Welcome to

ResistoXplorer

- a web-based tool for visualization and exploratory analysis of resistome data

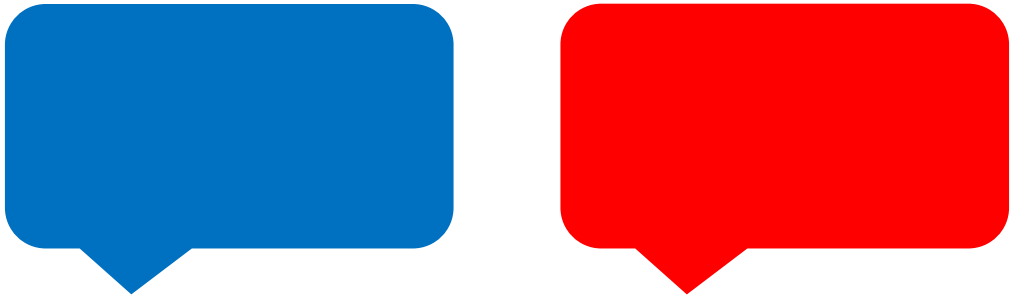
The key features include:

- Support for a wide array of common as well as advanced methods for composition profiling, visualization and exploratory data analysis;
- Comprehensive support for various data normalization methods coupled with standard as well as more recent statistical and machine learning algorithms;
- Support for a variety of methods for performing vertical data integrative analysis on paired metagenomic datasets (i.e. taxonomic and resistome abundance profiles);
- Comprehensive support for ARG functional annotations along with their microbe and phenotype associations based on data collected from more than 10 reference and curated databases;
- A powerful and fully featured network visualization for intuitive exploration of ARG-microbal host associations, incorporated with functional annotations enrichment analysis support.

In this manual, we will go through the analysis of resistome data using **INTEGRATION** data as input.

Please cite:
Dhariwal A, Junges R, Chen T, Petersen FC.
ResistoXplorer: a web-based tool for visualization and exploratory analysis of resistome data.

In this manual, you will encounter blue and red dialog boxes.



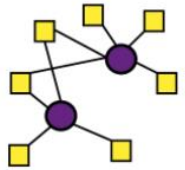
Blue dialogs indicate explanations and details for different functions in each page, while **red dialogs** indicate actions that will move forward with the analysis to a new screen or a download option for a visualization/analysis.

The question mark icons are available in ResistoXplorer.
If you hove over it, a short explanation about that item will appear.

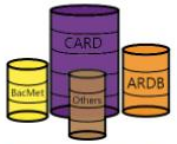
Throughout this manual you will also find additional explanations about the functionalities of ResistoXplorer following this icon.

In the front page of ResistoXplorer, you can select one of the three options for input data:
ARG list /// ARG table /// Integration

Features


ARG-MICROBE NETWORK EXPLORATION


INTERACTIVE EXPLORATION



COMPREHENSIVE DATABASE

MultiData Upload

1. Upload both your microbiome and resistome abundance profile

A. Taxonomic profile (Taxon abundance and taxonomy information)

1. Taxon abundance table: * (.txt or .csv format) Choose file No file chosen

2. Annotation file: (.txt or .csv format) Choose file No file chosen

B. Resistome profile (ARGs abundance and annotation information)

1. ARG abundance table: * (.txt or .csv format) Choose file No file chosen

2. Annotation file: (.txt or .csv format) Choose file No file chosen

C. Metadata (Sample metadata information)

Metadata file: (.txt or .csv format) Choose file No file chosen

Submit

Here you can upload your data in .txt or .csv formats, and you will need to upload the taxonomic, resistome, and metadata.

Optional: You can also upload the annotation files, if you have them.

If you click on the dataset, you will be able to download it as well.

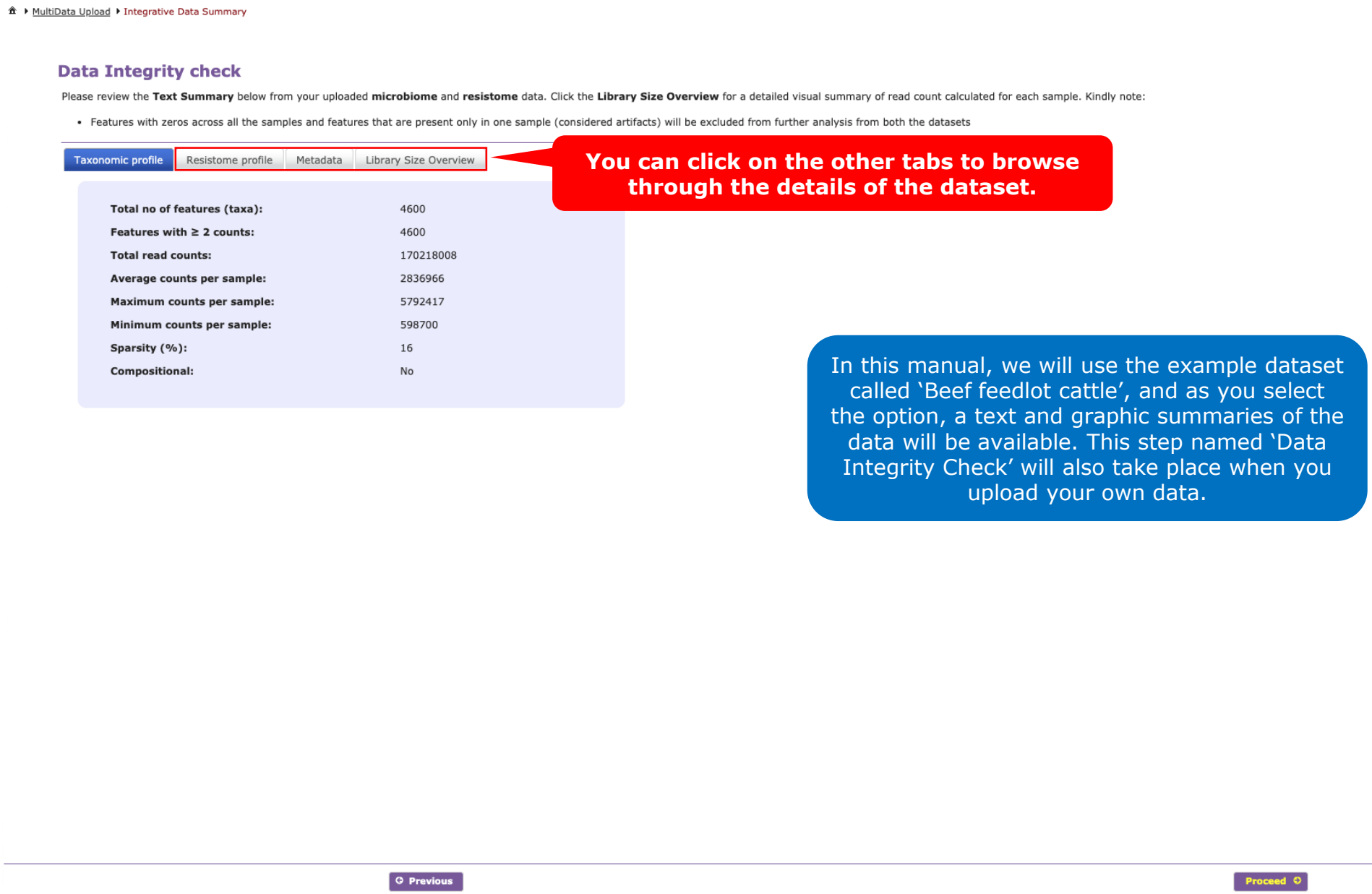
2. Try our example data set

Data Type	Description
Beef feedlot cattle	Gut microbiota (bacteriome) and resistome profile of 60 samples from commercial beef feedlot cattle treated with therapeutic doses of tulathromycin (E Doster, et al.). **Group label**: Untreated and Treated- indicating the treatment group; Day1 and Day11 - indicating the timepoint group.

Submit

Optional: Alternatively, you can use an example dataset to explore the functionalities of ResistoXplorer.

To proceed with the example dataset of your choice, click on 'Submit'. If you want to continue with your own data, click on the 'Submit' button above this one, after you have added your data.



You can click on the other tabs to browse through the details of the dataset.

In this manual, we will use the example dataset called 'Beef feedlot cattle', and as you select the option, a text and graphic summaries of the data will be available. This step named 'Data Integrity Check' will also take place when you upload your own data.

Data Integrity check

Please review the **Text Summary** below from your uploaded **microbiome** and **resistome** data. Click the **Library Size Overview** for a detailed visual summary of read count calculated for each sample. Kindly note:

- Features with zeros across all the samples and features that are present only in one sample (considered artifacts) will be excluded from further analysis from both the datasets

Taxonomic profile | **Resistome profile** | Metadata | Library Size Overview

Total no of features (ARGs):	134
Features with ≥ 2 counts:	134
Sparsity (%):	51
Compositional:	No
Total read counts:	3773873
Average counts per sample:	62897
Maximum counts per sample:	132644
Minimum counts per sample:	18177

Data Integrity check

Please review the **Text Summary** below from your uploaded **microbiome** and **resistome** data. Click the **Library Size Overview** for a detailed visual summary of read count calculated for each sample. Kindly note:

- Features with zeros across all the samples and features that are present only in one sample (considered artifacts) will be excluded from further analysis from both the datasets

Taxonomic profile | Resistome profile | **Metadata** | Library Size Overview

No. of samples in microbiome data:	60
No. of samples in resistome data:	60
No. of experimental factor:	3
No. of discrete experimental factor:	3
Sample names match (resistome vs microbiome):	Yes
No. of samples matched (microbiome vs resistome):	60
No. of samples that will be processed:	60

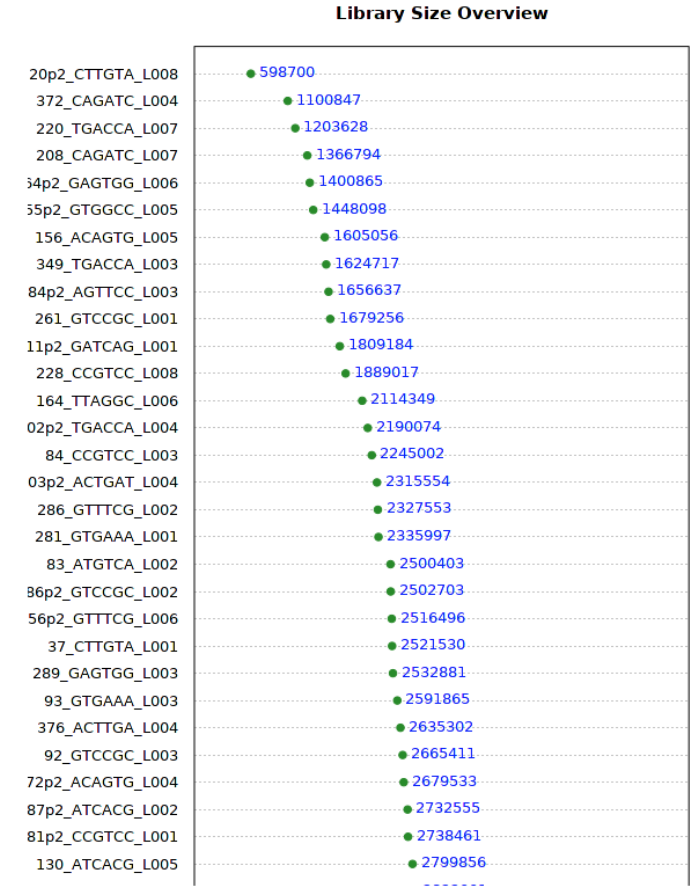
Data Integrity check

Please review the **Text Summary** below from your uploaded **microbiome** and **resistome** data. Click the **Library Size Overview** for a detailed visual summary of read count calculated for each sample. Kindly note:

- Features with zeros across all the samples and features that are present only in one sample (considered artifacts) will be excluded from further analysis from both the datasets

Taxonomic profile | Resistome profile | Metadata | Library Size Overview

Dataset: Taxonomic profile Update PDF SVG



If the details seem fine, go ahead and click on 'Proceed'.

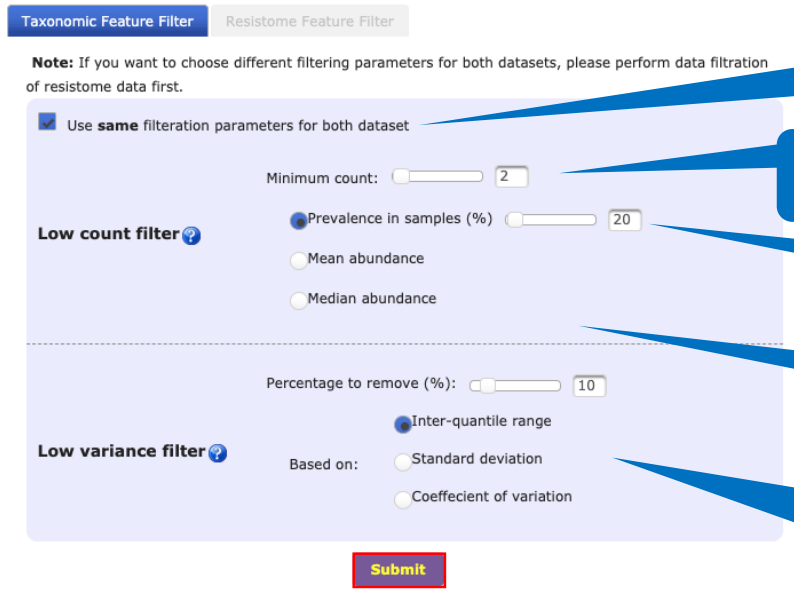
Previous Proceed

Data Filtration

Data Filtration aims at removing or filtering low quality or uninformative features from the data to improve the downstream statistical testing. Uninformative features can be filtered through two main ways:

- **Low count filter:** features that are present in a few samples with a very low read count are difficult to distinguish from sequencing errors. User need to set a minimum count (default value: 2). A 20% prevalence filter means that at least 20% of the values of a feature should contain at least 2 counts. You can also filter based on *mean* and *median* values.
- **Low variance filter:** features that do not vary or remain constant throughout the experiment conditions are not likely to be associated with the conditions under study. Feature variances can be calculated using *inter-quantile range (IQR)*, *standard deviation* or *coefficient of variation (CV)*.

We generally recommend users to use **more stringent fiteration** criteria for taxonomic data due to larger size (more no. of features) and more abundance (higher library sizes) as compared to resistome data.



For integrative data analysis, you can choose to either use the same filters for both taxonomic and resistome data, or you can set specific filters for each. If you would like to treat them separately, uncheck this box, and set the details by browsing in the tabs above.

Features with very small counts in very few samples are likely due to sequencing errors. You need to first specify a minimum count (default value is 2).

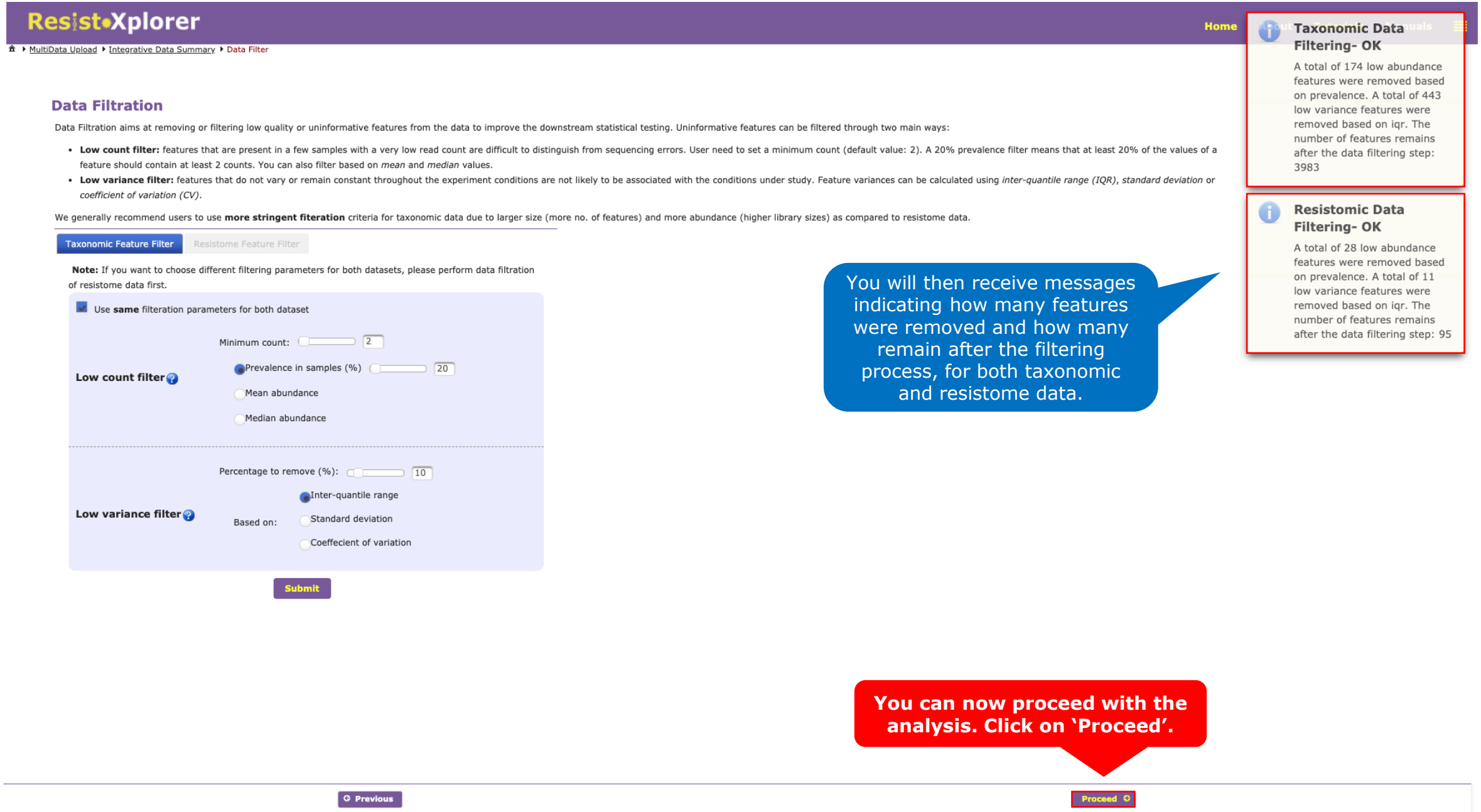
If you use 20% prevalence filter, meaning for any feature to be retained, at least 20% of its values should contain at least 2 counts.

In addition, you can also filter low abundance features based on their mean or median values below the minimum count.

Features that are close to constant throughout the experiment conditions can also be filtered, especially for comparative analysis. The variance can be measured using inter-quantile range (IQR), standard deviation or coefficient of variation (CV). The lowest percentage based on the cutoff will be excluded.

Feel free to set the parameters of your preference. In this manual, we will move forward with the default values. Click on 'Submit'.

Please notice that you cannot move forward with the analysis without clicking on 'Submit'.


ResistoXplorer
Home
MultiData Upload ▸ Integrative Data Summary ▸ Data Filter
Data Filtration
Data Filtration aims at removing or filtering low quality or uninformative features from the data to improve the downstream statistical testing. Uninformative features can be filtered through two main ways:
Low count filter: features that are present in a few samples with a very low read count are difficult to distinguish from sequencing errors. User need to set a minimum count (default value: 2). A 20% prevalence filter means that at least 20% of the values of a feature should contain at least 2 counts. You can also filter based on mean and median values.
Low variance filter: features that do not vary or remain constant throughout the experiment conditions are not likely to be associated with the conditions under study. Feature variances can be calculated using inter-quantile range (IQR), standard deviation or coefficient of variation (CV).
We generally recommend users to use more stringent fiteration criteria for taxonomic data due to larger size (more no. of features) and more abundance (higher library sizes) as compared to resistome data.
Taxonomic Feature Filter
Resistome Feature Filter
Note: If you want to choose different filtering parameters for both datasets, please perform data filtration of resistome data first.
Use same filteration parameters for both dataset
Low count filter
Minimum count: 2
Prevalence in samples (%) 20
Mean abundance
Median abundance
Low variance filter
Percentage to remove (%): 10
Based on:
Inter-quantile range
Standard deviation
Coeffecient of variation
Submit
Taxonomic Data Filtering- OK
A total of 174 low abundance features were removed based on prevalence. A total of 443 low variance features were removed based on iqr. The number of features remains after the data filtering step: 3983
Resistomic Data Filtering- OK
A total of 28 low abundance features were removed based on prevalence. A total of 11 low variance features were removed based on iqr. The number of features remains after the data filtering step: 95
You will then receive messages indicating how many features were removed and how many remain after the filtering process, for both taxonomic and resistome data.
You can now proceed with the analysis. Click on 'Proceed'.
Previous
Proceed

Why should I use the data filtering option?

Data filtering is important because features having very low counts or abundance across all the samples cannot be discriminated from sequencing errors, and they can interfere with some statistical and biological approximations. Thus, such features should be removed from the data before performing any downstream analysis.

Which category should I choose to perform data filtering?

ResistoXplorer automatically removes features that comprise of all zeros or that are only present in one sample.

However, for all other types of analysis, further data filtration is required. By default, features are filtered based on their sample prevalence and abundance levels. You can also choose to remove features having low variance across samples.

All analyses for integrative data are performed on filtered and normalized data.

Data Normalization

Data Normalization aims to address the high level of systematic variability (uneven sampling depth), sparsity and heterogeneity present in the metagenomic data to enable more biologically meaningful comparisons and interpretations. There are wide variety of methods available and their performance have been evaluated in terms of methods ability to identify differentially abundant genes (see MB Pereira et al.) in metagenomic count data. All these methods require "raw count data" as input. You can rarefy your data followed by either data scaling or data transformation. However, both data scaling and data transformation cannot be applied together, because scaled or transformed data is no longer valid count data. To account for compositionality, two CoDA recommended **log-ratio** transformations have been also implemented. Please note, zero have been replaced with a small **pseudocount** (*i.e., min (non zero value in table) *0.01*) before performing log-ratio transformations.

For integrative data analysis, both the datasets (taxonomic and resistome) are normalized using the **same** approach

All samples will be rarefied to even sequencing depth based on the sample having the lowest sequencing depth. If this sample contains extremely low reads, you may need to manually exclude this sample (using the Sample Editor) to avoid significant data loss. You can find out if this is the case from View Sample Size from the Data Summary page.

Data scaling aims to bring all samples to the same scale by dividing the samples by a scaling factor. Some common choices include total sum scaling (TSS), cumulative sum scaling (CSS), and upper-quantile scaling (UQ).

Variance stabilization transformation such as log-ratio transformation and its variations. Some common choices include centered log-ratio (CLR) transformation, relative log expression (RLE) normalization, or weighted trimmed mean of M-values (TMM).

Compositional data analysis (CoDA) recommended normalization approaches.

For this manual, we will go ahead with the default options as shown above. Click on 'Submit'.

Please notice that you cannot move forward with the analysis without clicking on 'Submit'.

Data Normalization - OK

No data rarefaction was performed. Performed count per million (CPM) normalization. No data transformation was performed.

You will then receive a message indicating the normalization procedures that were performed.

Data Normalization

Data Normalization aims to address the high level of systematic variability (uneven sampling depth), sparsity and heterogeneity present in the metagenomic data to enable more biologically meaningful comparisons and interpretations. There are wide variety of methods available and their performance have been evaluated in terms of methods ability to identify differentially abundant genes (see MB Pereira et al.) in metagenomic count data. All these methods require "raw count data" as input. You can rarefy your data followed by either data scaling or data transformation. However, both data scaling and data transformation cannot be applied together, because scaled or transformed data is no longer valid count data. To account for compositionality, two CoDA recommended **log-ratio** transformations have been also implemented. Please note, zero have been replaced with a small **pseudocount** (*i.e., min (non zero value in table) *0.01*) before performing log-ratio transformations.

For integrative data analysis, both the datasets (taxonomic and resistome) are normalized using the **same** approach

Data rarefying

- (•) Do not rarefy my data
- ( ) Rarefy to the minimum library size

Data scaling

- ( ) Do not scale my data
- (•) Count per Million (CPM)
- ( ) Log Count per Million (logCPM)
- ( ) Cumulative sum scaling (CSS)
- ( ) Upper-quantile normalization (UQ)
- ( ) Relative proportion

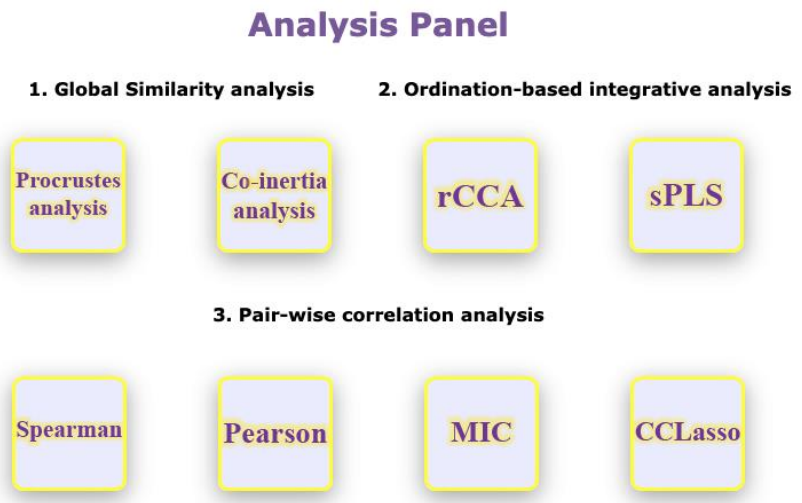
Data transformation

- (•) Do not transform my data
- ( ) Relative log expression (RLE)
- ( ) Trimmed mean of M-values (TMM)
- ( ) Hellinger transformation
- ( ) Centered log ratio (CLR)
- ( ) Additive log ratio (ALR)

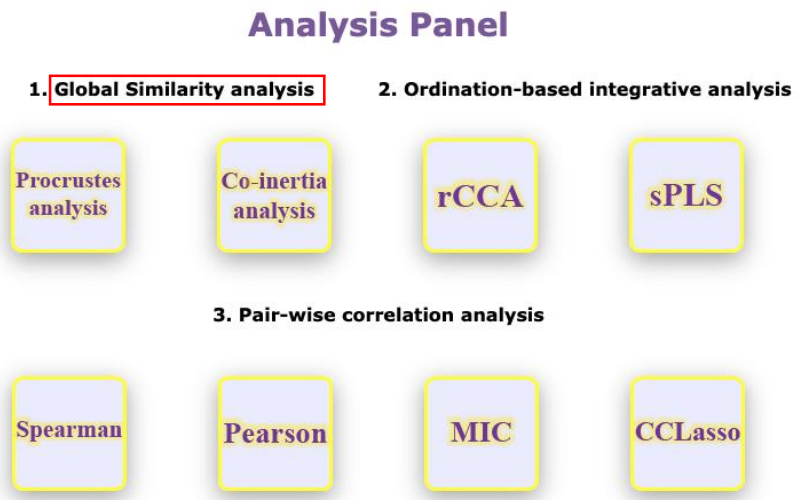
Submit

You can now proceed with the analysis. Click on 'Proceed'.

Here we have several options including Global Similarity analysis, Ordination-based integrative analysis, and pair-wise correlation analysis. In the next slides, we will go through important aspects of each, including some common questions.



The first type of analysis is 'Global Similarity analysis'.



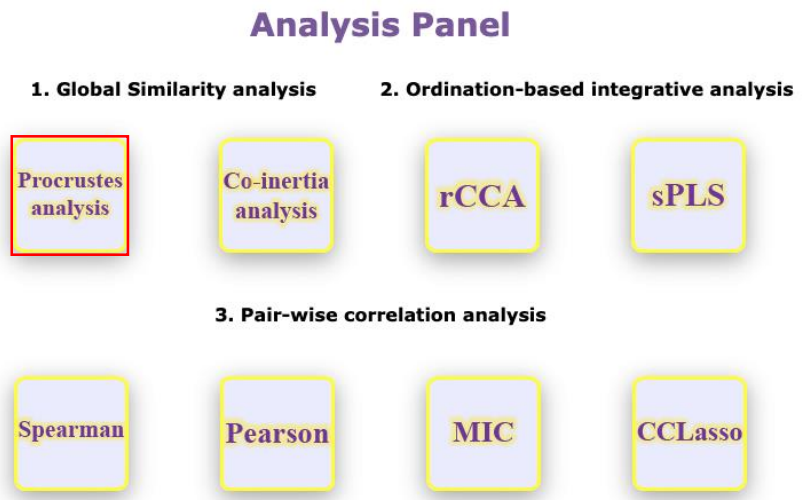
What are Global Similarity analysis methods and what are the differences between them?

The two methods for Global Similarity analysis employed in ResistoXplorer are:
- ‘Procrustes analysis’ and
- ‘Co-Inertia analysis’.

Both are symmetric methods which aim to analyze the covariance matrices. They differ in the form of presentation for the joint data ordination. Both are appropriate for community composition data. While Procrustes analysis creates a “compromise” ordination of two matrices measured on the same objects in order to visualize differences between two matrices, Co-inertia creates an ordination based on two covariance between two data matrices and plots both matrices in the same ordination space along with their variables.

Remember that you can always browse and go back to different steps of the process by utilizing the links provided here.

Let's start by looking at 'Procrustes analysis' – click here.



Procrustes analysis

General options:

Taxonomic level: Feature (Rownames) Profile level: Feature (Rownames)

Ordination method: PCoA Distance method: Bray-Curtis Index No. of Permutations: 999

Submit Downloads

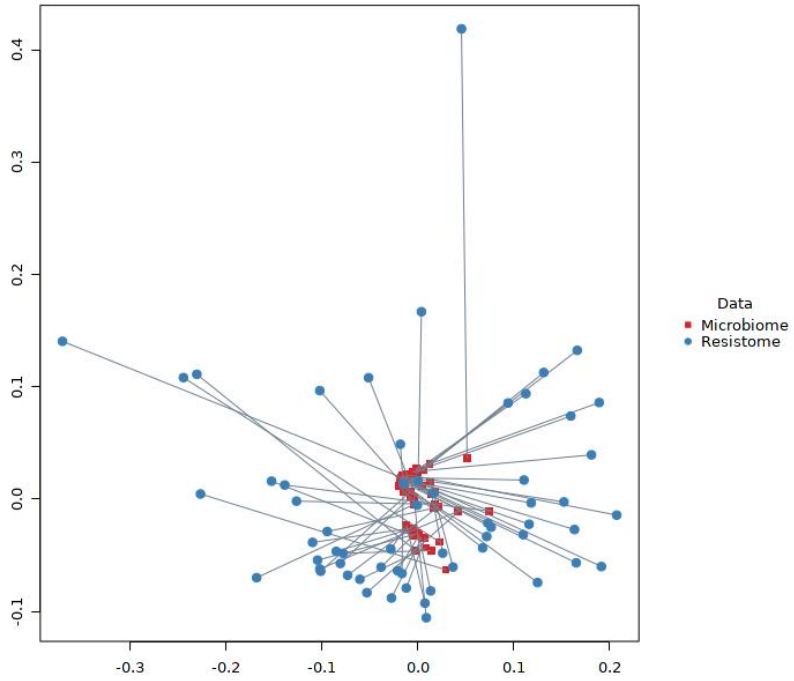
View options:

Color data points according to: Data (Omics) / Experimental factor Time Color Palette: Set1 Label samples by: None (for 2D plot only)

2D Graphics | 3D Graphics

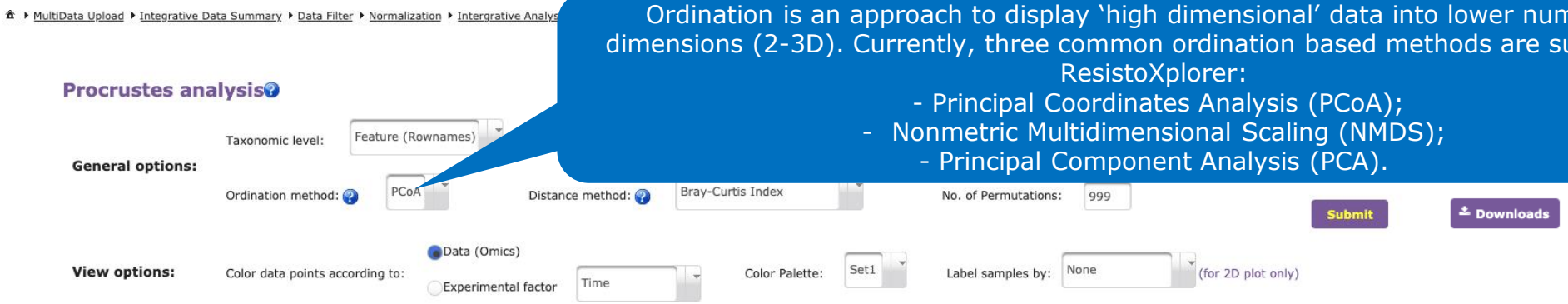
[Procrustes] Sum of Squares = 0.6014; Correlation coefficient (squared m12) = 0.6313; P-value < 0.506

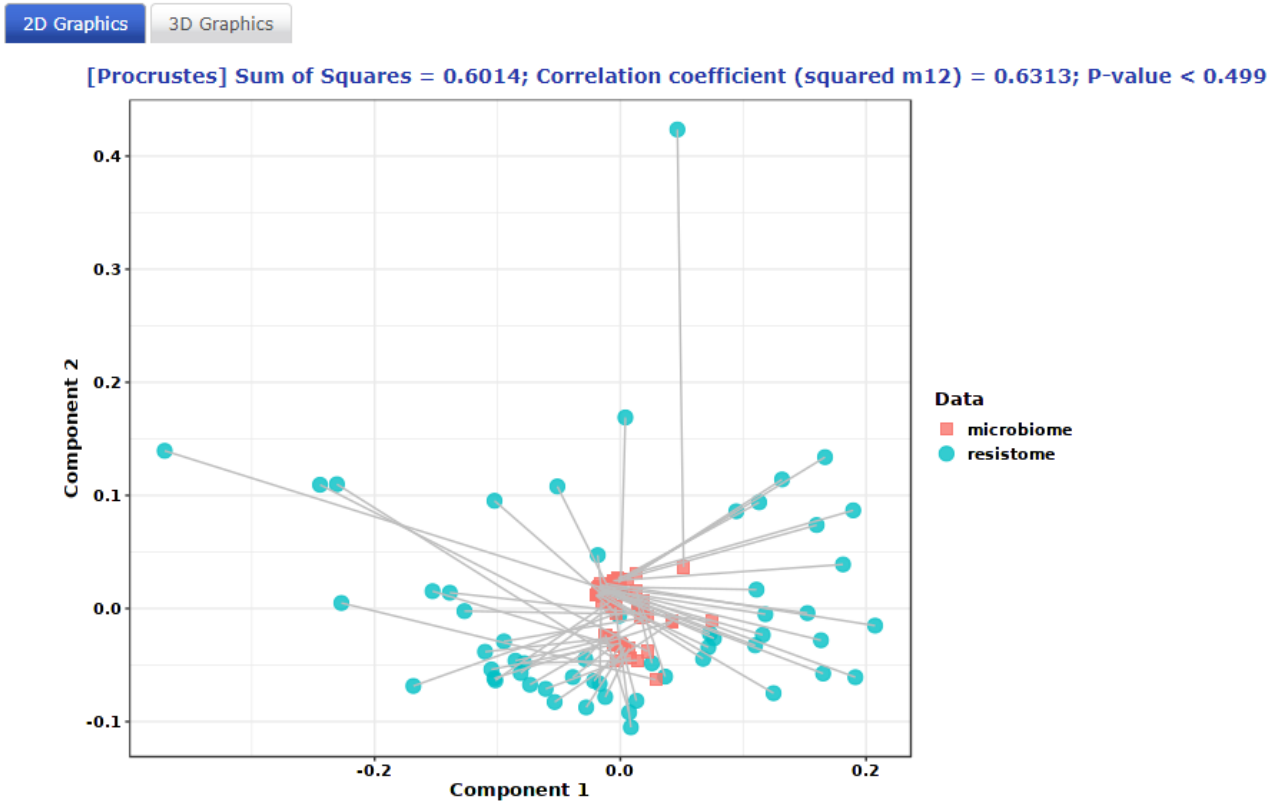
The statistical data is available here.



Procrustes analysis allows you to integrate omics data, and to analyze how these are spatially organized in dimensional shape analysis. Ordination and distance methods can be changed.

Ordination is an approach to display 'high dimensional' data into lower numbers of dimensions (2-3D). Currently, three common ordination based methods are supported in ResistoXplorer:
- Principal Coordinates Analysis (PCoA);
- Nonmetric Multidimensional Scaling (NMDS);
- Principal Component Analysis (PCA).





Procrustes analysis

General options:

Taxonomic level: Feature (Rownames) Profile level: Feature (Rownames)

Ordination method: PCoA Distance method: Bray-Curtis Index No. of Permutations: 999

Submit Downloads

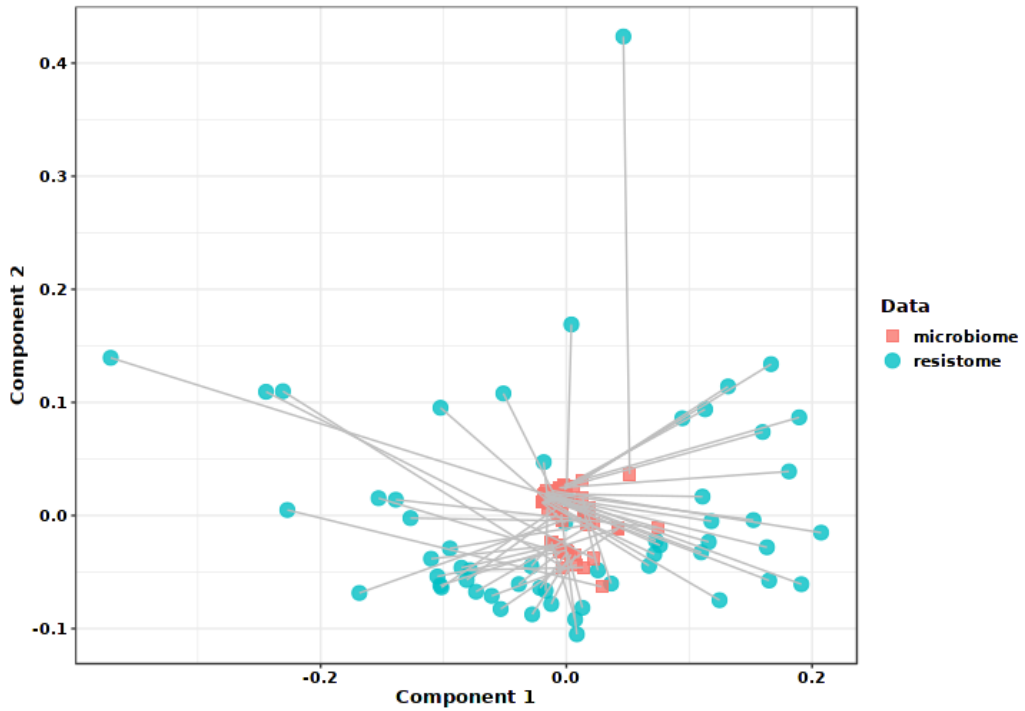
View options:

Color data points according to: Data (Omics) / Experimental factor Time Color Palette: Set1 Label samples by: None (for 2D plot only)

2D Graphics | 3D Graphics

If you click here, you will be able to see the same graph but in 3D.

[Procrustes] Sum of Squares = 0.6014; Correlation coefficient (squared m12) = 0.6313; P-value < 0.499



MultiData Upload ▸ Integrative Data Summary ▸ Data Filter ▸ Normalization ▸ Intergrative Analysis ▸ Procrustes Analysis

Procrustes analysis

General options:

Taxonomic level: Feature (Rownames) Profile level: Feature (Rownames)

Ordination method: PCoA Distance method: Bray-Curtis Index No. of Permutations: 999

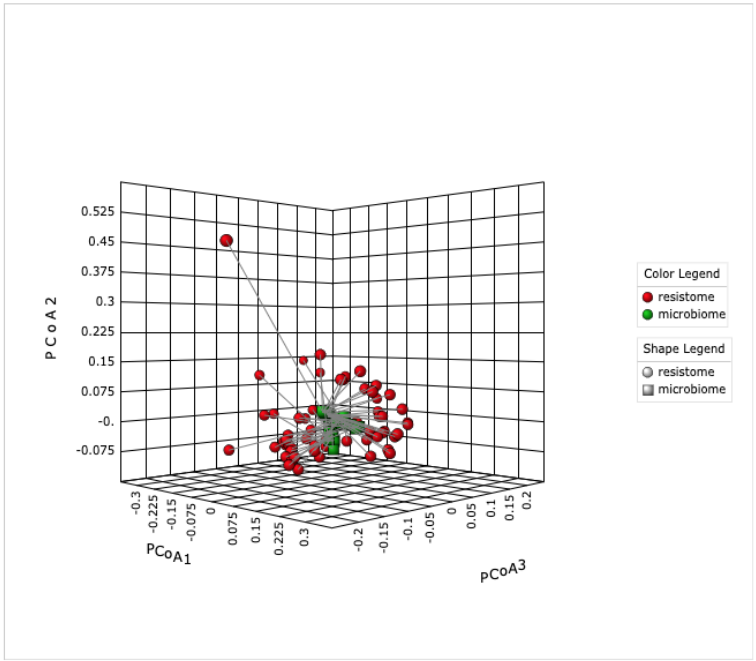
Submit Downloads

View options:

Color data points according to: Data (Omics) / Experimental factor Time Color Palette: Set1 Label samples by: None (for 2D plot only)

2D Graphics | 3D Graphics

Drag to rotate, **scroll** to zoom, **hover** on a data point to view



-0.3 -0.2 -0.1 0.0 0.1 0.2

You can click and drag to move the angle of the graph. If you hover over data points, information about them will be given. In addition, you can scroll to zoom in or out.

Procrustes analysis

General options:

Taxonomic level: Feature (Rownames) Profile level: Feature (Rownames)

Ordination method: PCoA Distance method: Bray-Curtis Index No. of Permutations: 999

Submit Downloads

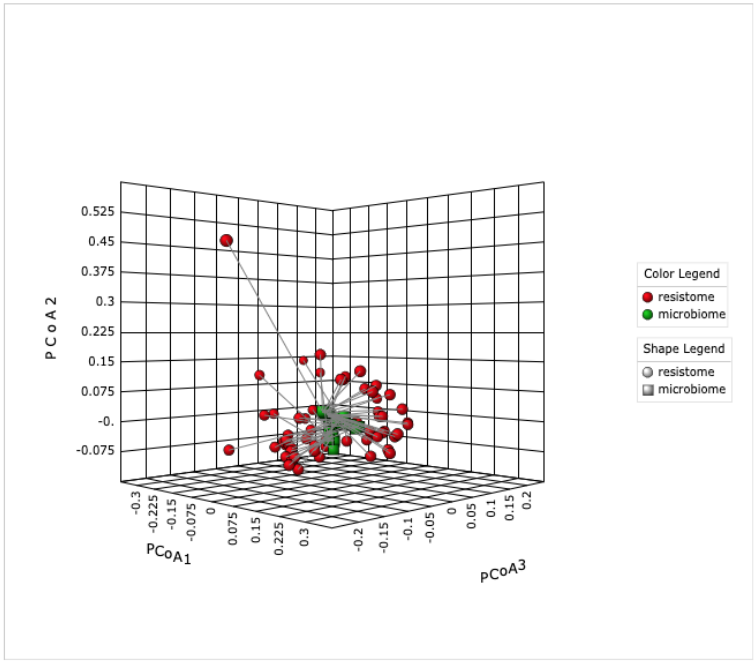
View options:

Color data points according to: Data (Omics) / Experimental factor Time Color Palette: Set1 Label samples by: None (for 2D plot only)

You can choose to download the analyses and/or graphs in a number of different formats by clicking here.

2D Graphics | 3D Graphics

Drag to rotate, **scroll** to zoom, **hover** on a data point to view



-0.3 -0.2 -0.1 0.0 0.1 0.2

Once you are finished, you can click on 'Integrative Analysis' to go back to the previous page.

Procrustes analysis

General options:

Taxonomic level: Feature (Rownames)

Profile level: Feature (Rownames)

Ordination method: PCoA

Distance method: Bray-Curtis Index

No. of Permutations: 999

Submit

Downloads

View options:

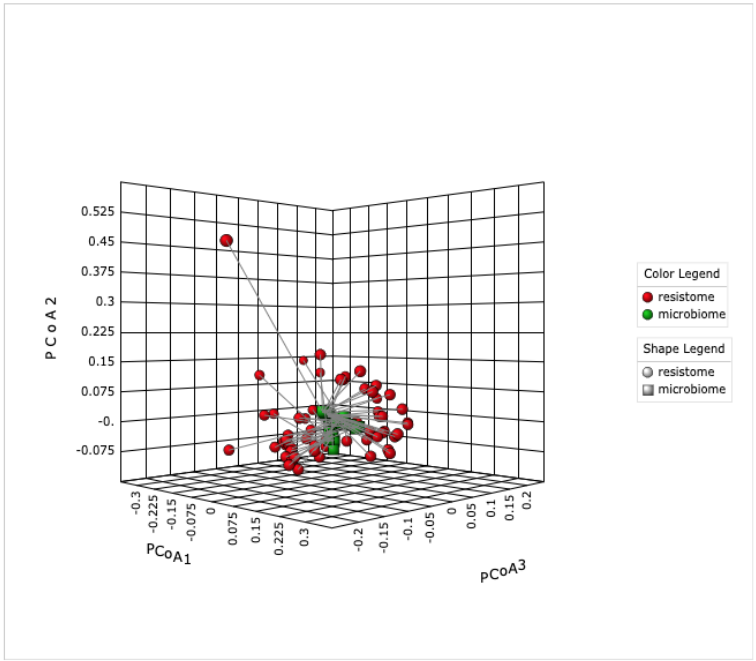
Color data points according to: Data (Omics) / Experimental factor — Time

Color Palette: Set1

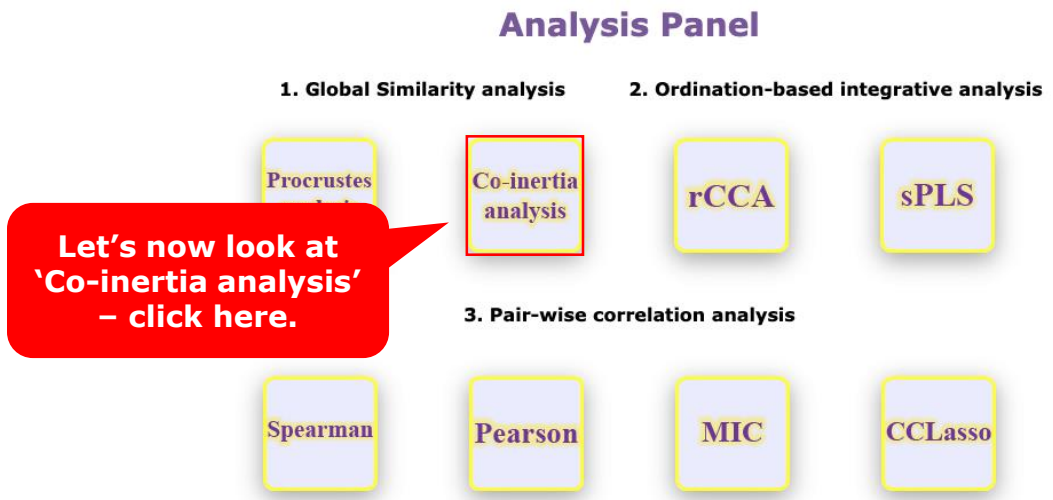
Label samples by: None (for 2D plot only)

2D Graphics | 3D Graphics

Drag to rotate, **scroll** to zoom, **hover** on a data point to view



-0.3 -0.2 -0.1 0.0 0.1 0.2


Analysis Panel
1. Global Similarity analysis
2. Ordination-based integrative analysis
Procrustes
Co-inertia analysis
rCCA
sPLS
Let's now look at 'Co-inertia analysis' – click here.
3. Pair-wise correlation analysis
Spearman
Pearson
MIC
CCLasso

MultiData Upload ▸ Integrative Data Summary ▸ Data Filter ▸ Normalization ▸ Intergrative Analysis ▸ Procrustes Analysis ▸ Coinertia Analysis

Co-inertia analysis (CIA)

General options:

Taxonomic level: Feature (Rownames) Profile level: Feature (Rownames)

Ordination method: PCoA Distance method: Bray-Curtis Index No. of Permutations: 99

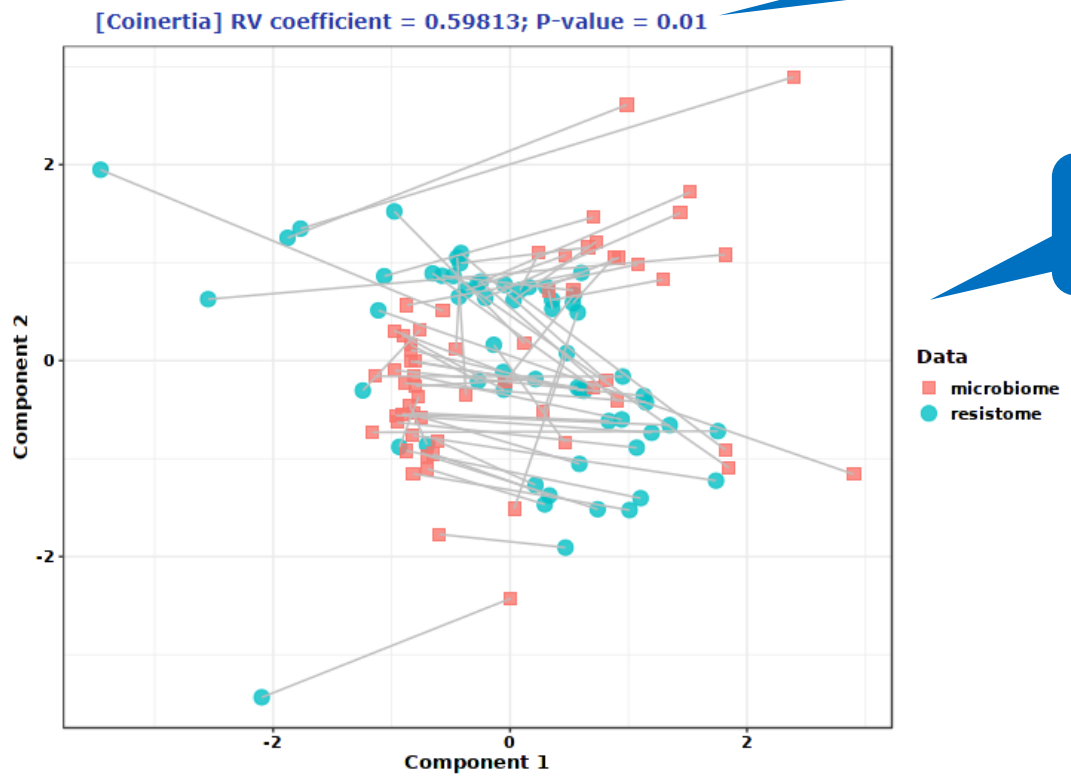
Submit Downloads

View options:

Color data points according to: Data (Omics) / Experimental factor Time Color Palette: Set1 Label samples by: None (for 2D plot only)

2D Graphics | 3D Graphics

The statistical data is available here.



Similarly to Procrustes analysis, Co-inertia aims to compare the distribution of a set of shapes based on the input data. As previously, we can also visualize this in 3D.

MultiData Upload ▸ Integrative Data Summary ▸ Data Filter ▸ Normalization ▸ Intergrative Analysis ▸ Procrustes Analysis ▸ Coinertia Analysis

Co-inertia analysis (CIA)

General options:

Taxonomic level: Feature (Rownames)

Profile level: Feature (Rownames)

Ordination method: PCoA

Distance method: Bray-Curtis Index

No. of Permutations: 99

View options:

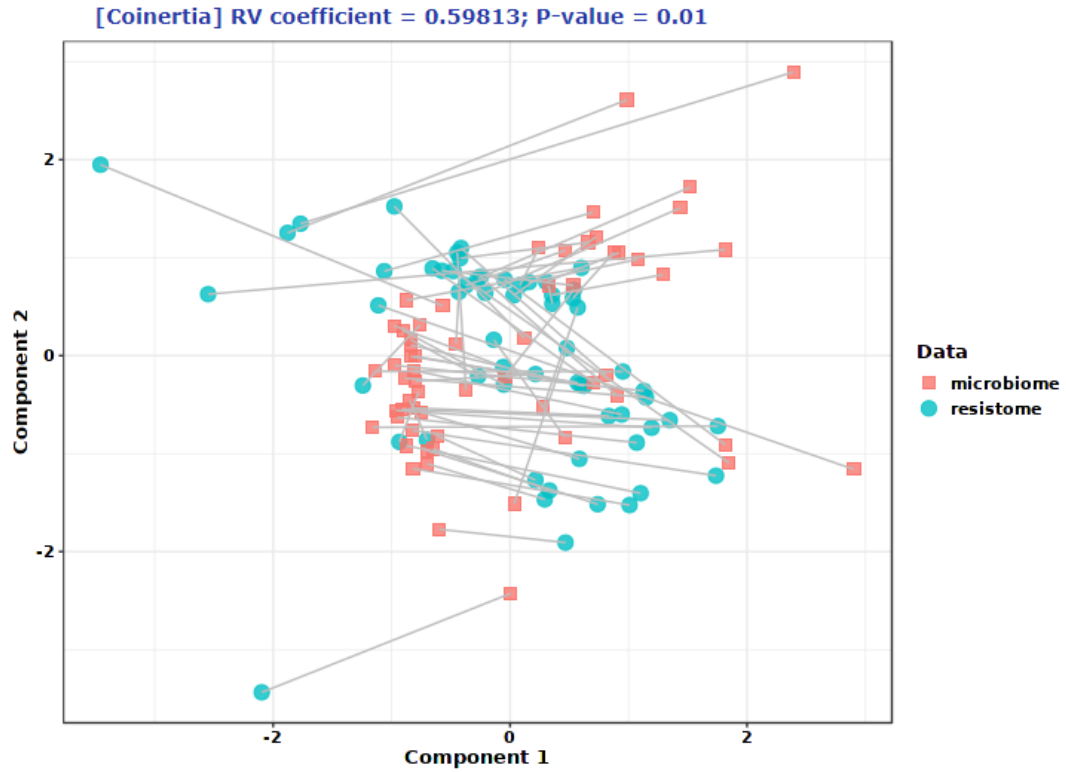
Color data points according to:

Data (

Experimental factor

Ordination is an approach to display 'high dimensional' data into lower numbers of dimensions (2-3D). Currently, the two common ordination based methods supported in ResistoXplorer for Co-inertia are:
- Principal Coordinates Analysis (PCoA);
- Principal Component Analysis (PCA).

2D Graphics | 3D Graphics



MultiData Upload ▸ Integrative Data Summary ▸ Data Filter ▸ Normalization ▸ Intergrative Analysis ▸ Procrustes Analysis ▸ Coinertia Analysis

Co-inertia analysis (CIA)

You can customize your options regarding the calculation such as ordination and distance methods. And you can also change the visualization.

General options:

Taxonomic level: Feature (Rownames)

Profile level: Feature (Rownames)

Ordination method: PCoA

Distance method: Bray-Curtis Index

No. of Permutations: 99

Submit

Downloads

View options:

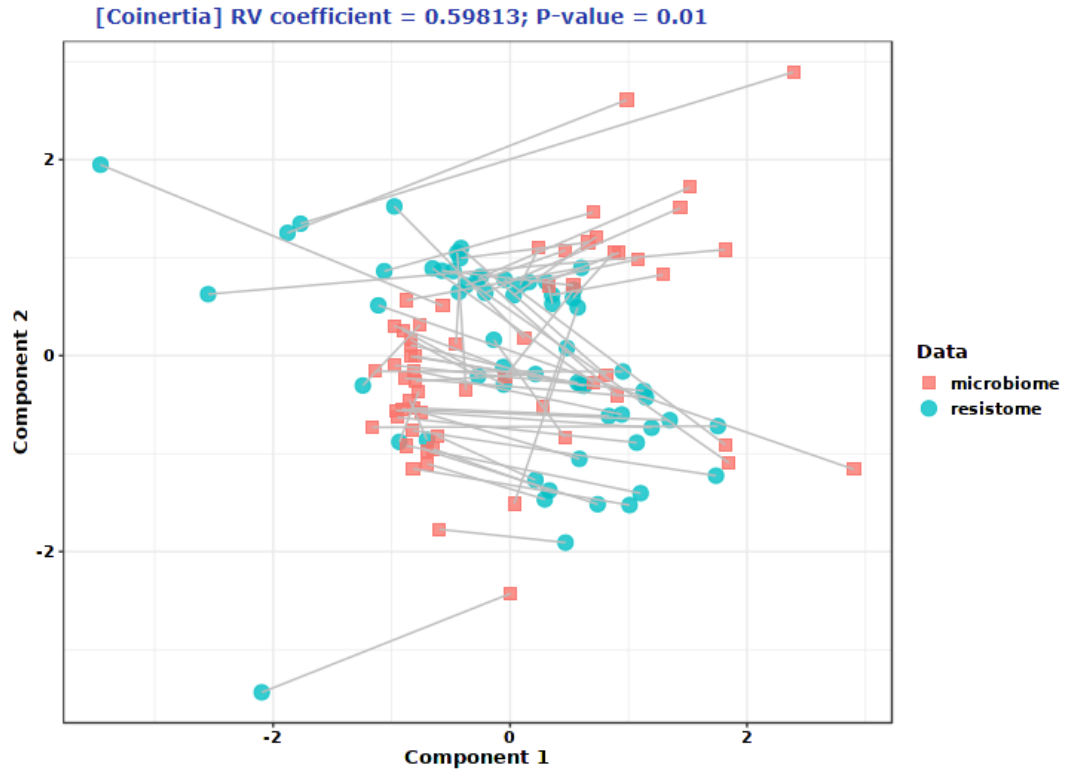
Color data points according to: Data (Omics) / Experimental factor — Time

Color Palette: Set1

Label samples by: None (for 2D plot only)

2D Graphics | 3D Graphics

If you click here, you will be able to see the same graph but in 3D.



Co-inertia analysis (CIA)

General options:

Taxonomic level: Feature (Rownames)

Profile level: Feature (Rownames)

Ordination method: PCoA

Distance method: Bray-Curtis Index

No. of Permutations: 99

Submit

Downloads

View options:

Color data points according to: Data (Omics) / Experimental factor — Time

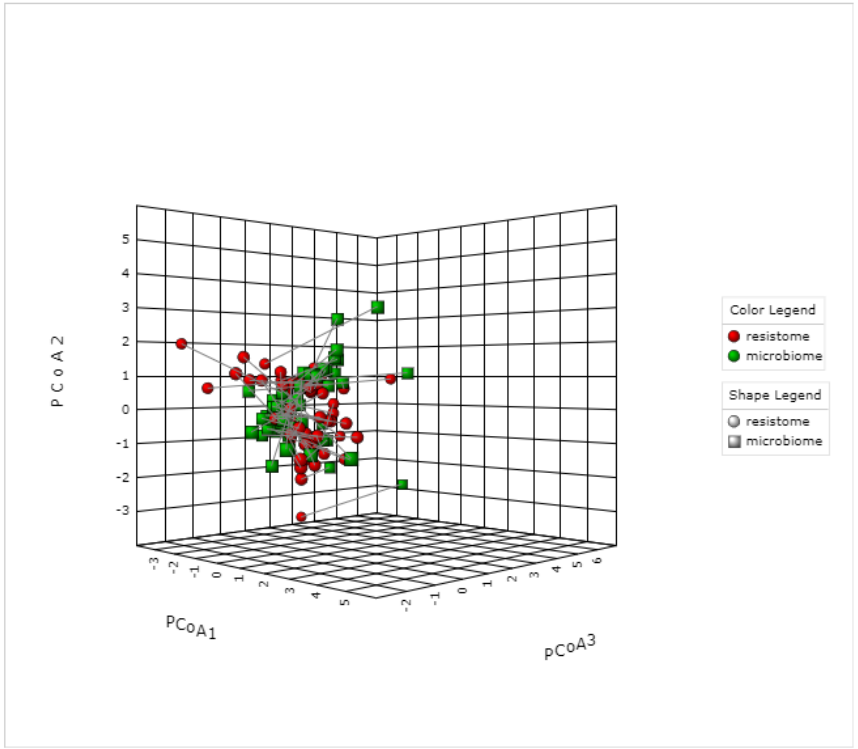
Color Palette: Set1

Label samples by: None (for 2D plot only)

You can choose to download the analyses and/or graphs in a number of different formats by clicking here.

2D Graphics | 3D Graphics

Drag to rotate, **scroll** to zoom, **hover** a data point to view



Once you are finished, you can click on 'Integrative Analysis' to go back to the previous page.

Co-inertia analysis (CIA)

General options:

Taxonomic level: Feature (Rownames)

Profile level: Feature (Rownames)

Ordination method: PCoA

Distance method: Bray-Curtis Index

No. of Permutations: 99

Submit

Downloads

View options:

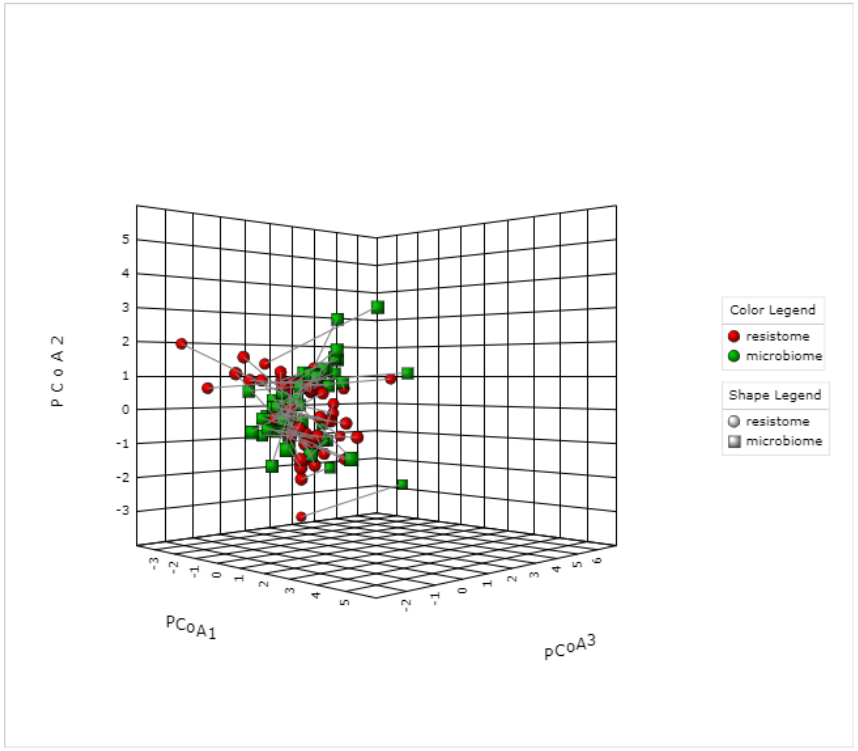
Color data points according to: Data (Omics) / Experimental factor — Time

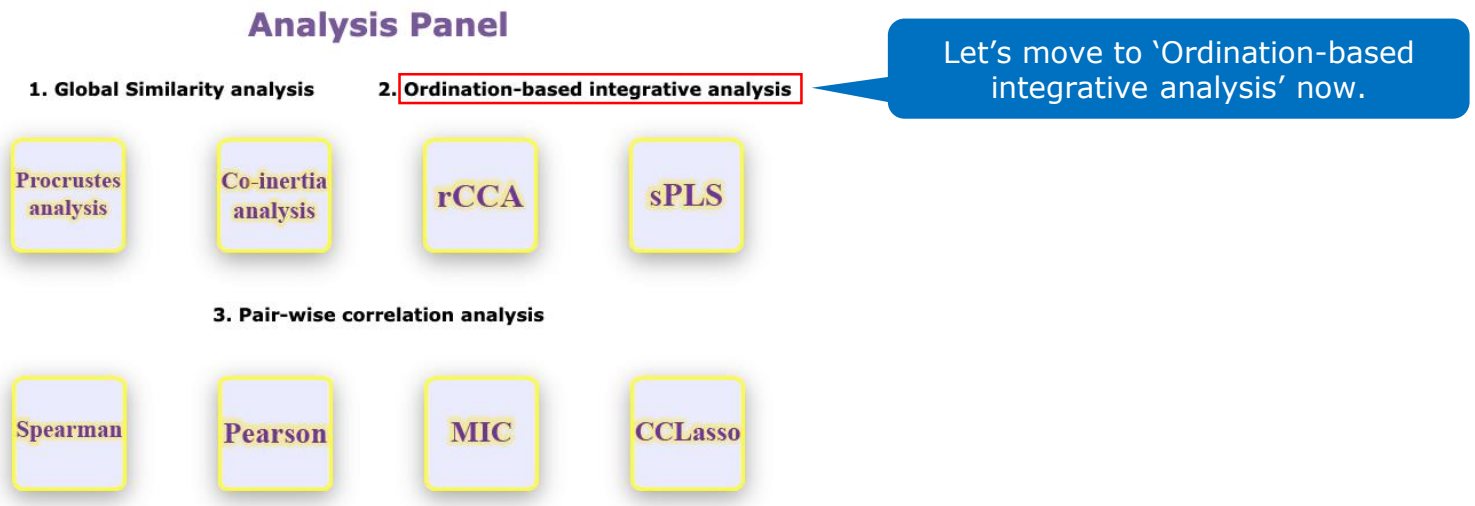
Color Palette: Set1

Label samples by: None (for 2D plot only)

2D Graphics | 3D Graphics

Drag to rotate, **scroll** to zoom, **hover** a data point to view





Let’s move to ‘Ordination-based integrative analysis’ now.

What are the options for ‘Ordination-based integrative analysis’ in ResistoXplorer?

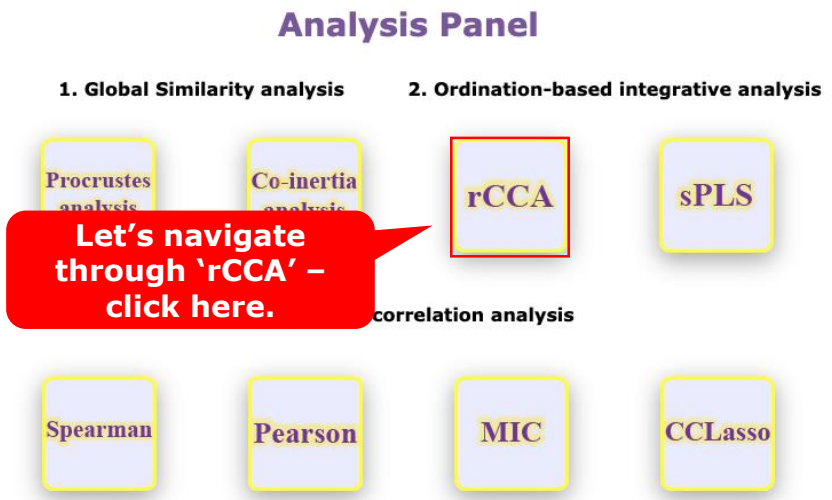
ResistoXplorer offers two options of techniques to perform Ordination-based integrative analysis:

- regularized Canonical Correlation Analysis (rCCA).
- Sparse Partial Least Squares regression (sPLS).

What are the main differences between rCCA and sPLS?

Canonical Correlation Analysis (CCA) and Partial Least Squares regression (PLS) are methods that focus on the integration and exploration of two datasets. While their purpose and application is very similar, they differ in the method utilized for such. While CCA maximizes the correlation between linear combinations regarding variables in each dataset, PLS focuses on maximizing the covariance. Despite the underlying differences in the algorithms, if the components are scaled, both methods should give similar results. Importantly, CCA will calculate all components directly without deflation, and PLS presents different deflation modes. The decision regarding which technique to use will depend on your preference and data.

For more details, please refer to the original mixOmics R package publication.


Analysis Panel
1. Global Similarity analysis
2. Ordination-based integrative analysis
Procrustes analysis
Co-inertia analysis
rCCA
sPLS
Let's navigate through 'rCCA' – click here.
correlation analysis
Spearman
Pearson
MIC
CCLasso

Regularized Canonical Correlation Analysis (rCCA)

General options:

Taxonomic level: Phylum

Profile level: Group

No of components: 5

Regularization method: Shrinkage

Validation method: 5-fold CV

Submit

Downloads

View options:

(for score plot only)

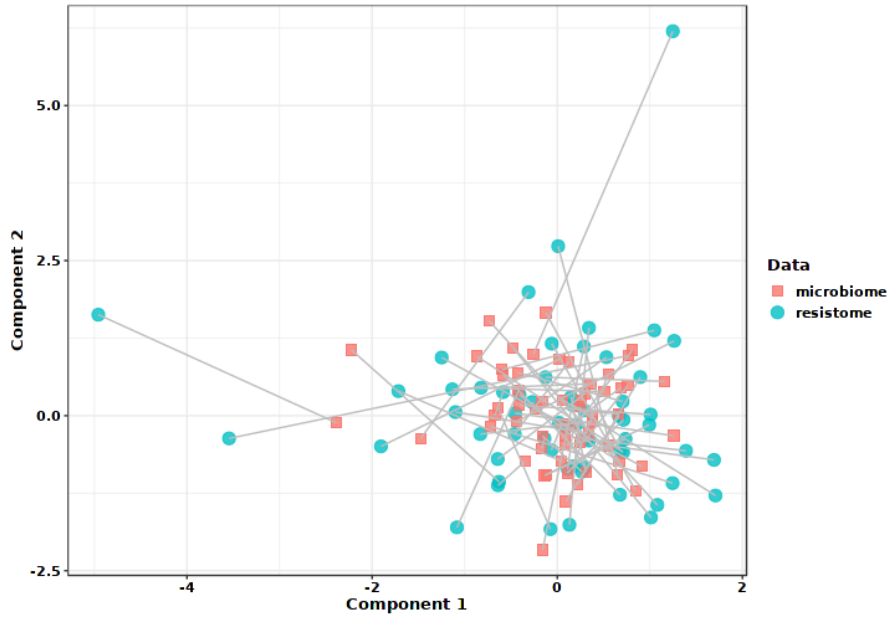
Color data points according to: Data (Omics) / Experimental factor — Time

Color Palette: Default

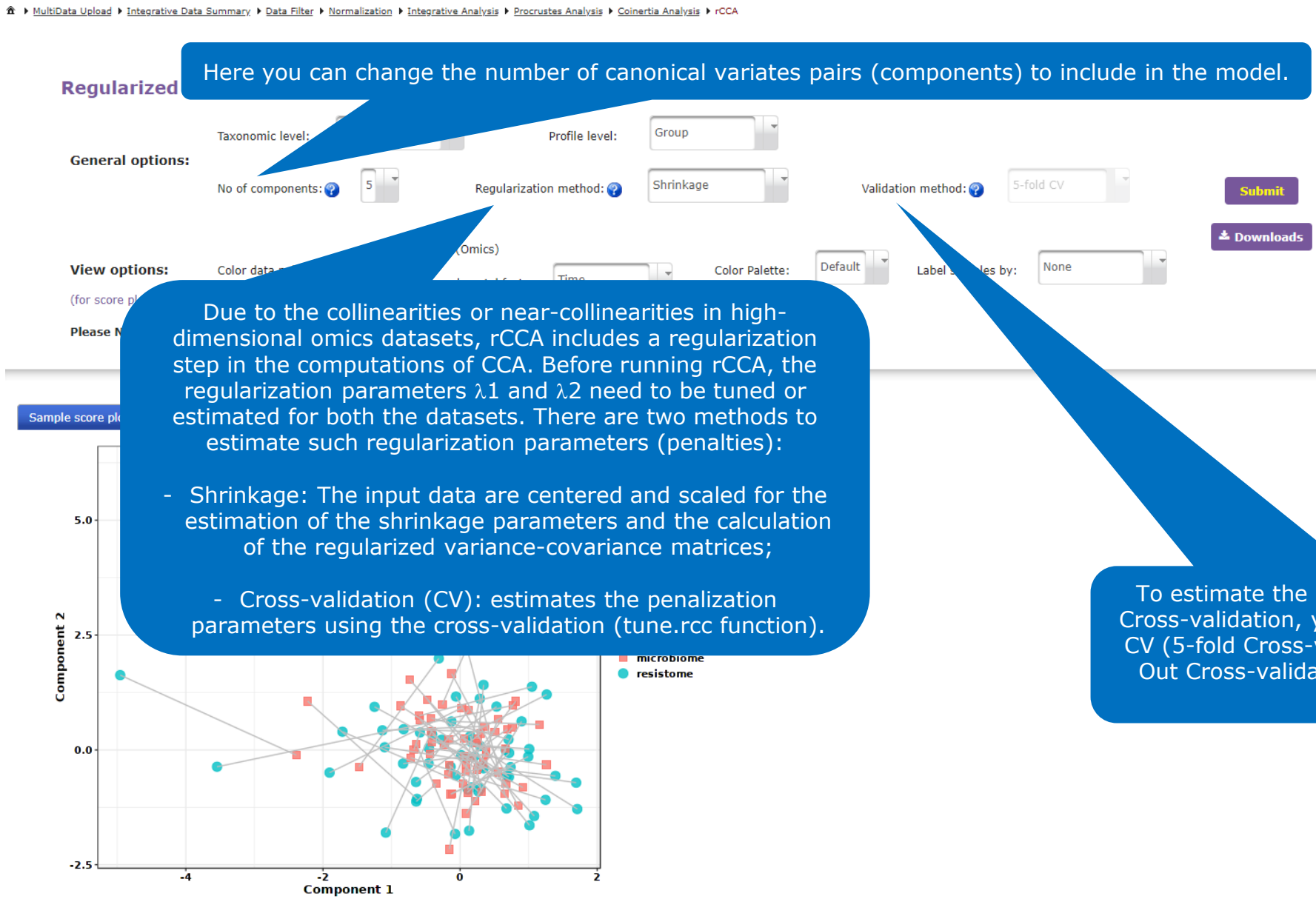
Label samples by: None

Please Note: The ***Color Palette*** and ***Label samples by*** options are applicable for sample score plot (2D) only.

Sample score plot (2D) | Sample score plot (3D) | Scree plot | Clustered Image Map (CIM) | Correlation Circle Plot



As we have seen, Regularized Canonical Correlation Analysis (RCCA) is a statistical technique that extracts correlations between two types of data to construct a homogeneous, uniform representation of heterogeneous data channels.


MultiData Upload
Integrative Data Summary
Data Filter
Normalization
Integrative Analysis
Procrustes Analysis
Coinertia Analysis
rCCA
Regularized
Here you can change the number of canonical variates pairs (components) to include in the model.
Taxonomic level:
Profile level:
Group
General options:
No of components:
5
Regularization method:
Shrinkage
Validation method:
5-fold CV
Submit
Downloads
View options:
Color Palette:
Default
None
Due to the collinearities or near-collinearities in high-dimensional omics datasets, rCCA includes a regularization step in the computations of CCA. Before running rCCA, the regularization parameters λ1 and λ2 need to be tuned or estimated for both the datasets. There are two methods to estimate such regularization parameters (penalties):
- Shrinkage: The input data are centered and scaled for the estimation of the shrinkage parameters and the calculation of the regularized variance-covariance matrices;
- Cross-validation (CV): estimates the penalization parameters using the cross-validation (tune.rcc function).
To estimate the regularization parameters using Cross-validation, you can choose from either 5-fold CV (5-fold Cross-validation) or LOOCV (Leave One Out Cross-validation) (internal) cross-validation method.
Component 2
Component 1
microbiome
resistome

Regularized Canonical Correlation Analysis (rCCA)

General options:

Taxonomic level: Phylum

Profile level: Group

No of components: 5

Regularization method: Shrinkage

Validation method: 5-fold CV

Submit

Downloads

View options:

(for score plot only)

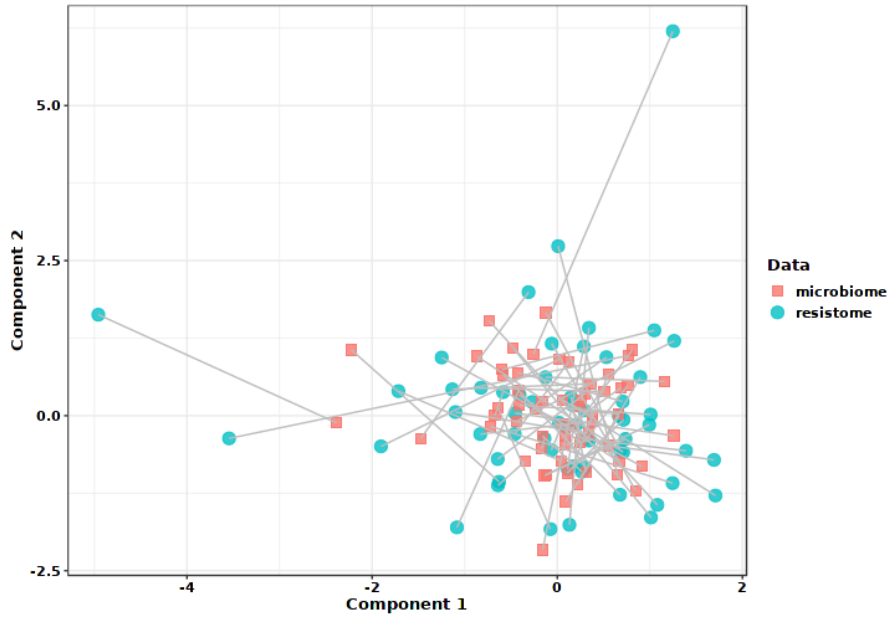
Color data points according to: Data (Omics) / Experimental factor — Time

Color Palette: Default

Label samples by: None

Please Note: The ***Color Palette*** and ***Label samples by*** options are applicable for sample score plot (2D) only.

Sample score plot (2D) | Sample score plot (3D) | Scree plot | Clustered Image Map (CIM) | Correlation Circle Plot



There are four options to visualize the data including Clustered Image Map (CIM), Correlation Circle Plot, Sample score plot (2D), and Sample score Plot (3D).

Regularized Canonical Correlation Analysis (rCCA)

General options:

Taxonomic level: Phylum

Profile level: Group

No of components: 5

Regularization method: Shrinkage

Validation method: 5-fold CV

Submit

Downloads

View options:

(for score plot only)

Color data points according to: Data (Omics) / Experimental factor — Time

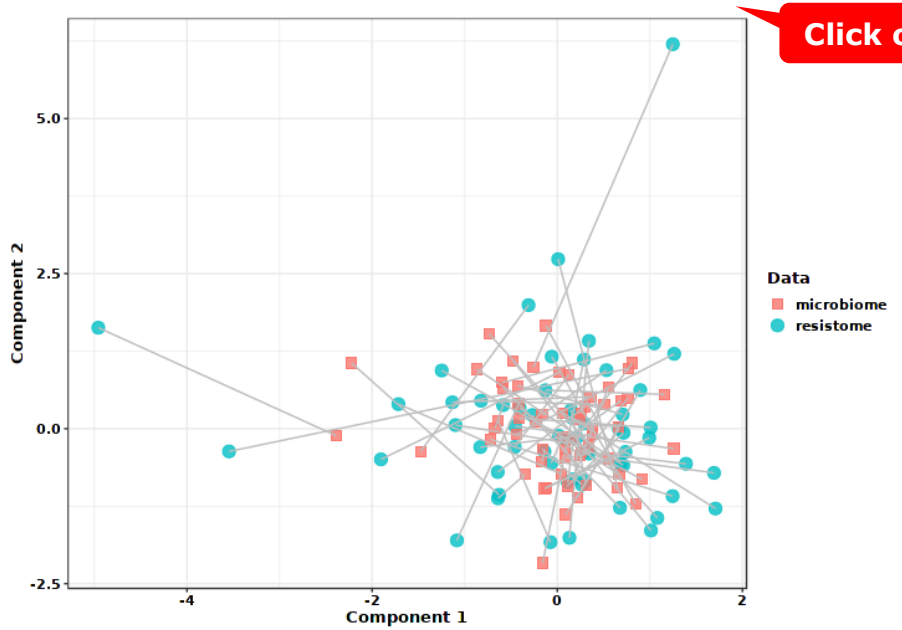
Color Palette: Default

Label samples by: None

Please Note: The ***Color Palette*** and ***Label samples by*** options are applicable for sample score plot (2D) only.

Sample score plot (2D) | Sample score plot (3D) | Scree plot | Clustered Image Map (CIM) | Correlation Circle Plot

Click on Clustered Image Map (CIM).



Regularized Canonical Correlation Analysis (rCCA)

General options:

Taxonomic level: Phylum

Profile level: Group

No of components: 5

Regularization method: Shrinkage

Validation method: 5-fold CV

Submit

Downloads

View options:

(for score plot only)

Color data points according to: Data (Omics) / Experimental factor — Time

Color Palette: Default

Label samples by: None

Please Note: The ***Color Palette*** and ***Label samples by*** options are applicable for sample score plot (2D) only.

Sample score plot (2D) | Sample score plot (3D) | Scree plot | **Clustered Image Map (CIM)** | Correlation Circle Plot

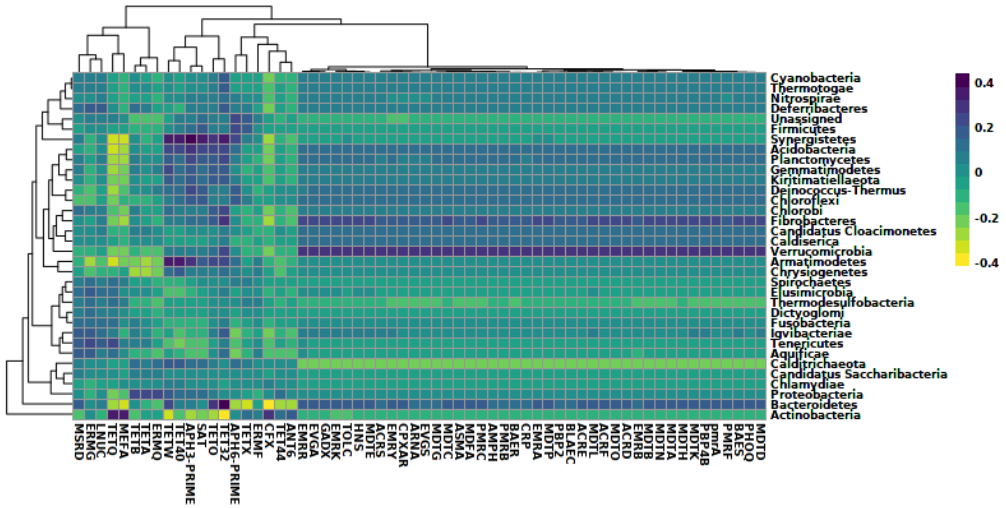
Distance measure: Euclidean

Color contrast: Viridis

Clustering algorithm: Complete

View mode: Overview

Update



Here we can see a heatmap showing the correlation between taxonomic and resistome features. Distance measures and clustering algorithm can be changed as well as the color scheme and visualizing mode. As the values range from -0.3 to 0.3, in this case, strong features were not observed.

Regularized Canonical Correlation Analysis (rCCA)

General options:

Taxonomic level: Phylum

Profile level: Group

No of components: 5

Regularization method: Shrinkage

Validation method: 5-fold CV

Submit

Downloads

View options:

(for score plot only)

Color data points according to: Data (Omics) / Experimental factor — Time

Color Palette: Default

Label samples by: None

Please Note: The ***Color Palette*** and ***Label samples by*** options are applicable for sample score plot (2D) only.

Sample score plot (2D) | Sample score plot (3D) | Scree plot | **Clustered Image Map (CIM)** | Correlation Circle Plot

Now click on Correlation Circle Plot.

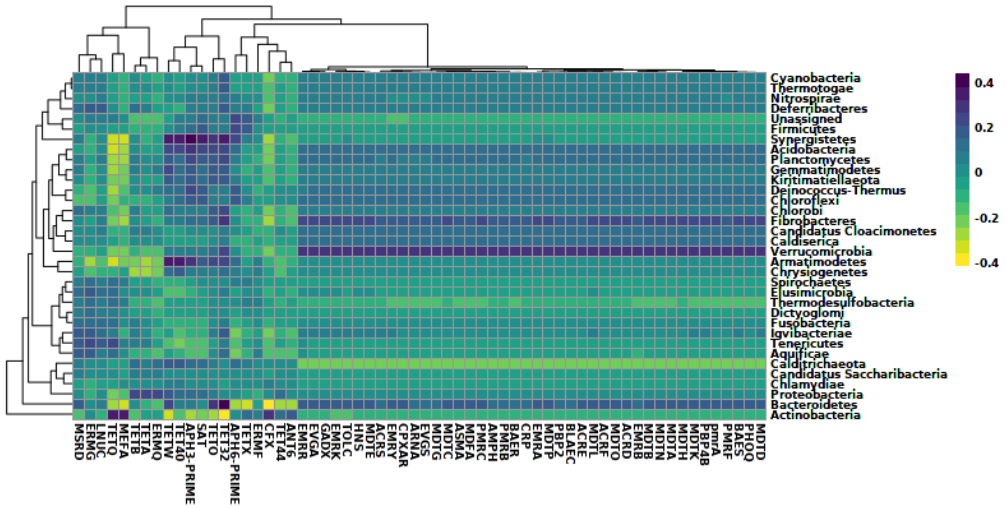
Distance measure: Euclidean

Color contrast: Viridis

Clustering algorithm: Complete

View mode: Overview

Update



Regularized Canonical Correlation Analysis (rCCA)

General options:

Taxonomic level: Phylum

Profile level: Group

No of components: 5

Regularization method: Shrinkage

Validation method: 5-fold CV

Submit

Downloads

View options:

(for score plot only)

Color data points according to: Data (Omics) / Experimental factor — Time

Color Palette: Default

Label samples by: None

Please Note: The ***Color Palette*** and ***Label samples by*** options are applicable for sample score plot (2D) only.

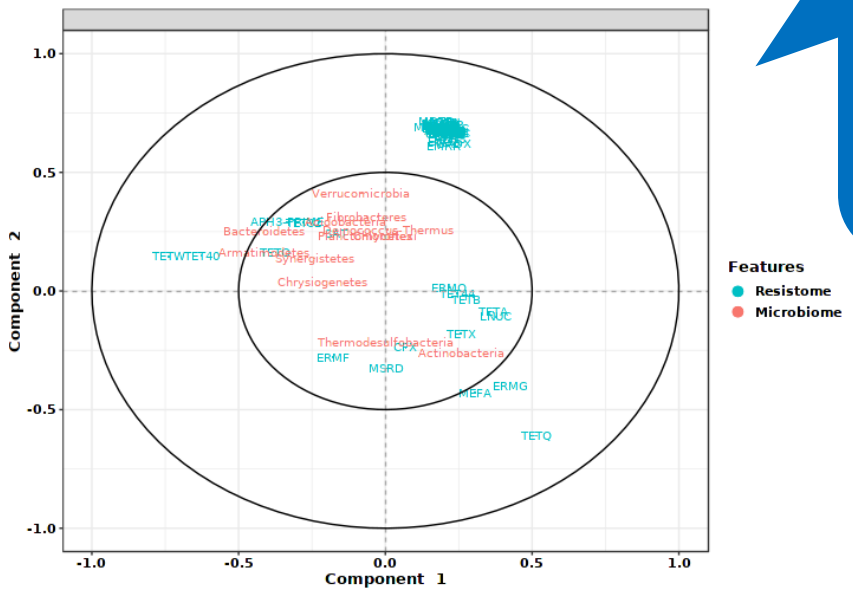
Sample score plot (2D) | Sample score plot (3D) | Scree plot | Clustered Image Map (CIM) | **Correlation Circle Plot**

Component on X-axis: 1

Component on Y-axis: 2

Update

Correlation threshold: 0.2



Here you can see a Principal Component Analysis (PCA) graph showing how samples are clustered based on the variables taken into account.

The correlation between original variables and each component defines the coordination of the variable. This facilitates the visualization of the correlation between variables established by the cosine angle between the center of the plot and the coordinate of the variable.

For more details please refer to the original publication.

Regularized Canonical Correlation Analysis (rCCA)

General options:

Taxonomic level: Phylum

Profile level: Group

No of components: 5

Regularization method: Shrinkage

Validation method: 5-fold CV

Submit

Downloads

View options:

(for score plot only)

Color data points according to: Data (Omics) / Experimental factor — Time

Color Palette: Default

Label samples by: None

Please Note: The ***Color Palette*** and ***Label samples by*** options are applicable for sample score plot (2D) only.

You can choose to download the analyses and/or graphs in a number of different formats by clicking here.

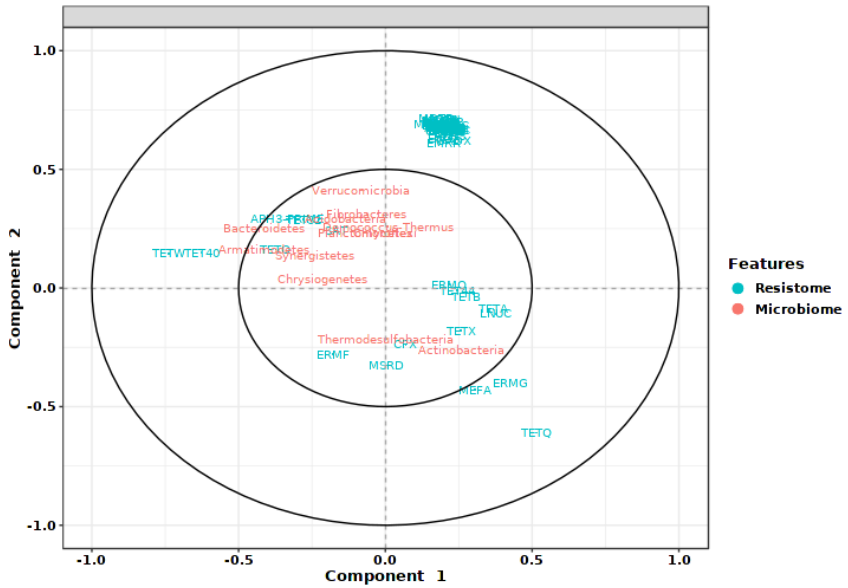
Sample score plot (2D) | Sample score plot (3D) | Scree plot | Clustered Image Map (CIM) | Correlation Circle Plot

Component on X-axis: 1

Component on Y-axis: 2

Update

Correlation threshold: 0.2



Once you are finished, you can click on 'Integrative Analysis' to go back to the previous page.

Regularized Canonical Correlation Analysis (rCCA)

General options:

Taxonomic level: Phylum

Profile level: Group

No of components: 5

Regularization method: Shrinkage

Validation method: 5-fold CV

Submit

Downloads

View options:

(for score plot only)

Color data points according to: Data (Omics) / Experimental factor — Time

Color Palette: Default

Label samples by: None

Please Note: The ***Color Palette*** and ***Label samples by*** options are applicable for sample score plot (2D) only.

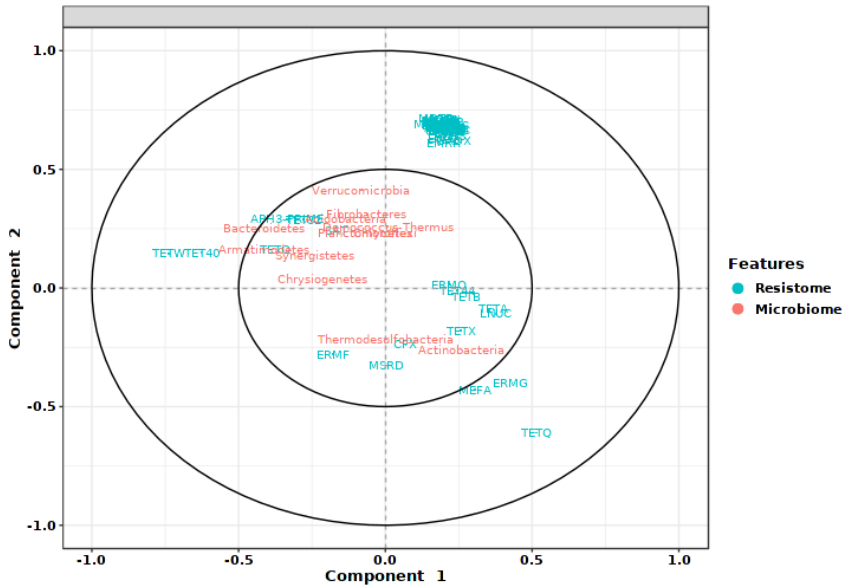
Sample score plot (2D) | Sample score plot (3D) | Scree plot | Clustered Image Map (CIM) | Correlation Circle Plot

Component on X-axis: 1

Component on Y-axis: 2

Update

Correlation threshold: 0.2



Analysis Panel

1. Global Similarity analysis

Procrustes analysis

Co-inertia analysis

2. Ordination-based integrative analysis

rCCA

sPLS

Moving forward to sPLS. Click here.

3. Pair-wise correlation analysis

Spearman

Pearson

MIC

CCLasso

Sparse Partial Least Squares (sPLS)

General options:

Taxonomic level: Phylum

Profile level: Group

No of components: 5

Mode: Regression

Submit

Downloads

View options:

(for score plot only)

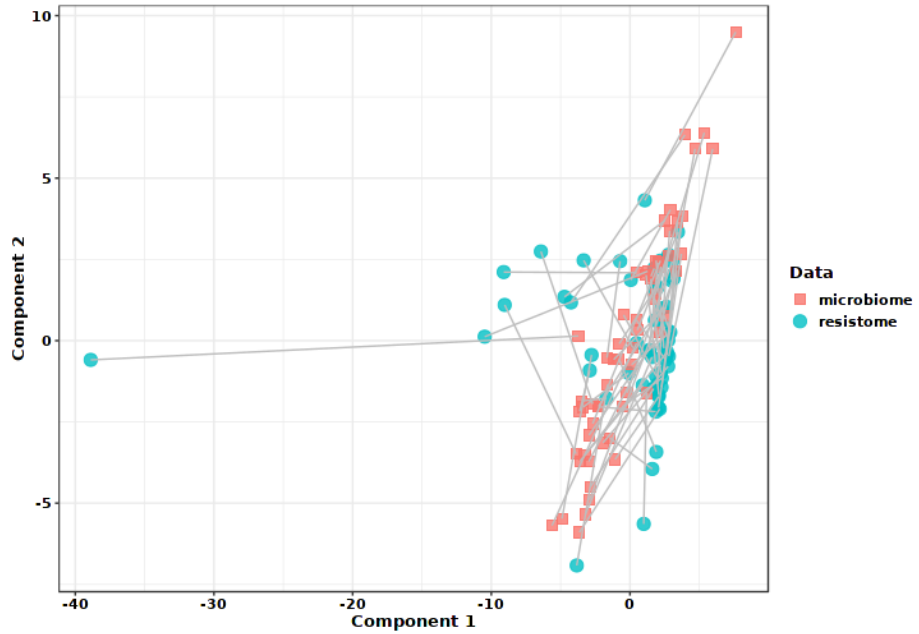
Color data points according to: Data (Omics) / Experimental factor — Time

Color Palette: Default

Label samples by: None

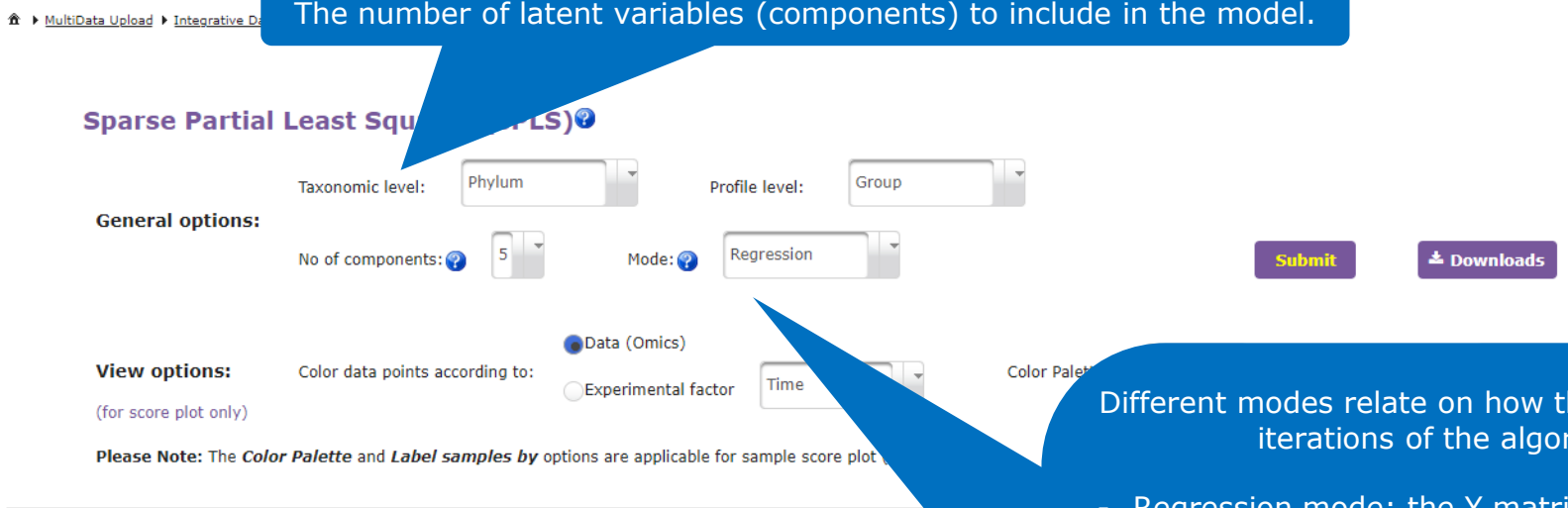
Please Note: The ***Color Palette*** and ***Label samples by*** options are applicable for sample score plot (2D) only.

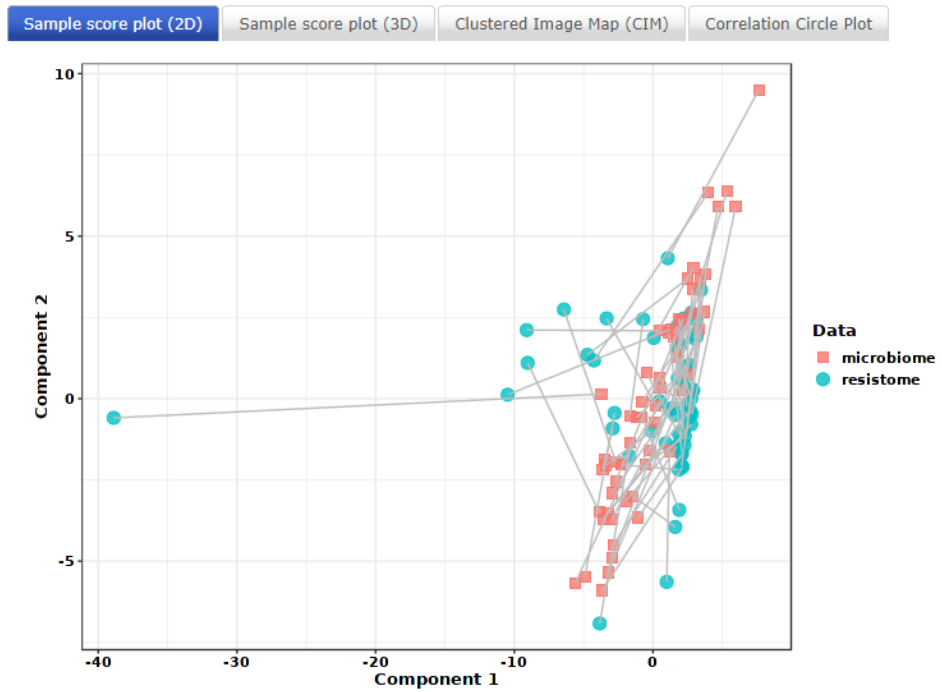
Sample score plot (2D) | Sample score plot (3D) | Clustered Image Map (CIM) | Correlation Circle Plot



Sparse Partial Least Squares (sPLS) is a multivariate methodology which integrates two omics datasets, in ResistoXplorer it is employed to combine metagenomics and resistome data. It acts by maximizing the covariance between components from these datasets. It combines both integration and variable selection simultaneously on two high-dimensional datasets in a one-step strategy.

The number of latent variables (components) to include in the model.





Different modes relate on how the Y matrix (taxonomic profile) is deflated across iterations of the algorithms – *i.e.* the different components.

- Regression mode: the Y matrix is deflated with respect to the information extracted/modelled from the local regression on X (resistome profile). The goal is to predict Y from X (Y and X play an assimetric role). Consequently, the latent variables computed to predict Y from X are different from those computed to predict X from Y.

- Canonical mode: the Y matrix is deflated to the information extracted/modelled from the local regression on Y. Here X and Y play a symetric role and the goal is similar to a Canonical Correlation type of analysis.

- Invariant mode: the Y matrix is not deflated and perform a redundancy analysis.

- Classic mode: similar to a regression mode. It gives identical results for the loading vectors associated to the Y dataset.

Sparse Partial Least Squares (sPLS)

General options:

Taxonomic level: Phylum

Profile level: Group

No of components: 5

Mode: Regression

Submit

Downloads

View options:

(for score plot only)

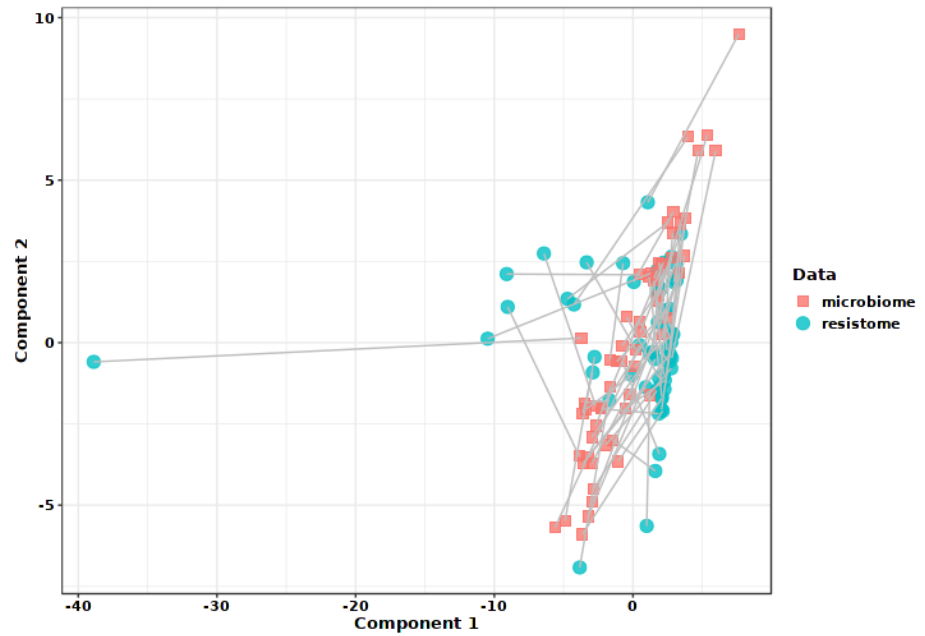
Color data points according to: Data (Omics) / Experimental factor — Time

Color Palette: Default

Label samples by: None

Please Note: The ***Color Palette*** and ***Label samples by*** options are applicable for sample score plot (2D) only.

Sample score plot (2D) | Sample score plot (3D) | Clustered Image Map (CIM) | Correlation Circle Plot



Same as with rCCA, data can be visualized with Clustered Image Map (CIM), Correlation Circle Plot, and Sample score Plots (2D and 3D).

Sparse Partial Least Squares (sPLS)

General options:

Taxonomic level: Phylum

Profile level: Group

No of components: 5

Mode: Regression

Submit

Downloads

You can choose to download the analyses and/or graphs in a number of different formats by clicking here.

View options:

(for score plot only)

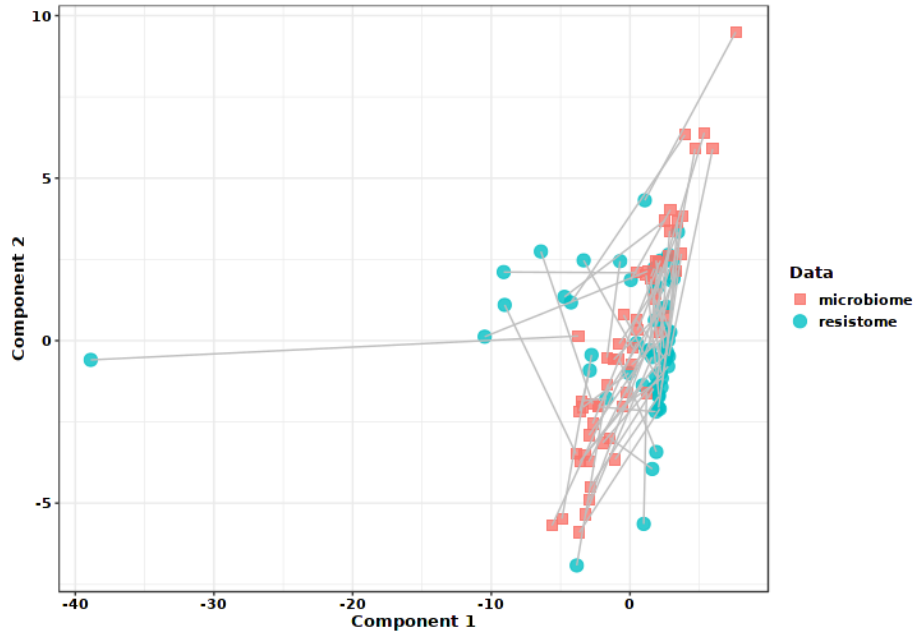
Color data points according to: Data (Omics) / Experimental factor — Time

Color Palette: Default

Label samples by: None

Please Note: The ***Color Palette*** and ***Label samples by*** options are applicable for sample score plot (2D) only.

Sample score plot (2D) | Sample score plot (3D) | Clustered Image Map (CIM) | Correlation Circle Plot



Once you are finished, you can click on 'Integrative Analysis' to go back to the previous page.

Sparse Partial Least Squares (sPLS)

General options:

Taxonomic level: Phylum

Profile level: Group

No of components: 5

Mode: Regression

Submit

Downloads

View options:

(for score plot only)

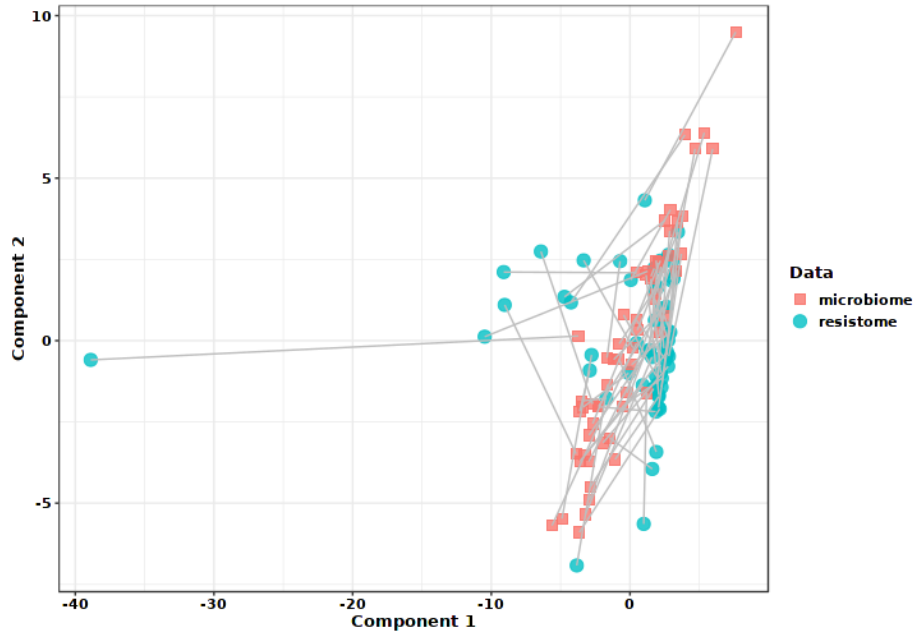
Color data points according to: Data (Omics) / Experimental factor — Time

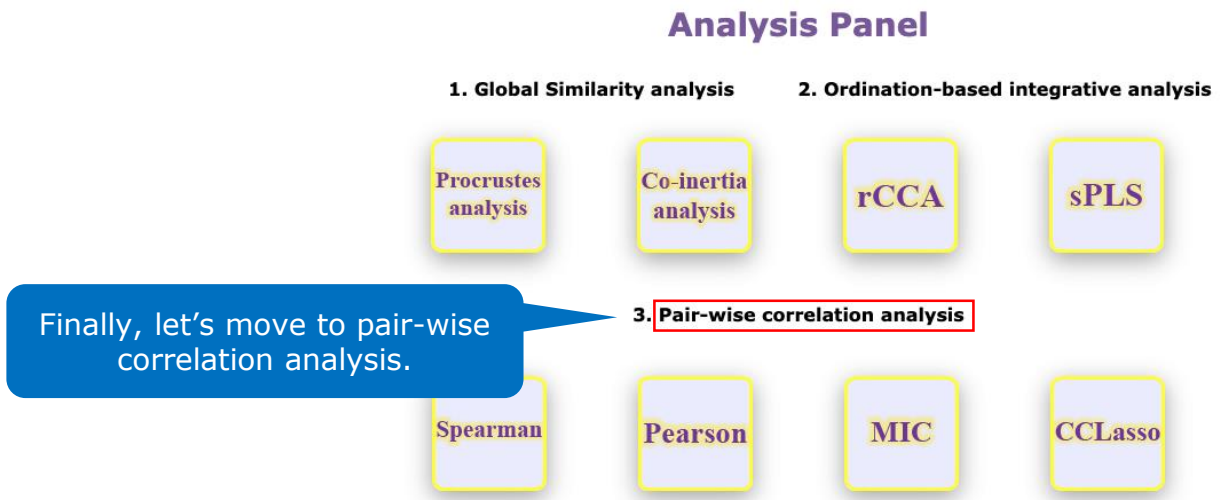
Color Palette: Default

Label samples by: None

Please Note: The ***Color Palette*** and ***Label samples by*** options are applicable for sample score plot (2D) only.

Sample score plot (2D) | Sample score plot (3D) | Clustered Image Map (CIM) | Correlation Circle Plot




Analysis Panel
1. Global Similarity analysis
2. Ordination-based integrative analysis
Procrustes analysis
Co-inertia analysis
rCCA
sPLS
Finally, let's move to pair-wise correlation analysis.
3. Pair-wise correlation analysis
Spearman
Pearson
MIC
CCLasso

What are the options for ‘Pair-wise correlation analysis’ in ResistoXplorer?

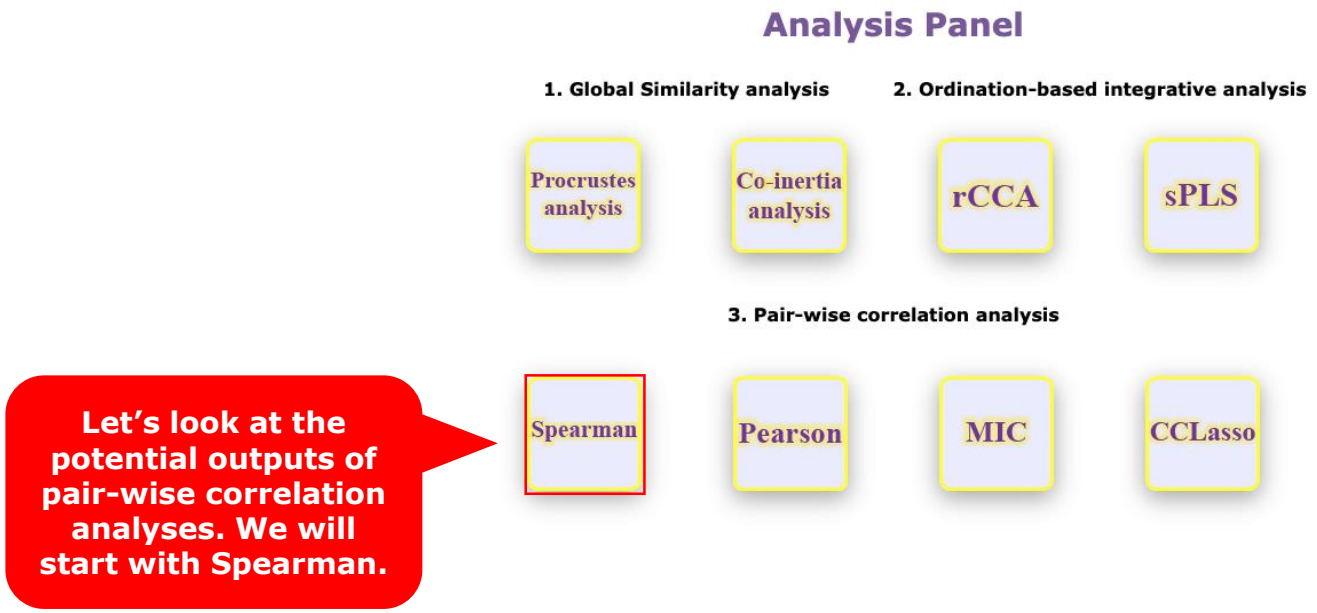
ResistoXplorer offers four options of to perform pair-wise correlation analysis between microbiome and resistome data. The methods are:

- Spearman correlation analysis;
- Pearson correlation analysis;
- Maximal Information Coefficient (MIC) analysis; and
- Correlation inference for compositional data through Lasso (CCLasso).

What information can the ‘Pair-wise correlation analysis’ methods provide?

ResistoXplorer provides four types of correlation analyses, namely Spearman, Pearson, Maximal Information Coefficient (MIC) and Correlation inference for compositional data through Lasso (CCLasso). These methods can be used to determine if there are any strong associations between individual taxa (microbiome) and ARGs (resistome).

Both sets of data (microbiome and resistome) can be analyzed at different taxonomic levels such as phylum, genus and species, and at different functional levels such as class and mechanism (based on available annotations). The level of significance to be considered can also be adjusted.



Let's look at the potential outputs of pair-wise correlation analyses. We will start with Spearman.

Here you will be able to set up the different options for the correlation analysis. Once selections are made, remember to click on 'Submit'.

Spearman rank correlation

It is the most widely used method to perform pair-wise correlation analysis between microbiome and resistome abundance profile. **Please Note:** The microbiome and resistome count [illegible] been transformed to their relative abundances (proportion) in order to perform such analysis. Also, features (i.e., ARGs and species) that are not found (zero count) in more than half of the samples are removed from both datasets before merging [illegible] correlation to alleviate the bias from potential joint-ranking of zero values.

General options:

Taxonomic level: Phylum

Profile level: Group

Corr. coeffecient cutoff: 0.6

Adjusted p-value cutoff: 0.05

p-value correction method: Bonferroni-Hochberg (FDR)

Submit

Downloads

Co-occurance Network

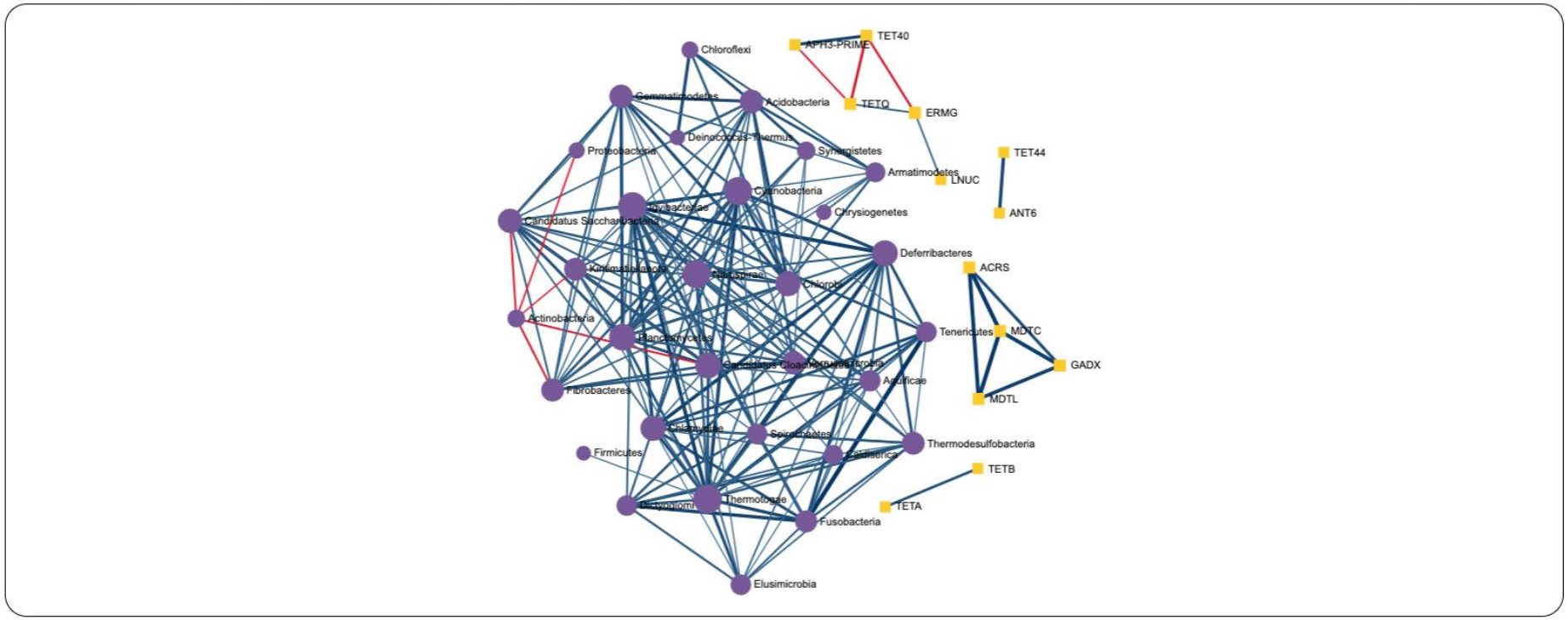
The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout

Node size based on: Degree

Submit

Refresh



MultiData Upload ▸ Integrative Data Summary ▸ Data Filter ▸ Normalization ▸ Integrative Analysis ▸ Procrustes Analysis ▸ Coinertia Analysis ▸ rCCA ▸ sPLS ▸ Spearman association

Spearman rank correlation

It is the most widely used method to perform pair-wise correlation analysis between microbiome and resistome abundance profile. **Please Note:** The microbiome and resistome count data (after data filtering) have been transformed to their relative abundances (proportion) in order to perform such analysis. Also, features (i.e., ARGs and species) that are not found (zero count) in more than half of the samples are removed from both datasets before merging and performing correlation to alleviate the bias from potential joint-ranking of zero values.

General options:

Taxonomic level: Phylum | Profile level: Group | Corr. coeffecient cutoff: 0.6

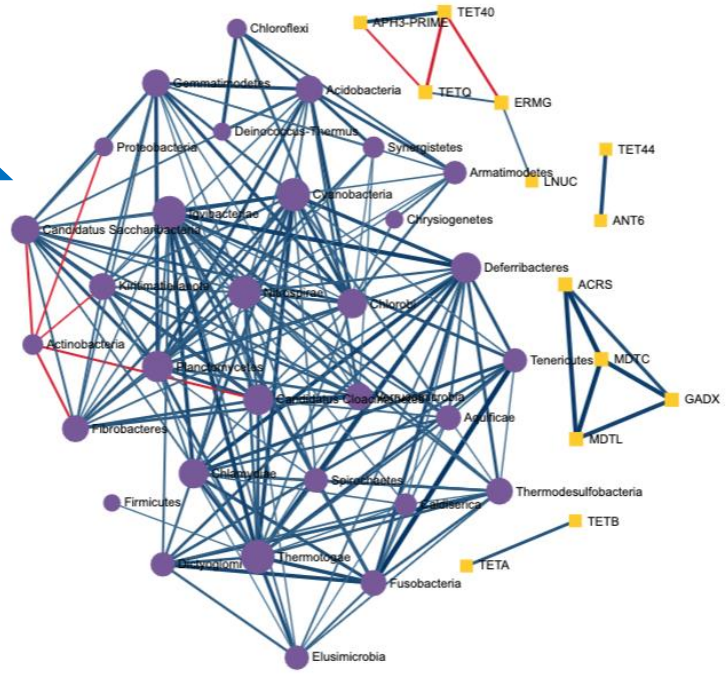
Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

Submit | Downloads

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



The size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them. If you double-click a node, it will highlight all the associated or correlated neighbour nodes.

The results can be visualized in the 'co-occurrence network'. Resistome data is shown in yellow squares, while microbiome data are purple dots. Positive correlations are marked as blue lines, while negative correlations are in red.

Spearman rank correlation

It is the most widely used method to perform pair-wise correlation analysis between microbiome and resistome abundance profile. **Please Note:** The microbiome and resistome count data (after data filtering) have been transformed to their relative abundances (proportion) in order to perform such analysis. Also, features (i.e., ARGs and species) that are not found (zero count) in more than half of the samples are removed from both datasets before merging and performing correlation to alleviate the bias from potential joint-ranking of zero values.

General options:

Taxonomic level: Phylum | Profile level: Group | Corr. coeffecient cutoff: 0.6

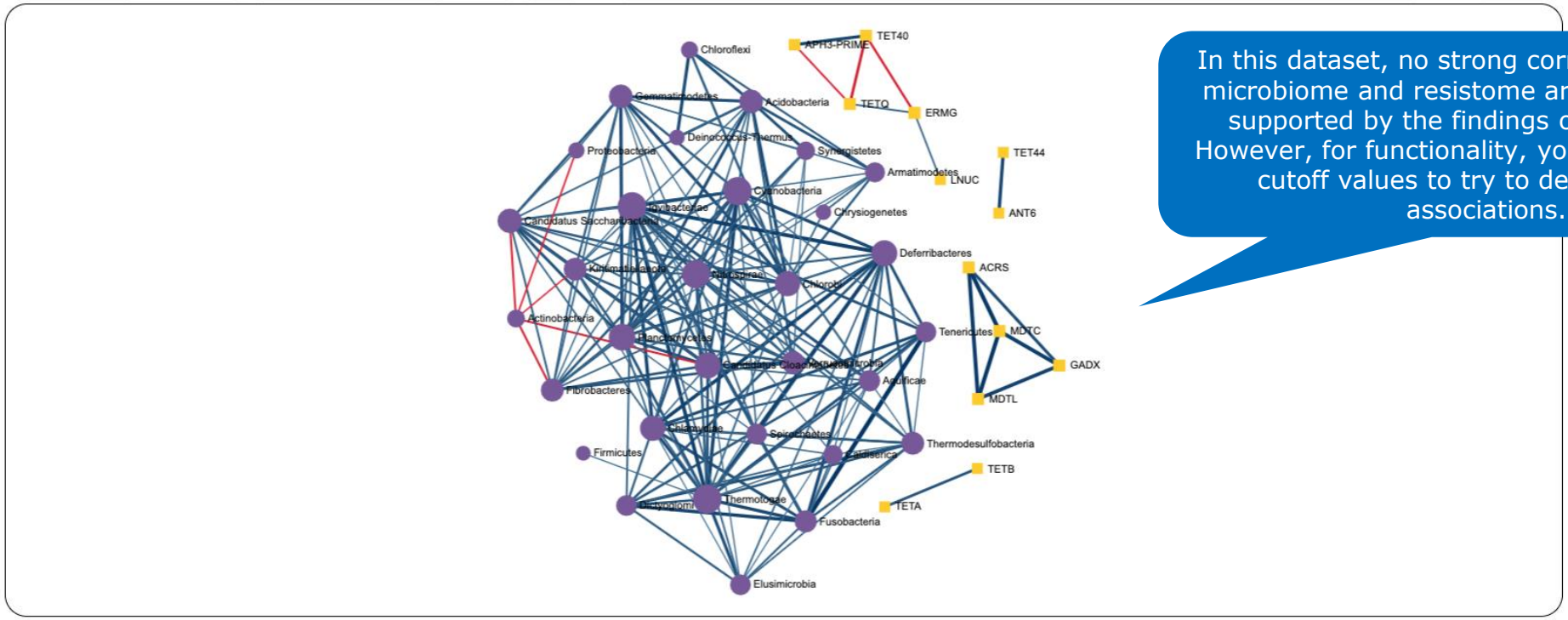
Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

Submit | Downloads

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



In this dataset, no strong correlation between microbiome and resistome are seen, which is supported by the findings of the authors. However, for functionality, you can reduce the cutoff values to try to detect weaker associations.

MultiData Upload ▸ Integrative Data Summary ▸ Data Filter ▸ Normalization ▸ Integrative Analysis ▸ Procrustes Analysis ▸ Coinertia Analysis ▸ rCCA ▸ sPLS ▸ Spearman association

To try to detect weaker correlations, you can select the cutoff to be 0.3.

Spearman rank correlation

It is the most widely used method to perform pair-wise correlation analysis between microbiome and resistome abundance profile. **Please Note:** The microbiome and resistome count data (af… …have been transformed to their relative abundances (proportion) in order to perform such analysis. Also, features (i.e., ARGs and species) that are not found (zero count) in more than half of the samples are removed from both datasets before merging and perf… …correlation to alleviate the bias from potential joint-ranking of zero values.

General options:

Taxonomic level: Phylum

Profile level: Group

Corr. coeffecient cutoff: 0.6

Adjusted p-value cutoff: 0.05

p-value correction method: Bonferroni-Hochberg (FDR)

Submit

Downloads

After you changed the cutoff, click on 'Submit'.

Co-occurance Network

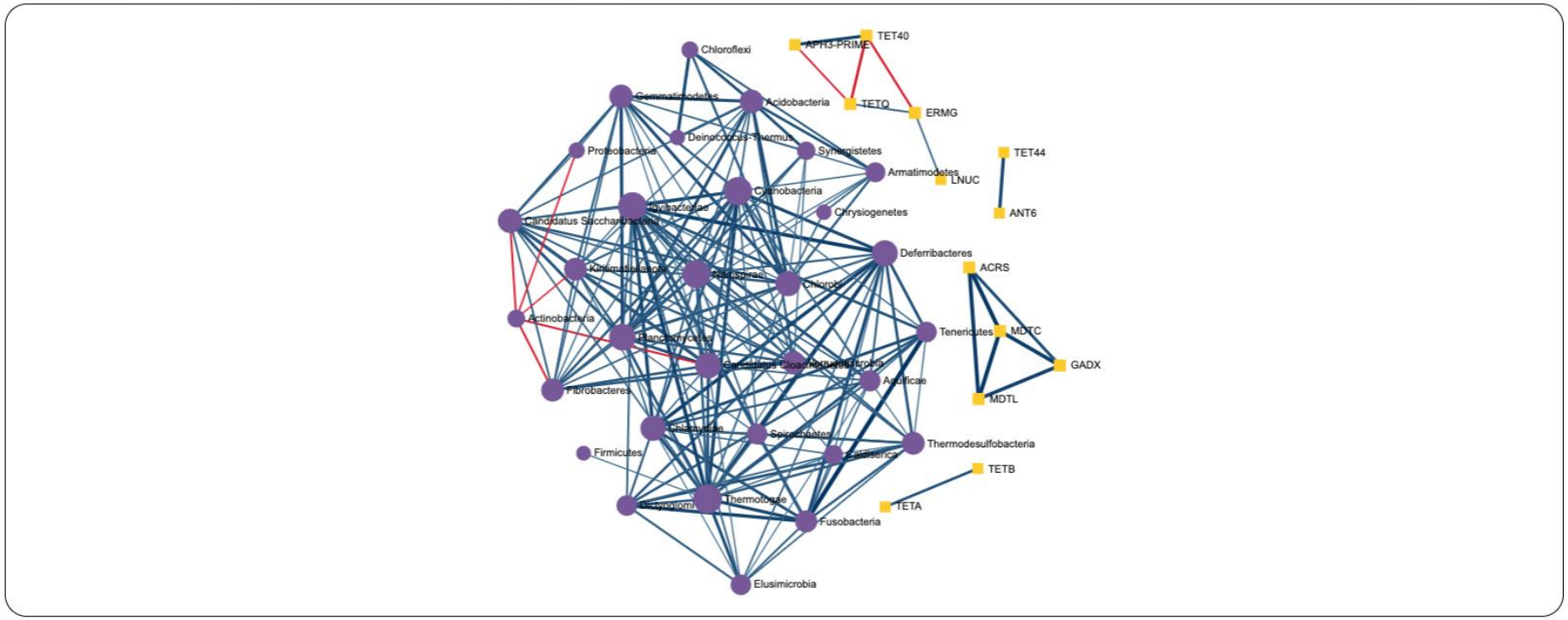
The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality… …while the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout

Node size based on: Degree

Submit

Refresh



Spearman rank correlation

It is the most widely used method to perform pair-wise correlation analysis between microbiome and resistome abundance profile. **Please Note:** The microbiome and resistome count data (after data filtering) have been transformed to their relative abundances (proportion) in order to perform such analysis. Also, features (i.e., ARGs and species) that are not found (zero count) in more than half of the samples are removed from both datasets before merging and performing correlation to alleviate the bias from potential joint-ranking of zero values.

General options:

Taxonomic level: Phylum | Profile level: Group | Corr. coeffecient cutoff: 0.3

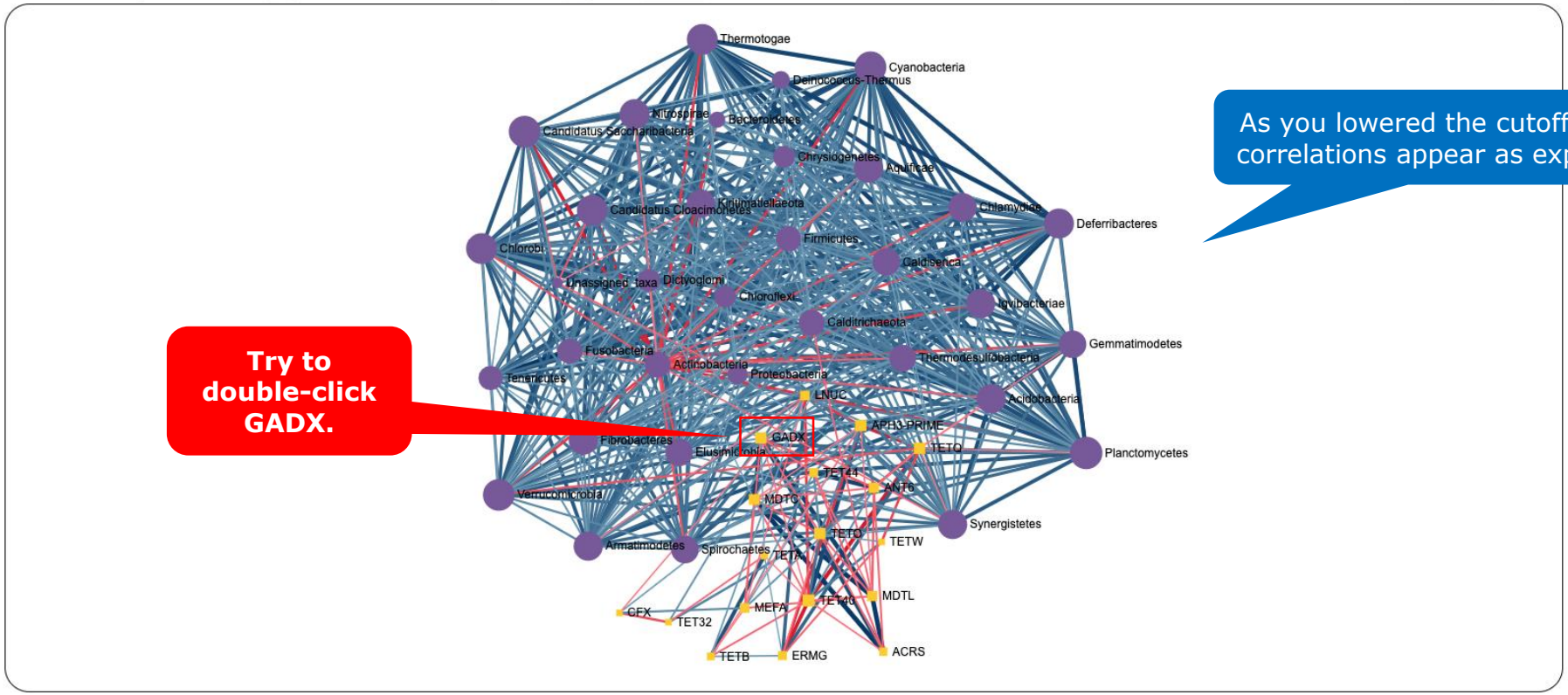
Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

Submit | Downloads

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



Try to double-click GADX.

As you lowered the cutoff, more correlations appear as expected.

Spearman rank correlation

It is the most widely used method to perform pair-wise correlation analysis between microbiome and resistome abundance profile. **Please Note:** The microbiome and resistome count data (after data filtering) have been transformed to their relative abundances (proportion) in order to perform such analysis. Also, features (i.e., ARGs and species) that are not found (zero count) in more than half of the samples are removed from both datasets before merging and performing correlation to alleviate the bias from potential joint-ranking of zero values.

General options:

Taxonomic level: Phylum | Profile level: Group | Corr. coeffecient cutoff: 0.3

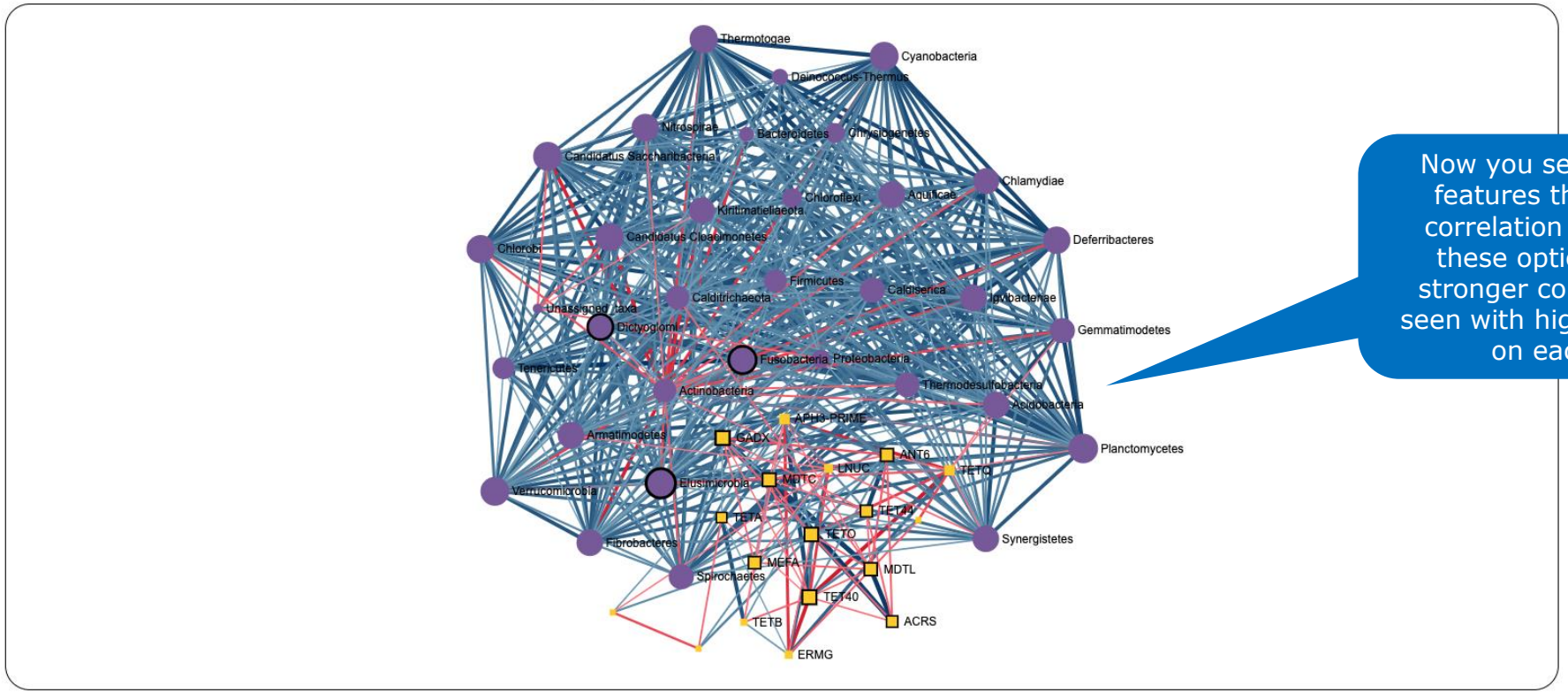
Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

Submit | Downloads

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



Now you see in highlight all features that presented a correlation with GADX with these options. Of course, stronger correlations will be seen with higher cutoffs based on each dataset.

Spearman rank correlation

It is the most widely used method to perform pair-wise correlation analysis between microbiome and resistome abundance profile. **Please Note:** The microbiome and resistome count data (after data filtering) have been transformed to their relative abundances (proportion) in order to perform such analysis. Also, features (i.e., ARGs and species) that are not found (zero count) in more than half of the samples are removed from both datasets before merging and performing correlation to alleviate the bias from potential joint-ranking of zero values.

General options:

Taxonomic level: Phylum | Profile level: Group | Corr. coeffecient cutoff: 0.3

Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

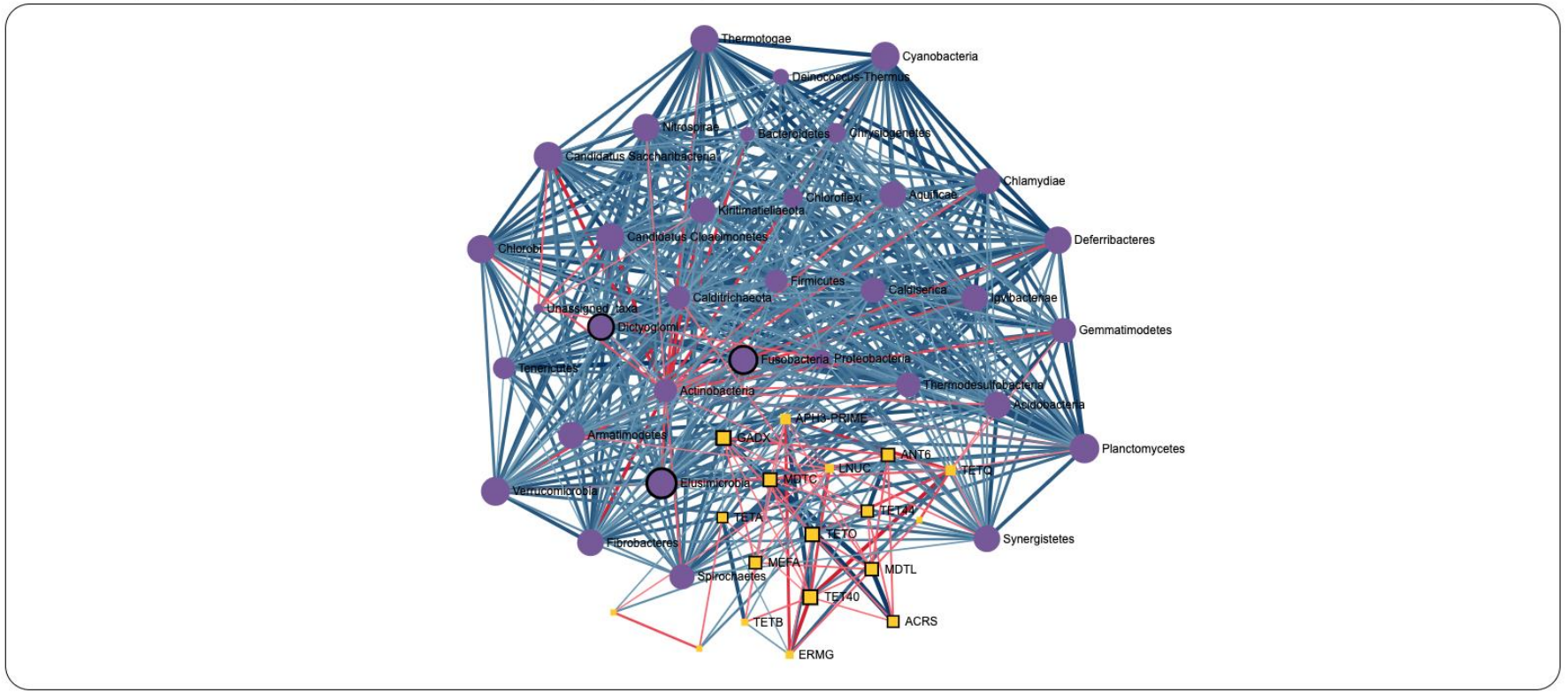
Submit | Downloads

You can choose to download the analyses and/or graphs in a number of different formats by clicking here.

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



Once you are finished, you can click on 'Integrative Analysis' to go back to the previous page.

Spearman rank correlation

It is the most widely used method to perform pair-wise correlation analysis between microbiome and resistome abundance profile. **Please Note:** The microbiome and resistome count data (after data filtering) have been transformed to their relative abundances (proportion) in order to perform such analysis. Also, features (i.e., ARGs and species) that are not found (zero count) in more than half of the samples are removed from both datasets before merging and performing correlation to alleviate the bias from potential joint-ranking of zero values.

General options:

Taxonomic level: Phylum | Profile level: Group | Corr. coeffecient cutoff: 0.3

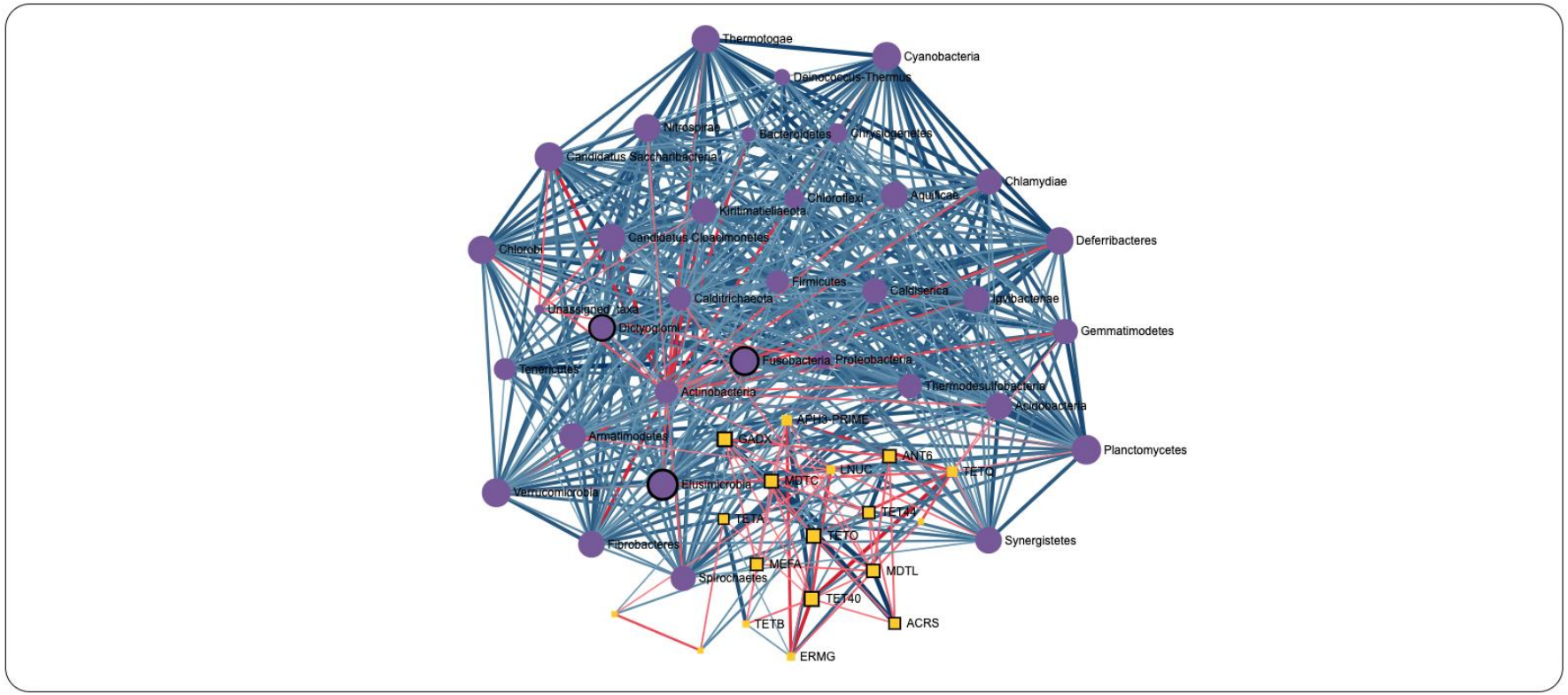
Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

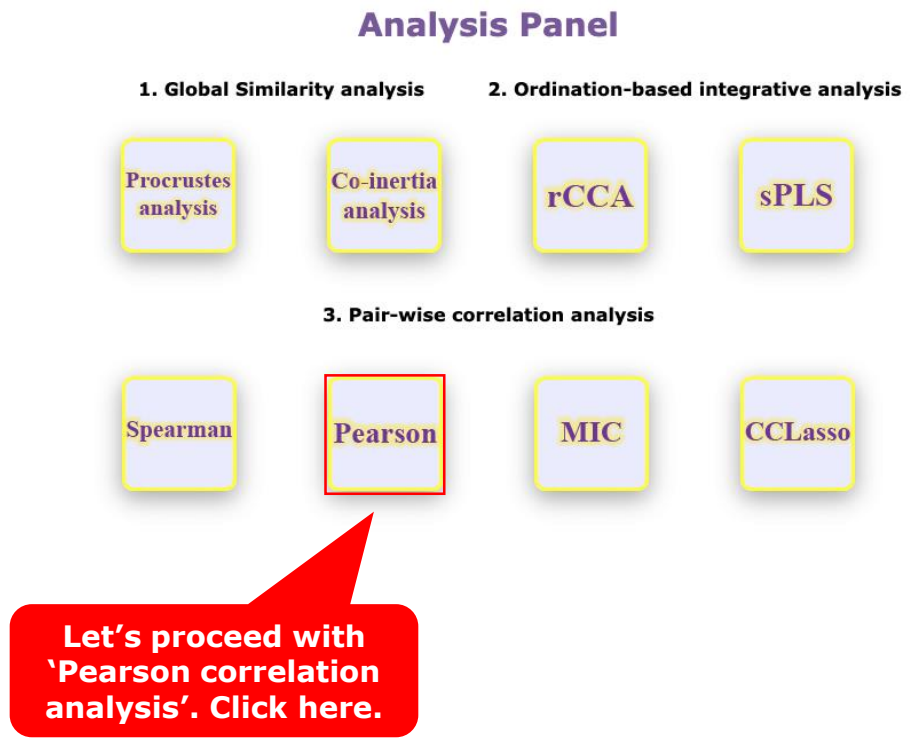
Submit | Downloads

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh




Analysis Panel
1. Global Similarity analysis
2. Ordination-based integrative analysis
Procrustes analysis
Co-inertia analysis
rCCA
sPLS
3. Pair-wise correlation analysis
Spearman
Pearson
MIC
CCLasso
Let's proceed with 'Pearson correlation analysis'. Click here.

Here you will be able to set up the different options for the correlation analysis. Once selections are made, remember to click on 'Submit'.

Pearson correlation

Please Note: The microbiome and resistome count data (after data filtering) have been transformed to their relative abundances (proportion) in order to perform pair-wise pearson correla… …filtering i.e., features (ARGs and species) that are missing (zero count) in more than half of the samples are removed from both datasets before merging and performing such analysis.

General options:

Taxonomic level: Phylum

Profile level: Group

Corr. coeffecient cutoff: 0.6

Adjusted p-value cutoff: 0.05

p-value correction method: Bonferroni-Hochberg (FDR)

Submit

Downloads

The main difference between Pearson and Spearman correlation coefficients is that Pearson assesses linear and continuous relationships, while Spearman addresses rank relationships.

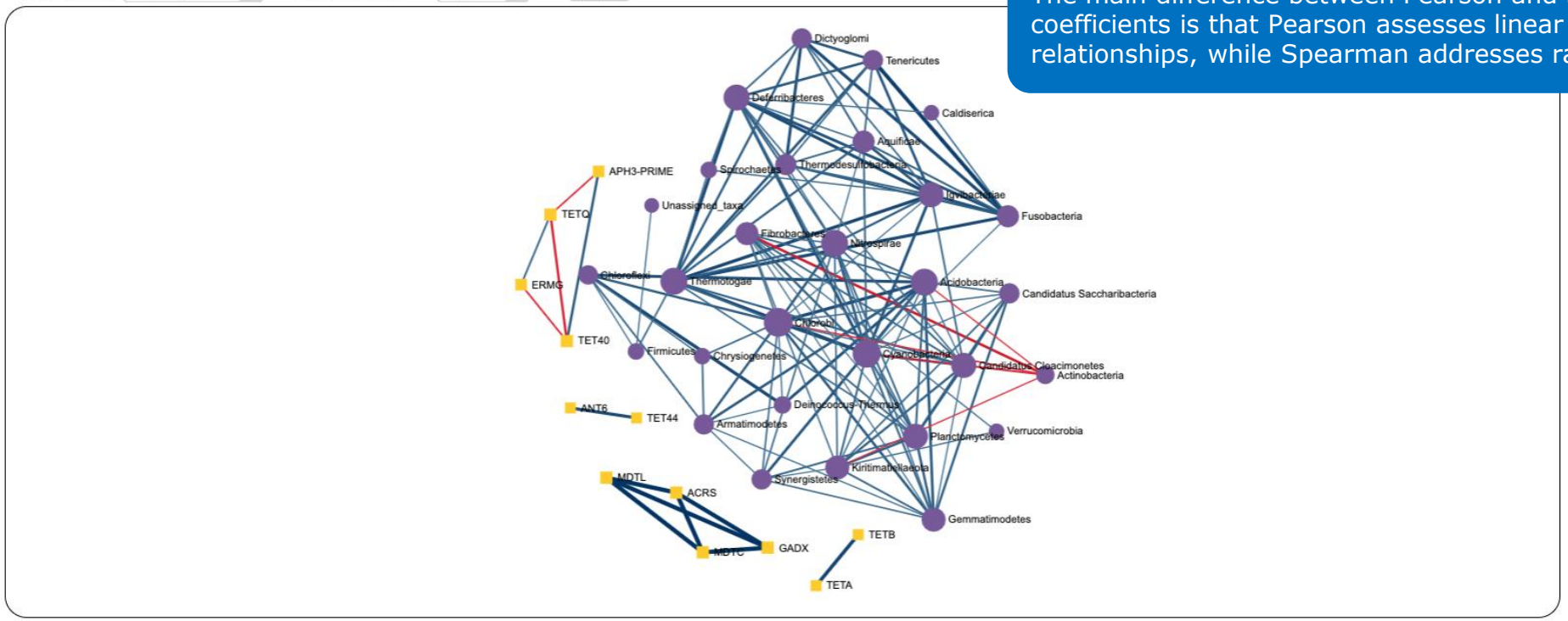
Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (… …the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout

Node size based on: Degree

Submit



MultiData Upload ▸ Integrative Data Summary ▸ Data Filter ▸ Normalization ▸ Integrative Analysis ▸ Procrustes Analysis ▸ Coinertia Analysis ▸ rCCA ▸ sPLS ▸ Spearman association ▸ Pearson association

Pearson correlation

Please Note: The microbiome and resistome count data (after data filtering) have been transformed to their relative abundances (proportion) in order to perform pair-wise pearson correlation analysis. Also, an additional data filtering i.e., features (ARGs and species) that are missing (zero count) in more than half of the samples are removed from both datasets before merging and performing such analysis.

General options:

Taxonomic level: Phylum | Profile level: Group | Corr. coeffecient cutoff: 0.6

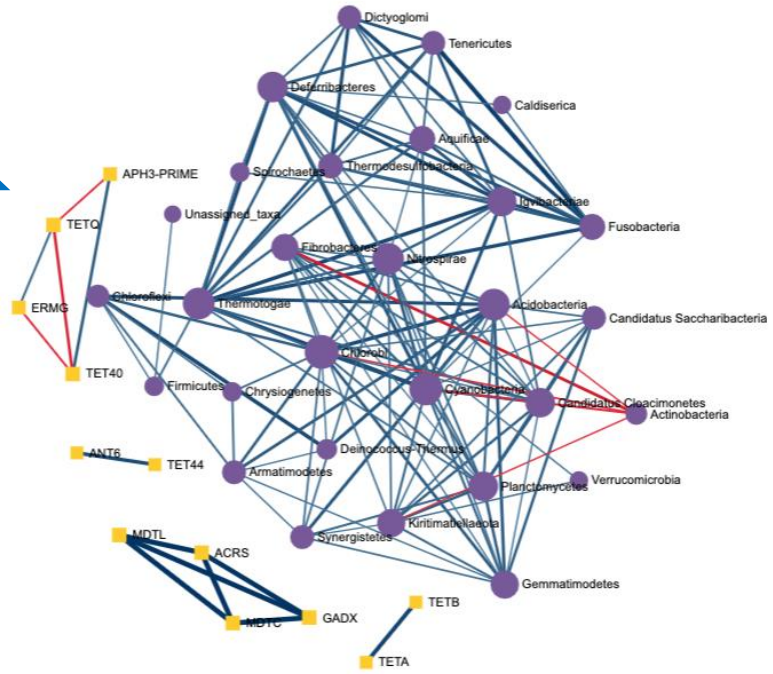
Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

Submit | Downloads

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



The size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them. If you double-click a node, it will highlight all the associated or correlated neighbour nodes.

Likewise, the results can be visualized in the 'co-occurrence network'. Resistome data is shown in yellow squares, while microbiome data are purple dots. Positive correlations are marked as blue lines, while negative correlations are in red.

Pearson correlation

Please Note: The microbiome and resistome count data (after data filtering) have been transformed to their relative abundances (proportion) in order to perform pair-wise pearson correlation analysis. Also, an additional data filtering i.e., features (ARGs and species) that are missing (zero count) in more than half of the samples are removed from both datasets before merging and performing such analysis.

General options:

Taxonomic level: Phylum | Profile level: Group | Corr. coeffecient cutoff: 0.6

Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

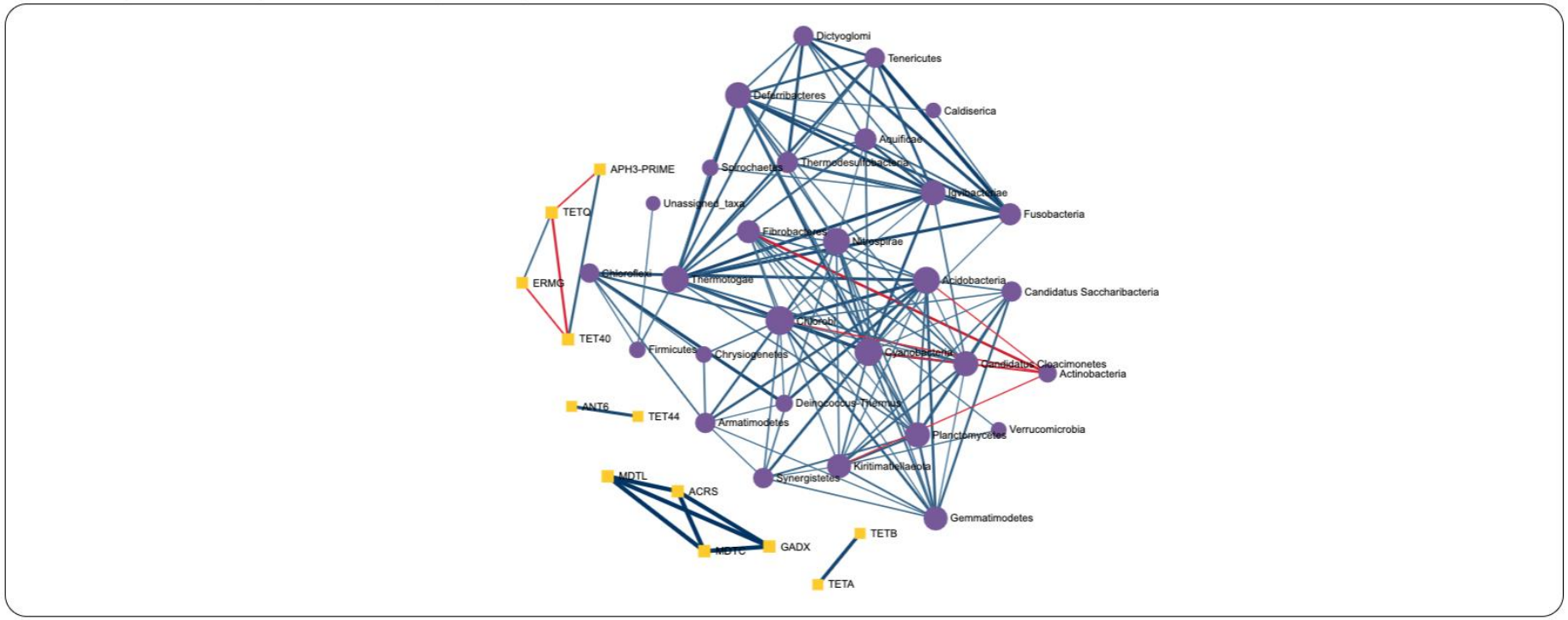
Submit | Downloads

You can choose to download the analyses and/or graphs in a number of different formats by clicking here.

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



Once you are finished, you can click on 'Integrative Analysis' to go back to the previous page.

Pearson correlation

Please Note: The microbiome and resistome count data (after data filtering) have been transformed to their relative abundances (proportion) in order to perform pair-wise pearson correlation analysis. Also, an additional data filtering i.e., features (ARGs and species) that are missing (zero count) in more than half of the samples are removed from both datasets before merging and performing such analysis.

General options:

Taxonomic level: Phylum | Profile level: Group | Corr. coeffecient cutoff: 0.6

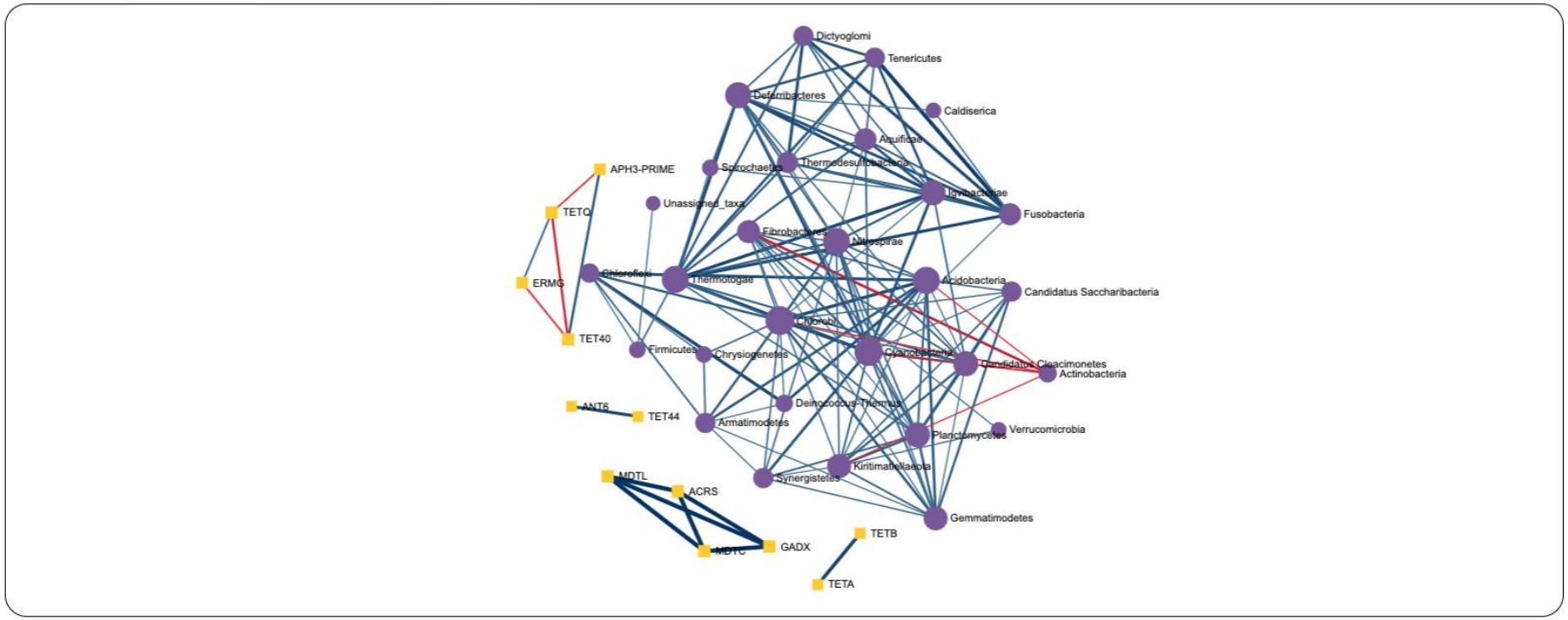
Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

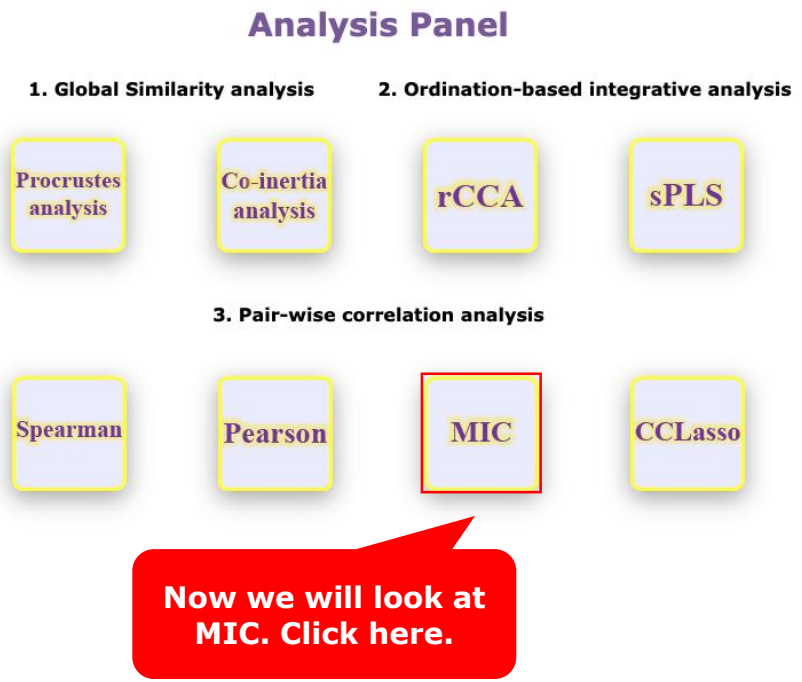
Submit | Downloads

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh




Analysis Panel
1. Global Similarity analysis
2. Ordination-based integrative analysis
Procrustes analysis
Co-inertia analysis
rCCA
sPLS
3. Pair-wise correlation analysis
Spearman
Pearson
MIC
CCLasso
Now we will look at MIC. Click here.

Maximal Information Coefficient (MIC) analysis

General options:

Taxonomic level: Phylum | Profile level: Class | No. of permutations: 1000

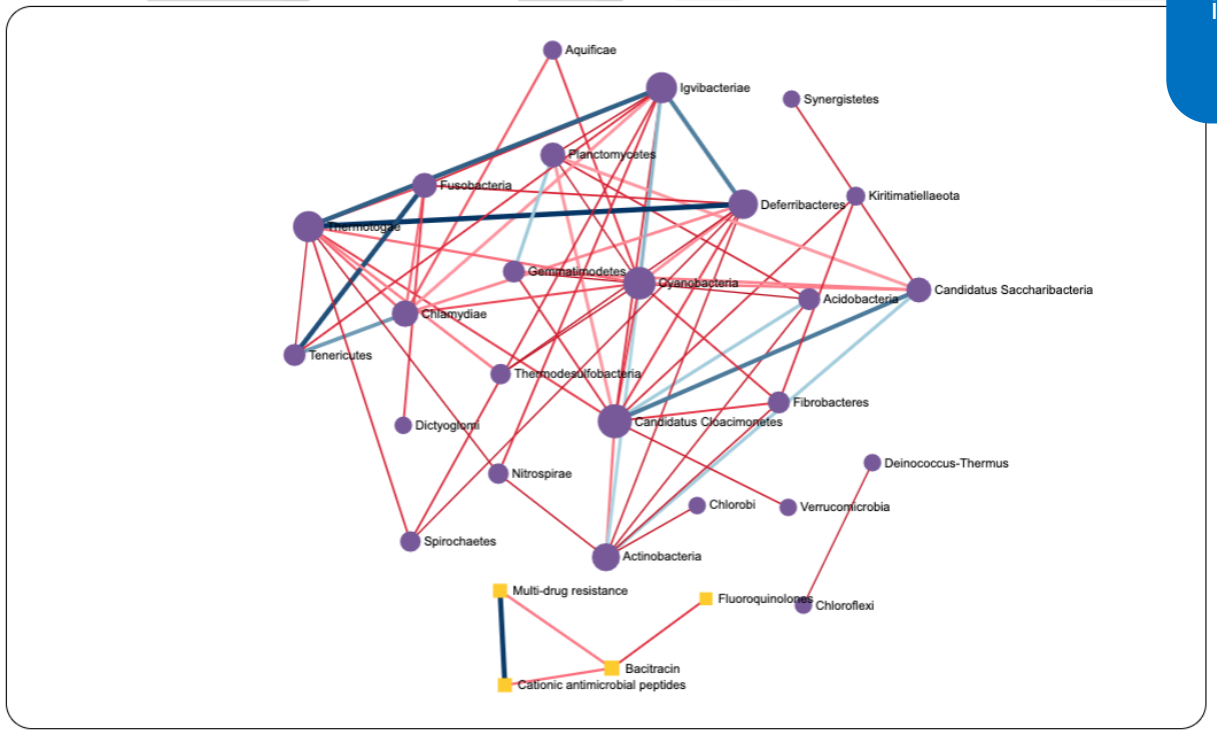
Corr. coeffecient cutoff: 0.6 | Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

Submit | Downloads

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the strength of correlation between them (**MIC**: varies from 0 to 1). User can **zoom** in and ou the network along with **dragging** nodes.Also, **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



Maximal Information Coefficient (MIC) is a measure of two-variable dependence designed specifically for rapid exploration of many dimensional data sets. MIC is part of a larger family of maximal-information-based nonparametric exploration (MINE) statistics, which can be used not only to identify important relationships in data sets but also to characterize them.

Maximal Information Coefficient (MIC) analysis

General options:

Taxonomic level: Phylum | Profile level: Class | No. of permutations: 1000

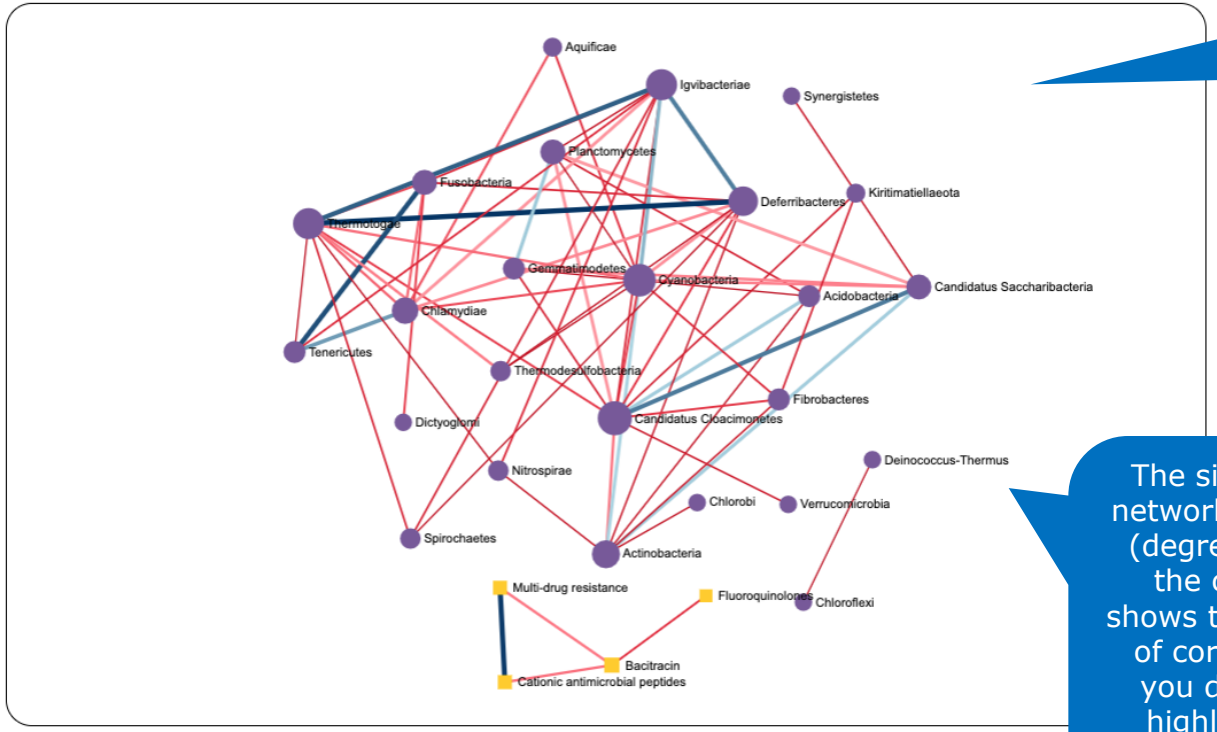
Corr. coeffecient cutoff: 0.6 | Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

Submit | Downloads

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the strength of correlation between them (**MIC**: varies from 0 to 1). User can **zoom** in and out the network along with **dragging** nodes.Also, **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



Likewise, the results can be visualized in the 'co-occurrence network'. Resistome data is shown in yellow squares, while microbiome data are purple dots. Positive correlations are marked as blue lines, while negative correlations are in red.

The size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them. If you double-click a node, it will highlight all the associated or correlated neighbour nodes.

Maximal Information Coefficient (MIC) analysis

General options:

Taxonomic level: Phylum | Profile level: Class | No. of permutations: 1000

Corr. coeffecient cutoff: 0.6 | Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

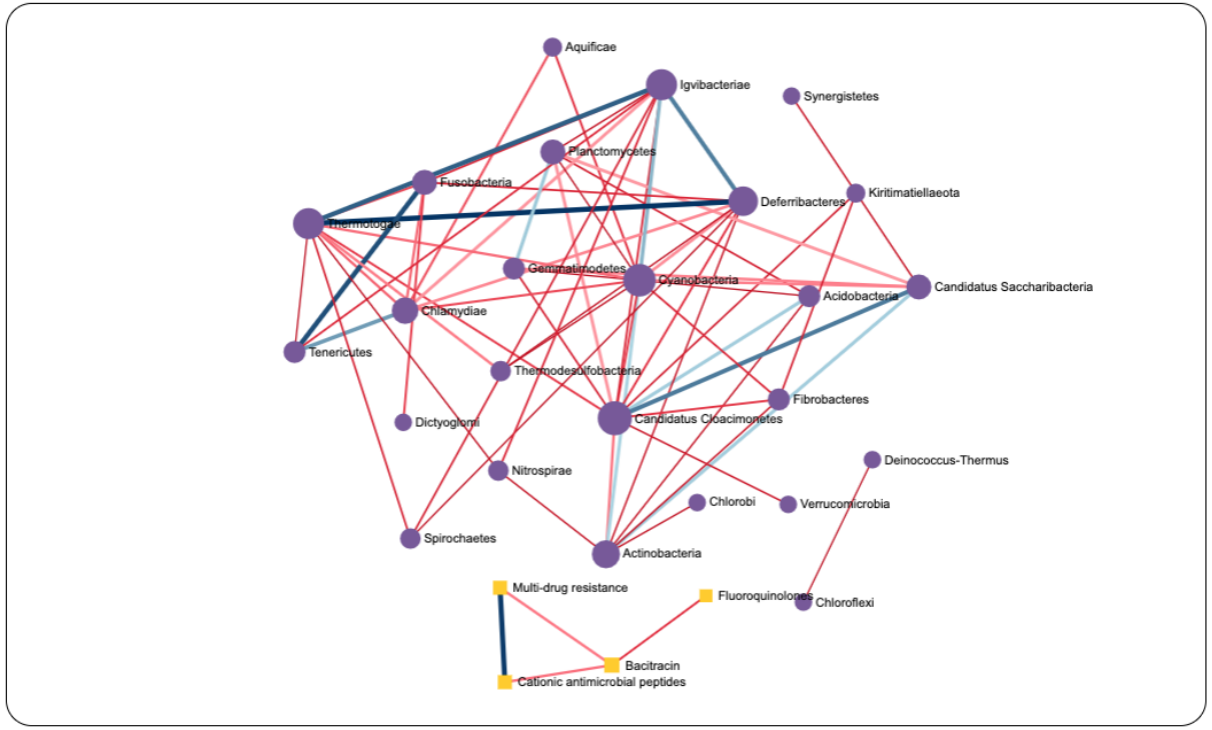
Submit | Downloads

You can choose to download the analyses and/or graphs in a number of different formats by clicking here.

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the strength of correlation between them (**MIC**: varies from 0 to 1). User can **zoom** in and out the network along with **dragging** nodes.Also, **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



Once you are finished, you can click on 'Integrative Analysis' to go back to the previous page.

Maximal Information Coefficient (MIC) analysis

General options:

Taxonomic level: Phylum | Profile level: Class | No. of permutations: 1000

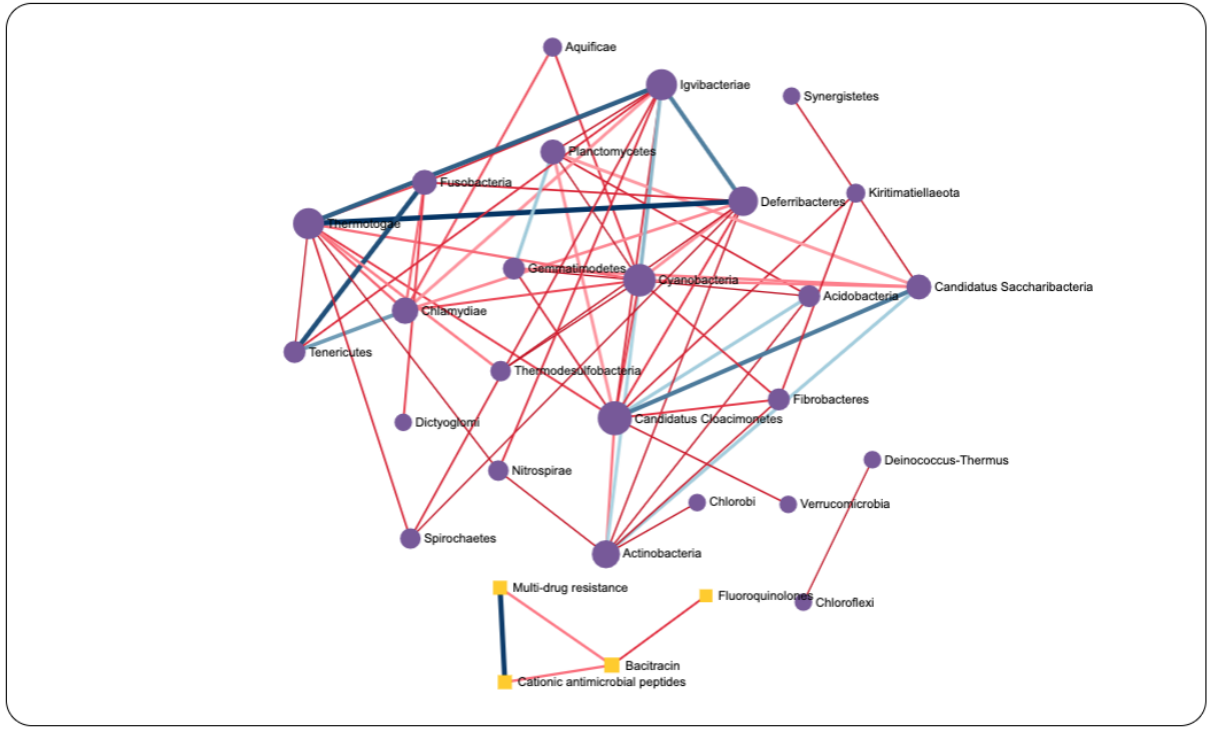
Corr. coeffecient cutoff: 0.6 | Adjusted p-value cutoff: 0.05 | p-value correction method: Bonferroni-Hochberg (FDR)

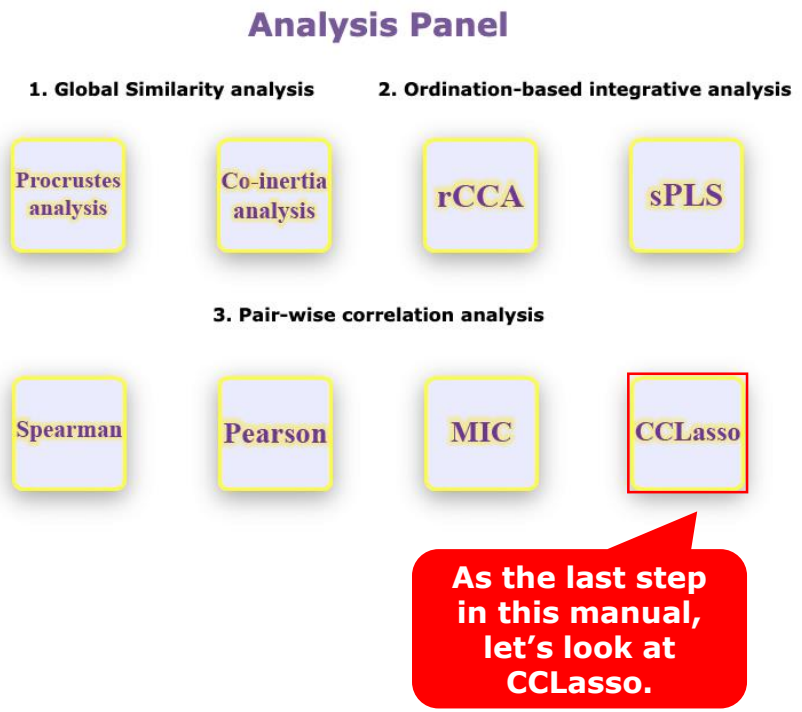
Submit | Downloads

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the strength of correlation between them (**MIC**: varies from 0 to 1). User can **zoom** in and out the network along with **dragging** nodes.Also, **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh




Analysis Panel
1. Global Similarity analysis
2. Ordination-based integrative analysis
Procrustes analysis
Co-inertia analysis
rCCA
sPLS
3. Pair-wise correlation analysis
Spearman
Pearson
MIC
CCLasso
As the last step in this manual, let's look at CCLasso.

Correlation inference for compositional data through Lasso (CCLasso)

General options:

Taxonomic level: Phylum

Profile level: Class

p-value cutoff: 0.05

Corr. coeffecient cutoff: 0.6

No. of max. iterations: 10

No. of Bootstraps: 10

Submit

Downloads

Co-occurance Network

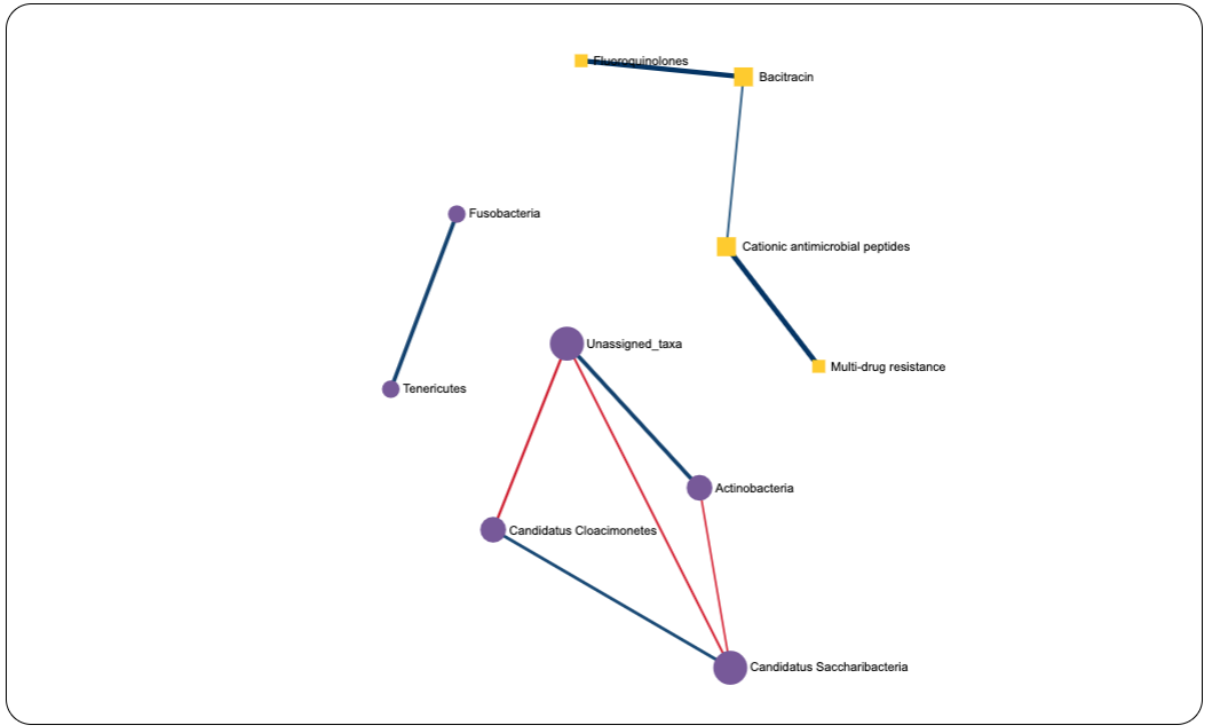
The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout

Node size based on: Degree

Submit

Refres



CCLasso is based on least squares with L1 penalty after log ratio transformation for compositional (metagenomic) data to infer the correlations among features through a latent variable model.

Correlation inference for compositional data through Lasso (CCLasso)

General options:

Taxonomic level: Phylum | Profile level: Class | p-value cutoff: 0.05

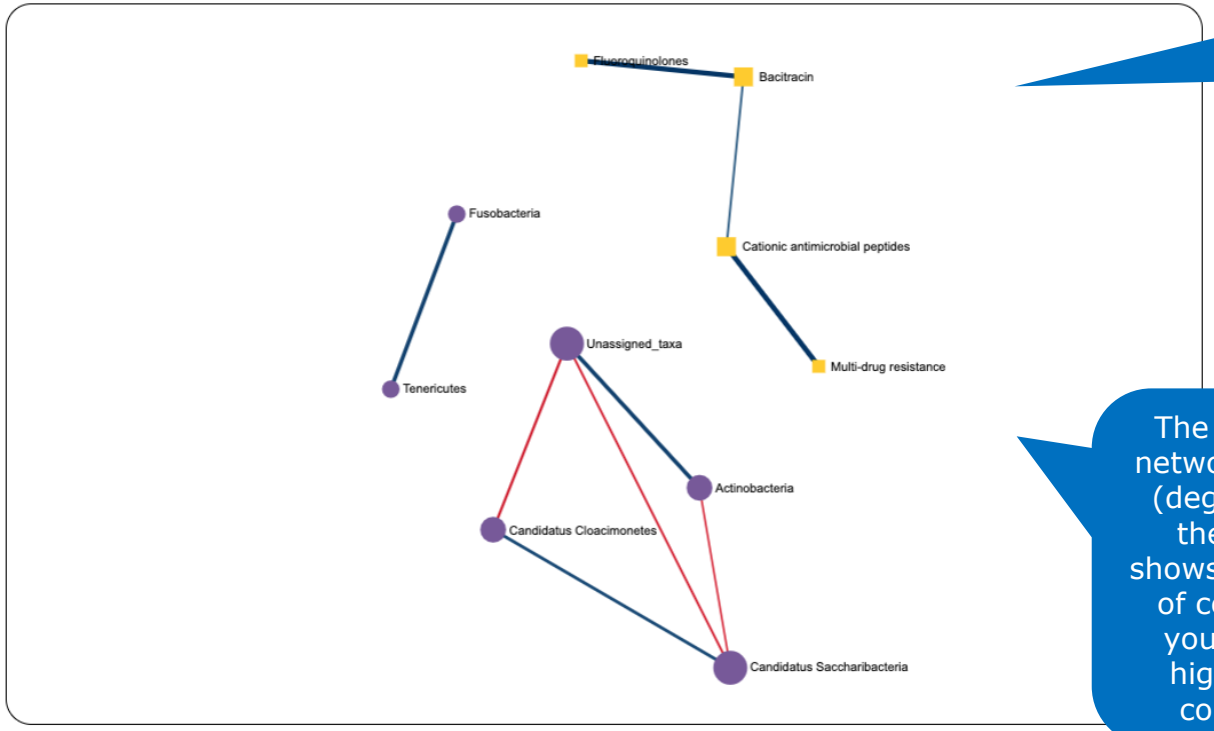
Corr. coeffecient cutoff: 0.6 | No. of max. iterations: 10 | No. of Bootstraps: 10

Submit | Downloads

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



Results can be visualized in the 'co-occurrence network'. Resistome data is shown in yellow squares, while microbiome data are purple dots. Positive correlations are marked as blue lines, while negative correlations are in red.

The size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them. If you double-click a node, it will highlight all the associated or correlated neighbour nodes.

Correlation inference for compositional data through Lasso (CCLasso)

General options:

Taxonomic level: Phylum

Profile level: Class

p-value cutoff: 0.05

Corr. coeffecient cutoff: 0.6

No. of max. iterations: 10

No. of Bootstraps: 10

Submit

Downloads

Let's change the coefficient cutoff to 0.3

After you have made the selection, click on 'Submit'.

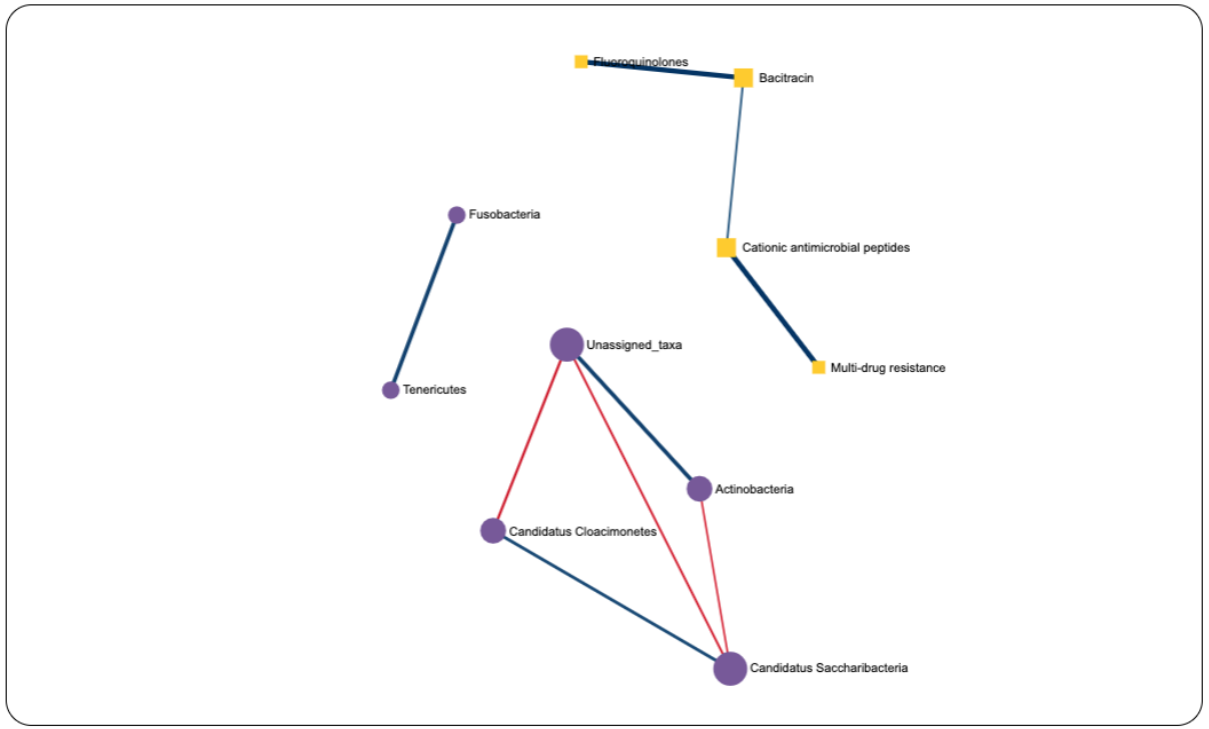
The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout

Node size based on: Degree

Submit

Refresh



Correlation inference for compositional data through Lasso (CCLasso)

General options:

Taxonomic level: Phylum | Profile level: Class | p-value cutoff: 0.05

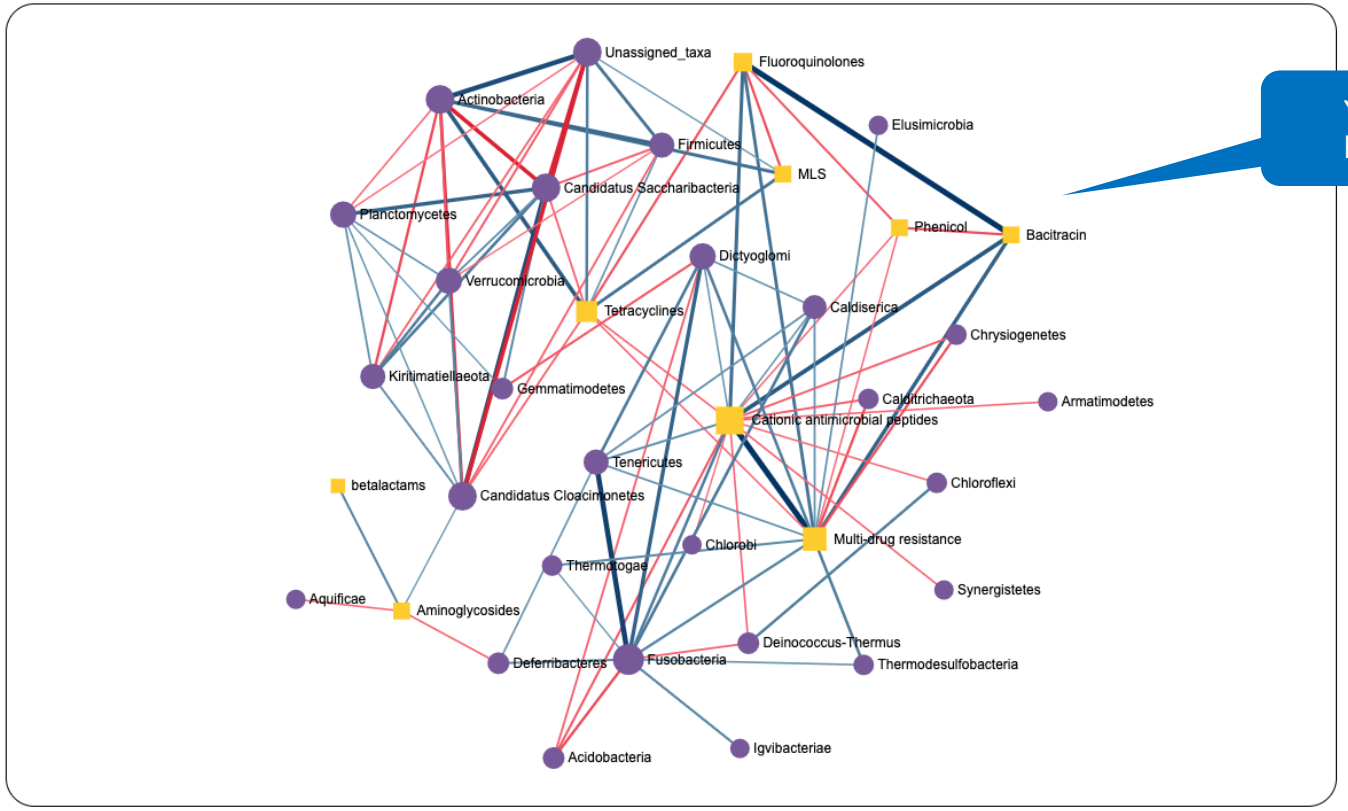
Corr. coeffecient cutoff: 0.3 | No. of max. iterations: 10 | No. of Bootstraps: 10

Submit | Downloads

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive).
Double click on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



You can now observe several correlations between microbiome and resistome data.

Correlation inference for compositional data through Lasso (CCLasso)

General options:

Taxonomic level: Phylum | Profile level: Class | p-value cutoff: 0.05

Corr. coeffecient cutoff: 0.3 | No. of max. iterations: 10 | No. of Bootstraps: 10

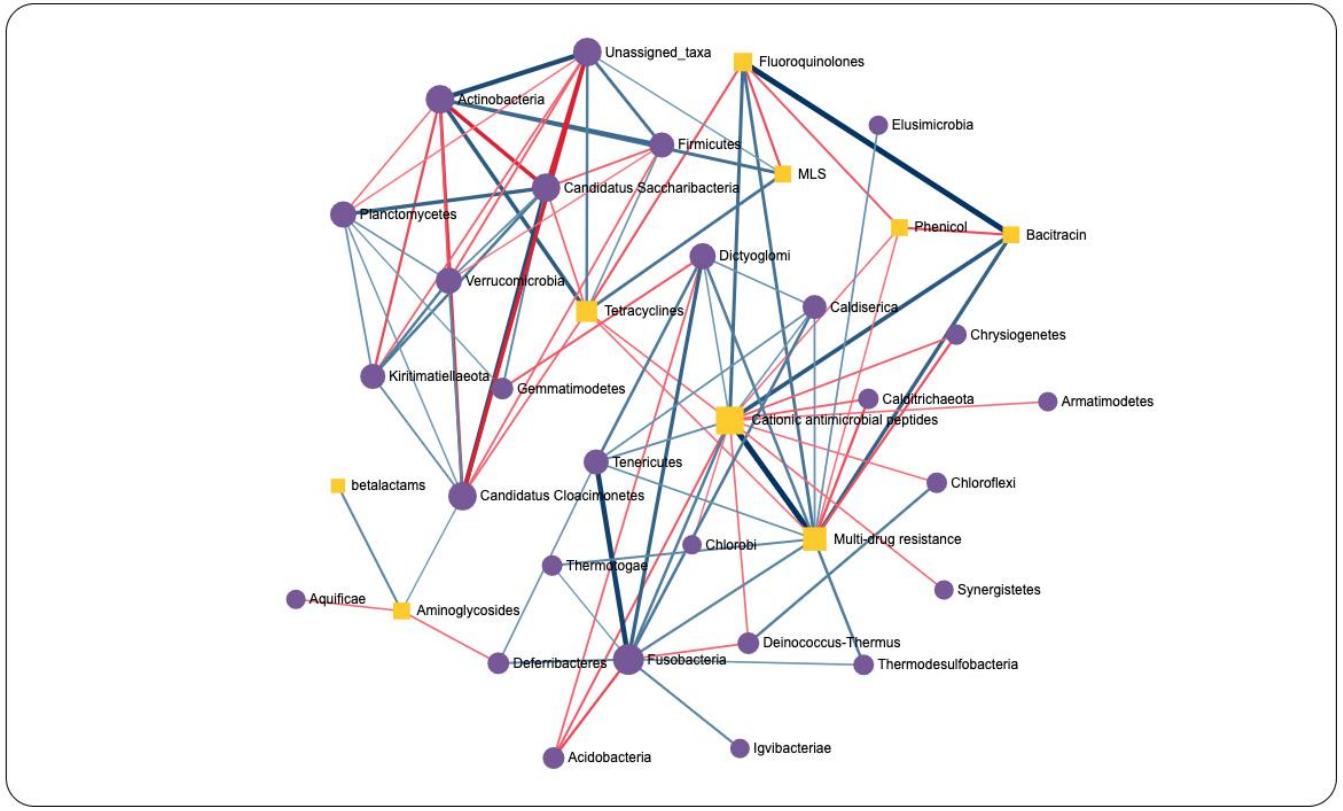
Submit | Downloads

You can choose to download the analyses and/or graphs in a number of different formats by clicking here.

Co-occurance Network

The color and shape of nodes are based on the type of data or profile (**Resistome**: yellow square; **Microbiome**: purple circle). In the network, the size of node represents the network centrality-based measure (degree or betweenness). While the color and width of edges shows the magnitude and strength of correlation between them (**Red**: negative; **Blue**: positive). **Double click** on the respective node will highlight all the associated or correlated neighbour nodes.

Network Layout: Default Layout | Node size based on: Degree | Submit | Refresh



THIS MANUAL IS FINISHED.

To explore more features of ResistoXplorer based on different input data, please check our manuals for **ARG list** and **ARG table**.

Thank you for using

ResistoXplorer

Please cite:

Dhariwal A, Junges R, Chen T, Petersen FC.
ResistoXplorer: a web-based tool for visualization and exploratory analysis of resistome data.