

Analysis of Gut Microbiome Structure Based on GMPR+Spectrum

Xin Xiong, Yuyan Ren and Jianfeng He *

Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; xiongxin@kust.edu.cn (X.X.); renyuyan@stc.kust.edu.cn (Y.R.)

* Correspondence: jfenghe@foxmail.com

Abstract: The gut microbiome is related to many major human diseases, and it is of great significance to study the structure of the gut microbiome under different conditions. Multivariate statistics or pattern recognition methods were often used to identify different structural patterns in gut microbiome data. However, these methods have some limitations. Minimal hepatic encephalopathy (MHE) datasets were taken as an example. Due to the physical lack or insufficient sampling of the gut microbiome in the sequencing process, the microbiome data contains many zeros. Therefore, the geometric mean of pairwise ratios (GMPR) was used to normalize gut microbiome data, then Spectrum was used to analyze the structure of the gut microbiome, and lastly, the structure of core microflora was compared with Network analysis. GMPR calculates the Intraclass correlation coefficient (ICC), whose reproducibility was significantly better than other normalization methods. In addition, running-time, Normalized Mutual Information (NMI), Davies-Boulding Index (DBI), and Calinski-Harabasz index (CH) of GMPR+Spectrum were far superior to other clustering algorithms such as M3C, iClusterPlus. GMPR+Spectrum can not only perform better but also effectively identify the structural differences of intestinal microbiota in different patients and excavate the unique critical bacteria such as *Akkermansia*, and *Lactobacillus* in MHE patients, which may provide a new reference for the study of the gut microbiome in disease.

Keywords: spectrum; normalization; gut microbiome; minimal hepatic encephalopathy; hepatic encephalopathy



Citation: Xiong, X.; Ren, Y.; He, J.

Analysis of Gut Microbiome Structure Based on GMPR+Spectrum. *Appl. Sci.* **2022**, *12*, 5895. <https://doi.org/10.3390/app12125895>

Academic Editor: Giovanna Traina

Received: 17 April 2022

Accepted: 6 June 2022

Published: 9 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Gut microbiome is associated with many major human diseases, such as obesity, diabetes, cirrhosis, autism, allergies, inflammatory bowel diseases, cardiovascular diseases, multiple cancers, and depression. Therefore, the gut microbiome may become a recent target for interventional therapies and play an essential role in diagnosing, analyzing, and treating these major diseases [1]. Microbiome studies are extensively used to analyze the microbial communities' composition and diversity of the flora. They are used to study one of the fundamental questions of microbial ecology: how many taxa or OTUs (operational taxonomic units) exist? Usually, multivariate statistical or pattern recognition methods are used to identify different structural patterns in microbial data, such as principal component analysis (PCA) [2–4], principal coordinate analysis (PCoA) [5–7], partitioning around medoid (PAM) clustering [8,9], etc. However, this standard multivariate technique does not applicable to highly diverse microbial data [10]. On the one hand, microbial data with high diversity tend to have sparse data sets, and on the other hand, most taxonomic units occur in only a few samples with low abundance. In addition, microbial genetic samples differ in reading length: small samples are inherently noisier than large samples.

Microbiome sequencing data contains many zeros due to physical deficiencies or under-sampling during the microbiome sequencing process. The complex processes involved in the sequencing process cause the depth of sequencing to vary with the sample, sometimes varying by several orders of magnitude. Therefore, the sequenced intestinal flora data are characterized by large data volume, a large number of OTUs, and sparse

distribution [11]. Normalization is crucial as it aims to correct or reduce bias caused by the sequencing depth and is an essential pre-processing step before any downstream statistical analysis of high-throughput sequencing experiments [12,13]. Several normalization methods are commonly used for sequencing data, especially for RNA-Seq data [12,13]. Other popular methods for normalization of microbiome data, besides the size-factor-based methods, such as the geometric mean of pairwise ratios (GMPR), Trimmed mean of M values (TMM), and Relative Log Expression (RLE), are all methods of sparsification. The above methods have disadvantages and advantages in specific applications. Sparsity, for example, leads to discarding most reads and may not be optimal from an information point of view. However, it is still widely used for microbiome data analysis, especially for α and β diversity analysis. In addition, it suffers from a significant power loss due to the discarding of a large number of reading operations [14]. Instead, size factors can be included as offsets in a count-based parametric model to address the problem of uneven sequencing depth [15]. In comparison, GMPR consistently showed the best level of variability in reducing OTUs at different prevalence levels and increased reproducibility among replicates normalized to the abundance of taxonomic units [16]. In addition, GMPR normalization has been studied for distance-based (weighted) statistical methods such as ranking, clustering, and PERMANOVA based on GMPR-normalized abundance data [17,18].

Clustering analysis plays an essential role in data mining and has many applications in image processing, data analysis, market research, pattern recognition and other fields [19–21]. In recent years, spectral clustering has become one of the widely used clustering algorithms [22]. Compared with traditional clustering methods, it is more adaptable to data distribution, especially for data sets with different densities, random complex shapes, and unstable sizes, and the computational effort in clustering is much smaller and less complicated to implement. It is also much less computationally intensive, not very complex to implement, and has higher performance. In contrast, Spectrum [23] used in this paper enhances the similarity between points sharing nearest neighbors using a self-adjusting density-aware kernel. The data integration and diffusion process through tensor product maps reduce noise, reveals the underlying structure, and automatically finds the optimal number of clusters K by analyzing the feature vector distribution. The algorithm can find clustering of arbitrary data shapes, noisy data in the dataset can be handled efficiently, the number of clusters K of Gaussian and non-Gaussian structures can be found automatically, the running time is short, and good clustering results can be shown for large data sets.

In this paper, we first normalized the gut flora data using GMPR [16] algorithm and then analyzed them using Spectrum. The gut microbial datasets of patients with minimal hepatic encephalopathy, hepatic encephalopathy, and healthy controls were used as examples. Minimal hepatic encephalopathy (MHE) is a very insidious stage in the pathogenesis of hepatic encephalopathy (HE), and studies [24–29] have shown that the prevalence of MHE reached 20–80% in patients with cirrhosis. MHE is a common complication of liver disease, typically characterized by altered neurocognitive function [30–33], with an unnoticeable onset and no obvious clinical manifestations of HE, and the cognitive dysfunction caused by MHE can consume many medical resources and impose a great financial burden on patients and their families. Due to the high prevalence of MHE, its harmful effects, and the complexity of clinical diagnosis, more clinical attention has been paid to early screening and diagnosis of MHE. In addition, more and more studies [34,35] have shown that dysbiosis of the gut microflora was associated with MHE and the occurrence of HE. Since the traditional screening and diagnostic methods used in clinical practice are time-consuming and subject to human factors, it is essential to identify structural changes in the gut flora data of MHE and HE patients.

2. Materials and Methods

2.1. Materials

The datasets used in this paper were obtained from 77 samples collected from the Department of Gastroenterology of the First People's Hospital of Yunnan Province, including

26 patients with minimal hepatic encephalopathy (Abbreviated as M), 25 patients with hepatic encephalopathy (Abbreviated as H), and 26 normal healthy controls (Abbreviated as N). The data collection process was as follows: (1) Sampling of fresh stool from samples; (2) Storage in liquid nitrogen within 2 h; (3) Storage in $-80\text{ }^{\circ}\text{C}$ refrigerator; (4) Extraction of fecal microbial DNA by kit method [36], and completion of 16SrRNA high-throughput sequencing according to standard operating instructions [37]. (5) After splicing the original sequences, performed quality control, selected representative sequences (OTU), clustered them, and then performed species annotation. (6) The OTUs count table after sequencing was obtained. The data collected in this study were approved by the ethics committee of the First People’s Hospital of Yunnan Province, and all subjects signed an informed consent form. As shown in Table A1 that a partial table of OTUs counts after sequencing, where rows (OTU_0, OTU_1, OTU_2, . . .) represent OTUs counts and columns (H1, H2, H3, . . .) represent sample ID numbers of patients with hepatic encephalopathy. The data used in the experiments were absolute abundance data.

2.2. Methods

The flow chart of the flora data processing in this study is shown in Figure 1. The experimental data were first normalized by GMPR, then clustering using the Spectrum algorithms, and compared with Spectrum without GMPR, M3C [38], and iClusterPlus [39] in terms of performance metrics including Normalized Mutual Information (NMI), Davies-Boulding Index (DBI), Calinski-Harabasz Index (CH) and algorithm running time and finally compared with the network analysis method for core flora.

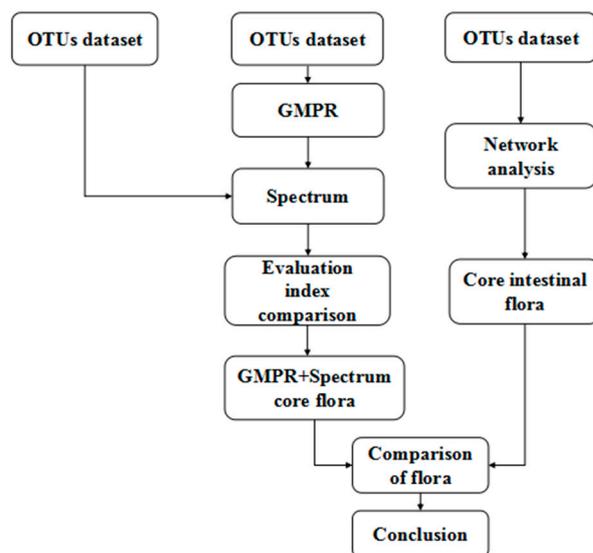


Figure 1. Schematic diagram of flora data processing in the paper.

2.2.1. Geometric Mean of Paired Ratios (GMPR)

GMPR [16] is a normalized method specially used to solve the problem of zero expansion of data. In principle, it can be applied to any sequencing data. It is mainly to solve some situations, such as many zeros in the data and different sequencing depths due to physical lack or insufficient microbial sampling.

The OTUs count table in the paper is the absolute abundance, which contains 77 samples, 1442 OTUs. GMPR is used to calculate the size factor of a given sample, and the size factor could estimate the library size of a given sample. The formula is as follows:

The first step is to calculate r_{ij} ,

$$r_{ij} = k \in \{1, \dots, q\} \left| \begin{matrix} \text{Median} \\ c_{ki} \cdot c_{kj} \neq 0 \end{matrix} \left\{ \frac{c_{ki}}{c_{kj}} \right\} \right. \quad (1)$$

where k is the number of OTUs, r_{ij} is the median count ratio representing the non-zero counts between sample i and sample j , and C_{ki} , C_{kj} are expressed as the abundance data of the k th OTU in the sample i and sample j .

Then calculate the size factor s_i of a given sample i ,

$$s_i = \left(\prod_{j=1}^n r_{ij} \right)^{1/n}, i = 1, \dots, n \tag{2}$$

In short, the basic step of GMPR was to first compare pairs of samples in the OTU count table, and then combined the paired comparison results to obtain the final estimated value.

2.2.2. Other Normalization Methods

Two popular normalization methods for RNA-Seq data include trimmed mean of M values (TMM) [40], and relative log expression normalization (RLE) [41]. TMM method selects a reference sample first, and all other samples are compared to the reference sample. The log ratios' trimmed (weighted) mean is then calculated as the TMM size factor (log scale). RLE method calculates the geometric means of all features as a reference, and all samples are compared to the reference to produce ratios (fold changes) for all features. The median ratio is then taken to be the RLE size factor.

2.2.3. Spectrum Algorithm

Spectrum [23] is a new spectral clustering method, its idea is to view the data analysis problem as an optimal partitioning problem of the graph, where all OTUs are viewed as vertices in the space, and the vertices are connected with edges with weights. The edges with weights can be regarded as the similarity in OTUs. The key to this algorithm is that the self-adjusting density-aware kernel is employed to construct the similarity matrix, with the advantage that the similarity between the nearest neighbors can be further enhanced, while it can reduce noise. Spectrum can find the optimal number of clusters (K) involving the distribution of feature vectors, regardless of Gaussian or non-Gaussian structure [23].

The similarity matrix A^* is computed using the adaptive density-aware kernel in the Spectrum algorithm. Starting with A^* , the Ng spectral clustering method is used in Spectrum. At the same time, the number of clusters is estimated using an eigenvalue heuristic. Finally, the eigenvector matrix is clustered using Gaussian mixture modeling (GMM) to obtain the final output, i.e., a delineation of the feature clusters represented by the OTUs.

1. The adaptive density-aware kernel is first used in Spectrum algorithm to compute the similarity matrix between different OTUs.

The adaptive density perception kernel is:

$$A_{ij} = \exp \left(\frac{-d^2(s_i, s_j)}{\sigma_i \sigma_j (CNN(s_i, s_j) + 1)} \right) \tag{3}$$

where $d(s_i, s_j)$ represents the Euclidean distance between point s_i and s_j , σ_i and σ_j are the local scaling parameter, $CNN(s_i, s_j)$ is the number of points in the connection area of the ϵ neighborhood around the point s_i and s_j , and the ϵ -neighborhood of the point represents the radius of the sphere around the point.

2. The diagonal matrix D is obtained from A^* , the diagonal matrix where (i,i) element is the i th row of the sum of A^* , and the normalized Laplacian matrix L is constructed using D .

$$L = D^{-1/2} A^* D^{-1/2} \tag{4}$$

3. Decompose eigenvalues of L and extract its eigenvectors X_1, X_2, \dots, X_{N+1} and eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_{N+1}$.

- Determine the difference in eigenvalues, start with the second eigenvalue, i.e., $n = 2$, and choose the optimal k , the difference in eigenvalues is maximized and denoted by k^* .

$$k^* = \underset{n}{\operatorname{argmax}} (\lambda_n - \lambda_{n+1}) \quad (5)$$

- Obtain the largest eigenvectors K^* and then form the matrix (each eigenvector is arranged in columns to form n vectors in a k^* -dimensional space), i.e., $X = [x_1, x_2, \dots, x_{k^*}] \in R_{N+k^*}$.
- Form the matrix Y from X by renormalizing each of X 's rows to have unit length.

$$Y_{ij} = \frac{X_{ij}}{\left(\sum_j x_{ij}^2\right)^{1/2}} \quad (6)$$

- Finally, each row of Y is considered as an OTU feature s_i , and finally all OTUs are clustered into k^* clusters using GMM. The obtained class labels are the class labels of the original OTUs.

2.2.4. Monte Carlo Reference-Based Consensus Clustering (M3C)

Genome-wide expression data are stratified using clustering algorithms to stratify patients for precision medicine. The Monti consensus clustering algorithm [42], a widely used method, determines the number of clusters (K) by the stability selection principle. The algorithm works by resampling and clustering the data in each cluster and calculating an $N \times N$ consensus matrix. Each element represents the proportion of time that two samples are clustered together. A fully stable matrix consisting entirely of zeros and ones represents whether all sample pairs are clustered or not in the resampling iteration. The next step is to compare the stability of these consensus matrices to determine K . The fuzzy clustering ratio (PAC) score [43] is used to evaluate the stability of the consensus matrix for each K . However, it is biased towards larger values of K . In contrast another widely used delta K metric is more subjective in finding K as it relies on finding an elbow point and is not as good as the PAC score. Monte Carlo Reference-based Consensus Clustering (M3C) [38] addresses these issues by comparing the true stability scores with the expected scores under a stochastic model. M3C uses Monte Carlo simulation to generate a distribution of stability scores along with a range of K by comparing it with actual stability scores to determine the optimal K and reject the null hypothesis $K = 1$.

2.2.5. IClusterPlus

IClusterPlus was developed for comprehensive cluster analysis of multi-type genomic data [39]. Multi-type genomic data such as array comparative gene hybridization (aCGH), gene expression microarray, RNA-seq, DNA-seq, etc. iClusterPlus samples a range of lambda values from the parameter space based on a unified design to search for the best model [44]. The number of points to sample (n .lambda) depends on the number of data types. If the number of clusters in the sample is known, the corresponding k (the number of latent variables) can be directly selected for cluster analysis. If the number of clusters is not known in advance, k can be tested from 1 to N (a reasonable number of clusters). For each k , Bayesian Informative Features (BIC) is used to select the best sparse model with the best combination of penalty parameters. To choose the best k , by calculating the deviation rate, i.e., \log -likelihood (fit)- \log -likelihood (null model) divided by \log -likelihood (full model)- \log -likelihood (null model) ratio. Deviation rate can be interpreted as EV percentage. Choose k where the percentage EV curve plateaus the optimal number of clusters.

2.2.6. Network Analysis

Microbial networks are an increasingly popular tool for studying the structure of microbial communities, as they integrate multiple types of information and may represent system-level behavior. The analysis of microbial networks allows one to predict pivotal

species and inter-species interactions. In recent years, various network methods have been successfully applied in different biological contexts. Among them, the correlation-based association network approach is the most commonly used method for analyzing microbial interaction networks due to the simplicity and robustness of the computational process. Network analysis in some disciplines, especially medical-related ones, provides more options for further data mining and analysis. Therefore, we used network analysis methods to validate the identified flora's reliability further.

The network analysis method is based on the concept of a network diagram in mathematics, and the microbial interaction network is constructed based on the Pearson correlation between all OTU species, and different correlation coefficients represent the difference in the relationship in different OTUs. Meanwhile, each network node corresponds to each OTU, i.e., colony species, and the edges between different species are determined by the pairwise Pearson correlations between species, i.e., the significant correlation between a certain bacterium and another bacterium. Ju et al. [45] ranked all the nodes in the network according to the degree from highest to lowest and selected the top ten nodes as the core nodes. Nodes, where the core module represent the core species in the global network. Therefore, the top ten OTU nodes corresponding to the degree of connectivity (Z_i) within the module were selected as our core nodes in this study. These nodes represent the key species that may play an essential role in maintaining the structural stability of the microbial community, i.e., the core gut flora.

2.2.7. Evaluation Index of Normalization Algorithm

Intraclass correlation coefficient (ICC) [46] is often used to evaluate the reproducibility or consistency of different measurement methods or raters for the same quantitative measurement results.

ICC is defined as:

$$ICC = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2} \quad (7)$$

where σ_b^2 represents the data variability between different normalization methods for the same sample type and σ_e^2 represents the variability between different sample types. ICC is calculated for the four types of sample data ("all samples", "M", "H" and "N"). The ICC was estimated by the R package "ICC", and its value is close to 1 indicates the better reproducibility of the method.

2.2.8. Evaluation Index of Clustering Algorithm

In this paper, NMI, DBI, CH, and running time are employed to evaluate the performance of the clustering algorithm. These metrics are defined and formulated as follows:

1. Normalized Mutual information (NMI)

NMI [47], which determines clustering quality, is a common method. The more significant NMI value means better performance. The joint distribution of random variables X and Y is $p(x,y)$, and the edge distribution is $p(x)$ and $p(y)$, respectively. The mutual information $I(X,Y)$ is the relative entropy of the joint distribution $p(x,y)$ and the product distribution $p(x)p(y)$:

$$I(X, Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (8)$$

$$H(X) = \sum_{i=1}^n p(x_i) I(x_i) = \sum_{i=1}^n p(x_i) \log_b \frac{1}{p(x_i)} = -\sum_{i=1}^n p(x_i) \log_b p(x_i) \quad (9)$$

$$NMI(X, Y) = 2R = 2 \frac{I(X, Y)}{H(X) + H(Y)} \quad (10)$$

2. Davies-Boulding Index (DBI)

DBI, also known as the classification appropriateness index [48], is the maximum value of the sum of the average distance $avg(C)$ between the samples of each of two clusters C_i, C_j divided by the distance between the centroids of the two clusters. The larger the inter-class distance, the better the clustering effect.

$$avg(C) = \frac{2}{|C|(|C| - 1)} \sum_{1 < i < j < |C|} dist(x_i, x_j) \quad (11)$$

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} \left(\frac{avg(C_i) + avg(C_j)}{dist(u_i, u_j)} \right) \quad (12)$$

where $avg(C)$ means the average distance of cluster class C , $|C|$ means the number of cluster classes C , and $dist(x_i, x_j)$ is the distance between two samples x_i, x_j , and u_i, u_j are the center of the cluster class C_i, C_j , respectively.

3. Calinski-Harabasz index (CH)

The CH index is the ratio of inter-cluster distance to intra-cluster distance [49]. The larger the value $CH(K)$, the better the clustering effect. The formula is as follows:

$$CH(K) = \frac{tr(B)/(K - 1)}{tr(W)/(N - K)} \quad (13)$$

where $tr(B) = \sum_{j=i}^k \|z_i - z\|^2$ represents the trace of the inter-cluster distance difference matrix, $tr(W) = \sum_{j=i}^k \sum_{x^i \in k} \|x_i - z_i\|^2$ represents the trace of the intra-cluster departure matrix, where z is the mean of the whole data set, z_j is the mean of the j th cluster c_j , N represents the number of clusters, and K is the current class.

3. Results

3.1. Reproducibility of GMPR

The normalization methods include GMPR, TMM, TMM+ (add a pseudocount for TMM) [50], RLE, RLE+ (add a pseudocount for RLE) [50] were employed to preprocess four different types of data of "all samples", "H", "N" and "M". It has been seen from the Figure 2 that in different normalization methods, the ICC of "all samples" is larger than that of "H", "N", "M". It shows that all samples achieve higher reproducibility in all the applied normalization methods. All samples obtained a larger sample size across all the sample styles, showing that reproducibility decreases as the number of samples decreases. The ICC of GMPR is higher than other methods in all normalization methods under any sample type. This indicates that the GMPR method is more robust and reproducibility than other normalization methods.

3.2. Cluster Number

To verify the performance of the used method, all samples is subjected to GMPR+Spectrum clustering. Since the algorithm performs an eigendecomposition of the constructed Laplacian matrix, solves for the eigenvectors and eigenvalues, and maximizes the difference of the eigenvalues (corresponding to the difference between the eigenvalues of two neighboring eigenvectors, i.e., the difference of the eigenvalues) [23]. Therefore, the optimal number of clusters for all samples represented in Figure 3 is 8.

GMPR+Spectrum clustering is performed for M, H, and N groups further to analyze the M, H, and N groups. As shown in Figure 4, the optimal number of clusters for the chronic cirrhotic patients is 2. The clustering of the remaining two groups is similar to Figure 4. Therefore, we can acquire the optimal number of clusters for the three groups.

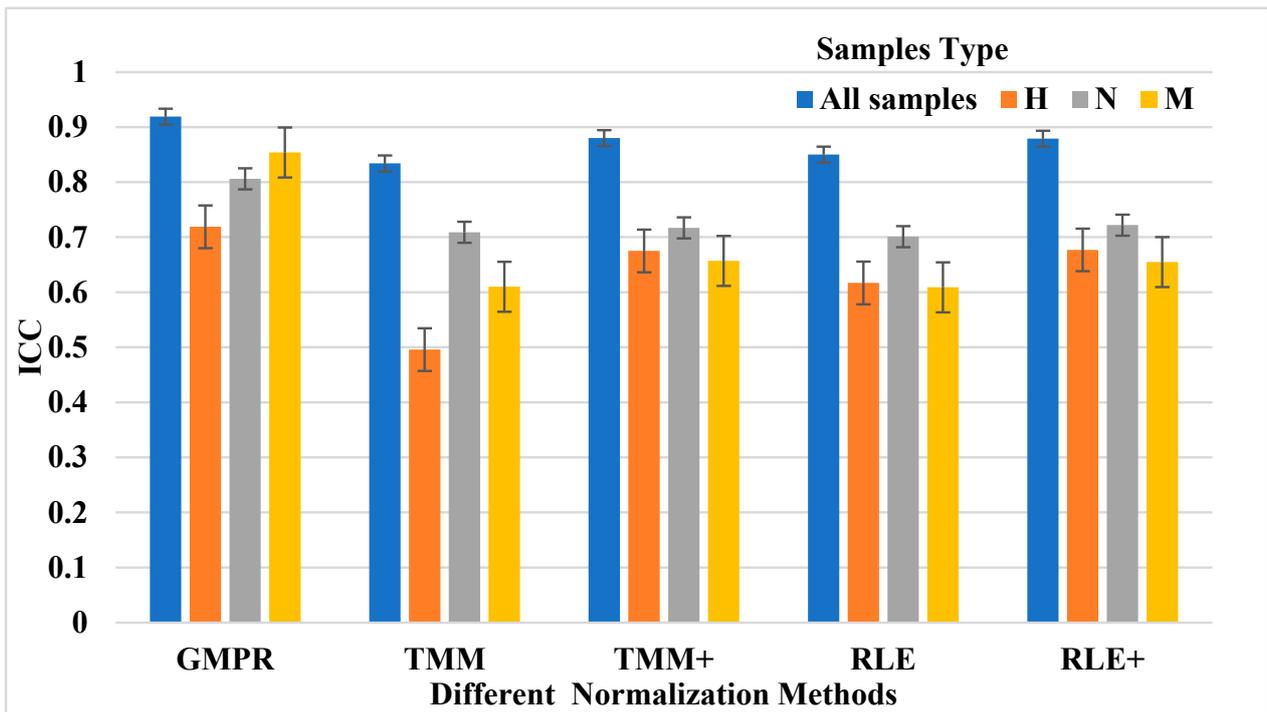


Figure 2. ICC of four samples under different normalization methods. The horizontal axis represents different normalization methods (Trimmed Mean of M values (TMM), add a pseudocount for TMM (TMM+), Relative Log Expression normalization (RLE), add a pseudocount for RLE (RLE+), and the vertical axis represents the ICC value.

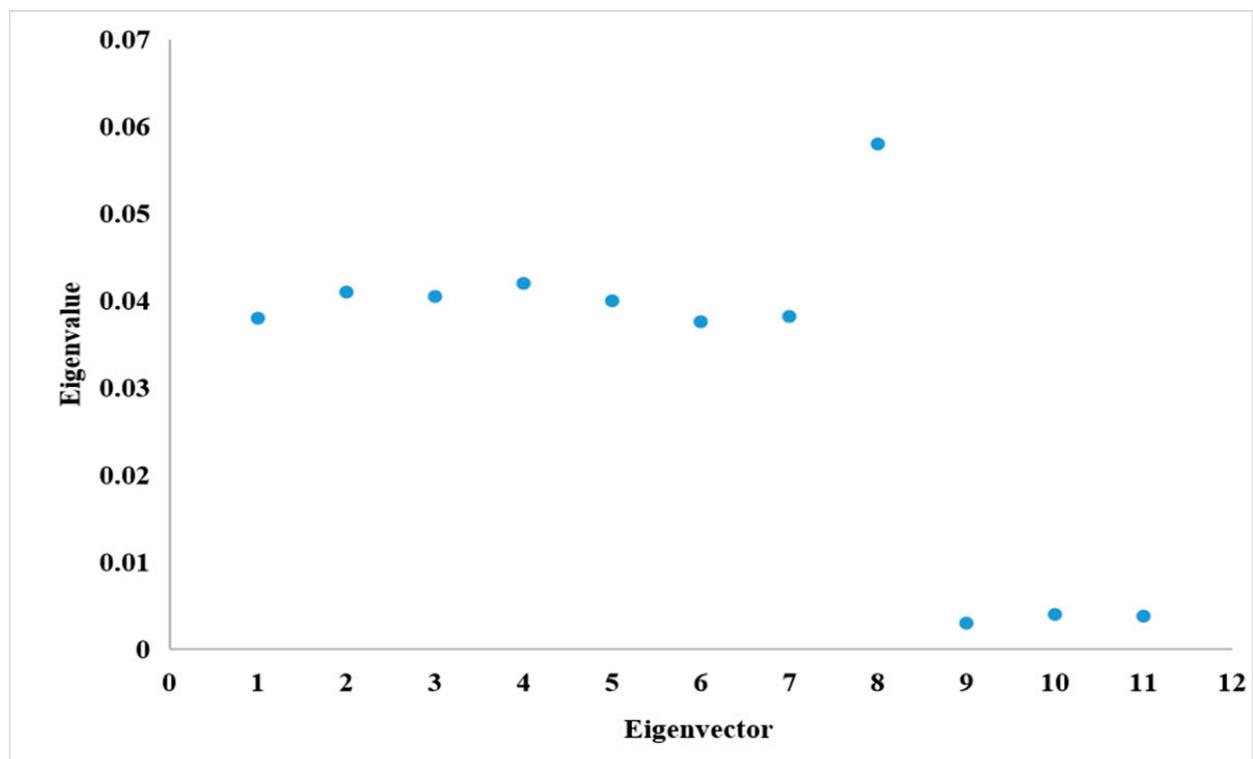


Figure 3. The optimal number of clusters corresponding to the eigenvectors of the Laplacian operator in all samples. The horizontal axis represents the eigenvectors, and the vertical axis represents the eigenvalues.

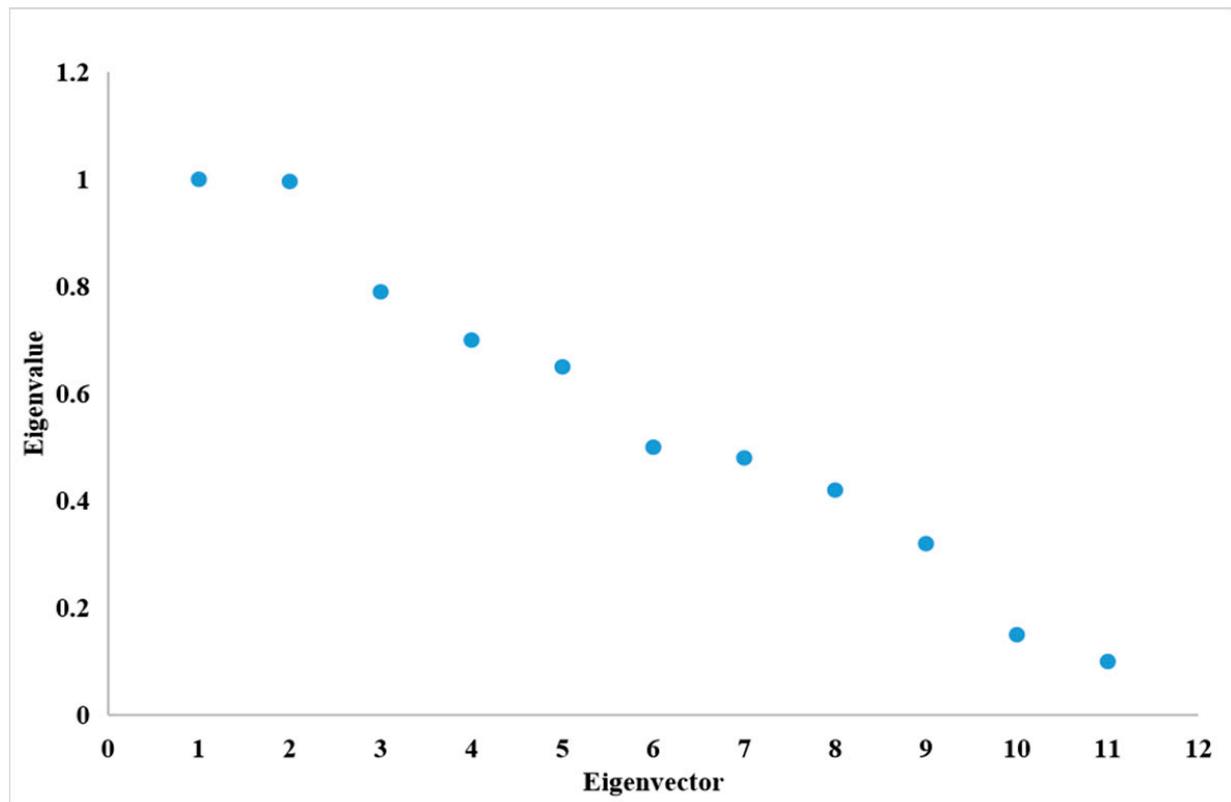


Figure 4. The optimal number of clusters corresponding to the eigenvectors of the Laplacian operator in hepatic encephalopathy. The horizontal axis represents the eigenvectors, and the vertical axis represents the eigenvalues.

3.3. Clustering Evaluation Indicators

GMPR+Spectrum classified all samples' data for 1442 OTUs into 8 classes with NMI of 0.3641, DBI of 4.2359, CH of 24.4724, and running time of 26.75 s. The Spectrum without GMPR divided these data into 3 classes, and all metrics except DBI are lower than the GMPR+Spectrum. In addition, as shown in Table 1, the performance of M3C and iClusterPlus are inferior to GMPR+Spectrum. As performances of N, H, and M, the clustering evaluation indicators are shown in Table A2.

Table 1. Clustering evaluation indicators of four different algorithms in all samples.

| Index | GMPR+Spectrum | Spectrum | M3C | iClusterPlus |
|-----------------|---------------|----------|---------|--------------|
| NMI | 0.3641 | 0.1932 | 0.0047 | 0.2623 |
| DBI | 4.2359 | 2.7343 | 3.2742 | 7.4851 |
| CH | 24.4724 | 14.4933 | 1.0000 | 1.0157 |
| Runtimes/second | 26.75 | 36.85 | 3096.19 | 117.31 |
| Cluster number | 8 | 3 | 4 | 3 |

3.4. Core Microflora by GMPR+Spectrum (Genus)

All samples dataset was clustered into 8 classes using GMPR+Spectrum. The OTUs of Cluster1 contain 24 different genera, the OTUs of Cluster2 contain 31 different genera, the OTUs of Cluster3 contain 54 different genera, the OTUs of Cluster4 contain 38 different genera, the OTUs of Cluster5 contain 25 different genera, the OTUs of Cluster6 contain 21 different genera, the OTUs of Cluster7 contain 18 different genera, the OTUs of Cluster8 contain 30 different genera, and the detailed bacteria contain in each cluster can be found in Table A3.

In addition, M, H, and N groups were clustered into 2 classes by GMPR+Spectrum, and the core OTUs in each category were identified according to the score value in the algorithm. The score value represents the proportion of the variance of a certain OTU to the total variance, which is actually the proportion of a certain feature value to the sum of all feature values [23]. The larger the score value, the larger the contribution rate, indicating the stronger information of the original variables contained in that OTU. Therefore, the size of the score value is used as a measure to determine whether a certain OTU is a core colony.

The score values were calculated in OTUs, including cluster1 and 2 for M, H, and N groups. Since many bacteria were unlabeled and there were many duplicate bacteria, the bacteria with the high score values were used as representative bacteria. Thus, we identified the special core bacteria of group H containing mainly OTU280 (*Herbaspirillum*), OTU340 (*Clostridium*), and OTU373 (*Ruminococcus*), corresponding to scores of 0.130, 0.309, and 0.158, in that order. In addition, the important core bacteria of the M group were found to include OTU2 (*Lactobacillus*), OTU359 (*Akkermansia*), OTU280 (*Herbaspirillum*), and OTU428 (*Acidaminococcus*), with scores of 0.085, 0.438, 0.413, and 0.179, respectively. The score values for each OTU in M, H, and N groups can be seen in Tables A4–A6. It was found that the core bacteria were concentrated in cluster3 of all samples in group N. The core flora of the M was concentrated in cluster7 of all samples except for the *Herbaspirillum* and the *Akkermansia* in the core flora of M were only distributed in cluster1 and cluster7. The core bacteria of H were all concentrated in cluster1 of all samples except for *Herbaspirillum*, and *Pyramidobacter* in this core group was only present in cluster1. In general, *Herbaspirillum* is only present in H, M but not in N. *Pyramidobacter* is only present in the key bacteria of H and *Akkermansia* is only present in the core bacteria of M. Furthermore, the bacteria in groups M, H, and N were all found in all samples, and the signature bacteria of each group were identified. Among them, all samples and M, H, and N groups were found to be clustered by GMPR+Spectrum to distinguish the similarity between the various populations in different OTUs, as well as to identify the differences in flora that exist between healthy individuals, patients with minimal hepatic encephalopathy and hepatic cirrhosis.

3.5. Network Analysis Core Flora (Genus Level)

In order to compare the structure with the core flora identified by GMPR+Spectrum, we also use the network analysis method to construct the gut flora interaction network among different OTUs and then take the MCODE method to identify and visualize the core gut microbiome contained in the interaction network for each group. The MCODE method calculates the adjacent subgraphs and graph densities contained in each node in the network graph, and the score value of a node reflects the density of the node and its surrounding nodes. Then the algorithm expands from the node with the maximum score value to the surrounding nodes, and the qualified nodes are added to this module and generate a module with similar clustering coefficients.

The intra-module connectivity (Z_i) value is a measure of the role of a node in its module, and the larger the Z_i value is the greater the role played by this node in that module, and then the top 10 OTUs in each module are considered as the core nodes within each key module according to the magnitude of Z_i value corresponding to each OTU. Thus, we obtained the core gut microbiome network for modules 1, 2, and 3 (containing many core modules, but we only chose 3 modules to show here) of all samples' group, as shown in Figure 5, where MCODE1 scored 5 and contained 5 nodes with 10 edges, each corresponding to a Z_i value of 0.935. MCODE2, with a score of 3, contained 3 nodes and 3 edges, and the Z_i values of nodes OTU1430, OTU1111, and OTU535 were 1.402, 1.351, and 1.351 in order. MCODE3 had a score of 5 and contained 5 nodes with 10 edges, and the Z_i values of nodes OTU101, OTU183, OTU907, OTU1153, and OTU582 were 2.293, 2.156, 2.020, 1.951, 1.951, etc. The details of the core colonies contained in each module can be found in Table A7.

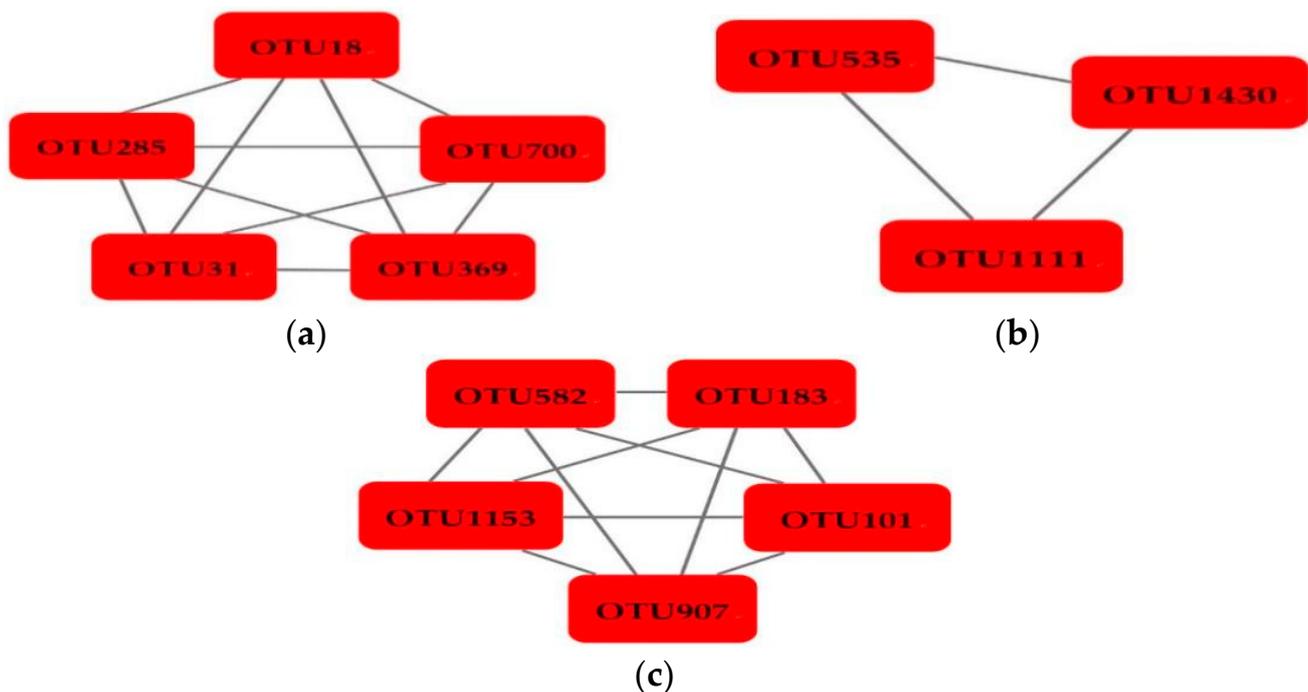


Figure 5. Network diagram of the core gut microbiome of the all samples group. (a) represents module1. (b) represents module2. (c) represents module3.

The details of core colonies and corresponding Zi values in patients with N, H, and M contained in modules 1, 2, and 3 can be found in Tables A8–A10. The Zi values of nodes OTU6, OTU683, OTU658, OTU944, OTU406, OTU378, and OTU440 in H were 0.933, 0.906, 0.906, 0.701, 0.574, 0.574, 1.232 representing *Coprococcus*, *Prevotella*, *Lachnospira*, *Parabacteroides*, *Streptococcus*, and *Clostridium*, respectively. The Zi values of nodes OTU201, OTU1063, OTU861, OTU1425, OTU1237, OTU225, OTU202, OTU238, OTU1440, OTU1250, and OTU1383 in M were 1.145, 1.115, 1.115, 1.054, 1.054, 1.029, 1.029, 0.984, 0.843, 0.843, and 0.843, representing *Ruminococcus*, *Bacteroides*, *Clostridium*, *Lachnospira*, *Faecalibacterium*, *Actinomyces*, *Coprococcus*, *Faecalibacterium*, *Veillonella*, *Sutterella*, *Oscillospira*, respectively.

In the network analysis, the mean scores of the core gut microbiota of the normal, minimal hepatic encephalopathy and cirrhotic groups were 8.33, 9.33, and 10, respectively, with higher scores indicating more complex networks. In the three networks, we also found many similar bacteria among different groups, but the intestinal flora of the M group was more complex, and *Prevotella*, *Lachnospira*, and *Veillonella* were the key bacteria in the intestinal flora of the M group, which were not included in the core flora of normal subjects. *Actinomyces*, *Sutterella*, and *Oscillospira* are key bacteria in the intestinal flora of patients with mild hepatic encephalopathy, which are not included in the core flora of N and H groups. *Streptococcus*, as critical bacteria in the intestinal flora of patients with cirrhosis, was equally absent in the other two groups.

3.6. GMPR+Spectrum and Network Analysis Flora Comparison (Genus)

When comparing the core bacteria identified by GMPR+Spectrum and network analysis, it was found that many core bacteria co-exist in both methods. However, since GMPR+Spectrum and network analysis were two different methods, it was not guaranteed that each cluster of GMPR+Spectrum matches exactly with each module of network analysis, and the following situation may occur, for example, the core bacteria in cluster4 of GMPR+Spectrum appeared in module8 of network analysis method at the same time. The core bacteria in cluster1, cluster2, cluster3, cluster4, and cluster5 of GMPR+Spectrum could be found in module 6 of network analysis. The specific relationship can be seen in Figure 6.

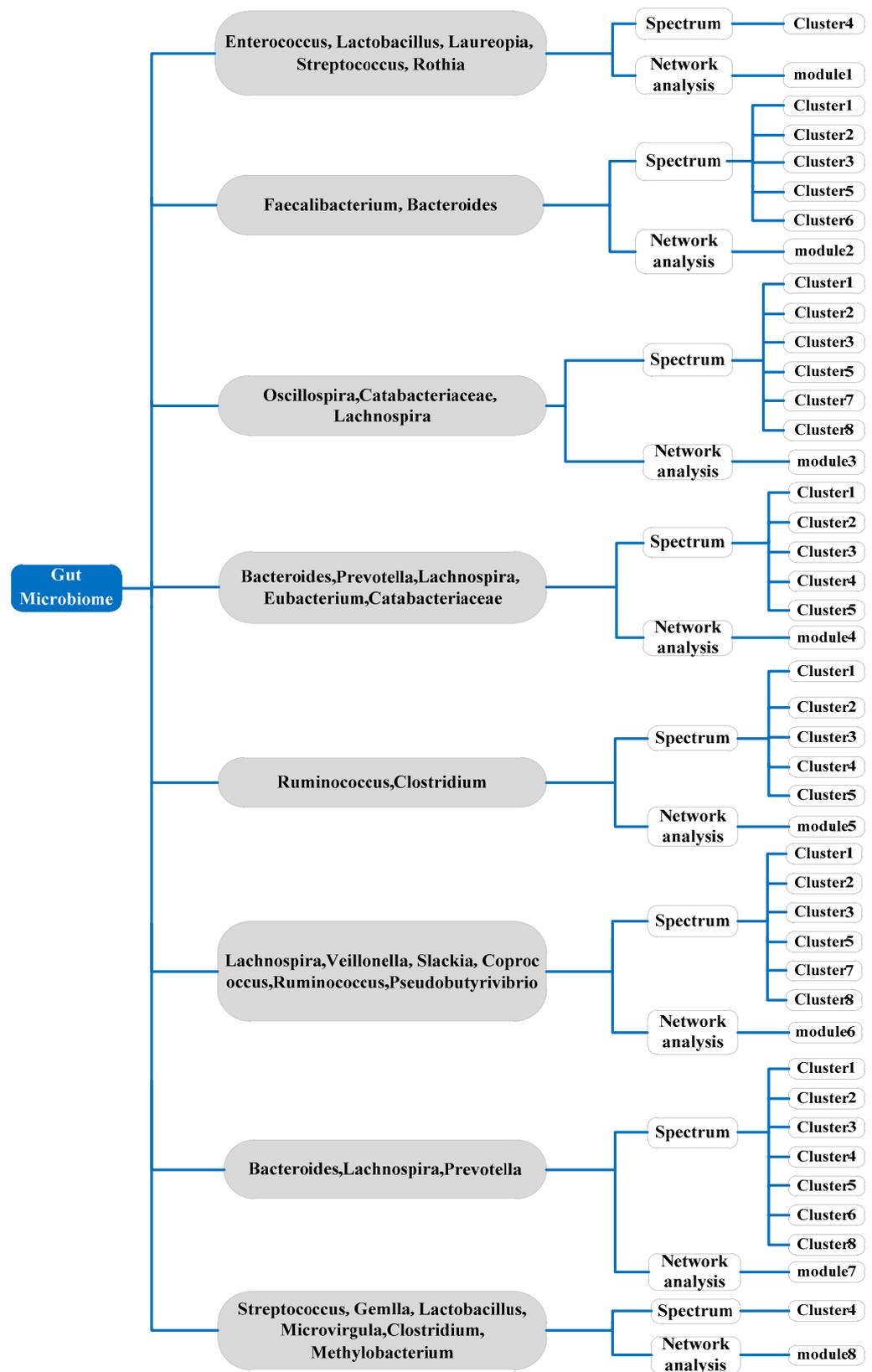


Figure 6. Correspondence between GMPR+Spectrum and network analysis of core flora in all samples group.

A comparative analysis of the core bacteria included in GMPR+Spectrum and network analysis revealed that some core bacteria could be found in both methods, while some

differences existed between the bacteria identified by the two methods. The common bacteria were *Coprococcus*, *Clostridium* of H, *Faecalibacterium*, *Bacteroides*, *Prevotella* in M, and *Clostridium*, *Faecalibacterium*, *Fusobacterium*, and *Bacteroides* in N. The difference was that *Lactobacillus*, *Akkermansia*, *Herbaspirillum* in M and *Oscillospira*, *Dialister* in H were found only in GMPR+Spectrum, etc.

4. Discussion

In this paper, GMPR+Spectrum was used to cluster the all samples dataset to analyze the structure of the intestinal flora. The sequencing data contains many zeros due to the missing or under-sampled intestinal flora in the sequencing process. Therefore, the GMPR method, which can effectively avoid the problem of zero inflation of the intestinal flora data, was first used to normalize the intestinal flora. Then Spectrum was used to analyze the structure of the intestinal flora. The results showed that the GMPR+Spectrum algorithm was the fastest compared with M3C and iClusterPlus on different groups and performed well. Moreover, most of the core clusters of the network analysis method were included in different clusters of GMPR+Spectrum.

In Spectrum, graph theory is used for algorithmic analysis, and the idea is to view the data analysis problem as a problem of optimal partitioning of graphs, while the network analysis method is based on the concept of network graphs in mathematics, where networks are also called “graphs”, and the idea is to view the data analysis problem as a problem of dividing a large network into smaller networks [51–53]. Therefore, the similarity between the two methods is that they both transform the data analysis problem into a graph, and the essence of both is to partition the graph, and the final result is to make the correlation between different subgraphs/subnetworks low and the correlation within the subgraphs/subnetworks high.

The differences are: (1) The way of calculating the similarity matrix is different. The Pearson correlation coefficient method is used for the network analysis method, while the adaptive density-aware kernel in Spectrum is used to calculate the similarity matrix. (2) The graph partitioning method is different. In Spectrum, the Laplace matrix is mainly used to turn the complete undirected graph into a subgraph. The score value of each node for subnetwork partitioning based on the MCODE is calculated by the network analysis method, which reflects the density of the node and the surrounding nodes. (3) In network analysis, the network is constructed based on the optimal threshold value, but the threshold value is artificially chosen. While for Spectrum in the clustering process, Ng spectral clustering method is used, and also the eigenvalue heuristic is used to estimate the number of clusters, and finally, the final eigenvector matrix of GMM clustering is utilized to obtain the optimal number of clusters. (4) In the Spectrum algorithm, the bacteria with the top ranking of score value are taken as the key bacteria. The score value represents the proportion of variance of a certain OTU to the total variance, which is actually the proportion of a certain eigenvalue to the total sum of all eigenvalues. Therefore, the larger score of an OTU, the greater contribution of that OTU to the total OTU. In the network analysis, the OTU with the highest Z_i ranking is used as the core bacteria of each module, and the Z_i value is a measure of the role of a node in the module where it is located. From the experimental results, it is clear that GMPR+Spectrum and network analysis can find the same bacteria in all types of populations, but given that there are still some differences between the two methods, it can also come out that different bacteria are found in the respective methods. Therefore, we performed another specific analysis for these common bacteria under the existing studies.

GMPR+Spectrum method identified all samples as well as in the flora of H, N, and M and found that the *Herbaspirillum* was only present in the core flora of H, M but not in N. In fact, *Herbaspirillum* belongs to Gram-negative bacilli and this bacterium can cause a decrease in the number of Bifidobacterium, further promoting chronic inflammation in the liver [54]. In addition, previous studies had mostly found *Herbaspirillum* in plants, and only in recent years had it been isolated in clinical patients [55–61]. In particular, in a study by Jia

et al. same bacteria of *Herbaspirillum* were found to be a potential opportunistic pathogen for cirrhotic patients and some immunocompromised elderly patients [62]. Although few studies had been conducted on *Herbaspirillum* in humans, some studies had shown that *Herbaspirillum* was a potential opportunistic pathogen, meaning that *Herbaspirillum* may be a crucial bacterium for appropriate disease screening and diagnosis of clinical patients with cirrhosis and minimal hepatic encephalopathy.

In addition, *Akkermansia* was present only in the core bacteria of M in the GMPR+Spectrum method. *Akkermansia* is oval-shaped Gram-negative bacteria that are “probiotic” in many diseases [63], and researchers had seen their potential as the next generation of probiotic drugs that could be potential targets for improving metabolic diseases such as liver diseases. At the same time, some studies have shown that *Akkermansia* may have some negative effects [2]. For example, when the liver degenerates, metabolism will be destroyed, resulting in changes in the abundance of *Akkermansia* abundance [64]. A recent study by Bajaj et al. [65] found that *Akkermansia* change in healthy individuals and MHE patients, specifically *Akkermansia* are higher in the absence of MHE. In contrast, we have only found seen that *Akkermansia* may serve as a critical bacterium to distinguish minimal hepatic encephalopathy from normal individuals, and the specific immunomodulatory mechanism of action still needs to be further investigated subsequently.

The similarities and differences of the core bacteria identified by GMPR+Spectrum and network analysis methods could be found when comparing core flora. The similarity lies in the fact that normal healthy controls (N) have a more abundant flora than patients with minimal hepatic encephalopathy and hepatic encephalopathy, as well as in the fact that at the genus level, both methods can identify some common core flora, such as *Clostridium*, *Ruminococcus* as critical bacteria in cirrhotic patients, the effect of changes in *Clostridium* and *Ruminococcus* on the fecal microbiota of HE patients was confirmed in a study by Bajaj et al. [65], which shows that changes in fecal microbial composition occur in healthy individuals and HE patients, especially in *Clostridium* and *Ruminococcus*, and that changes in these bacteria are associated with the severity of cirrhosis and worsening of the complications of cirrhosis, but in the present study, the specific mechanism of action of *Clostridium* and *Ruminococcus* in this study is not clear. In addition, the difference between the two methods is that the core bacteria in the hepatic encephalopathy patients were *Herbaspirillum*, *Pyramidobacter*, *Faecalibacterium*, *Fusobacterium*, *Dialister*, and *Bacteroides*, which were only the core bacteria identified by GMPR+Spectrum.

At the same time, the critical flora of the M group at the genus level included *Lactobacillus*, *Akkermansia*, *Herbaspirillum*, and *Acidaminococcus*, which were identified by the GMPR+Spectrum method as the specific key bacteria. The GMPR+Spectrum method found *Lactobacillus* as a critical bacterium in the intestinal flora of patients with minimal hepatic encephalopathy, it did not include essential bacteria of cirrhosis, while in the Bajaj study [66], it was shown that *Lactobacillus* in the stool of MHE patients had unique characteristics and that these bacteria could be used for MHE patients for diagnosis. It has even been demonstrated [67] that microecological inhibitors containing *Bifidobacteria* and *Lactobacillus* can regulate the structure of the intestinal flora, inhibiting the growth of ammonia-producing, urease producing bacteria, and have a role in reducing the growth of ammonia. Minimal hepatic encephalopathy as the beginning of the pathogenesis of hepatic encephalopathy [68], so the difference between the two in *Lactobacillus* may serve as an effective way to differentiate between them.

In addition, GMPR+Spectrum has identified *Coprococcus*, *Dialister* in the core flora of cirrhotic patients, *Prevotella*, *Acidaminococcus* in the core flora of MHE patients, and which are also absent from the core flora of normal healthy controls, but the mechanism of action of these bacteria in patients with minimal hepatic encephalopathy and hepatic encephalopathy has not been identified in the current study. It is to be followed up with more in-depth studies. It shows [69] that there is a close relationship between the occurrence of MHE and bacteria that can affect ammonia. Some bacteria containing urease are associated with increased ammonia in MHE. Still, other bacteria sometimes have other hidden effects

Table A2. Clustering evaluation indicators in N, H, M.

| | Index | Spectrum | GMPR+Spectrum | M3C | iClusterPlus |
|---|-----------------|----------|---------------|---------|--------------|
| N | NMI | 0.1524 | 0.1521 | 0.00046 | 0.2678 |
| | DBI | 2.7807 | 3.0169 | 2.8357 | 5.7933 |
| | CH | 2.5914 | 1.6433 | 0.0103 | 1.1882 |
| | Runtimes/second | 23.18 | 20.39 | 157.36 | 38.69 |
| | Cluster number | 2 | 2 | 4 | 3 |
| H | NMI | 0.1555 | 0.1558 | 0.00018 | 0.27001 |
| | DBI | 3.2233 | 3.0738 | 1.8580 | 6.1212 |
| | CH | 1.6681 | 2.3817 | 20.451 | 1.2989 |
| | Runtimes/second | 17.14 | 15.86 | 151.75 | 28.83 |
| | Cluster number | 2 | 2 | 2 | 3 |
| M | NMI | 0.1623 | 0.1647 | 0.00059 | 0.2705 |
| | DBI | 2.7388 | 2.8420 | 2.9583 | 5.9948 |
| | CH | 2.7118 | 2.6339 | 0.0102 | 1.4089 |
| | Runtimes/second | 18.00 | 15.29 | 154.06 | 33.76 |
| | Cluster number | 2 | 2 | 3 | 3 |

Table A3. OTUs contained in each Cluster after GMPR+Spectrum.

| Cluster | OTU ID |
|----------|---|
| Cluster1 | OTU8(<i>Clostridium</i>), OTU11(<i>Dialister</i>), OTU12, OTU13(<i>Megasphaera</i>), OTU16(<i>Blautia</i>), OTU20(<i>Coprococcus</i>), OTU86(<i>Ruminococcus</i>), OTU106(<i>Lachnospira</i>), OTU112(<i>Oscillospira</i>), OTU160(<i>Bacteroides</i>), OTU163(<i>Prevotella</i>), OTU203(<i>Streptococcus</i>), OTU281(<i>Eubacterium</i>), OTU306(<i>Alistipes</i>), OTU313(<i>Sutterella</i>), OTU325(<i>Epulopiscium</i>), OTU363(<i>Phascolarctobacterium</i>), OTU366(<i>Pyramidobacter</i>), OTU409(<i>Fusobacterium</i>), OTU509(<i>Megamonas</i>), OTU647(<i>Roseburia</i>), OTU701(<i>Lachnospira</i>), OTU1821(<i>Faecalibacterium</i>), OTU2170(<i>Enterobacter</i>), OTU3123(<i>Akkermansia</i>) |
| Cluster2 | OTU15(<i>Lactobacillus</i>), OTU60, OTU170(<i>Ruminococcus</i>), OTU252(<i>Faecalibacterium</i>), OTU299(<i>Propionibacterium</i>), OTU309(<i>Oscillospira</i>), OTU324(<i>Brachybacterium</i>), OTU373(<i>Parabacteroides</i>), OTU374(<i>Thermus</i>), OTU426(<i>Lachnospira</i>), OTU434(<i>Dialister</i>), OTU448(<i>Deinococcus</i>), OTU495(<i>Coprococcus</i>), OTU515(<i>Clostridium</i>), OTU530(<i>Megamonas</i>), OTU534(<i>Streptococcus</i>), OTU696(<i>Veillonella</i>), OTU731(<i>Roseburia</i>), OTU760(<i>Bacteroides</i>), OTU884(<i>Lachnospira</i>), OTU1041(<i>Blautia</i>), OTU1237(<i>Phascolarctobacterium</i>), OTU1244(<i>Desulfovibrio</i>), OTU1279(<i>Fusobacterium</i>), OTU1548(<i>Eubacterium</i>), OTU2149(<i>Alistipes</i>), OTU2563(<i>Bulleidia</i>), OTU2796(<i>Campylobacter</i>), OTU2898(<i>Brevundimonas</i>), OTU2936(<i>Leptotrichia</i>), OTU3265(<i>Methylobacterium</i>), OTU3379 (<i>Prevotella</i>) |
| Cluster3 | OTU79(<i>Clostridium</i>), OTU81(<i>Oscillospira</i>), OTU88(<i>Ruminococcus</i>), OTU142(<i>Anoxybacillus</i>), OTU153(<i>Staphylococcus</i>), OTU279(<i>Paludibacter</i>), OTU344(<i>Herbaspirillum</i>), OTU367 (<i>Comamonas</i>), OTU379(<i>Acinetobacter</i>), OTU388(<i>Lactococcus</i>), OTU490(<i>Coprococcus</i>), OTU499(<i>Dietzia</i>), OTU514(<i>Phascolarctobacterium</i>), OTU531(<i>Lactobacillus</i>), OTU543(<i>Eubacterium</i>), OTU546(<i>Micrococcus</i>), OTU586(<i>Dialister</i>), OTU605(<i>Roseburia</i>), OTU674(<i>Veillonella</i>), OTU691 (<i>Faecalibacterium</i>), OTU756(<i>Blautia</i>), OTU763(<i>Selenomonas</i>), OTU811(<i>Lachnospira</i>), OTU873 (<i>Brevundimonas</i>), OTU1104(<i>Streptococcus</i>), OTU1304(<i>Prevotella</i>), OTU1307(<i>Megamonas</i>), OTU1331(<i>Moryella</i>), OTU1602(<i>Odoribacter</i>), OTU1612(<i>Corynebacterium</i>), OTU1637(<i>Fusobacterium</i>), OTU1683(<i>Acidaminococcus</i>), OTU1695(<i>Bacteroides</i>), OTU1725(<i>Parabacteroides</i>), OTU1794(<i>Gemella</i>), OTU2000(<i>Alistipes</i>), OTU2197(<i>Porphyromonas</i>), OTU2419(<i>Escherichia</i>), OTU2426(<i>Sutterella</i>), OTU2520(<i>Brevibacterium</i>), OTU2570(<i>Morganella</i>), OTU2704(<i>Epulopiscium</i>), OTU2727(<i>Enterobacter</i>), OTU2731(<i>Variovorax</i>), OTU2797(<i>Klebsiella</i>), OTU2807(<i>Adlercreutzia</i>), OTU2843(<i>Atopobium</i>), OTU2948(<i>Chryseobacterium</i>), OTU2963(<i>Haloanella</i>), OTU3000, OTU3075(<i>Coprobacillus</i>), OTU3294(<i>Methylobacterium</i>), OTU3301(<i>Sphingomonas</i>), OTU3609(<i>Haemophilus</i>) |
| Cluster4 | OTU0(<i>Lactobacillus</i>), OTU14(<i>Enterococcus</i>), OTU48, OTU49(<i>Clostridium</i>), OTU110(<i>Megamonas</i>), OTU118(<i>Streptococcus</i>), OTU130(<i>Gemella</i>), OTU138(<i>Parabacteroides</i>), OTU143(<i>Bacteroides</i>), OTU145(<i>Prevotella</i>), OTU164(<i>Enterobacter</i>), OTU166(<i>Abiotrophia</i>), OTU195(<i>Lactococcus</i>), OTU263(<i>Actinomyces</i>), OTU305(<i>Neisseria</i>), OTU317(<i>Microbacterium</i>), OTU352(<i>Rothia</i>), OTU376(<i>Thermus</i>), OTU386(<i>Cetobacterium</i>), OTU467(<i>Escherichia</i>), OTU562(<i>Granulicatella</i>), OTU708(<i>Lachnospira</i>), OTU744(<i>Eubacterium</i>), OTU847(<i>Methylobacterium</i>), OTU1056(<i>Lautropia</i>), OTU1060(<i>Blautia</i>), OTU1413(<i>Oribacterium</i>), OTU1666(<i>Leuconostoc</i>), OTU2454(<i>Eikenella</i>), OTU2601(<i>Coprococcus</i>), OTU2632(<i>Aggregatibacter</i>), OTU2679(<i>Haemophilus</i>), OTU2862 (<i>Adlercreutzia</i>), OTU2909(<i>Eggerthella</i>), OTU2911(<i>Campylobacter</i>), OTU2953(<i>Microvirgula</i>), OTU2980(<i>Collinsella</i>), OTU3039(<i>Collinsella</i>), OTU3591(<i>Veillonella</i>) |
| Cluster5 | OTU4(<i>Weissella</i>), OTU42(<i>Clostridium</i>), OTU58, OTU64(<i>Coprococcus</i>), OTU127(<i>Oscillospira</i>), OTU190(<i>Veillonella</i>), OTU200(<i>Ruminococcus</i>), OTU218(<i>Prevotella</i>), OTU272(<i>Odoribacter</i>), OTU283(<i>Faecalibacterium</i>), OTU286(<i>Parabacteroides</i>), OTU272(<i>Odoribacter</i>), OTU334(<i>Slackia</i>), OTU339(<i>Selenomonas</i>), OTU361(<i>Bacteroides</i>), OTU403(<i>Eubacterium</i>), OTU428(<i>Dialister</i>), OTU624(<i>Lachnospira</i>), OTU635(<i>Anaeroglobus</i>), OTU664(<i>Roseburia</i>), OTU827(<i>Phascolarctobacterium</i>), OTU918(<i>Blautia</i>), OTU1412(<i>Megamonas</i>), OTU2119(<i>Alistipes</i>), OTU2500(<i>Leptotrichia</i>), OTU3057, OTU3097 (<i>Fusobacterium</i>) |

Table A3. Cont.

| Cluster | OTU ID |
|----------|---|
| Cluster6 | OTU2(<i>Lactobacillus</i>), OTU6(<i>Dialister</i>), OTU9(<i>Veillonella</i>), OTU22(<i>Lachnospira</i>), OTU24(<i>Roseburia</i>), OTU28(<i>Megasphaera</i>), OTU39(<i>Coprococcus</i>), OTU44, OTU47(<i>Ruminococcus</i>), OTU50(<i>Clostridium</i>), OTU76(<i>Phascolarctobacterium</i>), OTU146(<i>Bacteroides</i>), OTU150(<i>Prevotella</i>), OTU219(<i>Faecalibacterium</i>), OTU238(<i>Alistipes</i>), OTU248(<i>Parabacteroides</i>), OTU265(<i>Enterobacter</i>), OTU301(<i>Odoribacter</i>), OTU332(<i>Sutterella</i>), OTU362(<i>Asteroleplasma</i>), OTU419(<i>Fusobacterium</i>), OTU2047(<i>Leclercia</i>) |
| Cluster7 | OTU5(<i>Coprococcus</i>), OTU17(<i>Veillonella</i>), OTU23, OTU45(<i>Clostridium</i>), OTU70(<i>Eubacterium</i>), OTU105(<i>Oscillospira</i>), OTU119(<i>Ruminococcus</i>), OTU230(<i>Prevotella</i>), OTU253(<i>Faecalibacterium</i>), OTU259(<i>Parabacteroides</i>), OTU297(<i>Haemophilus</i>), OTU461(<i>Akkermansia</i>), OTU488(<i>Megamonas</i>), OTU553(<i>Lactobacillus</i>), OTU555(<i>Streptococcus</i>), OTU565(<i>Acidaminococcus</i>), OTU2712(<i>Fusobacterium</i>), OTU3479(<i>Escherichia</i>), OTU167(<i>Bacteroides</i>) |
| Cluster8 | OTU82(<i>Roseburia</i>), OTU83(<i>Lachnospira</i>), OTU114(<i>Clostridium</i>), OTU122, OTU229(<i>Holdemania</i>), OTU256(<i>Parabacteroides</i>), OTU481(<i>Megamonas</i>), OTU491(<i>Peptostreptococcus</i>), OTU528(<i>Faecalibacterium</i>), OTU557(<i>Coprococcus</i>), OTU561(<i>Blautia</i>), OTU567(<i>Veillonella</i>), OTU580(<i>Dialister</i>), OTU644(<i>Ruminococcus</i>), OTU693(<i>Oscillospira</i>), OTU1030(<i>Prevotella</i>), OTU1174(<i>Desulfovibrio</i>), OTU1429(<i>Actinomyces</i>), OTU1627(<i>Bacteroides</i>), OTU1793(<i>Streptococcus</i>), OTU1810(<i>Bilophila</i>), OTU1996(<i>Oxalobacter</i>), OTU2053(<i>Alistipes</i>), OTU2130(<i>Odoribacter</i>), OTU2597(<i>Raoultella</i>), OTU2599(<i>Epulopiscium</i>), OTU2714(<i>Fusobacterium</i>), OTU2719(<i>Sutterella</i>), OTU2894(<i>Sarcina</i>), OTU3056, OTU3062(<i>Coprobacillus</i>) |

Table A4. The score value and core flora of GMPR+Spectrum in N.

| N | OTUID | SCORE | Family | Genus |
|----------|--------|------------------------|----------------------------|-------------------------|
| Cluster1 | OTU45 | 0.528 | <i>Lachnospiraceae</i> | <i>Clostridium</i> |
| | OTU104 | 0.478 | <i>Streptococcaceae</i> | <i>Streptococcus</i> |
| | OTU230 | 0.475 | <i>Ruminococcaceae</i> | <i>Oscillospira</i> |
| | OTU125 | 0.330 | <i>Bacillaceae</i> | <i>Anoxybacillus</i> |
| | OTU2 | 0.068 | <i>Lactobacillaceae</i> | <i>Lactobacillus</i> |
| | OTU136 | 0.056 | <i>Staphylococcaceae</i> | <i>Staphylococcus</i> |
| | OTU274 | 0.018 | <i>Alcaligenaceae</i> | <i>Sutterella</i> |
| Cluster2 | OTU313 | 0.308 | <i>Ruminococcaceae</i> | <i>Faecalibacterium</i> |
| | OTU354 | 0.265 | <i>Ruminococcaceae</i> | <i>Oscillospira</i> |
| | OTU326 | 0.235 | <i>Clostridiaceae</i> | <i>Clostridium</i> |
| | OTU320 | 0.225 | <i>Ruminococcaceae</i> | <i>Eubacterium</i> |
| | OTU314 | 0.195 | <i>Erysipelotrichaceae</i> | <i>Clostridium</i> |
| | OTU329 | 0.187 | <i>Fusobacteriaceae</i> | <i>Fusobacterium</i> |
| | OTU367 | 0.185 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| OTU473 | 0.134 | <i>Lachnospiraceae</i> | <i>Clostridium</i> | |

Table A5. The score value and core flora of GMPR+Spectrum in H.

| H | OTUID | SCORE | Family | Genus |
|----------|--------|------------------------|--------------------------------|-------------------------|
| Cluster1 | OTU141 | 0.530 | <i>Erysipelotrichaceae</i> | |
| | OTU51 | 0.323 | <i>Lachnospiraceae</i> | |
| | OTU111 | 0.272 | <i>Ruminococcaceae</i> | <i>Oscillospira</i> |
| | OTU280 | 0.130 | <i>Oxalobacteraceae</i> | <i>Herbaspirillum</i> |
| | OTU296 | 0.116 | <i>Dethiosulfovibrionaceae</i> | <i>Pyramidobacter</i> |
| | OTU204 | 0.109 | <i>Veillonellaceae</i> | <i>Dialister</i> |
| Cluster2 | OTU313 | 0.593 | <i>Ruminococcaceae</i> | <i>Faecalibacterium</i> |
| | OTU407 | 0.348 | <i>Ruminococcaceae</i> | <i>Oscillospira</i> |
| | OTU323 | 0.320 | <i>Lachnospiraceae</i> | <i>Coprococcus</i> |
| | OTU340 | 0.309 | <i>Lachnospiraceae</i> | <i>Clostridium</i> |
| | OTU329 | 0.284 | <i>Fusobacteriaceae</i> | <i>Fusobacterium</i> |
| | OTU339 | 0.200 | <i>Veillonellaceae</i> | <i>Dialister</i> |
| | OTU458 | 0.173 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| OTU373 | 0.158 | <i>Lachnospiraceae</i> | <i>Ruminococcus</i> | |

Table A6. The score value and core flora of GMPR+Spectrum in M.

| M | OTUID | SCORE | Family | Genus |
|----------|--------|-------|----------------------------|-------------------------|
| Cluster1 | OTU2 | 0.085 | <i>Lactobacillaceae</i> | <i>Lactobacillus</i> |
| | OTU12 | 0.000 | <i>Veillonellaceae</i> | |
| Cluster2 | OTU168 | 0.754 | <i>Veillonellaceae</i> | |
| | OTU359 | 0.438 | <i>Verrucomicrobiaceae</i> | <i>Akkermansia</i> |
| | OTU280 | 0.413 | <i>Oxalobacteraceae</i> | <i>Herbaspirillum</i> |
| | OTU364 | 0.333 | <i>Ruminococcaceae</i> | |
| | OTU443 | 0.330 | <i>Ruminococcaceae</i> | <i>Faecalibacterium</i> |
| | OTU404 | 0.319 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| | OTU155 | 0.255 | <i>Prevotellaceae</i> | <i>Prevotella</i> |
| | OTU428 | 0.179 | <i>Veillonellaceae</i> | <i>Acidaminococcus</i> |

Table A7. Core flora and corresponding Zi values in each module of all samples group in network analysis.

| Spectrum | Zi | OTU ID | Family | Genus |
|----------|----------|---------|----------------------------|--------------------------|
| module1 | 0.934501 | OTU31 | <i>Enterococcaceae</i> | <i>Enterococcus</i> |
| | 0.934501 | OTU18 | <i>Lactobacillaceae</i> | <i>Lactobacillus</i> |
| | 0.934501 | OTU700 | <i>Burkholderiaceae</i> | <i>Lautropia</i> |
| | 0.934501 | OTU369 | <i>Streptococcaceae</i> | <i>Streptococcus</i> |
| | 0.934501 | OTU285 | <i>Micrococcaceae</i> | <i>Rothia</i> |
| module2 | 1.402386 | OTU1430 | <i>Ruminococcaceae</i> | <i>Faecalibacterium</i> |
| | 1.351226 | OTU1111 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| module3 | 2.292694 | OTU101 | <i>Ruminococcaceae</i> | <i>Oscillospira</i> |
| | 1.951359 | OTU1153 | <i>Catabacteriaceae</i> | |
| | 1.951359 | OTU582 | <i>Lachnospiraceae</i> | |
| module4 | 1.827142 | OTU1103 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| | 1.746558 | OTU180 | <i>Prevotellaceae</i> | <i>Prevotella</i> |
| | 1.742641 | OTU571 | <i>Lachnospiraceae</i> | <i>Lachnospira</i> |
| | 1.742641 | OTU235 | <i>Ruminococcaceae</i> | <i>Eubacterium</i> |
| | 1.678291 | OTU1083 | <i>Catabacteriaceae</i> | |
| module5 | 1.960498 | OTU372 | <i>Lachnospiraceae</i> | <i>Ruminococcus</i> |
| | 1.727345 | OTU1300 | <i>Clostridiaceae</i> | <i>Clostridium</i> |
| | 1.494192 | OTU431 | <i>Turicibacteraceae</i> | |
| module6 | 2.426804 | OTU493 | <i>Lachnospiraceae</i> | <i>Lachnospira</i> |
| | 2.310227 | OTU1420 | <i>Veillonellaceae</i> | <i>Veillonella</i> |
| | 2.310227 | OTU272 | <i>Coriobacteriaceae</i> | <i>Slackia</i> |
| | 1.960498 | OTU1187 | <i>Lachnospiraceae</i> | <i>Coprococcus</i> |
| | 1.960498 | OTU281 | <i>Ruminococcaceae</i> | |
| | 1.960498 | OTU90 | <i>Lachnospiraceae</i> | <i>Pseudobutyrvibrio</i> |
| module7 | 1.235633 | OTU355 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| | 0.813126 | OTU356 | <i>Lachnospiraceae</i> | |
| | 0.644123 | OTU854 | <i>Prevotellaceae</i> | <i>Prevotella</i> |
| module8 | 0.668220 | OTU674 | <i>Streptococcaceae</i> | <i>Streptococcus</i> |
| | 0.668220 | OTU113 | <i>Gemellaceae</i> | <i>Gemella</i> |
| | 0.420731 | OTU286 | <i>Lactobacillaceae</i> | <i>Lactobacillus</i> |
| | 0.420731 | OTU1326 | <i>Neisseriaceae</i> | <i>Microvirgula</i> |
| | 0.173242 | OTU658 | <i>Ruminococcaceae</i> | <i>Clostridium</i> |
| | 0.173242 | OTU593 | <i>Methylobacteriaceae</i> | <i>Methylobacterium</i> |

Table A8. Core colonies and corresponding Zi values in each module of N group in network analysis.

| N | OTUID | Zi | Family | Genus |
|--------|---------|------------------------|----------------------------|-------------------------|
| MCODE1 | OTU216 | 1.107 | <i>Ruminococcaceae</i> | <i>Faecalibacterium</i> |
| | OTU1110 | 1.052 | <i>Rikenellaceae</i> | <i>Alistipes</i> |
| | OTU222 | 1.052 | <i>Porphyromonadaceae</i> | <i>Parabacteroides</i> |
| | OTU1354 | 1.052 | <i>Fusobacteriaceae</i> | <i>Fusobacterium</i> |
| | OTU1335 | 1.052 | <i>Comamonadaceae</i> | <i>Brachymonas</i> |
| | OTU1334 | 1.052 | <i>Ruminococcaceae</i> | <i>Ruminococcus</i> |
| | OTU1293 | 1.052 | <i>Clostridiaceae</i> | <i>Clostridium</i> |
| | OTU1286 | 1.052 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| MCODE2 | OTU1046 | 1.382 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| | OTU268 | 1.382 | <i>Ruminococcaceae</i> | |
| | OTU106 | 1.074 | <i>Lachnospiraceae</i> | <i>Ruminococcus</i> |
| | OTU180 | 1.030 | <i>Clostridium</i> | |
| | OTU907 | 0.986 | <i>Ruminococcaceae</i> | |
| | OTU101 | 0.986 | <i>Lactobacillaceae</i> | <i>Lactobacillus</i> |
| | OTU66 | 0.986 | <i>Lachnospiraceae</i> | |
| | OTU1058 | 0.986 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| OTU682 | 0.986 | <i>Lachnospiraceae</i> | <i>Roseburia</i> | |
| MCODE3 | OTU202 | 1.285 | <i>Lachnospiraceae</i> | <i>Coproccoccus</i> |
| | OTU1380 | 1.243 | <i>Lachnospiraceae</i> | |
| | OTU1333 | 1.243 | <i>Erysipelotrichaceae</i> | <i>Clostridium</i> |
| | OTU1324 | 1.243 | <i>Comamonadaceae</i> | <i>Variovorax</i> |
| | OTU1310 | 1.243 | <i>Coriobacteriaceae</i> | <i>Adlercreutzia</i> |
| | OTU1233 | 1.243 | <i>Enterobacteriaceae</i> | |
| | OTU1184 | 1.243 | <i>Ruminococcaceae</i> | <i>Faecalibacterium</i> |
| | OTU1166 | 1.243 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |

Table A9. Core colonies and corresponding Zi values in each module of H group in network analysis.

| H | OTUID | Zi | Family | Genus |
|--------|---------|-------|---------------------------|------------------------|
| MCODE1 | OTU368 | 0.960 | | |
| | OTU6 | 0.933 | <i>Lachnospiraceae</i> | <i>Coproccoccus</i> |
| | OTU1023 | 0.933 | <i>Lachnospiraceae</i> | |
| | OTU692 | 0.906 | <i>Ruminococcaceae</i> | |
| | OTU683 | 0.906 | <i>Prevotellaceae</i> | <i>Prevotella</i> |
| | OTU658 | 0.906 | <i>Lachnospiraceae</i> | <i>Lachnospira</i> |
| | OTU632 | 0.906 | <i>Ruminococcaceae</i> | |
| | OTU628 | 0.906 | <i>Lachnospiraceae</i> | |
| | OTU619 | 0.906 | <i>Lachnospiraceae</i> | |
| | OTU559 | 0.906 | <i>Lachnospiraceae</i> | |
| MCODE2 | OTU1380 | 0.701 | <i>Lachnospiraceae</i> | |
| | OTU1317 | 0.701 | | |
| | OTU1036 | 0.701 | | |
| | OTU944 | 0.701 | <i>Porphyromonadaceae</i> | <i>Parabacteroides</i> |
| | OTU891 | 0.701 | | |
| | OTU665 | 0.701 | <i>Lachnospiraceae</i> | |
| | OTU434 | 0.574 | <i>Lachnospiraceae</i> | |
| | OTU431 | 0.574 | <i>Lachnospiraceae</i> | |
| | OTU406 | 0.574 | <i>Streptococcaceae</i> | <i>Streptococcus</i> |
| | OTU378 | 0.574 | <i>Lachnospiraceae</i> | <i>Coproccoccus</i> |

Table A9. Cont.

| H | OTUID | Zi | Family | Genus |
|--------|---------|------------------------|------------------------|--------------------|
| MCODE3 | OTU912 | 1.656 | <i>Prevotellaceae</i> | <i>Prevotella</i> |
| | OTU134 | 1.514 | <i>Prevotellaceae</i> | <i>Prevotella</i> |
| | OTU1009 | 1.408 | <i>Prevotellaceae</i> | <i>Prevotella</i> |
| | OTU935 | 1.373 | <i>Prevotellaceae</i> | <i>Prevotella</i> |
| | OTU494 | 1.302 | <i>Lachnospiraceae</i> | <i>Lachnospira</i> |
| | OTU911 | 1.302 | <i>Prevotellaceae</i> | <i>Prevotella</i> |
| | OTU888 | 1.302 | <i>Prevotellaceae</i> | <i>Prevotella</i> |
| | OTU446 | 1.267 | <i>Veillonellaceae</i> | <i>Veillonella</i> |
| | OTU440 | 1.232 | <i>Ruminococcaceae</i> | <i>Clostridium</i> |
| OTU422 | 1.232 | <i>Lachnospiraceae</i> | <i>Coprococcus</i> | |

Table A10. Core colonies and corresponding Zi values in each module of M group in network analysis.

| M | OTUID | Zi | Family | Genus |
|---------|---------|------------------------|----------------------------|-------------------------|
| MCODE1 | OTU201 | 1.145 | <i>Ruminococcaceae</i> | <i>Ruminococcus</i> |
| | OTU1063 | 1.115 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| | OTU861 | 1.115 | <i>Ruminococcaceae</i> | <i>Clostridium</i> |
| | OTU64 | 1.085 | <i>Lachnospiraceae</i> | <i>Clostridium</i> |
| | OTU1392 | 1.054 | <i>Ruminococcaceae</i> | |
| | OTU1425 | 1.054 | <i>Lachnospiraceae</i> | <i>Lachnospira</i> |
| | OTU1304 | 1.054 | <i>Lachnospiraceae</i> | <i>Clostridium</i> |
| | OTU1378 | 1.054 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| | OTU1237 | 1.054 | <i>Ruminococcaceae</i> | <i>Faecalibacterium</i> |
| OTU1184 | 1.054 | <i>Ruminococcaceae</i> | <i>Faecalibacterium</i> | |
| MCODE2 | OTU1069 | 1.075 | <i>Ruminococcaceae</i> | |
| | OTU1036 | 1.075 | | |
| | OTU225 | 1.029 | <i>Actinomycetaceae</i> | <i>Actinomyces</i> |
| | OTU202 | 1.029 | <i>Lachnospiraceae</i> | <i>Coprococcus</i> |
| | OTU551 | 1.029 | <i>Lachnospiraceae</i> | |
| | OTU465 | 1.029 | <i>Erysipelotrichaceae</i> | |
| | OTU322 | 1.029 | <i>Lachnospiraceae</i> | <i>Coprococcus</i> |
| OTU238 | 0.984 | <i>Ruminococcaceae</i> | <i>Faecalibacterium</i> | |
| MCODE3 | OTU956 | 0.843 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| | OTU1440 | 0.843 | <i>Veillonellaceae</i> | <i>Veillonella</i> |
| | OTU1400 | 0.843 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| | OTU1409 | 0.843 | <i>Ruminococcaceae</i> | <i>Ruminococcus</i> |
| | OTU1250 | 0.843 | <i>Alcaligenaceae</i> | <i>Sutterella</i> |
| | OTU1383 | 0.843 | <i>Ruminococcaceae</i> | <i>Oscillospira</i> |
| | OTU1183 | 0.843 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| | OTU1166 | 0.843 | <i>Bacteroidaceae</i> | <i>Bacteroides</i> |
| | OTU1113 | 0.843 | <i>Prevotellaceae</i> | <i>Prevotella</i> |
| | OTU1021 | 0.843 | <i>Ruminococcaceae</i> | <i>Faecalibacterium</i> |

References

1. Qin, J.; Li, R.; Raes, J.; Arumugam, M.; Burgdorf, K.S.; Manichanh, C.; Nielsen, T.; Pons, N.; Levenez, F.; Yamada, T.; et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **2010**, *464*, 59–65. [\[CrossRef\]](#)
2. Qin, J.; Li, Y.; Cai, Z.; Li, S.; Zhu, J.; Zhang, F.; Liang, S.; Zhang, W.; Guan, Y.; Shen, D.; et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **2012**, *490*, 55–60. [\[CrossRef\]](#)
3. Lambeth, S.M.; Carson, T.; Lowe, J.; Ramaraj, T.; Leff, J.W.; Luo, L.; Bell, C.J.; Shah, V.O. Composition, Diversity and Abundance of Gut Microbiome in Prediabetes and Type 2 Diabetes. *J. Diabetes Obes.* **2015**, *2*, 1–7. [\[CrossRef\]](#)
4. Larsen, N.; Vogensen, F.K.; van den Berg, F.W.; Nielsen, D.S.; Andreasen, A.S.; Pedersen, B.K.; Al-Soud, W.A.; Sorensen, S.J.; Hansen, L.H.; Jakobsen, M. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS ONE* **2010**, *5*, e9085. [\[CrossRef\]](#)
5. Rajpal, D.K.; Klein, J.L.; Mayhew, D.; Boucheron, J.; Spivak, A.T.; Kumar, V.; Ingraham, K.; Paulik, M.; Chen, L.; Van Horn, S.; et al. Selective Spectrum Antibiotic Modulation of the Gut Microbiome in Obesity and Diabetes Rodent Models. *PLoS ONE* **2015**, *10*, e0145499. [\[CrossRef\]](#)

6. Stewart, C.J.; Ajami, N.J.; O'Brien, J.L.; Hutchinson, D.S.; Smith, D.P.; Wong, M.C.; Ross, M.C.; Lloyd, R.E.; Doddapaneni, H.; Metcalf, G.A.; et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **2018**, *562*, 583–588. [[CrossRef](#)]
7. Vatanen, T.; Franzosa, E.A.; Schwager, R.; Tripathi, S.; Arthur, T.D.; Vehik, K.; Lernmark, A.; Hagopian, W.A.; Rewers, M.J.; She, J.X.; et al. The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. *Nature* **2018**, *562*, 589–594. [[CrossRef](#)]
8. Arumugam, M.; Raes, J.; Pelletier, E.; Le Paslier, D.; Yamada, T.; Mende, D.R.; Fernandes, G.R.; Tap, J.; Bruls, T.; Batto, J.M.; et al. Enterotypes of the human gut microbiome. *Nature* **2011**, *473*, 174–180. [[CrossRef](#)]
9. Wu, G.D.; Chen, J.; Hoffmann, C.; Bittinger, K.; Chen, Y.Y.; Keilbaugh, S.A.; Bewtra, M.; Knights, D.; Walters, W.A.; Knight, R.; et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science* **2011**, *334*, 105–108. [[CrossRef](#)]
10. Holmes, I.; Harris, K.; Quince, C. Dirichlet multinomial mixtures: Generative models for microbial metagenomics. *PLoS ONE* **2012**, *7*, e30126. [[CrossRef](#)]
11. Davey, J.W.; Hohenlohe, P.A.; Etter, P.D.; Boone, J.Q.; Catchen, J.M.; Blaxter, M.L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* **2011**, *12*, 499–510. [[CrossRef](#)]
12. Dillies, M.A.; Rau, A.; Aubert, J.; Hennequet-Antier, C.; Jeanmougin, M.; Servant, N.; Keime, C.; Marot, G.; Castel, D.; Estelle, J.; et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinform.* **2013**, *14*, 671–683. [[CrossRef](#)]
13. Li, P.; Piao, Y.; Shon, H.S.; Ryu, K.H. Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data. *BMC Bioinform.* **2015**, *16*, 347. [[CrossRef](#)]
14. McMurdie, P.J.; Holmes, S. Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **2014**, *10*, e1003531. [[CrossRef](#)]
15. Chen, J.; King, E.; Deek, R.; Wei, Z.; Yu, Y.; Grill, D.; Ballman, K.; Stegle, O. An omnibus test for differential distribution analysis of microbiome sequencing data. *Bioinformatics* **2018**, *34*, 643–651. [[CrossRef](#)]
16. Chen, L.; Reeve, J.; Zhang, L.; Huang, S.; Wang, X.; Chen, J. GMPR: A robust normalization method for zero-inflated count data with application to microbiome sequencing data. *PeerJ* **2018**, *6*, e4600. [[CrossRef](#)]
17. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Pena, A.G.; Goodrich, J.K.; Gordon, J.I.; et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010**, *7*, 335–336. [[CrossRef](#)]
18. Chen, J.; Bittinger, K.; Charlson, E.S.; Hoffmann, C.; Lewis, J.; Wu, G.D.; Collman, R.G.; Bushman, F.D.; Li, H. Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* **2012**, *28*, 2106–2113. [[CrossRef](#)]
19. Jain, A.K.; Law, M. Data Clustering: A User's Dilemma. In Proceedings of the International Conference on Pattern Recognition & Machine Intelligence, Kolkata, India, 20–22 December 2005.
20. Larose, D.T.; Larose, C.D. Data preprocessing. In *Discovering Knowledge in Data (An Introduction to Data Mining)*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014; pp. 16–50.
21. Wedding, D.K. Discovering knowledge in data, an introduction to data mining. *Inf. Processing Manag.* **2005**, *41*, 1307–1309. [[CrossRef](#)]
22. von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **2007**, *17*, 395–416. [[CrossRef](#)]
23. John, C.R.; Watson, D.; Barnes, M.R.; Pitzalis, C.; Lewis, M.J. Spectrum: Fast density-aware spectral clustering for single and multi-omic data. *Bioinformatics* **2020**, *36*, 1159–1166. [[CrossRef](#)]
24. Groeneweg, M.; Moerland, W.; Quero, J.C.; Hop, W.C.; Krabbe, P.F.; Schalm, S.W. Screening of subclinical hepatic encephalopathy. *J. Hepatol.* **2000**, *32*, 748–753. [[CrossRef](#)]
25. Saxena, N.; Bhatia, M.; Joshi, Y.K.; Garg, P.K.; Tandon, R.K. Auditory P300 event-related potentials and number connection test for evaluation of subclinical hepatic encephalopathy in patients with cirrhosis of the liver: A follow-up study. *J. Gastroenterol. Hepatol.* **2001**, *16*, 322–327. [[CrossRef](#)]
26. Schomerus, H.; Hamster, W. Quality of life in cirrhotics with minimal hepatic encephalopathy. *Metab. Brain Dis.* **2001**, *16*, 37–41. [[CrossRef](#)]
27. Sharma, P.; Sharma, B.C.; Puri, V.; Sarin, S.K. Critical flicker frequency: Diagnostic tool for minimal hepatic encephalopathy. *J. Hepatol.* **2007**, *47*, 67–73. [[CrossRef](#)]
28. Bajaj, J.S. Management options for minimal hepatic encephalopathy. *Expert Rev. Gastroenterol. Hepatol.* **2008**, *2*, 785–790. [[CrossRef](#)]
29. Romero-Gomez, M.; Cordoba, J.; Jover, R.; del Olmo, J.A.; Ramirez, M.; Rey, R.; de Madaria, E.; Montoliu, C.; Nunez, D.; Flavia, M.; et al. Value of the critical flicker frequency in patients with minimal hepatic encephalopathy. *Hepatology* **2007**, *45*, 879–885. [[CrossRef](#)]
30. Bajaj, J.S.; Saeian, K.; Verber, M.D.; Hirschke, D.; Hoffmann, R.G.; Franco, J.; Varma, R.R.; Rao, S.M. Inhibitory control test is a simple method to diagnose minimal hepatic encephalopathy and predict development of overt hepatic encephalopathy. *Am. J. Gastroenterol.* **2007**, *102*, 754–760. [[CrossRef](#)]
31. Ford, J.M.; Gray, M.; Whitfield, S.L.; Turken, A.U.; Glover, G.; Faustman, W.O.; Mathalon, D.H. Acquiring and inhibiting prepotent responses in schizophrenia: Event-related brain potentials and functional magnetic resonance imaging. *Arch. Gen. Psychiatry* **2004**, *61*, 119–129. [[CrossRef](#)]

32. Schiff, S.; Vallesi, A.; Mapelli, D.; Orsato, R.; Pellegrini, A.; Umilta, C.; Gatta, A.; Amodio, P. Impairment of response inhibition precedes motor alteration in the early stage of liver cirrhosis: A behavioral and electrophysiological study. *Metab. Brain Dis.* **2005**, *20*, 381–392. [[CrossRef](#)]
33. Weissenborn, K.; Ennen, J.C.; Schomerus, H.; Ruckert, N.; Hecker, H. Neuropsychological characterization of hepatic encephalopathy. *J. Hepatol.* **2001**, *34*, 768–773. [[CrossRef](#)]
34. Ortiz, M.; Jacas, C.; Cordoba, J. Minimal hepatic encephalopathy: Diagnosis, clinical significance and recommendations. *J. Hepatol.* **2005**, *42* (Suppl. 1), S45–S53. [[CrossRef](#)]
35. Allampati, S.; Duarte-Rojo, A.; Thacker, L.R.; Patidar, K.R.; White, M.B.; Klair, J.S.; John, B.; Heuman, D.M.; Wade, J.B.; Flud, C.; et al. Diagnosis of Minimal Hepatic Encephalopathy Using Stroop EncephalApp: A Multicenter US-Based, Norm-Based Study. *Am. J. Gastroenterol.* **2016**, *111*, 78–86. [[CrossRef](#)]
36. Lim, M.Y.; Song, E.-J.; Kim, S.H.; Lee, J.; Nam, Y.-D. Comparison of DNA extraction methods for human gut microbial community profiling. *Syst. Appl. Microbiol.* **2018**, *41*, 151–157. [[CrossRef](#)]
37. Caporaso, J.G.; Lauber, C.L.; Walters, W.A.; Berg-Lyons, D.; Lozupone, C.A.; Turnbaugh, P.J.; Fierer, N.; Knight, R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* **2011**, *108* (Suppl. 1), 4516–4522. [[CrossRef](#)]
38. John, C.R.; Watson, D.; Russ, D.; Goldmann, K.; Ehrenstein, M.; Pitzalis, C.; Lewis, M.; Barnes, M. M3C: Monte Carlo reference-based consensus clustering. *Sci. Rep.* **2020**, *10*, 1816. [[CrossRef](#)]
39. Shen, R.; Olshen, A.B.; Ladanyi, M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* **2009**, *25*, 2906–2912. [[CrossRef](#)]
40. Robinson, M.D.; Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **2010**, *11*, R25. [[CrossRef](#)]
41. Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **2010**, *11*, R106. [[CrossRef](#)]
42. Monti, S.; Tamayo, P.; Mesirov, J.P.; Golub, T.R. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Mach. Learn.* **2003**, *52*, 91–118. [[CrossRef](#)]
43. Senbabaoglu, Y.; Michailidis, G.; Li, J.Z. Critical limitations of consensus clustering in class discovery. *Sci. Rep.* **2014**, *4*, 6207. [[CrossRef](#)]
44. Fang, K.T.; Wang, Y. Number-theoretic Methods in Statistics. In *Number-theoretic Methods in Statistics*; Chapman & Hall: London, UK, 1994.
45. Ju, F.; Zhang, T. Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *ISME J.* **2015**, *9*, 683–695. [[CrossRef](#)]
46. Sinha, R.; Chen, J.; Amir, A.; Vogtmann, E.; Shi, J.; Inman, K.S.; Flores, R.; Sampson, J.; Knight, R.; Chia, N. Collecting Fecal Samples for Microbiome Analyses in Epidemiology Studies. *Cancer Epidemiol. Biomark. Prev.* **2016**, *25*, 407–416. [[CrossRef](#)]
47. Zhang, P. Evaluating accuracy of community detection using the relative normalized mutual information. *J. Stat. Mech.—Theory Exp.* **2015**, *2015*, P11006. [[CrossRef](#)]
48. Theodoridis, S.; Pikrakis, A.; Koutroumbas, K.; Cavouras, D. *Introduction to Pattern Recognition: A Matlab Approach*; Elsevier Inc.: Amsterdam, The Netherlands, 2010.
49. Cengizler, C.; Kerem-Un, M. Evaluation of Calinski-Harabasz Criterion as Fitness Measure for Genetic Algorithm Based Segmentation of Cervical Cell Nuclei. *Br. J. Math. Comput. Sci.* **2017**, *22*, 1–13. [[CrossRef](#)]
50. Mandal, S.; Van Treuren, W.; White, R.A.; Eggesbo, M.; Knight, R.; Peddada, S.D. Analysis of composition of microbiomes: A novel method for studying microbial composition. *Microb. Ecol. Health Dis.* **2015**, *26*, 27663. [[CrossRef](#)]
51. Kim, S.; Thapa, I.; Lu, G.; Zhu, L.; Ali, H.H. A systems biology approach for modeling microbiomes using split graphs. In Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, USA, 13–16 November 2017; pp. 2062–2068.
52. Pascoe, E.L.; Haufler, H.C.; Marchesi, J.R.; Perkins, S.E. Network analysis of gut microbiota literature: An overview of the research landscape in non-human animal studies. *ISME J.* **2017**, *11*, 2644–2651. [[CrossRef](#)]
53. Fattorusso, A.; Di Genova, L.; Dell’Isola, G.B.; Mencaroni, E.; Esposito, S. Autism spectrum disorders and the gut microbiota. *Nutrients* **2019**, *11*, 521. [[CrossRef](#)]
54. Baldani, J.I.; Baldani, V.L.D.; Seldin, L.; Dobereiner, J. Characterization of *Herbaspirillum seropedicae* gen. nov., sp. nov., a Root-Associated Nitrogen-Fixing Bacterium. *Int. J. Syst. Bacteriol.* **1986**, *36*, 86–93. [[CrossRef](#)]
55. Ziga, E.D.; Druley, T.; Burnham, C.A. *Herbaspirillum* species bacteremia in a pediatric oncology patient. *J. Clin. Microbiol.* **2010**, *48*, 4320–4321. [[CrossRef](#)]
56. Chen, J.; Su, Z.; Liu, Y.; Sandoghchian, S.; Zheng, D.; Wang, S.; Xu, H. *Herbaspirillum* species: A potential pathogenic bacteria isolated from acute lymphoblastic leukemia patient. *Curr. Microbiol.* **2011**, *62*, 331–333. [[CrossRef](#)] [[PubMed](#)]
57. Regunath, H.; Kimball, J.; Smith, L.P.; Salzer, W. Severe Community-Acquired Pneumonia with Bacteremia Caused by *Herbaspirillum aquaticum* or *Herbaspirillum huttiense* in an Immune-Competent Adult. *J. Clin. Microbiol.* **2015**, *53*, 3086–3088. [[CrossRef](#)] [[PubMed](#)]
58. Chemaly, R.F.; Dantes, R.; Shah, D.P.; Shah, P.K.; Pascoe, N.; Ariza-Heredia, E.; Perego, C.; Nguyen, D.B.; Nguyen, K.; Modarai, F.; et al. Cluster and sporadic cases of *herbaspirillum* species infections in patients with cancer. *Clin. Infect. Dis.* **2015**, *60*, 48–54. [[CrossRef](#)] [[PubMed](#)]

59. Suwantararat, N.; Adams, L.L.; Romagnoli, M.; Carroll, K.C. Fatal case of *Herbaspirillum seropedicae* bacteremia secondary to pneumonia in an end-stage renal disease patient with multiple myeloma. *Diagn. Microbiol. Infect. Dis.* **2015**, *82*, 331–333. [[CrossRef](#)] [[PubMed](#)]
60. Tan, M.J.; Oehler, R.L. Lower Extremity Cellulitis and Bacteremia with *Herbaspirillum seropedicae* Associated with Aquatic Exposure in a Patient with Cirrhosis. *Infect. Dis. Clin. Pract.* **2005**, *13*, 277–279. [[CrossRef](#)]
61. Spilker, T.; Uluer, A.Z.; Marty, F.M.; Yeh, W.W.; Levison, J.H.; Vandamme, P.; Lipuma, J.J. Recovery of *Herbaspirillum* species from persons with cystic fibrosis. *J. Clin. Microbiol.* **2008**, *46*, 2774–2777. [[CrossRef](#)]
62. Marques, A.C.; Paludo, K.S.; Dallagassa, C.B.; Surek, M.; Pedrosa, F.O.; Souza, E.M.; Cruz, L.M.; LiPuma, J.J.; Zanata, S.M.; Rego, F.G.; et al. Biochemical characteristics, adhesion, and cytotoxicity of environmental and clinical isolates of *Herbaspirillum* spp. *J. Clin. Microbiol.* **2015**, *53*, 302–308. [[CrossRef](#)]
63. Routy, B.; Le Chatelier, E.; Derosa, L.; Duong, C.P.M.; Alou, M.T.; Daillere, R.; Fluckiger, A.; Messaoudene, M.; Rauber, C.; Roberti, M.P.; et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* **2018**, *359*, 91–97. [[CrossRef](#)]
64. Anhe, F.F.; Nachbar, R.T.; Varin, T.V.; Vilela, V.; Dudonne, S.; Pilon, G.; Fournier, M.; Lecours, M.A.; Desjardins, Y.; Roy, D.; et al. A polyphenol-rich cranberry extract reverses insulin resistance and hepatic steatosis independently of body weight loss. *Mol. Metab.* **2017**, *6*, 1563–1573. [[CrossRef](#)]
65. Bajaj, J.S.; Heuman, D.M.; Hylemon, P.B.; Sanyal, A.J.; White, M.B.; Monteith, P.; Noble, N.A.; Unser, A.B.; Daita, K.; Fisher, A.R.; et al. Altered profile of human gut microbiome is associated with cirrhosis and its complications. *J. Hepatol.* **2014**, *60*, 940–947. [[CrossRef](#)]
66. Bajaj, J.S.; Fagan, A.; White, M.B.; Wade, J.B.; Hylemon, P.B.; Heuman, D.M.; Fuchs, M.; John, B.V.; Acharya, C.; Sikaroodi, M.; et al. Specific Gut and Salivary Microbiota Patterns Are Linked with Different Cognitive Testing Strategies in Minimal Hepatic Encephalopathy. *Am. J. Gastroenterol.* **2019**, *114*, 1080–1090. [[CrossRef](#)] [[PubMed](#)]
67. Felipo, V.; Urios, A.; Montesinos, E.; Molina, I.; Garcia-Torres, M.L.; Civera, M.; Olmo, J.A.; Ortega, J.; Martinez-Valls, J.; Serra, M.A.; et al. Contribution of hyperammonemia and inflammatory factors to cognitive impairment in minimal hepatic encephalopathy. *Metab. Brain Dis.* **2012**, *27*, 51–58. [[CrossRef](#)] [[PubMed](#)]
68. Karanfilian, B.V.; Park, T.; Senatore, F.; Rustgi, V.K. Minimal hepatic encephalopathy. *Clin. Liver Dis.* **2020**, *24*, 209–218. [[CrossRef](#)] [[PubMed](#)]
69. Zhang, Z.; Zhai, H.; Geng, J.; Yu, R.; Ren, H.; Fan, H.; Shi, P. Large-scale survey of gut microbiota associated with MHE Via 16S rRNA-based pyrosequencing. *Am. J. Gastroenterol.* **2013**, *108*, 1601–1611. [[CrossRef](#)] [[PubMed](#)]