the American Gut Project. Multinomial logistic regression was used, and technical batch effects were explicitly controlled for; among the lifestyle factors, *alcohol_frequency*, *exercise frequency*, and *sleep duration* were found to be the most independent predictors of dominant genus classification. The above results extend previous research that has mainly studied the effect of lifestyle in isolation and focused on continuous diversity indices; now, more attention is being paid to how all daily behaviours work together to influence the gut ecosystem.

Alcohol consumption was identified as a typical modifiable risk factor for changes in the gut microbiota in all analyses. Among all predictors in the multivariate model, *alcohol_frequency* was the highest-ranked modifiable factor for dominant genus assignment; PERMANOVA analysis showed that it was significantly associated with changes in the global community structure ($F = 2.03$, $p = 0.01$, FDR = 0.04). The increasing proportion of alcohol-associated *Bacteroides* in the different consumption groups (10% for non-drinkers and 39% for daily drinkers) is in line with the results of a 2025 meta-analysis that general ethanol exposure lowers microbial diversity and reduces beneficial commensals such as *Faecalibacterium* and *Akkermansia*, while expanding pro-inflammatory groups [23]. Species of *Bacteroides* are able to break down ethanol and may increase in number following prolonged exposure to high concentrations of alcohol in the gut, thus aggravating endotoxemia due to impaired intestinal permeability [12]. A systematic review of 50 clinical studies further demonstrated that gut dysbiosis is a frequent manifestation of alcohol use disorder [16], and a 2025 review elucidated the intricate connection between gut and brain dysfunction under alcohol-induced dysbiosis that promotes neuroinflammation [17]. The non-linear pattern of *Enterobacter*, which is highest among occasional drinkers, may be due to dose-dependent ecological changes or unadjusted confounding. Importantly, these compositional changes could be identified at the level of dominant genus classification; therefore, even this coarse categorical phenotype showed biologically significant variations in the gut ecosystem.

The second highest-risk factor for lifestyle was *exercise_frequency* in the multivariate model. *Rothia* was positively correlated with *exercise_frequency* (FDR = 0.039) in the differential abundance analysis, and it had a relatively high abundance among people who exercised 3-5 times a week. Although the specific function of *Rothia* in the gut ecosystem has not been determined, its enrichment in individuals who exercise regularly is in line with other research that links exercise to changes in gut microbial communities via alterations in intestinal transit, bile acid metabolism and mucosal immune regulation [24]. A 2024 meta-analysis of 25 studies further demonstrated that exercise alters the diversity of the gut microbiota [9]. However, *exercise_frequency* was not statistically significant in PERMANOVA; thus, it is suggested that the impact of

exercise frequency on the global community structure may be relatively small, or perhaps requires a larger sample size and more objective measures of physical activity to reveal such differences.

Sleep_duration ranked third in multivariate analysis and aligns with the rising recognition of bidirectional gut-brain axis communication in sleep regulation. Among the descriptive analyses, the group reporting 7-8 hours of sleep had a relatively high level of overall family diversity, while the group sleeping 5-6 hours had a low level. Modifications in the gut microbiome and its byproducts can affect the endocrine and nervous systems to alter neurotransmitter secretion and central nervous system function [25]. In 2025, further studies have shown that changes in the gut microbiota are connected to problems with sleep through complex neuroendocrine and immune pathways, and they may be treated [15]. *Sleep_duration* did not result in genus-level associations after FDR correction, and since objective sleep measures were unavailable, further studies employing objective sleep tracking are required.

Dietary Pattern ranked below alcohol, exercise and sleep in multivariate importance and did not reach statistical significance in PERMANOVA. The above results probably indicate that the coarse granularity of the dietary metadata in the American Gut Project (e.g., "*Omnivore*", "*Vegetarian*") does not include specific details on fibre intake and macronutrient composition. A deficiency has been noted in recent critical assessments of the diet-microbiota study field [26]. Research in the same group of people has shown that after the fact, dietary patterns based on food groups are more closely linked to changes in the gut microbiome than individual dietary characteristics [27]. Additionally, a dominant genus phenotype may conceal dietary alterations in subordinate taxa; furthermore, the length of exposure to a new diet is also required for microbial shifts to manifest [28].

The third is that technical batch effects are widely distributed. The dataset identifier was consistently the strongest predictor in all the analyses, surpassing the combined effect of all lifestyle variables and thus providing a recent reason for demanding strict batch correction [29, 30]. In the sensitivity analysis, the direction of association between alcohol and exercise frequency remained the same within the strata of sample sources, thus confirming that the observed lifestyle signal was not a technical artifact due to confounding.

The following are some defects. Cross-sectional design does not support causal inference; the cohort is mainly Western and may not be generalizable to other regions [31]; lifestyle data were self-reported; only the dominant genus phenotype was used, thus omitting information on sub-dominant taxa; and unmeasured confounding factors still exist, such as unknown factors inherent in microbiome research [32]. However, this study has the following advantages: it covers multiple areas of life under a single system, shows the magnitude of batch effects, and adds a dominant genus classification as an additional phenotype to traditional diversity indicators.

The above results show that alcohol consumption, exercise and sleep need to be improved first in the upcoming intervention studies to promote good gut microbiota.

5. Conclusion

In a multi-factor analysis of 336 adults from the American Gut Project, after controlling for technical batch effects, modifiable lifestyle factors such as the frequency of alcohol consumption, exercise frequency and sleep duration were independently associated with dominant intestinal bacterial genus classification. Only alcohol consumption was associated with global community structure at the level of lifestyle factors, and *Bacteroides* showed an increase in relative abundance with higher levels of consumption. Exercise Frequency was significantly correlated with *Rothia* abundance. Sleep duration did not show genus-level associations after multiple comparison correction and was ranked third in multivariate importance. The dataset-of-origin

identifier remained the strongest predictor of dominant genus assignment, and therefore, explicit batch correction was required in the aggregated microbiome analysis. Based on the above results, we have obtained an all-encompassing view of how people's living habits affect the gut ecosystem and have identified high-priority targets for subsequent intervention research to improve the health of gut microbiota.

References

[1] Krautkramer, K. A., Fan, J., & Bäckhed, F. (2021). Gut microbial metabolites as multi-kingdom intermediates. *Nature Reviews Microbiology, 19*(2), 77-94.

[2] Fan, Y., & Pedersen, O. (2021). Gut microbes and metabolic health in humans. *Nature Reviews Microbiology, 19*(1), 55-71.

[3] Qin, Z., Feng, X., Hu, Y., & Su, S. (2025). The human gut microbiota is associated with host lifestyle: An overview. *Frontiers in Microbiology, 16*, 1549160.

[4] Rothschild, D., Weissbrod, O., & Barkan, E. (2018). Environment is more influential than host genetics in determining the human gut microbiota. *Nature, 555*(7695), 210-215.

[5] Sanz, Y. (2025). The gut microbiome links diet and health. *Nature Reviews Gastroenterology & Hepatology, 22*(8), 534-555.

[6] Hindle, V. K. (2025). Microbiota-oriented dietotherapy for health support: A systematic review. *The Journal of Nutrition, 155*(2), 381-401.

[7] Yu, J., Wu, Y., Zhu, Z., & Lu, H. (2025). Diet's impact on gut microbiota for the primary and secondary prevention of cardiovascular diseases: A systematic review. *Nutrition Journal, 24*(1), 17.

[8] Jiang, H., Wang, Y., & Liu, X. (2025). Exercise-induced modulation of gut microbiota in obese and type 2 diabetic individuals: A systematic review and meta-analysis. *Frontiers in Microbiology, 16*, 1671975.

[9] Min, L., Ablip, A., Wang, R., Luciana, T., Wei, M., & Ma, X. (2024). Effects of exercise on the gut microbiota in adults: A systematic review and meta-analysis. *Nutrients, 16*(7), 1070.

[10] Kim, Y. I., Choi, W., Seo, M., Ka, S., & Park, J. (2025). Effect of exercise on the human gut microbiota in overweight and obese individuals: A systematic review and meta-analysis of randomized controlled trials. *Physical Activity and Nutrition, 29*(2), 49–61.

[11] Matenchuk, B. A., Mandhane, P. J., & Kozyrskyj, A. L. (2020). Sleep, circadian rhythm and the gut microbiome. *Sleep Medicine Reviews, 53*, 101340.

[12] Sosnowski, K., & Przybyłkowski, A. (2024). Ethanol-induced alterations in the gut microbiome disrupt intestinal homeostasis: A review. *Gut Microbes, 16*(1), 2393272.

[13] Szajewska, H., Scott, K. P., de Meij, T., Forslund-Startceva, S. K., Knight, R., & Koren, O. (2025). Antibiotic-perturbed microbiota and the role of probiotics. *Nat Rev Gastroenterol Hepatol, 22*(3), 155–172.

[14] Vignesh, S. D., Vijayakumar, T. M., & Siddhu, N. S. (2024). Effect of food intake and sleep disturbance on gut microbiota. *Cureus, 16*(10), e70846.

[15] Lu, L. (2025). Brain-gut-microbiota interactions in sleep disorders. *Brain Med, 1*(2), 100023.

[16] Grodin, E. N. (2024). Gut microbiota in alcohol use disorder and alcohol-related liver disease: A systematic review of clinical studies. *Alcohol Clin Exp Res (Hoboken), 48*(8), 1466–1468.

[17] Skryabin, V. (2025). Gut microbiota and alcohol use disorder: New directions for treatment and recovery. *BJPsych Bull, 50*(2), 175–182.

[18] Babouee Flury, B., Andrey, D., & Kohler, P. (2024). Antibiotics' collateral effects on the gut microbiota in the selection of ESKAPE pathogens. *CMI Commun, 1*(2), 100012.

[19] Warren, M., O'Connor, C., & Lee, J. E. (2024). Dangers of the chronic stress response in the context of the microbiota-gut-immune-brain axis and mental health: A narrative review. *Frontiers in Immunology, 15*, 1365871.

[20] Gaskell, S. K. (2026). Gut microbiota and exercise: A systematic review of interventions and evidence limitations. *International Journal of Sports Medicine, 47*(4), 255–286. https://doi.org/10.1055/a-2705-2210
[21] Scanu, M., Del Chierico, F., & Marsiglia, R. (2024). Correction of batch effect in gut microbiota profiling of ASD cohorts from different geographical origins. *Biomedicines, 12*(10), 2350.
[22] Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology, 26*(1), 32–46.
[23] Alexandrescu, L., Tofolean, I. T., & Tofolean, D. E. (2025). Ethanol-induced dysbiosis and systemic effects: A meta-analysis of human and animal studies. *Microorganisms, 13*(9), 2000.
[24] Quaresma, M. V. L. D. S. (2024). The interplay of the gut microbiome and exercise in athletes. *Current Opinion in Clinical Nutrition and Metabolism, 27*(5), 428–434.
[25] Li, L. (2024). Gut microbiota: Candidates for a new strategy in treating sleep disorders. *Critical Reviews in Food Science and Nutrition, 64*(25), 9150–9167.
[26] Lotankar, M., Houttu, N., Mokkala, K., & Laitinen, K. (2024). Diet-gut microbiota relationships: Critical appraisal of evidence from metagenomic studies. *Nutrition Reviews, 82*(10), 1388–1410.
[27] Cotillard, A. (2022). A posteriori dietary patterns explain variations in the gut microbiome better than individual markers in the American Gut Project. *American Journal of Clinical Nutrition, 116*(4), 1015–1026.
[28] Leeming, E. R., Johnson, A. J., Spector, T. D., & Le Roy, C. I. (2019). Effect of diet on gut microbiota: Rethinking intervention duration. *Nutrients, 11*(12), 2862.
[29] Wang, B., & Luan, Y. (2024). Evaluation of normalisation methods for predicting quantitative phenotypes in metagenomic data analysis. *Frontiers in Genetics, 15*, 1369628.
[30] Gibbons, S. M., Duvallet, C., & Alm, E. J. (2018). Correcting for batch effects in case-control microbiome studies. *PLoS Computational Biology, 14*(4), e1006102.
[31] He, Y., Wu, W., & Zheng, H. M. (2018). Regional differences limit the application of healthy gut microbiome reference ranges and disease models. *Nature Medicine, 24*(10), 1532–1535.
[32] Thomas, A. M., & Segata, N. (2019). Multiple levels of the unknown in microbiome research. *BMC Biology, 17*(1), 48.