Innovative Metabolomics Insights for Better Health

Conjoint metabolome and microbial amplicon analysis report

Metware Biotechnology Inc.

www.metwarebio.com

Contents

MWXS-001Conjoint metabolome and microbial amplicon analysis report

1 Abstract

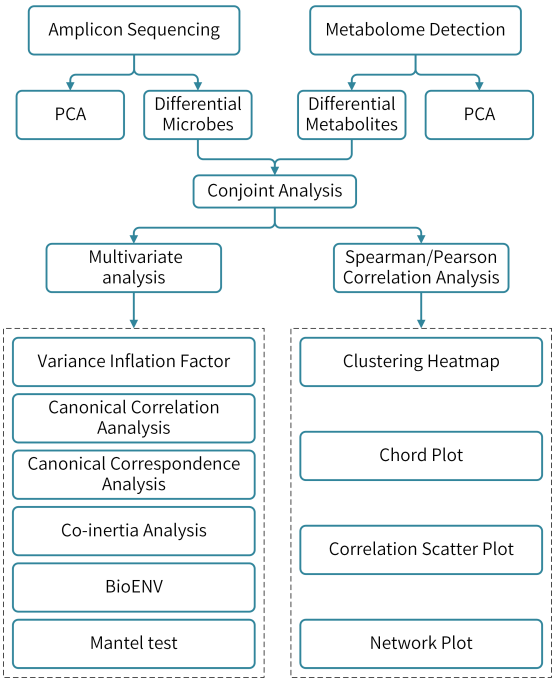
Metabolomics: Metabolites is a direct readout of an organism's phenotype, providing an intuitive and effective means to comprehend biological processes and their underlying mechanisms. By qualitatively and quantitatively analyzing metabolites, metabolomics enables the investigation of metabolic pathways or networks related to macro-phenotypic phenomena across diverse organisms, the evaluation of economically important traits in animals, the development of methdos for disease diagnosis and prevention, the safety assessment of animal products, and the characterizing of animal models in medical research.

Microbiome: The "microbial community" serves as a key driving force in the biogeochemical cycling of essential elements that sustain life on Earth. Comprised of multiple populations, these communities exhibit intricate relationships encompassing symbiosis, mutualism, coexistence, and competition among different microbial populations. Complex microbial communities influence various environments, including the gastrointestinal tract of mammals, soil ecosystems, and aquatic bodies. Recent studies have revealed the profound association between gut bacteria and the overall health and disease status of animals.

To maintain a stable ecological niche, microorganisms exert influence over multiple metabolic pathways in the host through the exchange of metabolic signals with the host's microbiota. This strong interaction significantly impacts the host's metabolic phenotype, contributes to the host's evolutionary adaptations, and fosters a mutually beneficial relationship between the host and microbiota. In the realm of microbiota research, two of the pivotal technical tools are microbiome analysis (amplicon/metagenome) and metabolomics. The microbiome enables the identification of structural and abundance variations within the microbial community and aids in predicting or annotating their functional disparities. Conversely, the metabolome provides a direct reflection of the functional activities of the microbiota and their interactions with the host. Both techniques are complementary and essential. Thus, integrating microbiomics and metabolomics offers valuable insights into understanding how environmental microbiota influence the metabolic state of the environment or host through colony metabolism and co-metabolism with the host.

2 Bioinformatics Analysis Process

Metabolomics and microbial amplicon sequencing are conducted on the same set of samples, followed by correlation analysis of the differential metabolites and microbial communities.



Bioinformatics Analysis Process

3 Sample Information

Sample information in the coanalysis

Table 1 Sample Information Sheet

Sample	Group	Meta_sample	Meta_group	Micro_sample	Micro_group
AA1	AA	AA-1	AA	AA1	AA
AA2	AA	AA-2	AA	AA2	AA
AA3	AA	AA-3	AA	AA3	AA
BB1	BB	BB-1	BB	BB1	BB
BB2	BB	BB-2	BB	BB2	BB
BB3	BB	BB-3	BB	BB3	BB

Table 1 Sample Information Sheet Continued table

Sample	Group	Meta_sample	Meta_group	Micro_sample	Micro_group

File path: 1.Data/sample.xlsx

- Sample, Group;
- Metabolome_sample, Metabolome_group：Sample names and grouping information used in metabolome analysis;
- Microbiome_sample, Microbiome_group：Sample names and grouping information used in microbiome analysis;

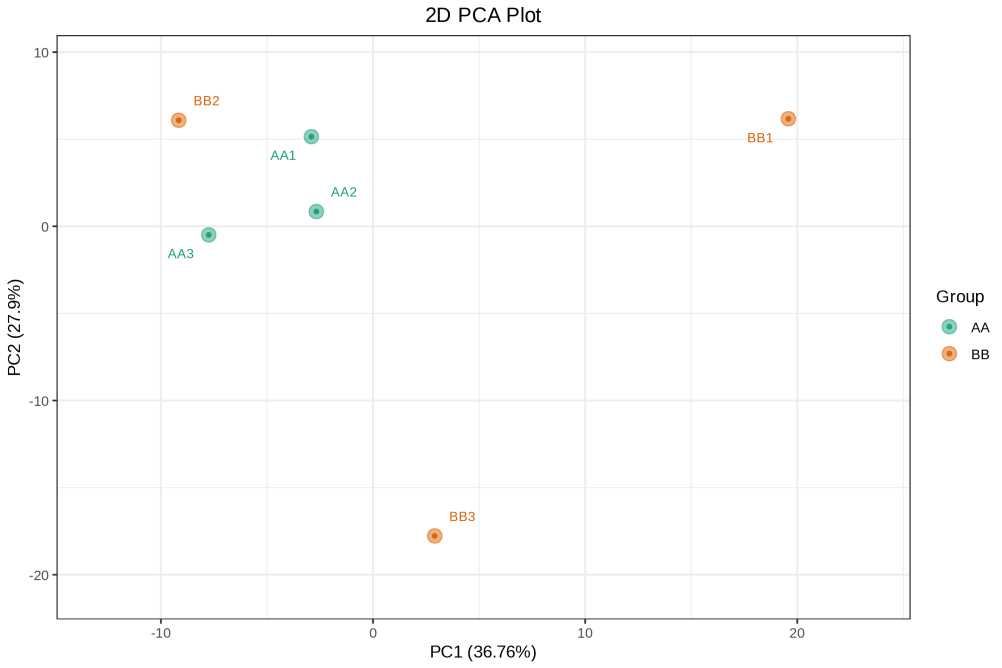
4 Principal Component Analysis

Principal component analysis is an unsupervised statistical analysis method that converts a set of potentially correlated variables into a set of linearly uncorrelated variables by orthogonal transformation. The converted set of variables are called principal components. This analysis is often used to study how to reveal the internal structure among multiple variables through a few principal components, i.e., to derive a few principal components from the original variables so that they retain as much information as possible about the original variables and are uncorrelated with each other. The usual mathematical processing is to make a linear combination of the original multiple indicators as a new composite indicator.

PCA analysis on the metabolome and microbiome separately visualizes whether there are differences between sample groups in the metabolome and microbiome, respectively. PCA is performed using the prcomp function of the R software (www.r-project.org) (version 4.1.2, same below). Data are UV normalized before analysis, i.e., each metabolite or microorganism is subtracted from the mean and divided by the standard deviation across all samples.

4.1 Principal Component Analysis of Metabolites

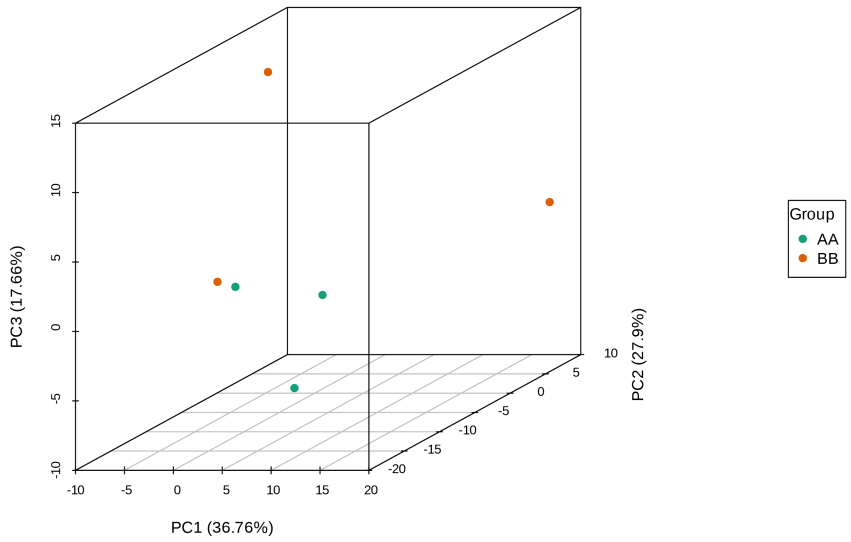
The results of the first two PCA principal components are plotted below:



2D PCA of Metabolites

PC1 denotes the first principal component, PC2 denotes the second principal component. Percentages indicate the contribution of the principal components to the differences in the samples. Each point in the graph represents one sample, and samples from the same group are represented by the same color. If there are more than three samples per group, an elliptical confidence interval will be added. Confidence intervals can also be added if desired for groups containing 3 samples.

The results of the first three PCA principal components are plotted below:



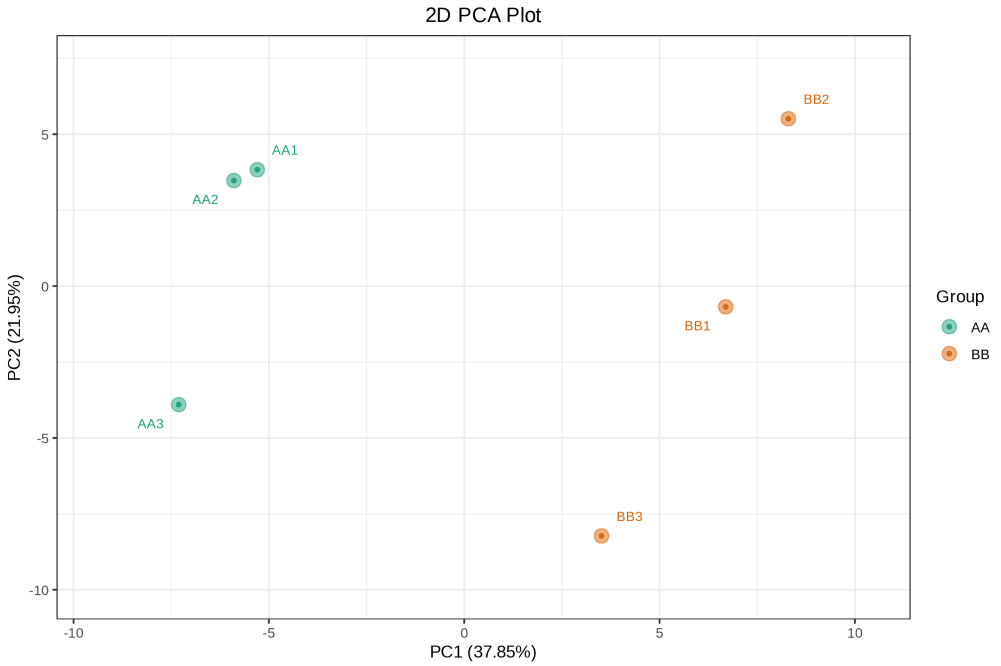
3D PCA of Metabolites

PC1 denotes the first principal component, PC2 denotes the second principal component. Percentages indicate the contribution of the principal components to the differences in the samples. Each point in the graph represents one sample, and samples from the same group are represented by the same color.

File path: 2.Basic_analysis/metabolome_PCA

4.2 Principal Component Analysis of Microbiota

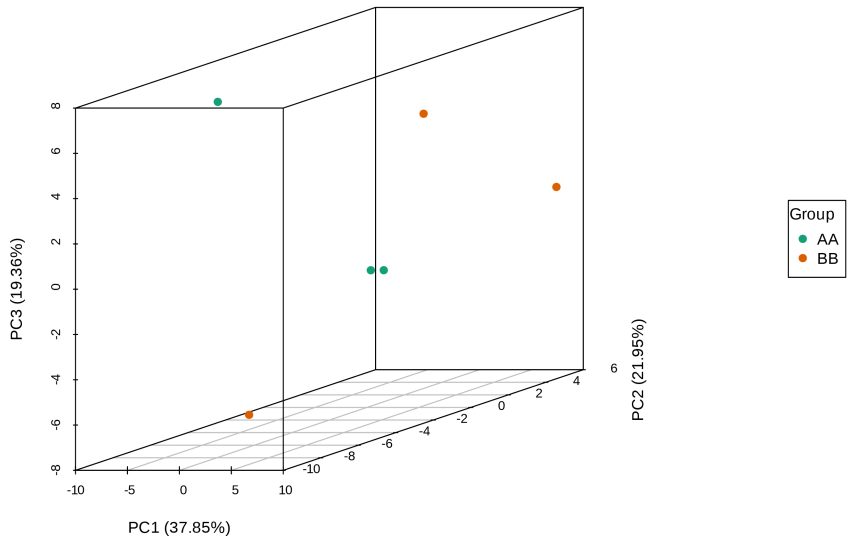
The results of the first two PCA principal components are plotted below:



2D PCA of Microbes

PC1 denotes the first principal component, PC2 denotes the second principal component. Percentages indicate the contribution of the principal components to the differences in the samples. Each point in the graph represents one sample, and samples from the same group are represented by the same color. If there are more than three samples per group, an elliptical confidence interval will be added. Confidence intervals can also be added if desired for groups containing 3 samples.

The results of the first three PCA principal components are plotted below:



3D PCA of Microbes

PC1 denotes the first principal component, PC2 denotes the second principal component, PC2 denotes the third principal component. Percentages indicate the contribution of the principal components to the differences in the samples. Each point in the graph represents one sample, and samples from the same group are represented by the same color.

File path: 2.Basic_analysis/microbiome_PCA

5 Spearman Correlation Analysis

Spearman rank correlation analysis describes the correlation between two aggregates using the Spearman correlation coefficient as the indicator and the rank correlation test to determine if there is a statistically significant correlation between the two aggregates. Unlike the Pearson correlation test, which generally requires the aggregate to follow a normal distribution, the Spearman rank correlation test requires a rank statistic with no requirement for the distribution form of the aggregates. Spearman correlation coefficient ranges from [-1, 1], with positive numbers indicating positive correlation and negative numbers indicating negative correlation, and the larger the absolute value, the greater the correlation. An absolute value of 1 indicates perfect correlation. The correlation analysis was calculated using the cor function of R software, and the significance test of correlation was calculated using the corPvalueStudent function of the WGCNA package of R software.

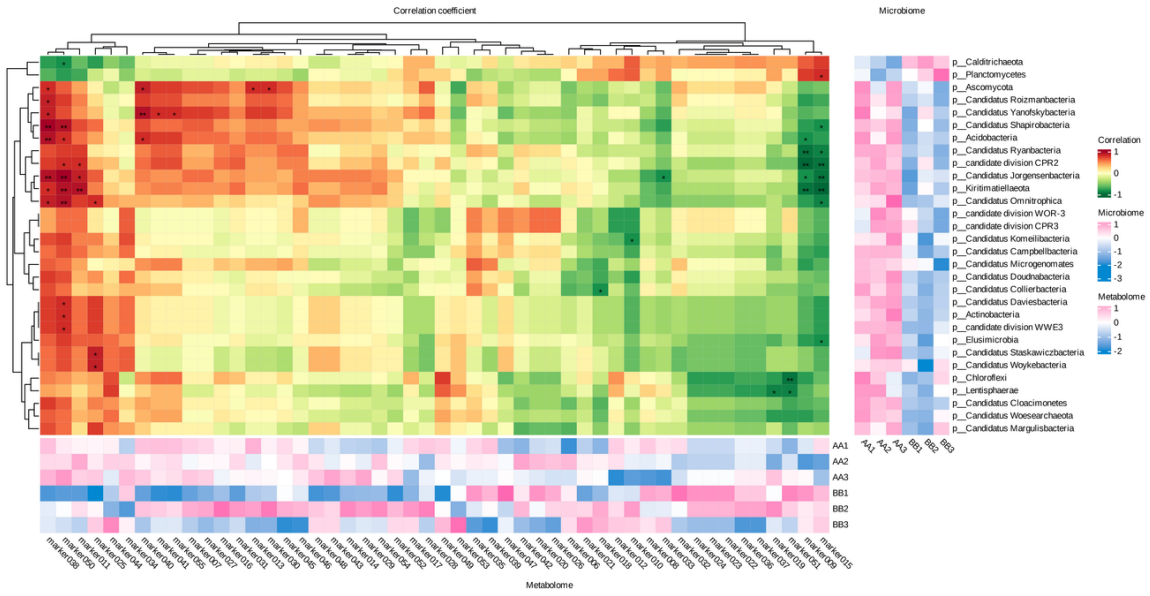
Both Spearman correlation analysis and Pearson correlation analysis are commonly used measures of correlation in multi-omics studies. Due to the complexity of organism regulation and high-throughput data, it is impossible to determine which correlation calculation method is the best. Therefore, both analysis strategies are provided in the final report, which can be selected as needed.

Note:

- For combined correlation heatmap and hierarchical clustering heatmap, partial data is extracted by default, and the default value is 50.
  - If the input differential metabolite file contains VIP (Variable Importance in Projection),then the metabolites will be sorted by VIP from largest to smallest and the top ranked metabolite will be taken. If it does not, the order of input will prevail.
  - Microbes are sorted by the sum of the relative quantitative values in all samples, from largest to smallest.
- When plotting the heatmap of correlation between top 20 differential metabolites and microbes, the top 20 metabolites were taken and the top 50 microbes were taken by default, and the ordering method was the same as described before.
- When plotting chord diagrams and correlation scatter plots, the 50 most correlated data are plotted by default.

5.1 Correlation Clustering Heatmap

Correlation analysis of differential microbes and differential metabolites was performed to calculate the Spearman correlation coefficients of microbes and metabolites. The correlation between differential microbes and differential metabolites at different taxonomic levels was demonstrated by clustering heatmaps. The heatmaps were drawn using the ComplexHeatmap package of R software.



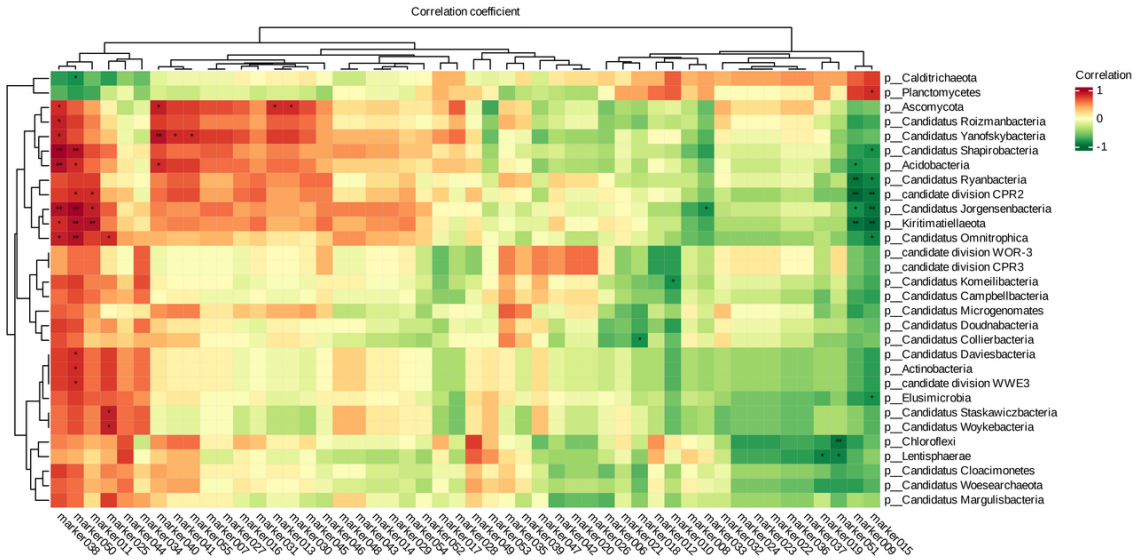
Spearman correlation clustering heatmap of differential microbes against differential metabolites

The heatmap in the center shows the magnitude of Spearman correlation. Horizontal coordinates represent metabolites and vertical coordinates represent microbes, where * denotes a P value < 0.05 for the significance test of the correlation coefficient, and ** denotes a P value < 0.01. The heatmap on the right shows the abundance of microbes at different taxonomic levels, and the lower heatmap shows the abundance of metabolites. Both microbe and metabolite abundance data were standardized using Z-score.

File path: 3.Advanced_analysis/*/1.spearman/combine_heatmap (* means differential group)

5.2 Hierarchical Clustering Heatmap of Correlation

In order to visualize the similarities and differences in the expression patterns of differential microbes and differential metabolites, Sperman correlation hierarchical clustering analysis was performed on differential microbes and differential metabolites. The closer the branches are, the more similar the expression patterns of the microbes or metabolites are. The heatmaps were drawn using the ComplexHeatmap package of R software.



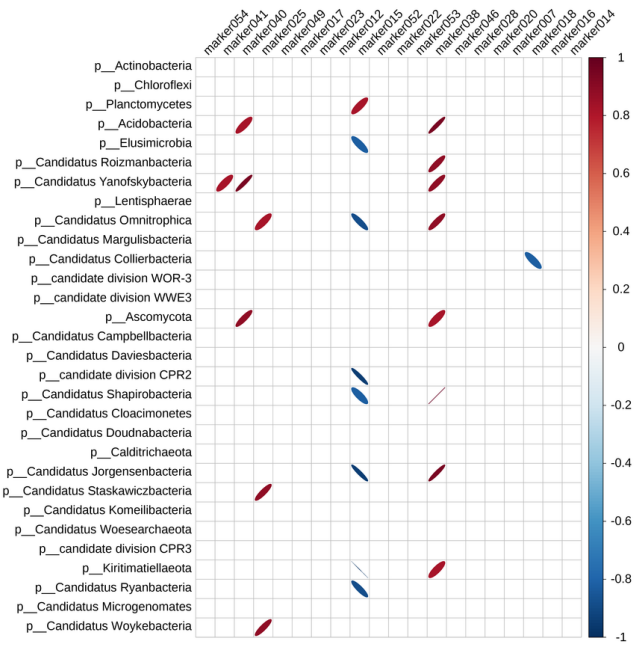
Spearman Correlation Hierarchical Clustering Heatmap of differential microbes against differential metabolites of differential microbes against differential metabolites

Microbes are listed per row and metabolites per column. The evolutionary tree on the left represents the results of hierarchical clustering of microbes, and the evolutionary tree on the top represents the results of hierarchical clustering of metabolites. Red color indicates a positive correlation and green color indicates a negative correlation. Significant differences are indicated by ”*” for P-value < 0.05 and highly significant differences are indicated by ”**” for P-value < 0.01 for the significance test of correlation coefficients.

File path: 3.Advanced_analysis/*/1.spearman/correlation_heatmap (* means differential group)

5.3 Heatmap of correlation between top 20 differential metabolites and microbes

The correlation data of the top 20 differential metabolites with differential microbes were extracted to plot heatmaps, or if the number of differential metabolites was less than 20, all the data were used to plot heatmaps. The heatmaps were drawn using the corrplot package of R software.



Spearman correlation heatmap of differential metabolites against differential microbes

Microbes are listed per row and metabolites per column. Red ellipses indicate positive correlations and blue ellipses indicate negative correlations. The larger the absolute value of the correlation, the thinner the ellipse. A blank grid indicates a significance test P-value greater than 0.05.

File path: 3.Advanced_analysis/*/1.spearman/ellipse_heatmap (* means differential group)

5.4 Significantly correlated microbes and metabolites

Differential microbes and metabolites were significantly correlated if the correlation |r| >= 0.8 and the P-value of the correlation coefficient significance test < 0.05. The Spearman correlation coefficient and correlation test results for the differential microbes and differential metabolites are shown in the table below (only one differential grouping is shown here, please see the the final report for full results).

Table 2 Spearman Correlation Coefficient of differential microbes against differential metabolites

Index	Taxonomy	Correlation	P-value	Log2FC
marker041	k__Bacteria;p__Candidatus Yanofskybacteria	0.8285714	0.0415627	0.9624409
marker040	k__Bacteria;p__Acidobacteria	0.8285714	0.0415627	3.1390399
marker040	k__Bacteria;p__Candidatus Yanofskybacteria	0.9428571	0.0048047	3.1390399
marker040	k__Eukaryota;p__Ascomycota	0.8857143	0.0188455	3.1390399
marker025	k__Bacteria;p__Candidatus Omnitrophica	0.8285714	0.0415627	0.6264040
marker025	k__Bacteria;p__Candidatus Staskawiczbacteria	0.8857143	0.0188455	0.6264040

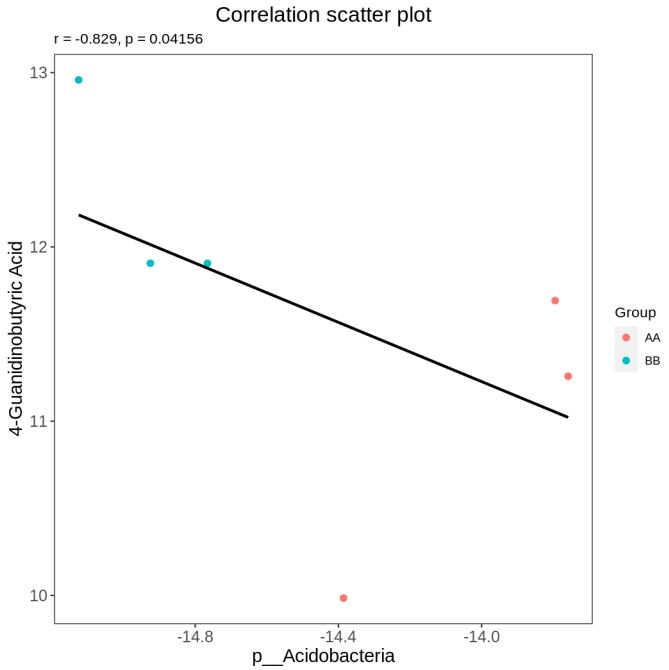
File path: 3.Advanced_analysis/*/1.spearman/*.spearman_correlation.filtered.xlsx

- Index: metabolite ID
- Taxonomy: Taxonomic information of microbes
- Correlation: Spearman correlation coefficient
- P-value: P-value of the significance test
- Compounds: name of the metabolite

5.5 Correlation Scatter Plot

Correlation scatterplot is a commonly used visual analysis method in correlation analysis to visualize the trend of the relationship between two variables. The scatterplot reflects the correlation between a particular significantly different microbe and metabolite. If the two are completely linearly correlated, all the data dots fall on the fitted straight line; if they are partially linearly correlated, the data dots fall on both sides of the straight line; and if they are linearly uncorrelated, the distribution of data dots for the two variables is scattered without any pattern.

The correlation data were filtered by correlation $|r| >= 0.8$ and significance test P-value < 0.05, and the 50 with the highest correlation was selected for plotting. Correlation scatter plots were generated using the ggplot2 package of R software. Only one classification is shown in the report, please see the the final report for full results.



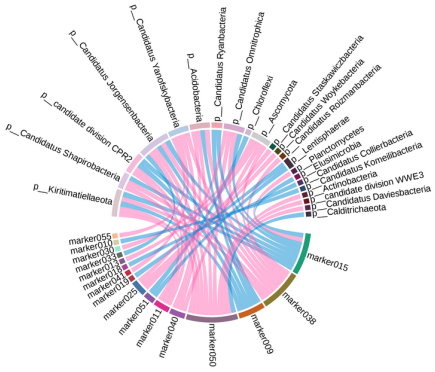
Spearman Correlation Scatter Plot of differential microbes against differential metabolites

The horizontal coordinates indicate the relative abundance of microbes; the vertical coordinates indicate the relative expression of metabolites; the dots represent samples, with different colors representing different groups; the r in the upper left corner is the correlation coefficient, and the P value is the significance level of the correlation.

File path: 3.Advanced_analysis/*/1.spearman/scatter (* means differential group)

5.6 Correlation Chord Diagram

A chord diagram is a visualization method to show the inter-relationships between data, where the node data are arranged radially along the circumference of a circle and the nodes are connected with arcs with weights (with width). Data with correlation |r| >= 0.8 and P-value < 0.05 for correlation coefficient significance test were selected for plotting, and only up to 50 correlation data were shown. Chord diagrams were plotted using the circlize package of the R software.



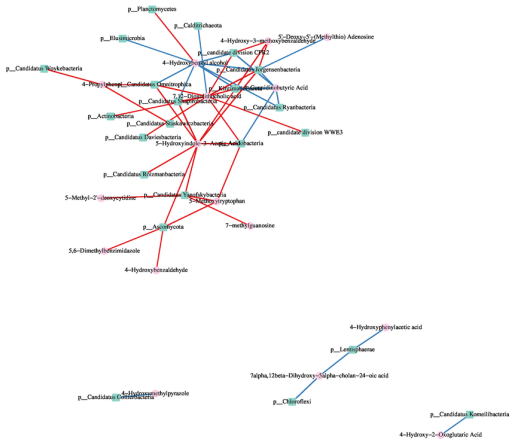
Spearman Correlation Chord Diagram of differential microbes against differential metabolites

The width of the link indicates the magnitude of the correlation between the two objects, the wider the link, the greater the absolute value of the correlation. Pink color indicates a positive correlation and blue color indicates a negative correlation.

File path: 3.Advanced_analysis/*/1.spearman/circos (* means differential group)

5.7 Correlation Network Plot

Network plots can be used to show correlations between microbes and metabolites, providing a new perspective for studying correlations between microbes and metabolites with significant differences. Correlation data with correlation |r| >= 0.8 and significance test P-value < 0.05 were selected for plotting. The 2D network diagrams were plotted using the igraph package of the R software; the 3D network diagrams were plotted using the networkD3 package of the R software; and the dynamic network diagrams are shown in the final report.



Spearman Correlation Network Plot of differential microbes against differential metabolites

Metabolites are shown in pink and microbes in light green. Connections between microbes and metabolites indicate correlations, with red indicating positive correlations and blue indicating negative correlations, and thicker lines indicating stronger correlations. The size of nodes in a dynamic network graph indicates the magnitude of the degree, i.e. the more edges a node is connected to, the larger it is.

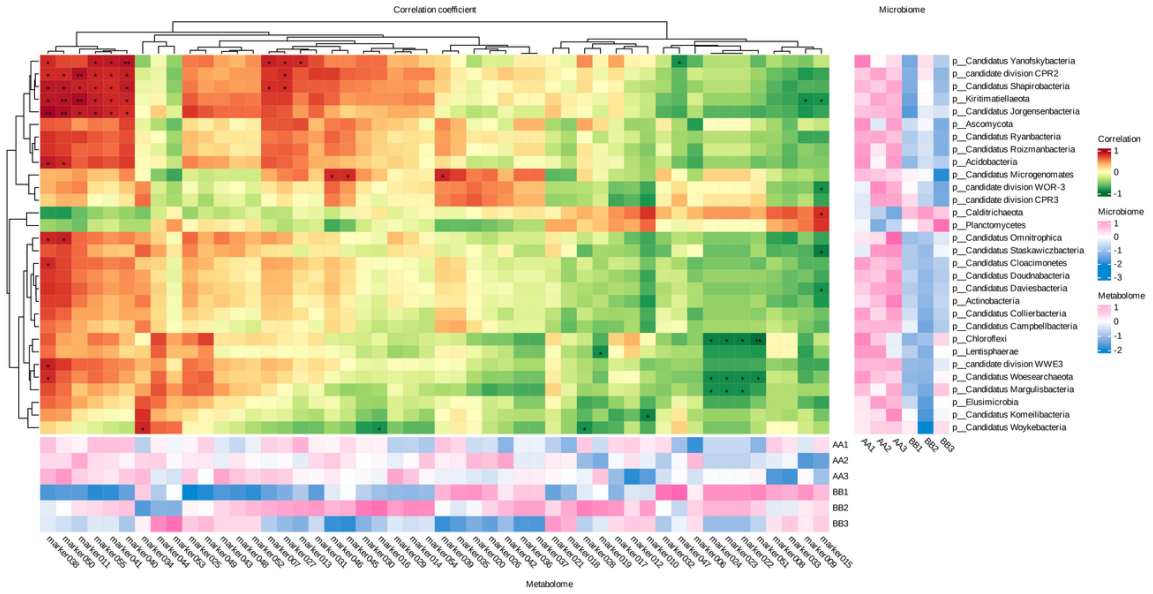
File path: 3.Advanced_analysis/*/1.spearman/network (* means differential group)

6 Pearson Correlation Analysis

Pearson correlation coefficient measures the degree of linear correlation between two variables X and Y. It is equal to the covariance of X and Y divided by the standard deviation of X and Y. Pearson correlation coefficient ranges from [-1, 1], with positive numbers indicating positive correlation and negative numbers indicating negative correlation, and the larger the absolute value, the greater the correlation. An absolute value of 1 indicates perfect linear correlation. The correlation analysis was calculated using the cor function of R software, and the significance test of correlation was calculated using the corPvalueStudent function of the WGCNA package of R software.

6.1 Correlation Clustering Heatmap

Correlation analysis of differential microbes and differential metabolites was performed to calculate the Pearson correlation coefficients of microbes and metabolites. The correlation between differential microbes and differential metabolites at different taxonomic levels was demonstrated by clustering heatmaps. The heatmaps were drawn using the ComplexHeatmap package of R software.



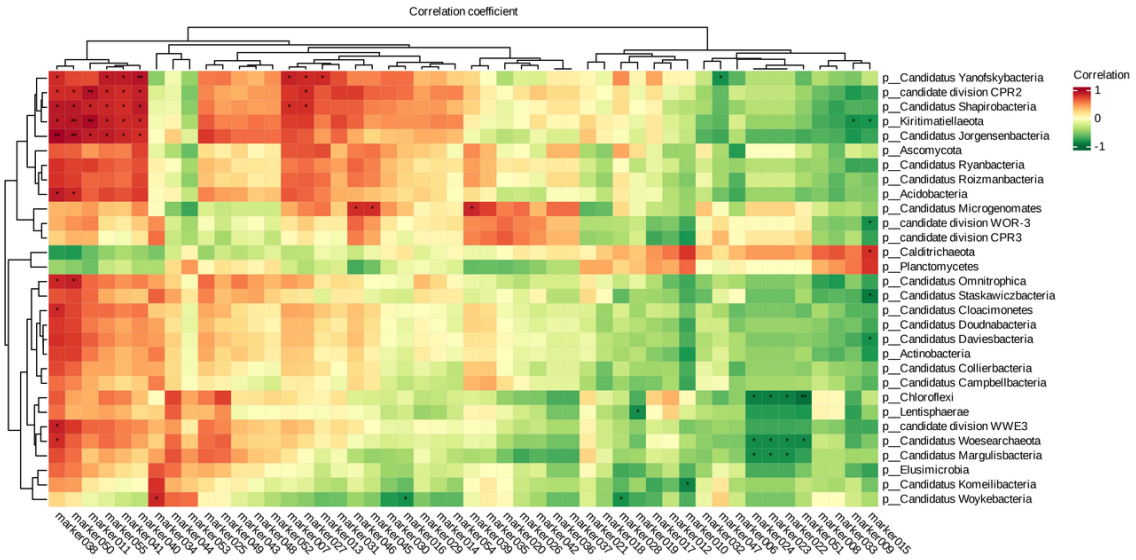
Pearson Correlation Clustering Heatmap of differential microbes against differential metabolites

The heatmap in the center shows the magnitude of Pearson correlation. Horizontal coordinates represent metabolites and vertical coordinates represent microbes, where * denotes a P value < 0.05 for the significance test of the correlation coefficient, and ** denotes a P value < 0.01. The heatmap on the right shows the abundance of microbes at different taxonomic levels, and the lower heatmap shows the abundance of metabolites. Both microbe and metabolite abundance data were standardized using Z-score.

File path: 3.Advanced_analysis/*/2.pearson/combine_heatmap (* means differential group)

6.2 Hierarchical Clustering Heatmap of Correlation

In order to visualize the similarities and differences in the expression patterns of differential microbes and differential metabolites, Pearson correlation hierarchical clustering analysis was performed on differential microbes and differential metabolites. The closer the branches are, the more similar the expression patterns of the microbes or metabolites are. The heatmaps were drawn using the ComplexHeatmap package of R software.



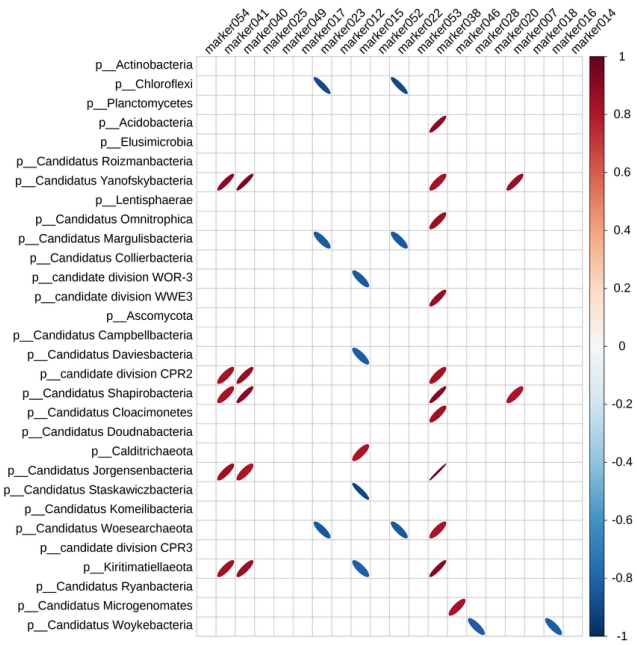
Pearson Correlation Hierarchical Clustering Heatmap of differential microbes against differential metabolites

Microbes are listed per row and metabolites per column. The evolutionary tree on the left represents the results of hierarchical clustering of microbes, and the evolutionary tree on the top represents the results of hierarchical clustering of metabolites. Red color indicates a positive correlation and green color indicates a negative correlation. Significant differences are indicated by "*" for P-value < 0.05 and highly significant differences are indicated by "**" for P-value < 0.01 for the significance test of correlation coefficients.

File path: 3.Advanced_analysis/*/2.pearson/correlation_heatmap (* means differential group)

6.3 Heatmap of correlation between top 20 differential metabolites and microbes

The correlation data of the top 20 differential metabolites with differential microbes were extracted to plot heatmaps, or if the number of differential metabolites was less than 20, all the data were used to plot heatmaps. The heatmaps were drawn using the corrplot package of R software.



Pearson correlation heatmap of differential metabolites against differential microbes

Microbes are listed per row and metabolites per column. Red ellipses indicate positive correlations and blue ellipses indicate negative correlations. The larger the absolute value of the correlation, the thinner the ellipse. A blank grid indicates a significance test P-value greater than 0.05.

File path: 3.Advanced_analysis/*/2.pearson/ellipse_heatmap (* means differential group)

6.4 Significantly correlated microbes and metabolites

Differential microbes and metabolites were significantly correlated if the correlation |r| >= 0.8 and the P-value of the correlation coefficient significance test < 0.05. The Pearson correlation coefficient and correlation test results for the differential microbes and differential metabolites are shown in the table below (only one differential grouping is shown here, please see the the final report for full results).

Table 3 Pearson Correlation Coefficient of differential microbes against differential metabolites

Index	Taxonomy	Correlation	P-value	Log2FC
marker041	k__Bacteria;p__Candidatus Yanofskybacteria	0.9044780	0.0132509	0.9624409
marker041	k__Bacteria;p__candidate division CPR2	0.8323950	0.0397830	0.9624409
marker041	k__Bacteria;p__Candidatus Shapirobacteria	0.8205697	0.0454045	0.9624409
marker041	k__Bacteria;p__Candidatus Jorgensenbacteria	0.8515811	0.0314075	0.9624409
marker041	k__Bacteria;p__Kiritimatiellaeota	0.8431403	0.0349777	0.9624409
marker040	k__Bacteria;p__Candidatus Yanofskybacteria	0.9308526	0.0070067	3.1390399

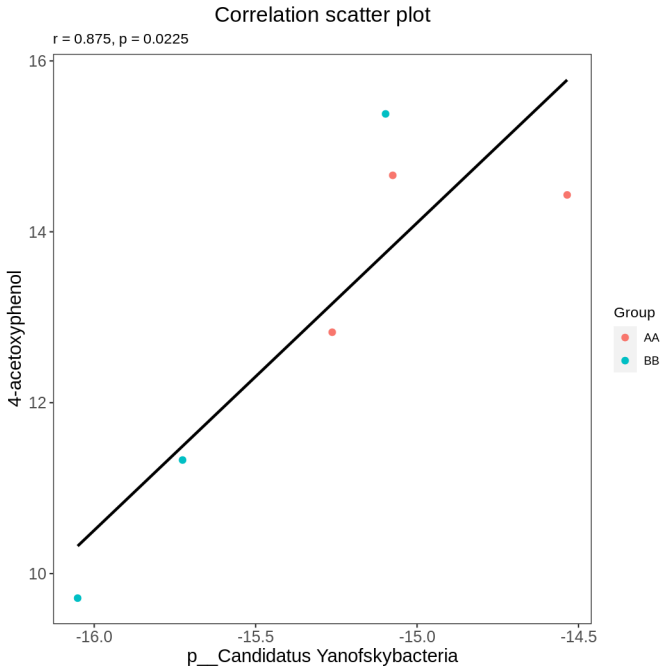
File path: 3.Advanced_analysis/*/2.pearson/*.pearson_correlation.filtered.xlsx

- Index: metabolite ID
- Taxonomy: Taxonomic information of microbes
- Correlation: Pearson correlation coefficient
- P-value: P-value of the significance test
- Compounds: name of the metabolite

6.5 Correlation Scatter Plot

Correlation scatterplot is a commonly used visual analysis method in correlation analysis to visualize the trend of the relationship between two variables. The scatterplot reflects the correlation between a particular significantly different microbe and metabolite. If the two are completely linearly correlated, all the data dots fall on the fitted straight line; if they are partially linearly correlated, the data dots fall on both sides of the straight line; and if they are linearly uncorrelated, the distribution of data dots for the two variables is scattered without any pattern.

The correlation data were filtered by correlation |r| >= 0.8 and significance test P-value < 0.05, and the 50 with the highest correlation was selected for plotting. Correlation scatter plots were generated using the ggplot2 package of R software. Only one classification is shown in the report, please see the the final report for full results.



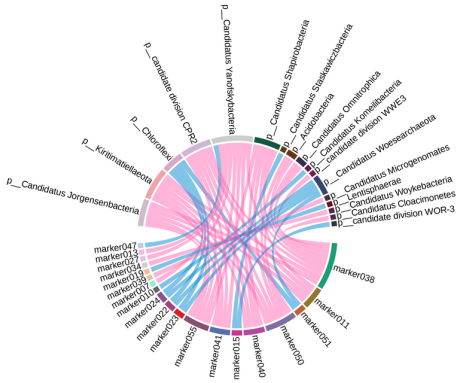
Pearson Correlation Scatter Plot of differential microbes against differential metabolites

The horizontal coordinates indicate the relative abundance of microbes; the vertical coordinates indicate the relative expression of metabolites; the dots represent samples, with different colors representing different groups; the r in the upper left corner is the correlation coefficient, and the P value is the significance level of the correlation.

File path: 3.Advanced_analysis/*/2.pearson/scatter (* means differential group)

6.6 Correlation Chord Diagram

A chord diagram is a visualization method to show the inter-relationships between data, where the node data are arranged radially along the circumference of a circle and the nodes are connected with arcs with weights (with width). Data with correlation |r| >= 0.8 and P-value < 0.05 for correlation coefficient significance test were selected for plotting, and only up to 50 correlation data were shown. Chord diagrams were plotted using the circlize package of the R software.



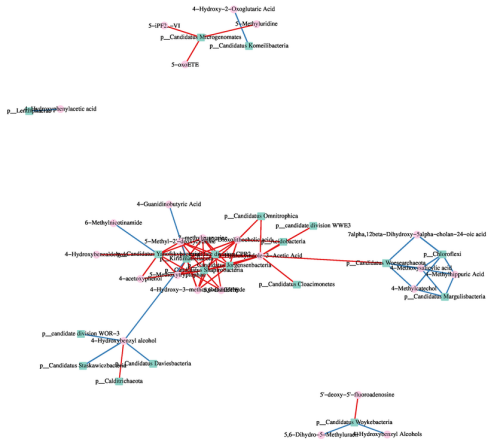
Pearson Correlation Chord Diagram of differential microbes against differential metabolites

The width of the link indicates the magnitude of the correlation between the two objects, the wider the link, the greater the absolute value of the correlation. Pink color indicates a positive correlation and blue color indicates a negative correlation.

File path: 3.Advanced_analysis/*/2.pearson/circos (* means differential group)

6.7 Correlation Network Plot

Network plots can be used to show correlations between microbes and metabolites, providing a new perspective for studying correlations between microbes and metabolites with significant differences. Correlation data with correlation |r| >= 0.8 and significance test P-value < 0.05 were selected for plotting. The 2D network diagrams were plotted using the igraph package of the R software; the 3D network diagrams were plotted using the networkD3 package of the R software; and the dynamic network diagrams are shown in the final report.



Pearson Correlation Network Plot of differential microbes against differential metabolites

Metabolites are shown in pink and microbes in light green. Connections between microbes and metabolites indicate correlations, with red indicating positive correlations and blue indicating negative correlations, and thicker lines indicating stronger correlations. The size of nodes in a dynamic network graph indicates the magnitude of the degree, i.e. the more edges a node is connected to, the larger it is.

File path: 3.Advanced_analysis/*/2.pearson/network (* means differential group)

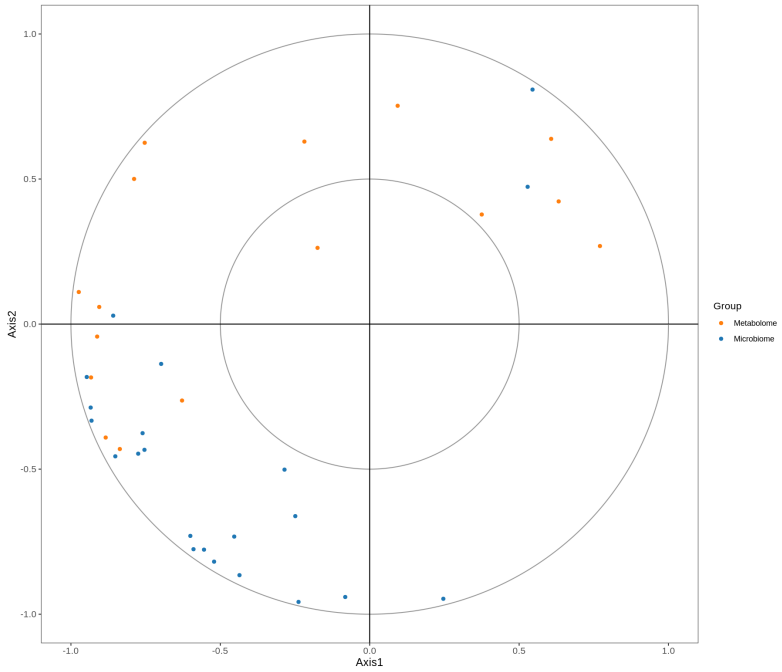
7 Multivariate analysis

Multivariate analysis was performed using the R software (v4.2.0).

7.1 Canonical Correlation Analysis

Canonical correlation analysis (CCA) is a multivariate statistical analysis method used to assess the overall correlation by calculating the correlation between two sets of data by using composite variables. CCA analysis was implemented using the CCA package (v1.2.1) in R.

The results of Spearman correlation |r| >= 0.8 and P-value < 0.05 of correlation coefficient significance test were screened first for differential metabolites and differential microbes at different classification levels, and then typical correlation analysis was performed on the screened metabolites and microbes. The results are shown in the figure below (here shows the plotting results without labels; please see the final report for the complete results).



Canonical correlation analysis of differential microbes and differential metabolites

The metabolites are indicated in orange, while the microbes are in blue. The graph is divided into four quadrants by the cross. The further from the origin and the closer to each other within the same quadrant means the higher the typical correlation. From left to right are the different classification levels of microbes, including Phylum, Class, Order, Family, Genus, and Species.

File path: 3.Advanced_analysis/*/3.multivariate/1.canonical_correlation_analysis (* means differential group)

7.2 VIF screening of Metabolites

There is often multicollinearity between metabolites. Multicollinearity means that metabolites are highly correlated with each other, which can lead to model distortion or difficulty in making accurate estimates.

Therefore, metabolites need to be filtered prior to analysis, and only metabolites that are independent of each other are retained.

The variance inflation factor (VIF) is a measure of the severity of covariance of variables. A VIF value is calculated for each candidate metabolite. Usually metabolites with a VIF value greater than 10 are considered to be colinear. Multiple calculations are performed, eliminating factors with VIF greater than 10 one by one, until all remaining metabolites are with VIF values less than 10. The VIF analysis was performed using the vegan package (v2.6.2) of the R software. The metabolites after eliminating co-linearity are shown below:

Table 4 Rejecting colinear differential metabolites

Environment	VIF
marker017	1.473857
marker040	2.321934
marker054	2.586543
marker023	4.957472
marker025	7.510599

File path: 3.Advanced_analysis/*/3.multivariate/VIF.xlsx

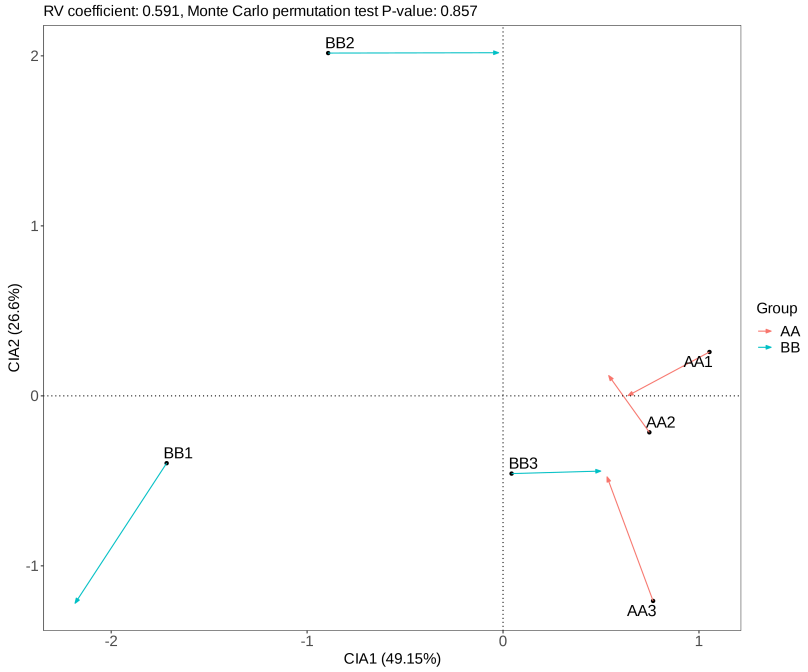
- Environment: metabolite
- VIF: variance inflation factor

7.3 Coinertia Analysis

Coinertia analysis (CIA) is a symmetric multivariate statistical method for measuring the agreement between two datasets. The basic principle of CIA is to find out synergistic structures that exist in the space of two datasets based on the covariance matrices of the two sets of variables and project them to the same space to reflect the shared trends or co-relationships of the datasets. There is no distinguishing between explanatory and response variables between the two datasets for which the CIA was conducted, and there is no explanatory or interpretive relationship between the two. The RV coefficient and the P-value of the Monte Carlo permutation test (999 times) are used to assess consistency, with the RV taking values in a range of [0, 1]. The CIA was performed using the ade4 package (v1.7.19) of the R software.

Both the CIA and the canonical correspondence analysis in the next section are sorting methods. In general, canonical correspondence analysis is used if there are a small number of metabolite variables or if

their correlation with each other is low. This is because at this point the multiple linear regression becomes an extended form similar to the one-dimensional linear regression, which makes canonical correspondence analysis more effective. In contrast, coinertia analysis is used when there are a large number of metabolites and a significant co-linear effect, as it does not require regression calculations, which avoids the associated co-linearity problem.



CIA analysis of differential microbes and differential metabolites

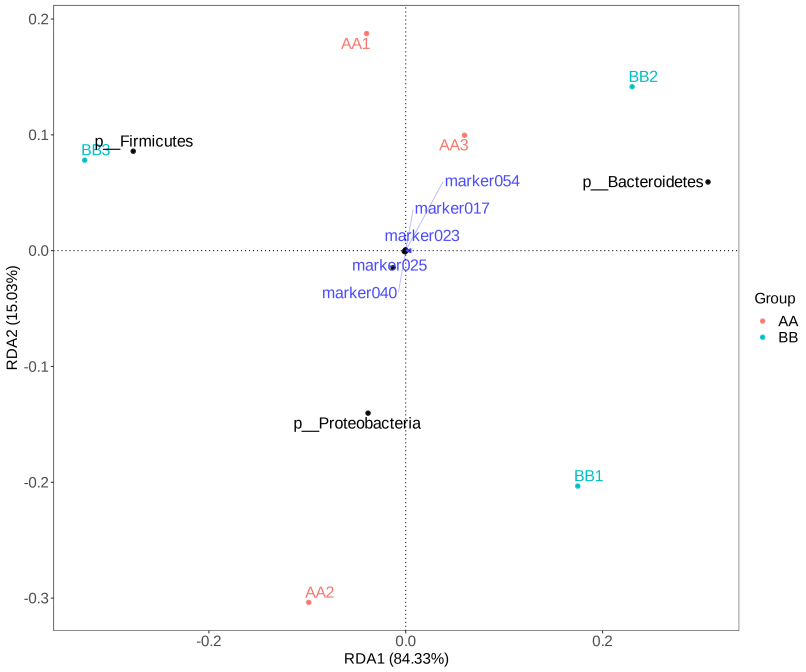
Each arrow indicates a sample, starting with a microbial sample and ending with a metabolite sample. The color indicates the grouping of the samples. The length of the line segment indicates the sample's difference between the two datasets, with the shorter the line segment the smaller the difference. The RV coefficient and the p-value of the Monte Carlo permutation test were used to assess consistency, with the RV taking values in a range of [0, 1].

File path: 3.Advanced_analysis/*/3.multivariate/2.coinertia_analysis (* means differential group)

7.4 Canonical Correspondence Analysis

Canonical correspondence analysis (CCA) is a ranking method developed based on correspondence analysis (CA), which is mainly used to reflect the relationship between colonies and metabolites. It can detect the relationship between metabolites, samples, and flora or the two-by-two relationship between them,

and can obtain the important driving factors affecting the distribution of samples. CCA analysis (also known as multivariate direct gradient analysis) combines correspondence analysis with multiple regression analysis, where each step of the calculation is regressed against the metabolite. Canonical correspondence analysis is based on two methods, CCA based on single-peak modeling and RDA based on linear modeling. Detrended correspondence analysis (DCA) is used to evaluate whether CCA or RDA should be used for a dataset. The program automatically selects the appropriate method based on the results of the DCA analysis, choosing RDA for DCA < 3, and CCA otherwise. The canonical correspondence analysis was performed using the vegan package (v2.6.2) of the R software.



Canonical correspondence analysis of differential microbes and differential metabolites

Black labels and dots indicate microbes; blue labels and arrows indicate environmental factors; colored labels indicate samples. The length of the arrow indicates the strength of the effect of the environmental factor on the microbial change. The longer the arrow, the greater the effect of the environmental factor on the microbial change. The vertical distance from the sample node to the environmental factor line segment and its extension line indicates the intensity of the environmental factor's impact on the sample - the closer the distance, the greater the impact of the environmental factor on the sample. Starting from the center origin, if the microbes are in the same direction as the arrows, it means that the environmental factors are positively correlated with the changes in the microbes, and vice versa indicates a negative correlation.

File path: 3.Advanced_analysis/*/3.multivariate/3.canonical_correspondence_analysis (* means differential group)

The envfit function was used to test the significance of each metabolite, and the results are displayed as follows:

Table 5 Envfit analysis for regression fitting of differential microbes to differential metabolites

Environment	RDA1	RDA2	R2	P-value
marker044	-0.9973592	-0.0726269	0.9973862	0.0013889
marker036	0.9991388	-0.0414930	0.9875284	0.0013889
marker037	0.9991388	-0.0414930	0.9875284	0.0013889
marker020	0.5657887	-0.8245503	0.9688163	0.0111111
marker026	0.7549542	-0.6557775	0.9349795	0.0152778
marker053	-0.9894140	-0.1451207	0.8213685	0.0402778
marker028	0.6806489	0.7326098	0.8360663	0.0777778
marker023	0.9901270	-0.1401730	0.8314438	0.0972222
marker022	0.9901270	-0.1401730	0.8314438	0.0972222
marker024	0.9901270	-0.1401730	0.8314438	0.0972222

File path: 3.Advanced_analysis/*/3.multivariate/3.canonical_correspondence_analysis

- CCA1/RDA1: the value is the cosine of the angle between the metabolite and the sorting axis, indicating the correlation between the metabolite and the sorting axis.
- CCA2/RDA2: same as CCA1/RDA1
- R2: It indicates the determination coefficient of the metabolite on the distribution of the species, the smaller the value, the smaller the impact of the metabolite on the distribution of the species.
- P-value: P value of the significance test. $P < 0.05$ indicates statistical significance.

7.5 BioENV combination of metabolites

Metabolites obtained from VIF screening were subjected to BioENV analysis. This analysis allows to obtain multiple sets of metabolite combinations and to calculate the correlation values of each set of metabolites with the microbial community. A combination of metabolites with the greatest correlation to the microbial

community can be screened out based on the correlation value, which has the greatest impact on the microbial community. The BioENV analysis was performed using the vegan package (v2.6.2) of the R software.

Table 6 BioENV analysis of differential metabolite combinations

Environment	Size	Correlation
marker017 + marker040	2	0.3250000
marker040 + marker023 + marker025	3	0.1607143
marker017 + marker040 + marker023 + marker025	4	0.1214286
marker017 + marker040 + marker054 + marker023 + marker025	5	-0.0142857

File path: 3.Advanced_analysis/*/3.multivariate/4.BioENV

- Environment: metabolite combinations
- Size: the number of metabolites in the combination
- Correlation: correlation between metabolites and microbes.

7.6 Mantel test

The Mantel test is a test of correlation between two matrices and is mostly used in ecology to calculate the correlation of microbial community data with metabolites. The metabolite combinations obtained from the BioENV screen will be used to calculate the overall correlation with the community data. The Mantel test was performed using the vegan package (v2.6.2) of the R software, and the results are shown below:

Table 7 Mantel test of differential microbes and differential metabolite combinations

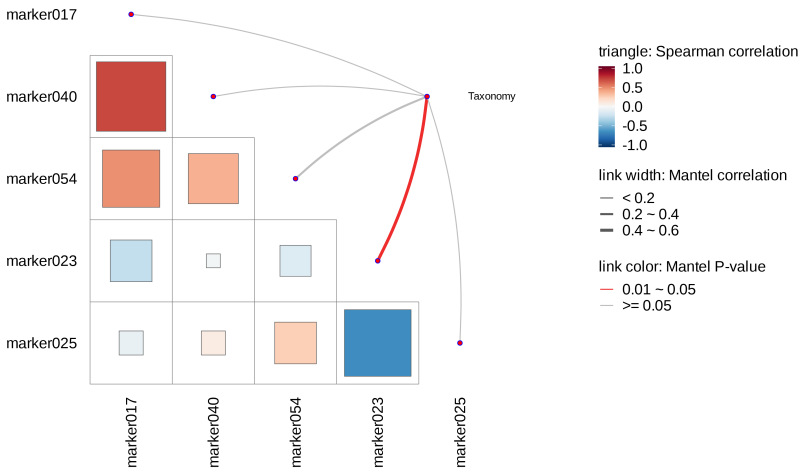
Environment	Correlation	P-value
marker017 + marker040	-0.0714286	0.5750000
marker040 + marker023 + marker025	0.3107143	0.1916667
marker017 + marker040 + marker023 + marker025	0.1428571	0.2958333
marker017 + marker040 + marker054 + marker023 + marker025	-0.0107143	0.4333333

File path: 3.Advanced_analysis/*/3.multivariate/5.Mantel-test

- Environment: metabolite combinations

- Correlation: indicates the Spearman correlation coefficient of the metabolite combination with the microbe. The larger the absolute value, the greater the correlation between the metabolite group and the species abundance.
- P-value: P value of the significance test. P < 0.05 indicates statistical significance.

The results of the Mantel test analysis of the microbial community against individual metabolites are shown below:



Mantel test of differential microbes and individual metabolites

The lower triangle is the Spearman correlation between metabolites. The size of the color block in the cell indicates the correlation coefficient, with red indicating a positive correlation and blue indicating a negative correlation. the significance test P-value is marked with , and , which indicate P-value < 0.05, P-value < 0.01, and P-value < 0.001, respectively, while the non-significant ones are without these markers. The connecting lines in the upper right area indicate the Mantel test results for communities and metabolites. The length of the line indicates the overall correlation coefficient, and the color of the connecting line indicates the significance test result of the correlation coefficient.

File path: 3.Advanced_analysis/*/3.multivariate/5.Mantel-test (* means differential group)

Reference

Franzosa, Eric A., Alexandra Sirota-Madi, Julian Avila-Pacheco, Nadine Fornelos, Henry J. Haiser, Stefan Reinker, Tommi Vatanen, et al. 2019. "Gut Microbiome Structure and Metabolic Activity in Inflammatory Bowel Disease." *Nature Microbiology* 4 (2): 293–305. https://doi.org/10.1038/s41564-018-0306-4.

Hess, Matthias, Alexander Sczyrba, Rob Egan, Tae-Wan Kim, Harshal Chokhawala, Gary Schroth, Shujun Luo, et al. 2011. "Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen." *Science* 331 (6016): 463. https://doi.org/10.1126/science.1200387.

Ilhan, Zehra Esra, John K DiBaise, Nancy G Isern, David W Hoyt, Andrew K Marcus, Dae-Wook Kang, Michael D Crowell, Bruce E Rittmann, and Rosa Krajmalnik-Brown. 2017. "Distinctive Microbiomes and Metabolites Linked with Weight Loss After Gastric Bypass, but Not Gastric Banding." *The ISME Journal* 11 (9): 2047–58. https://doi.org/10.1038/ismej.2017.71.

Lu, Kun, Ryan Phillip Abo, Katherine Ann Schlieper, Michelle E. Graffam, Stuart Levine, John S. Wishnok, James A. Swenberg, Steven R. Tannenbaum, and James G. Fox. 2014. "Arsenic Exposure Perturbs the Gut Microbiome and Its Metabolic Profile in Mice: An Integrated Metagenomics and Metabolomics Analysis." *Environmental Health Perspectives* 122 (3): 284–91. https://doi.org/10.1289/ehp.1307429.