Welcome to

ResistoXplorer

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

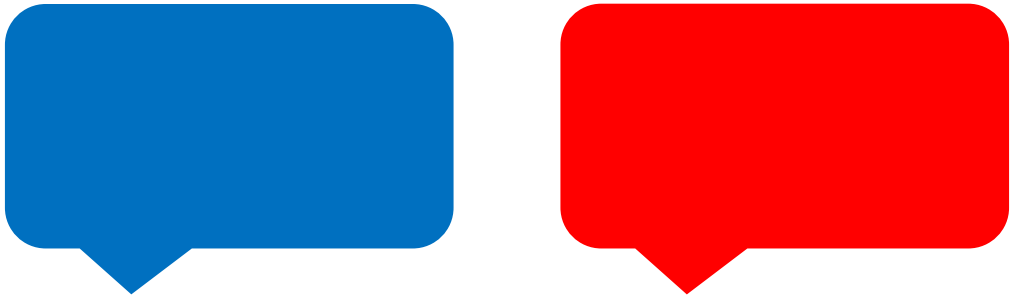
The key features include:

- 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 an **ARG table** 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.

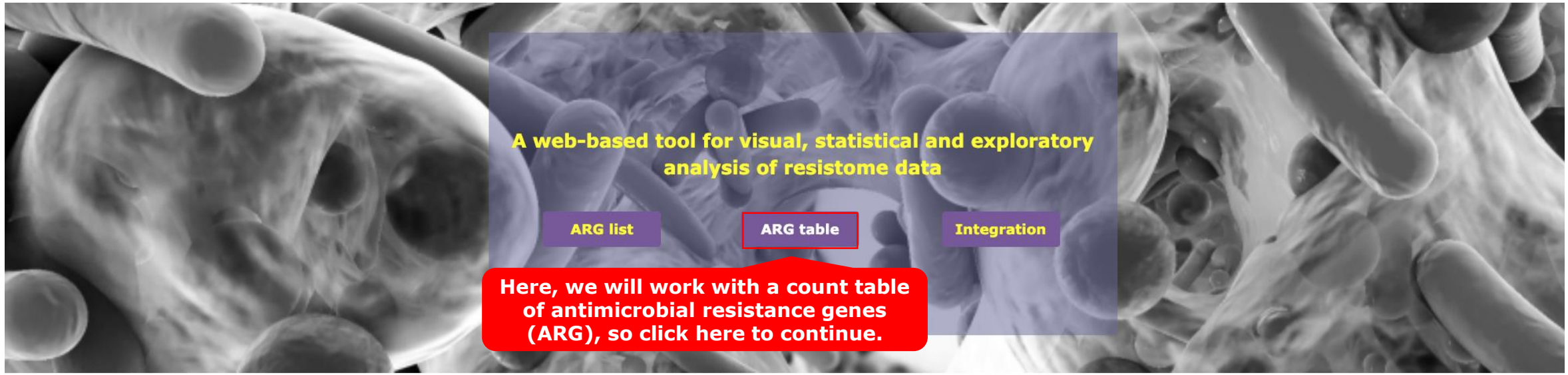


Blue dialogs indicate explanations and details for different functions in each page, while **red dialogs** indicate actions that will move forward with the analysis to a new screen or a download option for a visualization/analysis.

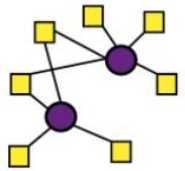
The question mark icons are available in ResistoXplorer.
If you hove over it, a short explanation about that item will appear.

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

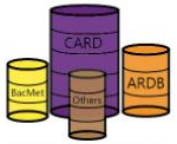


Features



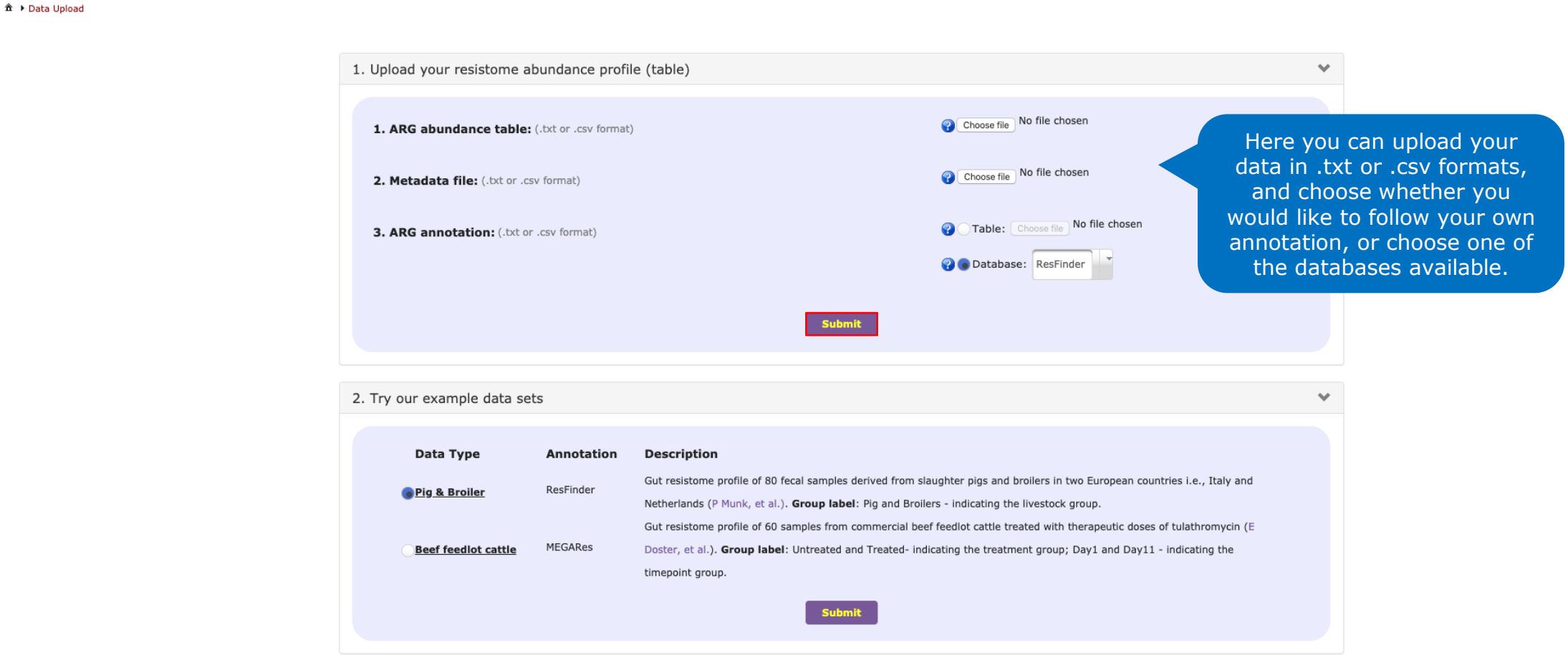
ARG-MICROBE NETWORK EXPLORATION

INTERACTIVE EXPLORATION



COMPREHENSIVE DATABASE

You will be then redirected to the upload screen where you can add your data.



You will be then redirected to the upload screen where you can add your data.

Data format

Each database can present a different data format, so please make sure that your gene table is in accordance with the database you will use. Otherwise you might not get any hits. In addition, the functional annotation level at which the resistome profile can be analyzed will be dependent on the information available in the respective database. For more information, please refer to the 'Data format' section in the homepage of ResistoXplorer.

Here below and in the next slide, we show how some of the formatting can look like for features or rownames:

ResFinder

aac(2')-Ia
aac(2')-Ib
aac(2')-Ic
aac(2')-Id
aac(2')-Ie
aac(2')-IIa
aac(3)-I
aac(3)-Ia

CARD

vanC
vanRA
vanSA
vanHA
vanA
vanXA
vanYA
vanZA

ARDB

aac2i
aac2i
aac2i
aac2ia
aac2ib
aac2ic
aac2ic

BacMet

acrA
acrB
acrD/yffA
acrE/envC
acrF/envD
actA
actR

AMP

acrB
almE
almF
almG
amiA
amiA
amiC

MEGARes 2 - drugs

MEG_372|Multi-compound|Drug_and_biocide_resistance|Drug_and_biocide_MATE_efflux_pumps|ABEM
MEG_373|Multi-compound|Drug_and_biocide_resistance|Drug_and_biocide_MATE_efflux_pumps|ABEM
MEG_374|Multi-compound|Drug_and_biocide_resistance|Drug_and_biocide_SMR_efflux_pumps|ABES
MEG_375|Multi-compound|Drug_and_biocide_resistance|Drug_and_biocide_SMR_efflux_pumps|ABES
MEG_376|Multi-compound|Drug_and_biocide_resistance|Drug_and_biocide_SMR_efflux_pumps|ABES
MEG_377|Multi-compound|Drug_and_biocide_resistance|Drug_and_biocide_SMR_efflux_pumps|ABES
MEG_378|Multi-compound|Drug_and_biocide_resistance|Drug_and_biocide_SMR_efflux_pumps|ABES
MEG_379|Multi-compound|Drug_and_biocide_resistance|Drug_and_biocide_SMR_efflux_pumps|ABES

SARG

0910185A
AAA19915
AAA26700
AAA88337
AAA92254
AAB00446
AAB20441
AAB20442
AAB34257

Resfams

AAC3
AAC3-I
AAC6-I
AAC6-Ib
AAC6-II
ABCAntibioticEffluxPump
adeA-adeI
adeB
adeC-adeK-oprM

ARG-ANNOT

aac
aac2-Ia
aac2-Ib
aac2-Ic
aac(2'''')-Id
aac2-Ie
aac3-I
aac-IIIa

ARGminer

BAE78082.1
WP_024565805.1
ALX99516.1
YP_186749.1
CAA64891.1
BAC11911.1
WP_000725529.1
AAC75138.1
AIA08936.1

deepARG

pmrA
novA
2
A1OC_02073
A464_1655
A670_03335
A989_16023
AA309_10385

MEGARes 1

Bla|OXA-223|JN248564|1-825|825|betalactams|Class_D_betalactamases|OXA
gi|698174209|gb|KM087859.1|betalactams|Class_C_betalactamases|MIR
1172|AF317511.1|AF317511|betalactams|Class_B_betalactamases|VIM
959|M97297.1|TRNVAN|Glycopeptides|VanA-type_accessory_protein|VANZA
Gly|VanY-A|M97297|9052-9963|912|Glycopeptides|VanA-type_accessory_protein|VANYA
Mdr|AY769962.1|gene1|Multi-drug_resistance|Multi-drug_efflux_pumps|ADEAI
617|HQ875016.1|HQ875016|Phenicol|Phenicol_efflux_pumps|CML
Bla|SHV-65|DQ174305|5-865|861|betalactams|Class_A_betalactamases|SHV
Bla|OXA-183|HQ111474|1057-1857|801|betalactams|Class_D_betalactamases|OXA

MEGARes 2

MEG_1|Drugs|Aminoglycosides|Aminoglycoside-resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation
MEG_2|Drugs|Aminoglycosides|Aminoglycoside-resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation
MEG_3|Drugs|Aminoglycosides|Aminoglycoside-resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation
MEG_4|Drugs|Aminoglycosides|Aminoglycoside-resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation
MEG_5|Drugs|Aminoglycosides|Aminoglycoside-resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation
MEG_6|Drugs|Aminoglycosides|Aminoglycoside-resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation
MEG_7|Drugs|Aminoglycosides|Aminoglycoside-resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation
MEG_8|Drugs|Aminoglycosides|Aminoglycoside-resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation
MEG_9|Drugs|Aminoglycosides|Aminoglycoside-resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation
MEG_10|Drugs|Aminoglycosides|Aminoglycoside-resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation
MEG_11|Drugs|Aminoglycosides|Aminoglycoside-resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation
MEG_12|Drugs|Aminoglycosides|Aminoglycoside-resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation

AMRFinder

1567214_ble
aac(2')-Ia
aac(2')-Ib
aac(2')-Ic
aac(2')-Id
aac(2')-Ie
aac(2')-IIa
aac(2')-IIb

Data Upload ▸ Data Inspection

Data Integrity check

Please review the **Text Summary** below from your uploaded 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 will be excluded from further analysis

Text Summary | Library Size Overview

Total no of features (ARGs) in abundance file:	1432
Features present in ≥ 2 samples:	1038
Annotation format:	Database
Database:	ResFinder
No. of experimental factors:	2 (Discrete: 2; Continous: 0)
No. of functional annotation levels:	3
Sparsity (%):	79
Compositional:	No
Total read counts:	2042387
Average counts per sample:	25529
Maximum counts per sample:	56829
Minimum counts per sample:	6518
No. of samples in abundance table:	80
No. of samples in metadata:	80
No. of sample names matched (abundance vs metadata table):	80
No. of feature names matched (abundance vs annotation database):	634
No. of samples that will be processed:	80
No. of features that will be processed:	634

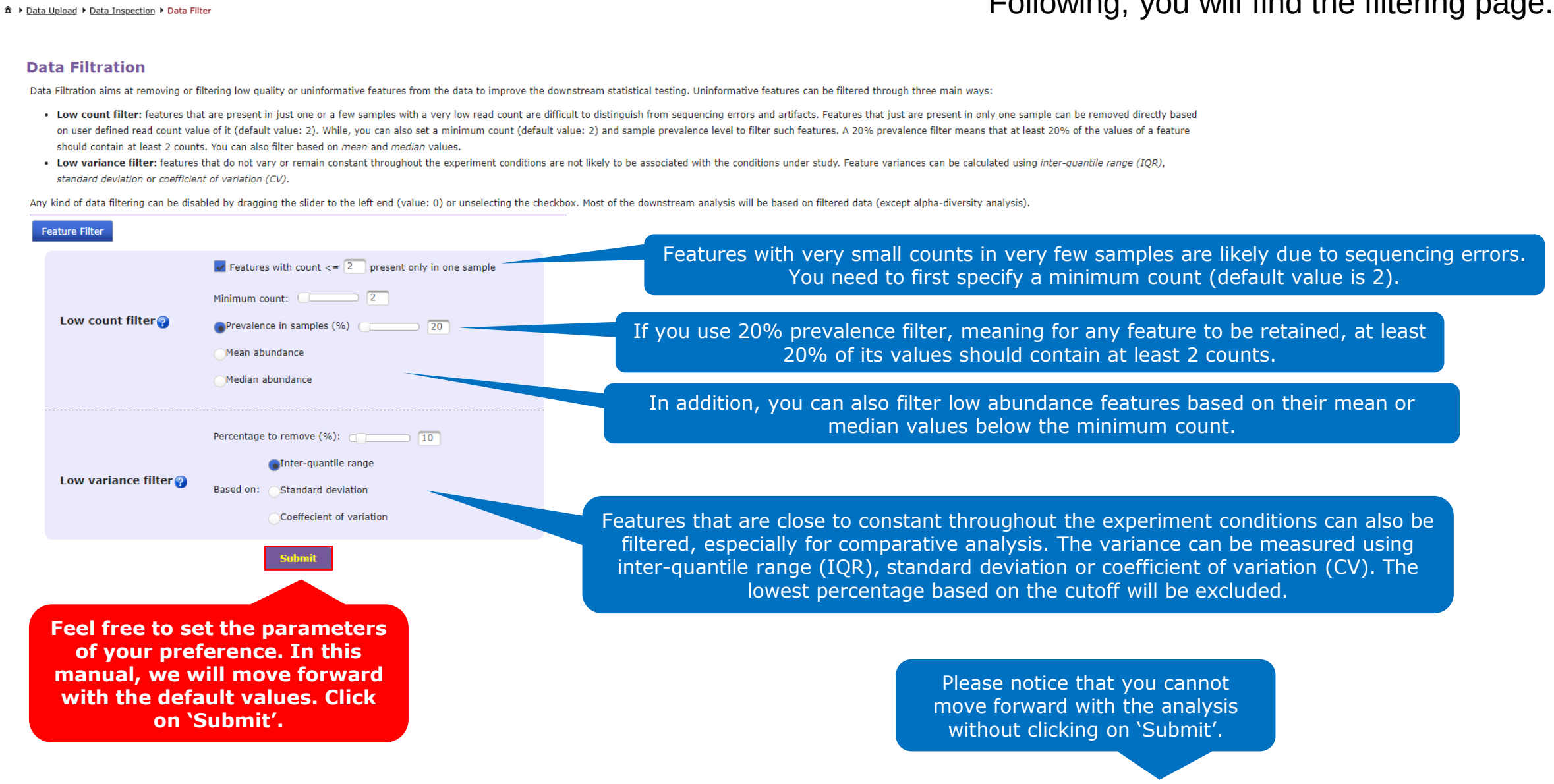
In this manual, we will use the first example dataset called 'Pig & Broiler', and as you select the option, a text and graphic summaries of the data will be available. This step named 'Data Integrity Check' will also take place when you upload your own data.

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

Previous

Proceed

Following, you will find the filtering page.



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).

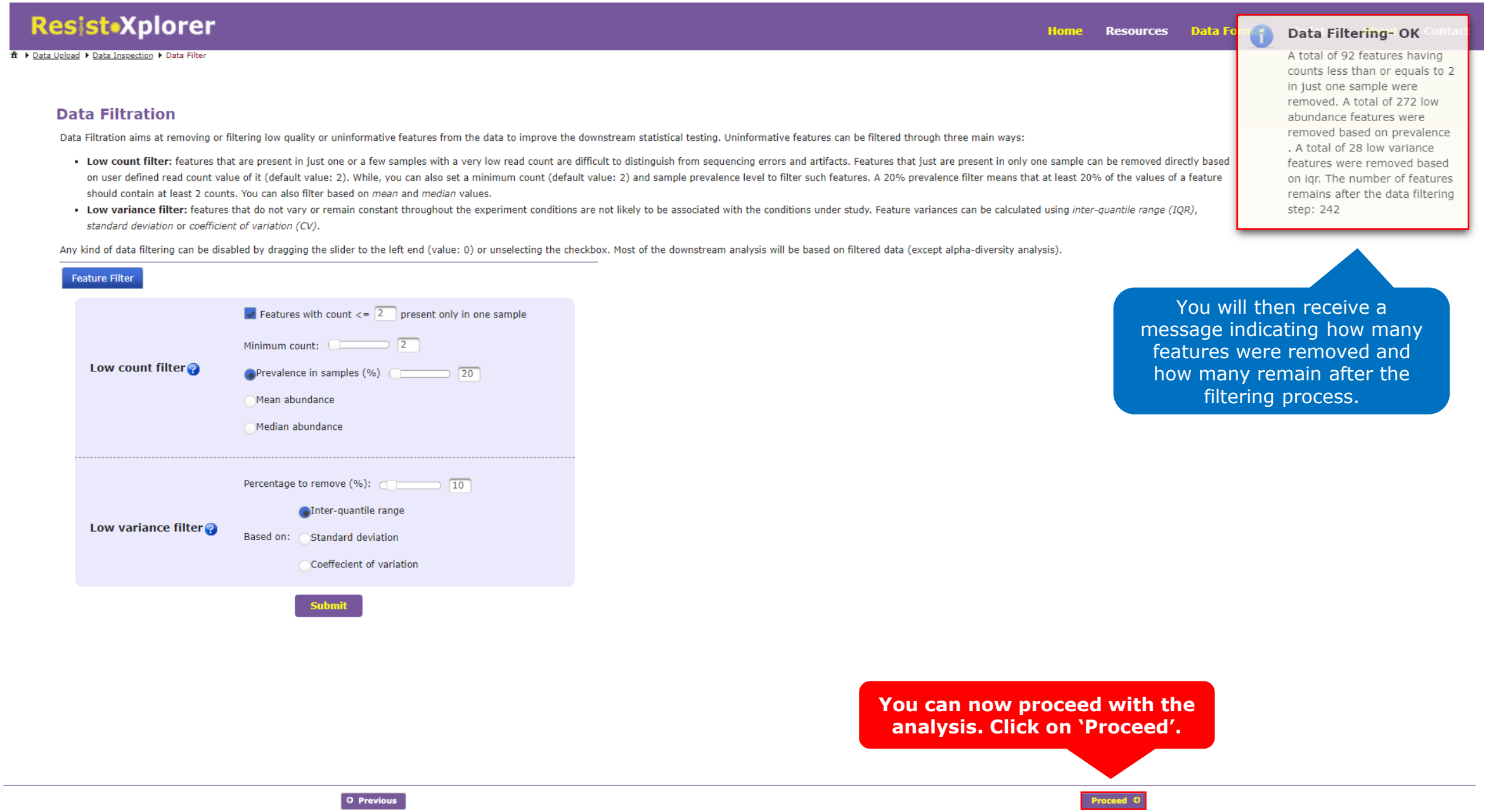
If you use 20% prevalence filter, meaning for any feature to be retained, at least 20% of its values should contain at least 2 counts.

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
Resources
Data Upload
Data Inspection
Data Filter
Data Filtering- OK
A total of 92 features having counts less than or equals to 2 in just one sample were removed. A total of 272 low abundance features were removed based on prevalence . A total of 28 low variance features were removed based on iqr. The number of features remains after the data filtering step: 242
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 three main ways:
Low count filter: features that are present in just one or a few samples with a very low read count are difficult to distinguish from sequencing errors and artifacts. Features that just are present in only one sample can be removed directly based on user defined read count value of it (default value: 2). While, you can also set a minimum count (default value: 2) and sample prevalence level to filter such features. 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).
Any kind of data filtering can be disabled by dragging the slider to the left end (value: 0) or unselecting the checkbox. Most of the downstream analysis will be based on filtered data (except alpha-diversity analysis).
Feature Filter
Low count filter
Features with count <= 2 present only in one sample
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
You will then receive a message indicating how many features were removed and how many remain after the filtering process.
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.

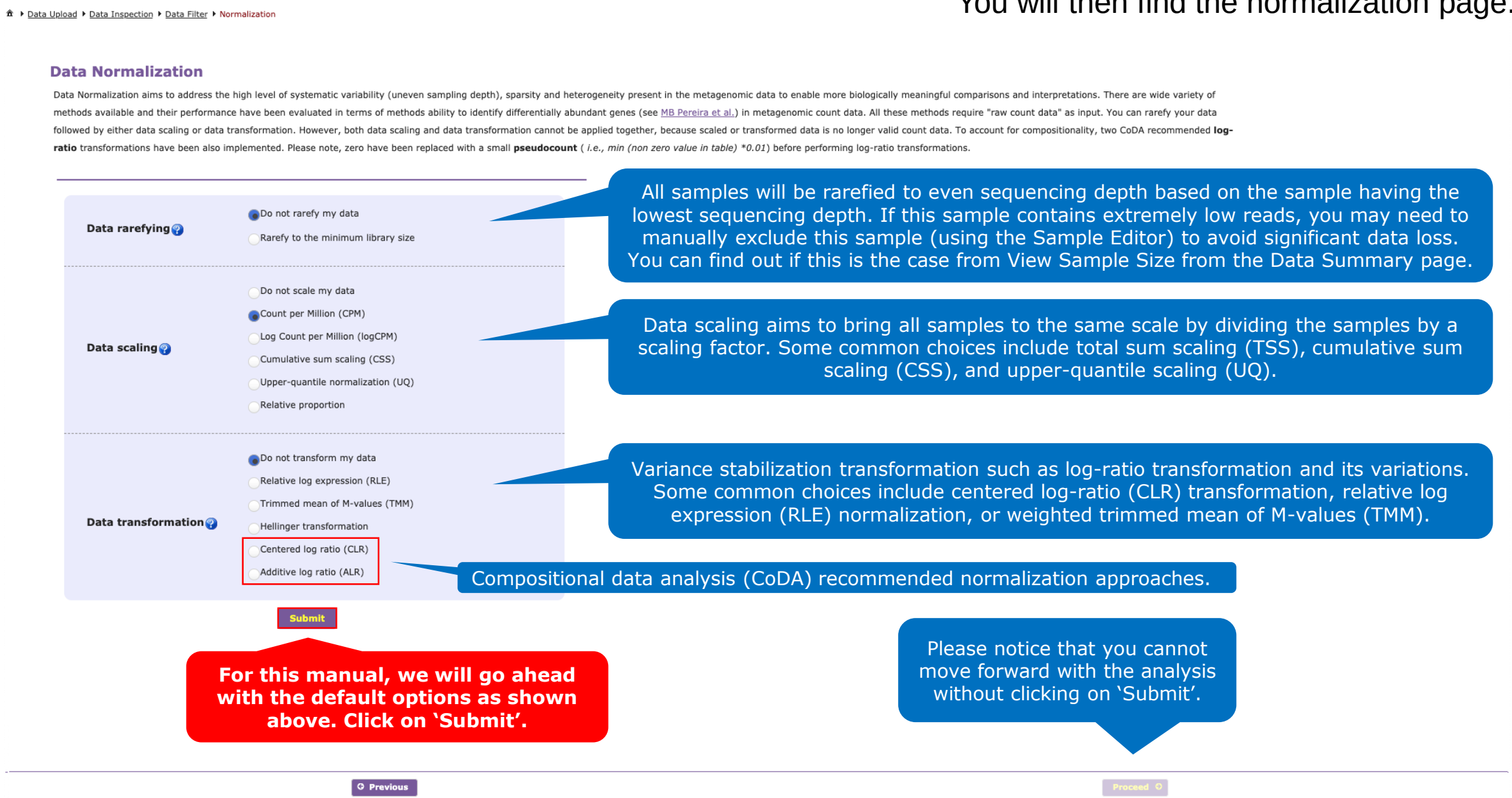
Which category should I choose to perform data filtering?

ResistoXplorer automatically removes features that comprise of all zeros or that are only present in one sample. This type of filtering is used for alpha diversity analysis.

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.

The best approach for filtering the data depends on the type of analysis. For instance, if the primary objective is to perform comparative analysis, then you should remove features that exhibit low variance based on their inter-quantile ranges, standard deviations or the coefficient of variations. These features are unlikely to be significant in comparative analysis. In case of integrative data analysis, you can also choose to apply different data filtration criteria for both microbiome and resistome count data.

You will then find the normalization page.



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.

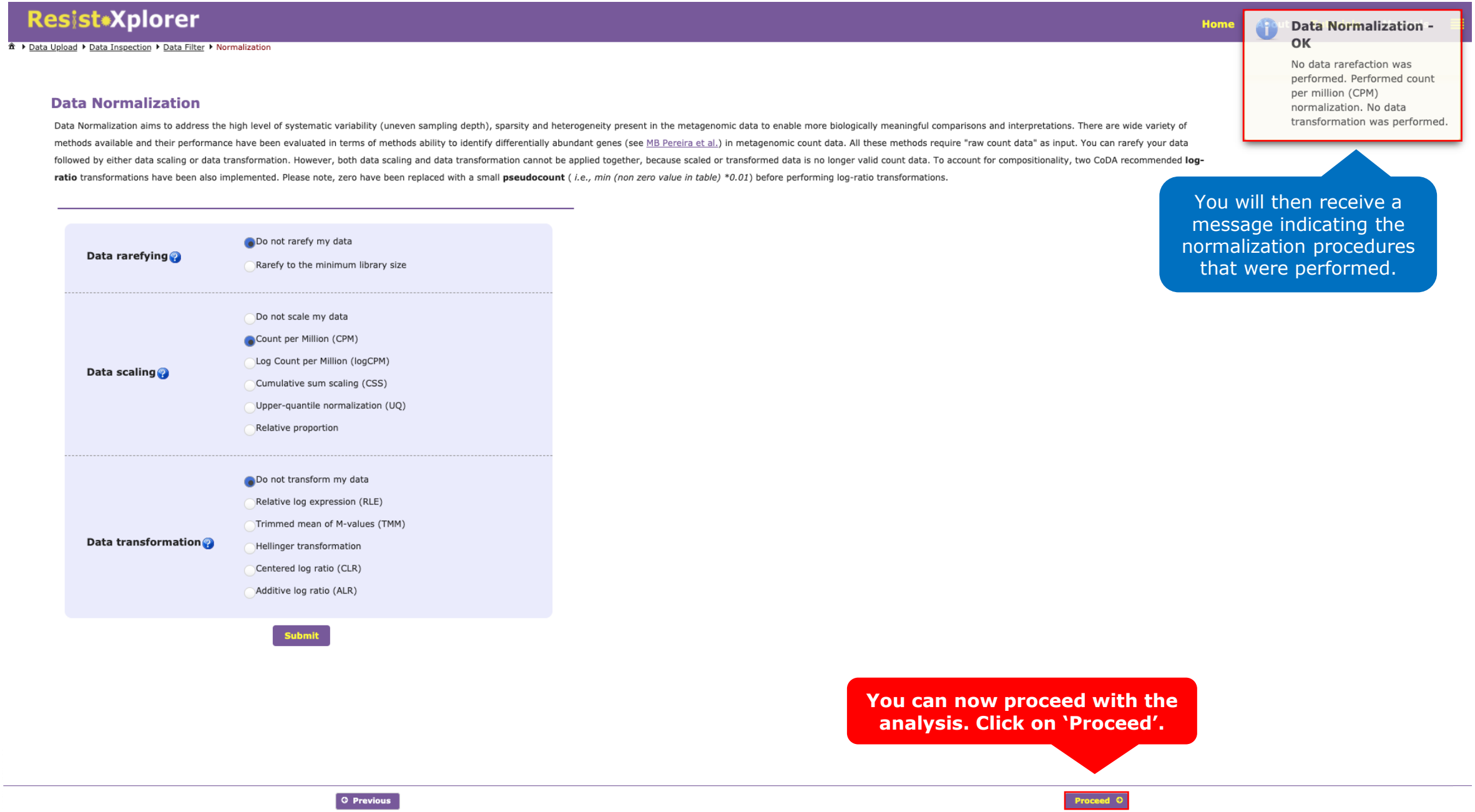
Data scaling aims to bring all samples to the same scale by dividing the samples by a scaling factor. Some common choices include total sum scaling (TSS), cumulative sum scaling (CSS), and upper-quantile scaling (UQ).

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'.


ResistoXplorer
Home
About
Data Upload
Data Inspection
Data Filter
Normalization
Data Normalization - OK
No data rarefaction was performed. Performed count per million (CPM) normalization. No data transformation was 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.
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 will then receive a message indicating the normalization procedures that were performed.
You can now proceed with the analysis. Click on 'Proceed'.
Previous
Proceed

Why should I normalize my data?

Metagenomic data possess some unique characteristics such as vast differences in sequencing depth, sparsity, skewed distributions, over-dispersion and compositionality. Such unique characteristics have made it unsuitable to directly use approaches designed in other omics fields to perform comparative analysis on metagenomic data. To deal with these issues, ResistoXplorer supports various normalization approaches such as:

Rarefaction: this method deals with uneven sequencing depths by randomly removing reads in the different samples until the sequencing depth is equal in all samples.

Scaling-based: these methods account for uneven sequencing depths by deriving a sample-specific scaling factor for bringing samples to the same scale for comparison.

Transformation-based: it includes approaches to deal with sparsity, compositionality, and large variations within the count data.

Applying suitable normalization methods can significantly improve the statistical power and reduce the false positive rate while identifying deferentially abundant resistance gene.

What are the various normalization methods and which one to choose?

ResistoXplorer provides a variety of widely used methods for normalizing the metagenomic count data. A brief description is provided below:

Count per million (CPM) normalization: also known as Total Sum Scaling (TSS). This method removes systematic variability related to uneven sequencing depth in different samples through simply dividing each feature count with the total read counts (library size) to provide relative proportion of counts for that feature. For convenience, we can multiply it by 1,000,000 (scaling factor) to get the number of reads corresponding to that feature per million reads. LefSe algorithm utilizes this kind of strategy.

Log Count per million (CPM) normalization: this method perform log transformation on count per million normalized data in order to deal with large variance in count distributions in addition to library size differences. This kind of approach is been used by R packages such as edgeR and voom which are designed for identifying deferentially abundant genes in RNA-Seq count data.

Relative proportion: this approach computes the relative proportion of a feature by dividing each feature count by the total number of counts (library size) per sample.

Cumulative Sum Scaling (CSS) normalization: this method corrects for differences in library size by calculating the scaling factors as the cumulative sum of gene abundances up to a data-derived threshold to remove the variability in data caused by highly abundant genes. By default, metagenomeSeq utilizes this approach for differential analysis.

Upper-quantile normalization: this approach calculates the scaling factors from the 75th percentile of the gene count distribution for each library, after removing genes which are zero in all libraries. This method is derived from edgeR package proposed by Bullard et al. (2010).

What are the various normalization methods and which one to choose?

Relative log expression (RLE) normalization: this method estimates the median library from the geometric mean of the gene-specific abundances over all samples. The median ratio of each sample to the median library is used as the scaling factor. By default, DESeq2 utilizes this approach for differential abundance testing. This method was initially proposed by Anders and Huber (2010).

Trimmed mean of M-values (TMM) normalization: this method is proposed by Robinson and Oshlack (2010), where the scaling factor is derived using a weighted trimmed mean over the differences of the log-transformed gene-count fold-change (relative abundance) between the samples. By default, edgeR utilizes this approach for differential analysis in ResistoXplorer.

Log-Ratio (CLR and ALR) Transformation: these methods are specifically designed to normalize compositional data. They transforms the relative abundances of each element, or the values in the table of counts for each element, to ratios between all parts by using either geometric mean of the sample or single element as the reference. Further, taking the logarithm of these ratios, brings the data in a Euclidean (real) space, such that standard statistical methods can be applied.

Hellinger Transformation: this method computes the relative proportion of a feature by dividing each feature count by the total number of counts (library size) per sample, and then taking the square root of it.

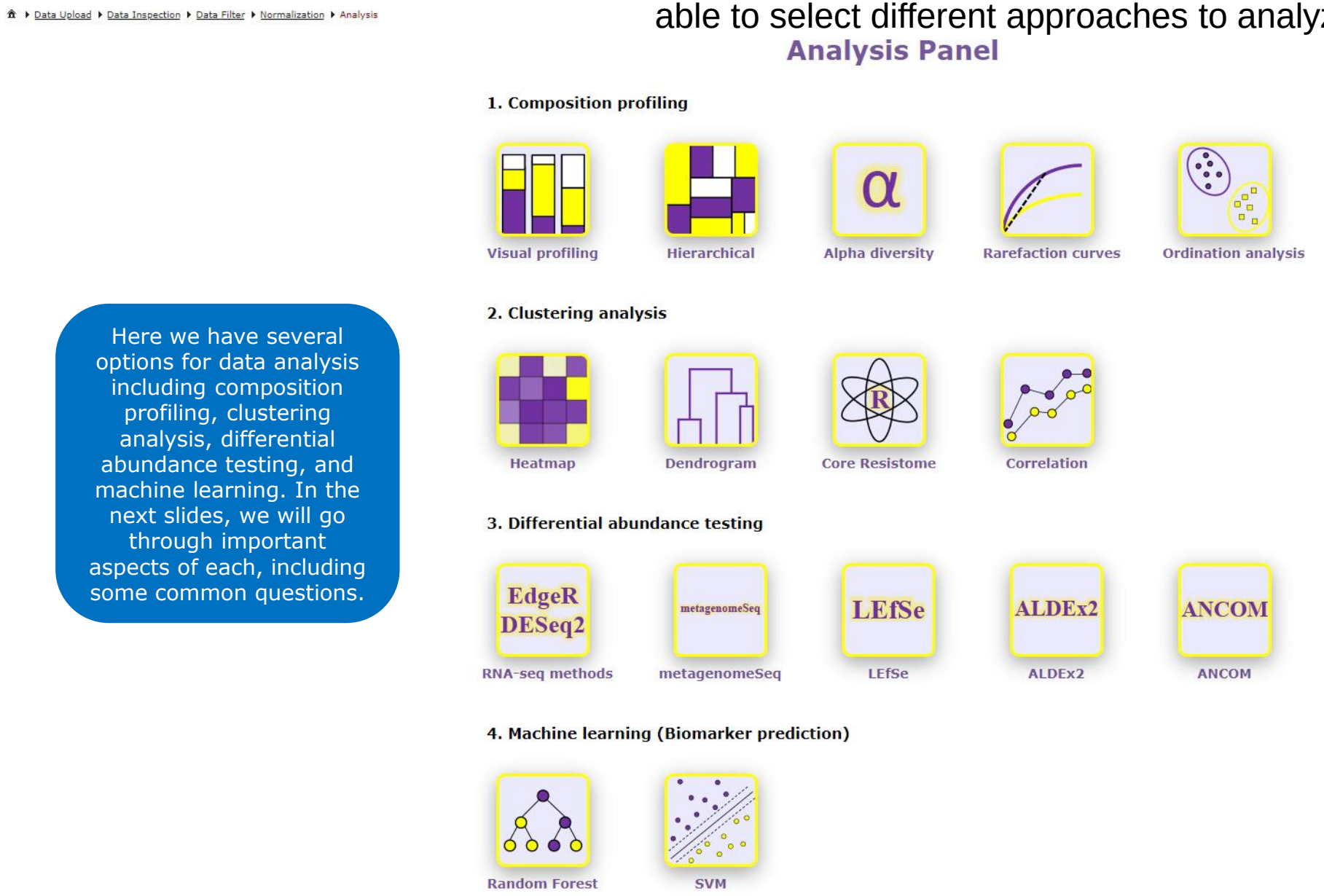
Rarefaction: this method deals with uneven sequencing depths by randomly removing reads in the different samples until the library size of all the samples are same as sample with lowest sequencing depth. Whenever the library size of the samples varies too much (i.e. >10X), it is recommended to perform rarefaction before normalizing your data.

What are the various normalization methods and which one to choose?

Data normalization is mainly intended for visual exploration such as ordination and clustering analysis. Also, differential abundance testing using different approaches are performed on normalized data. However, each of these methods will use their own specific normalization procedure. For example, relative log expression (RLE) normalization is used for DESeq2, and trimmed mean of M-values (TMM) is applied for edgeR.

Currently, there is no consensus with regard to which normalization should be used. We recommend users to explore different approaches and then visually examine the separation patterns (i.e. ordination and clustering analysis) to determine the effects of different normalization procedures with regards to experimental factor of interest. For detailed discussion about these methods, users can referred to these recent papers Paul J. McMurdie et al. and Mariana Buongermino Pereira et al.

Once filtering and normalization are selected, you will find the ‘Analysis panel’ home page, where you will be able to select different approaches to analyze and visualize the data.



Here we have several options for data analysis including composition profiling, clustering analysis, differential abundance testing, and machine learning. In the next slides, we will go through important aspects of each, including some common questions.

What are the different visualization options available for composition profiling?

Currently, for compositional profiling, ResistoXplorer supports stacked bar and area plots, which are found in the visual profiling section. In addition, under hierarchical analysis, you can create graphs such as Sankey, Sunburst, and Treemap.

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis

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 'Visual profiling' of the sample – click here.

Analysis Panel

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

RNA-seq methods

metagenomeSeq

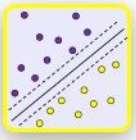
LEfSe

ALDEx2

ANCOM

4. Machine learning (Biomarker prediction)


Random Forest


SVM

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition

Composition Profiling

General options: Profile level: Mechanism

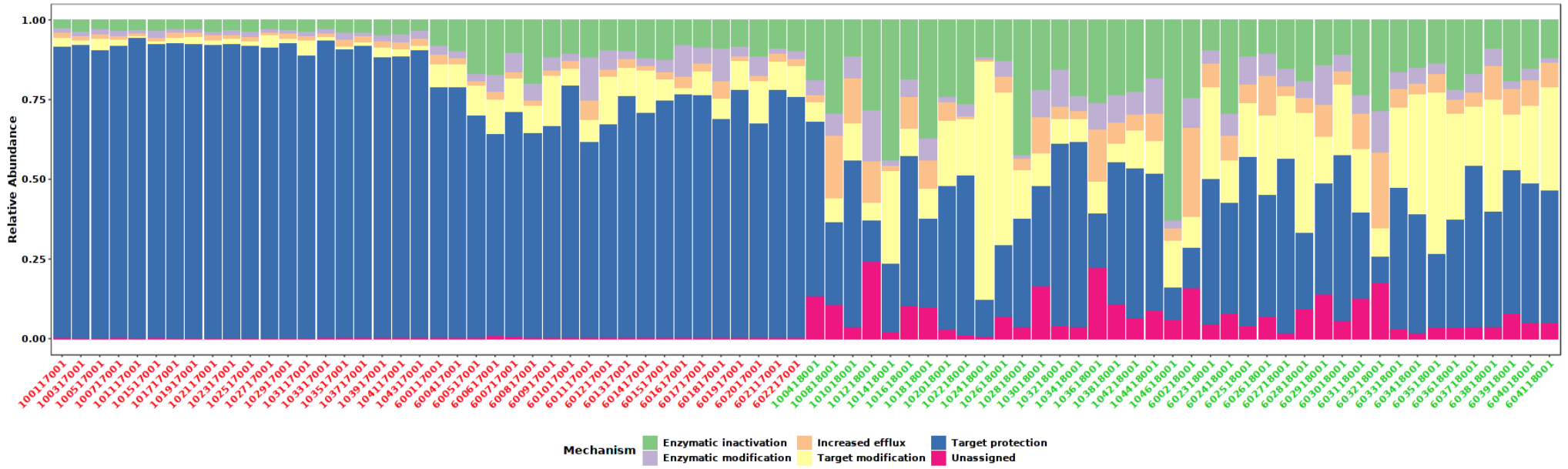
Graph type: Stacked Bar [Percentage Abundance] Color scheme: Palette_custom

View options: View type: All samples / By groups (Species) / Per sample (100117001)

Merge features with counts < 10 based on sum / median

Here you can select how to profile the genes, followed by the graph type and color scheme. Finally, you can also select how to group the samples.

In this example, each column represents a different sample and the 'y axis' is showing the relative abundance of mechanism of action of genes from different antibiotic classes, as shown in the legend below.



Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition

Composition Profiling

Let's try to make some changes, click on 'Profile level' and change to 'Class'.

Then click on 'Graph type' and change to 'Stacked Area Plot'.

General options: Profile level: Mechanism

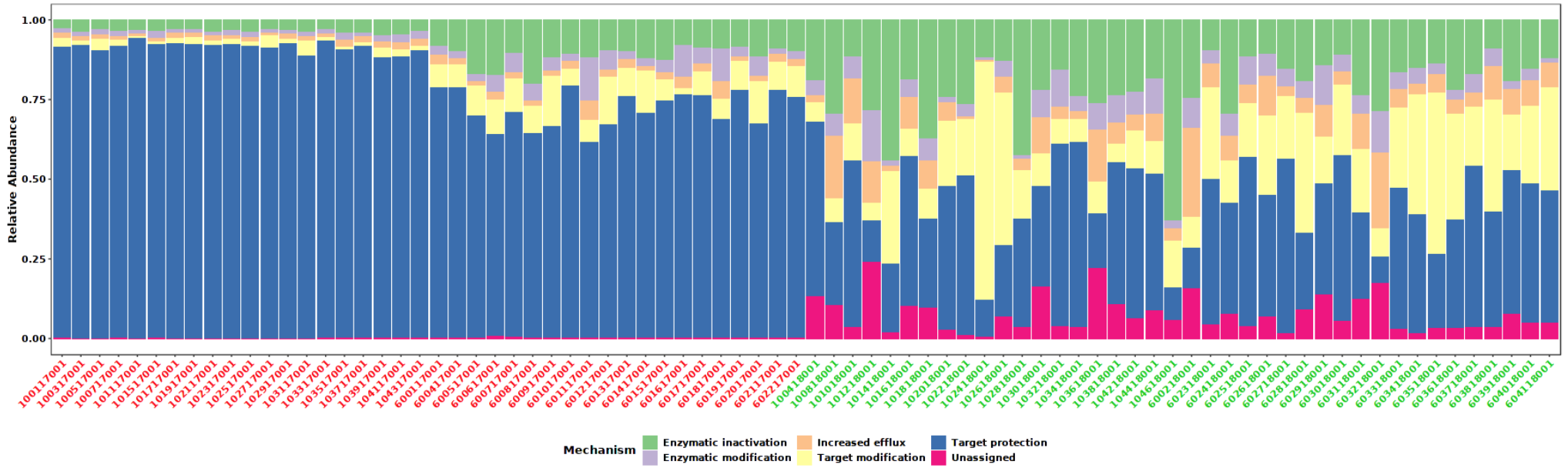
Graph type: Stacked Bar [Percentage Abundance] Color scheme: Palette_custom

View options: View type: All samples / By groups: Species / Per sample: 100117001

Merge features with counts < 10 based on sum / median

Submit

To apply the changes, click on 'Submit'.



Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition

Composition Profiling

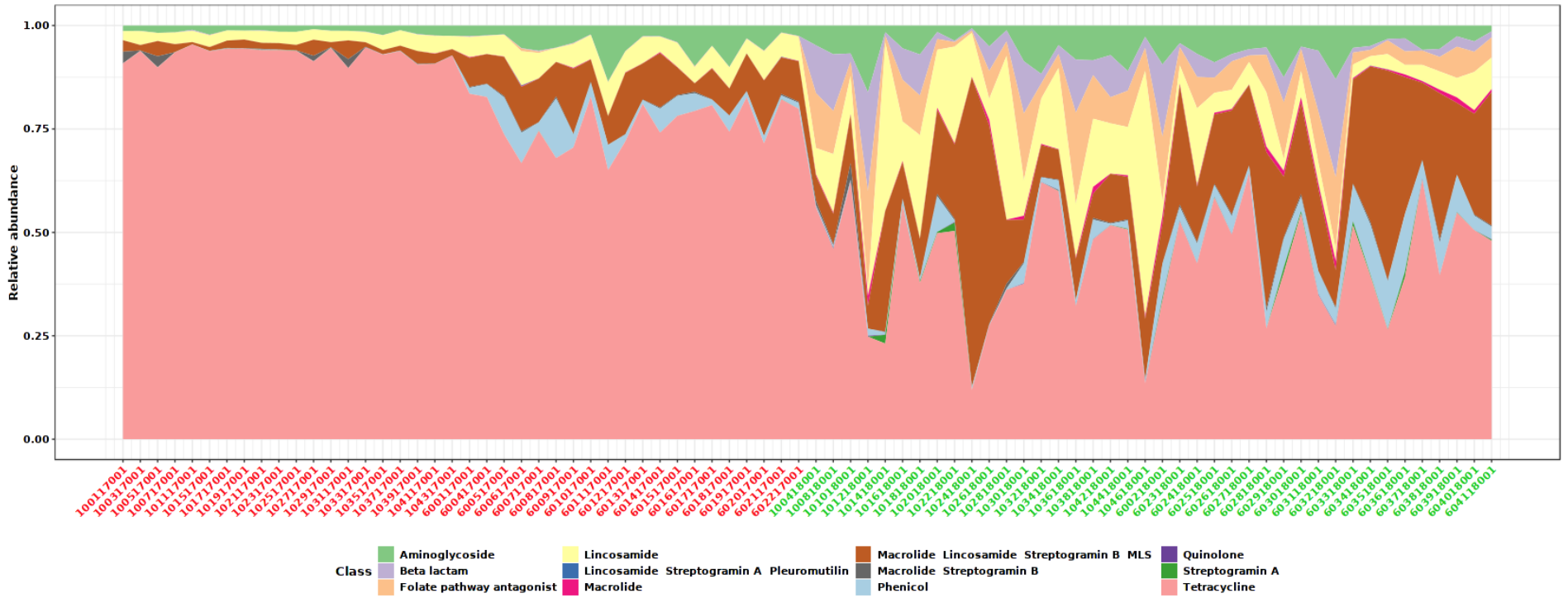
General options: Profile level: Class

Graph type: Stacked Area Plot Color scheme: Palette_custom

View options: View type: All samples / By groups (Species) / Per sample (100717001)

Merge features with counts < 10 based on sum / median

Now you are looking at abundance profile based on class. You are also seing a different visualization option with the area plot instead of the bar plot. Feel free to play around and analyze your data in different ways.



Composition Profiling

General options: Profile level: Class

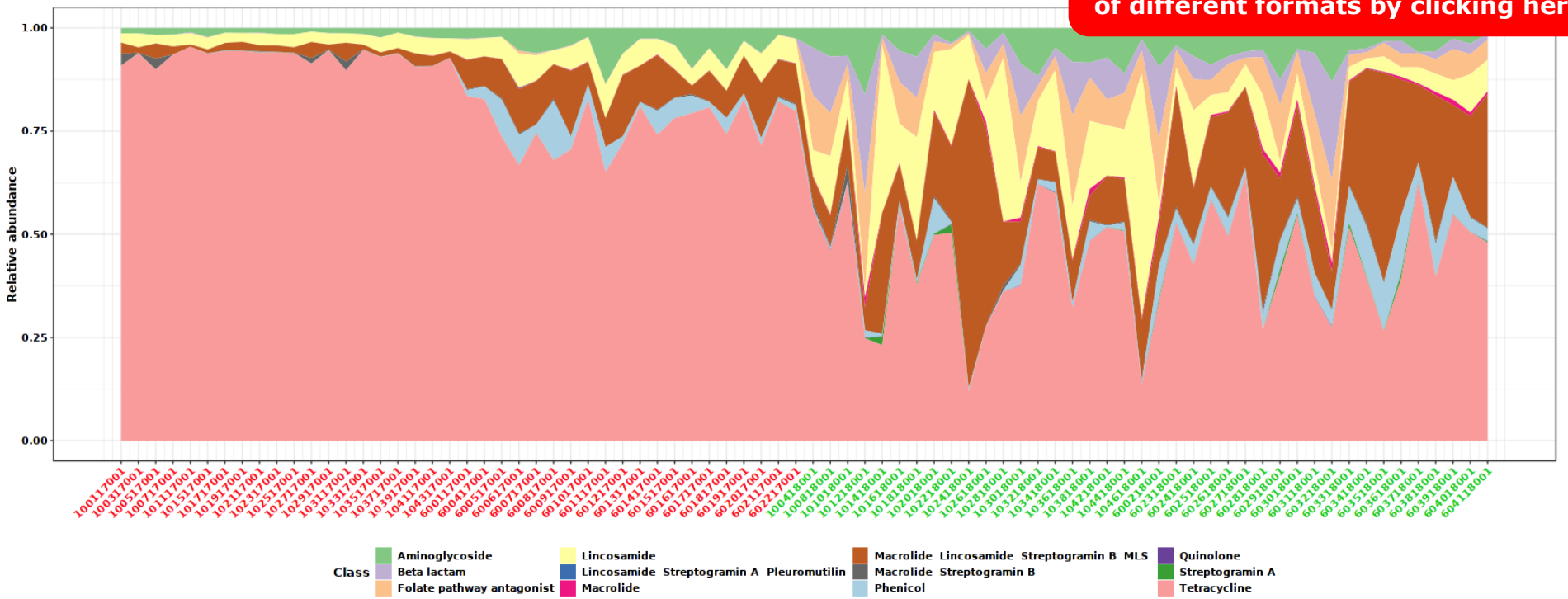
Graph type: Stacked Area Plot Color scheme: Palette_custom

View options: View type: All samples / By groups (Species) / Per sample (100717001)

Merge features with counts < 10 based on sum / median

Submit Downloads

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



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

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition

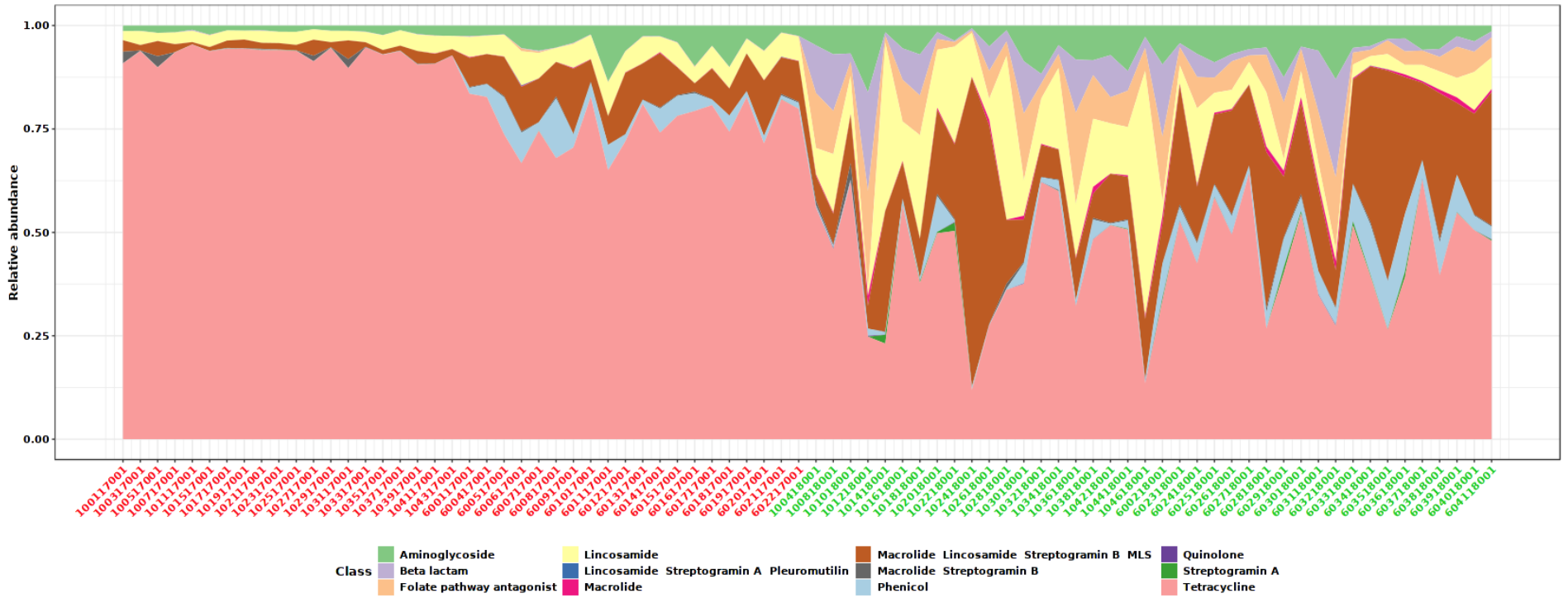
Composition Profiling

General options: Profile level: Class

Graph type: Stacked Area Plot Color scheme: Palette_custom Submit Downloads

View options: View type: All samples / By groups (Species) / Per sample (100717001)

Merge features with counts < 10 based on sum / median

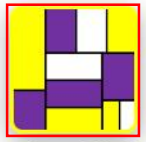


Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis

Let's move to 'Hierarchical analysis' now – click here.

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

RNA-seq methods

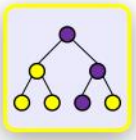
metagenomeSeq

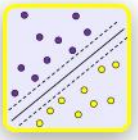
LEfSe

ALDEx2

ANCOM

4. Machine learning (Biomarker prediction)


Random Forest


SVM

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition

Hierarchical Composition Profiling

Here you can select the graph type and how to calculate gene counts. Also, you can choose to visualize samples individually or in a group.

View options:

Chart type: Sankey / Treemap / Sunburst

View type: All samples / By Metadata (Species, Group: Pig) / Per sample (100117001)

Submit

Calculate feature count based on: sum / mean

Show abundance value as: Absolute count / Relative proportion (%)

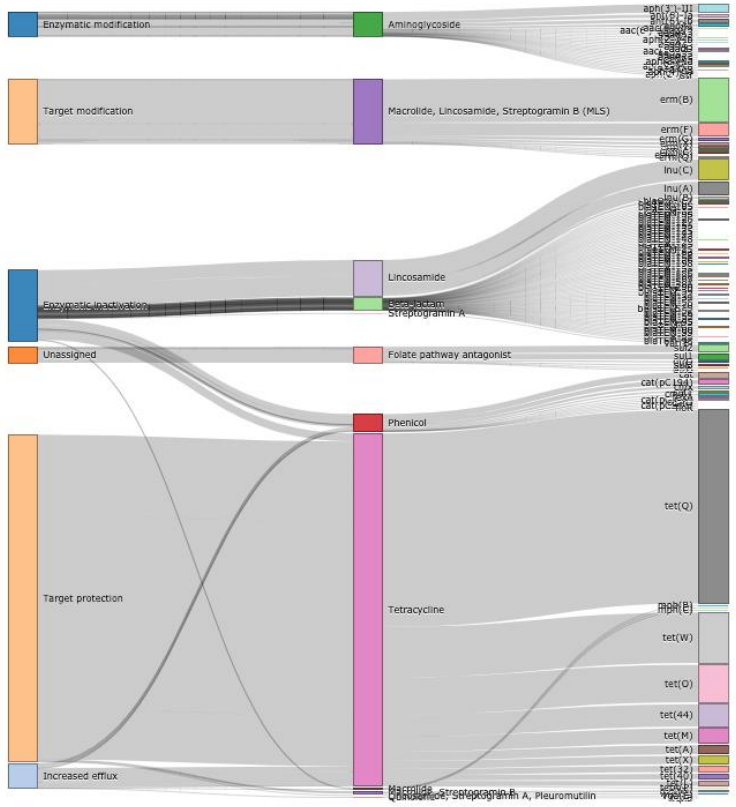
Filter features with counts < 10 based on sum / median

Downloads

Each of the charts have certain interactive functionality associated with it so that user can explore, represent and visualize their data in more better and detailed manner. For instance,

- **Sankey Diagram:** user can drag nodes horizontally and vertically using ***mouse***
- **Treemap:** user can ***zoom in*** to lower hierarchy levels by click on any block and ***zoom back out*** using the top horizontal bar (one level of the hierarchy is displayed at once)
- **Sunburst:** user can click on any arc to ***zoom in*** and click on the center circle to ***zoom out***

Also, kindly use feature filtration option for large datasets (with too many features) in order to enhance the chart layout and representation (avoid overlap)



Here you can see a Sankey graph. To the most right you will see the genes, in the middle the classification based on class, and to the left based on mechanism. You can also click and drag any of these nodes.

Hierarchical Composition Profiling

View options:

Chart type: Sankey / Treemap / Sunburst

View type: All samples / By Metadata (Species, Group: Pig) / Per sample (100117001)

Submit

Calculate feature count based on: sum / mean

Show abundance value as: Absolute count / Relative proportion (%)

Filter features with counts < 10 based on sum / median

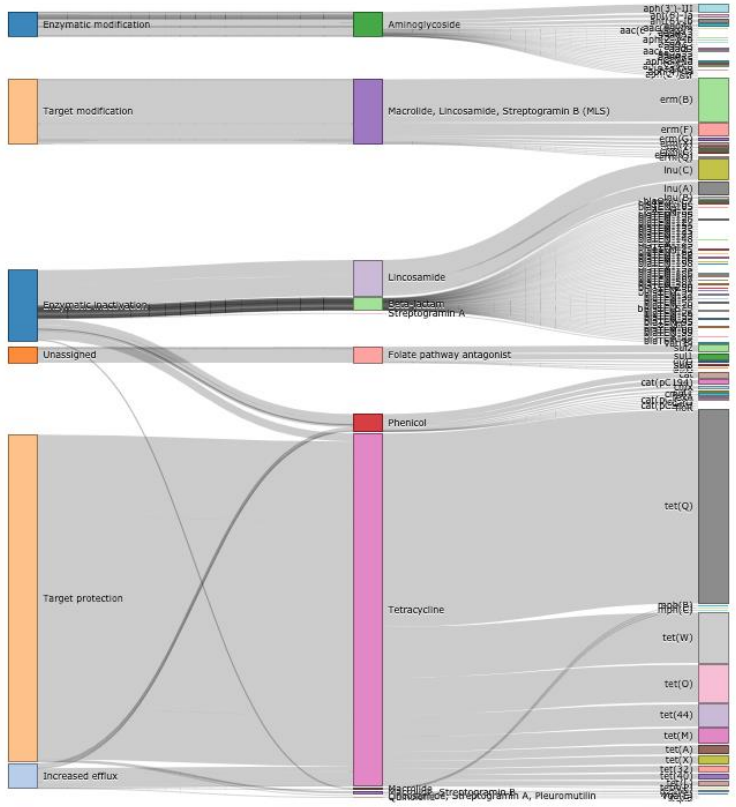
Downloads

Let's try to change to a 'Treemap', so click on its icon on the 'Chart type' option. To apply the changes, click on 'Submit'.

Each of the charts have certain interactive functionality associated with it so that user can explore, represent and visualize their data in more better and detailed manner. For instance,

- **Sankey Diagram:** user can drag nodes horizontally and vertically using ***mouse***
- **Treemap:** user can ***zoom in*** to lower hierarchy levels by click on any block and ***zoom back out*** using the top horizontal bar (one level of the hierarchy is displayed at once)
- **Sunburst:** user can click on any arc to ***zoom in*** and click on the center circle to ***zoom out***

Also, kindly use feature filtration option for large datasets (with too many features) in order to enhance the chart layout and representation (avoid overlap)



Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition

Hierarchical Composition Profiling

View options:

Chart type: Sankey / Treemap / Sunburst

View type: All samples / By Metadata (Species, Group: Pig) / Per sample (100117001)

Submit

Calculate feature count based on: sum / mean

Show abundance value as: Absolute count / Relative proportion (%)

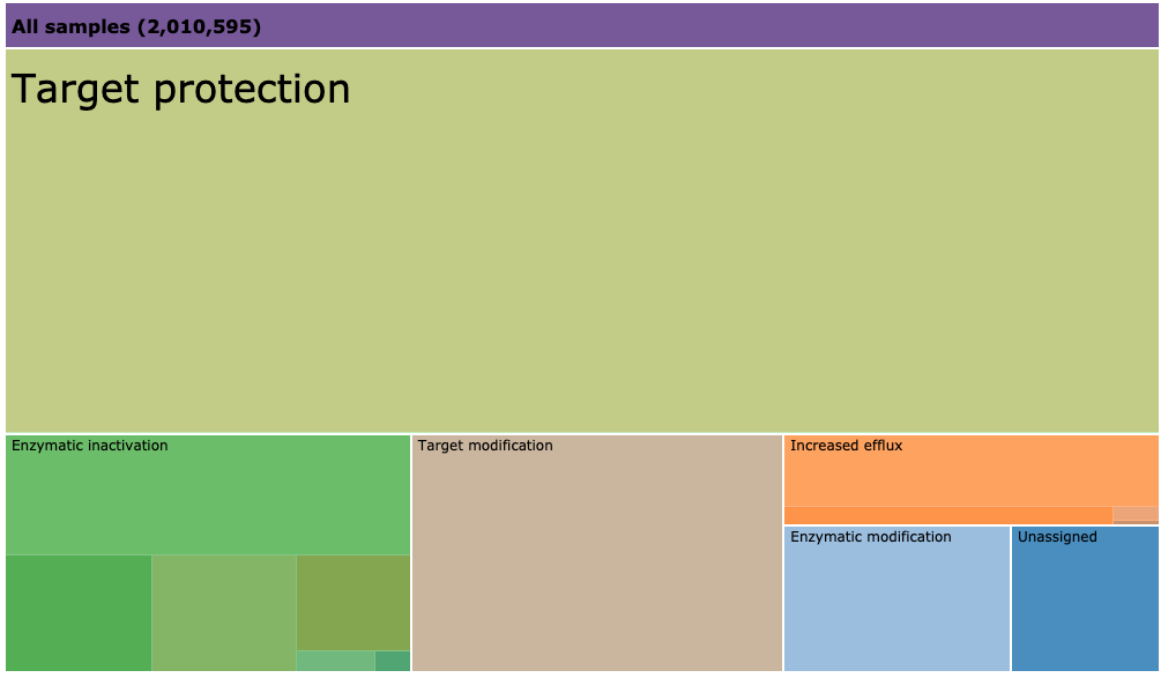
Filter features with counts < 10 based on sum / median

Downloads

Each of the charts have certain interactive functionality associated with it so that user can explore, represent and visualize their data in more better and detailed manner. For instance,

- **Sankey Diagram:** user can drag nodes horizontally and vertically using ***mouse***
- **Treemap:** user can ***zoom in*** to lower hierarchy levels by click on any block and ***zoom back out*** using the top horizontal bar (one level of the hierarchy is displayed at once)
- **Sunburst:** user can click on any arc to ***zoom in*** and click on the center circle to ***zoom out***

Also, kindly use feature filtration option for large datasets (with too many features) in order to enhance the chart layout and representation (avoid overlap)



Now you see a 'Treemap' graph. You can zoom in to lower hierarchy levels by click on any block and zoom back out using the top horizontal bar (one level of the hierarchy is displayed at once).

Hierarchical Composition Profiling

View options:

Chart type: Sankey / Treemap / Sunburst

View type: All samples / By Metadata (Species) Group (Pig) / Per sample (100117001)

Submit

Calculate feature count based on: sum / mean

Show abundance value as: Absolute count / Relative proportion (%)

Filter features with counts < 10 based on sum / median

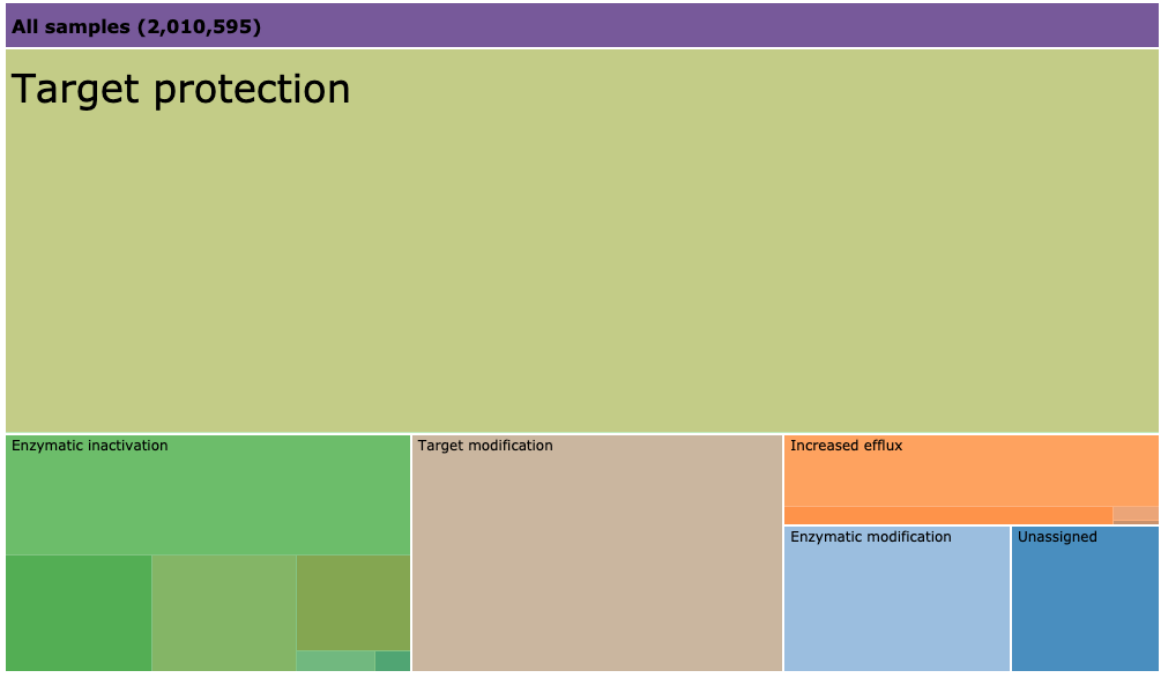
Downloads

Let's try to change to a 'Sunburst' now, so click on its icon on the 'Chart type' option. Then, to apply the changes, click on 'Submit'.

Each of the charts have certain interactive functionality associated with it so that user can explore, represent and visualize their data in more better and detailed manner. For instance,

- **Sankey Diagram:** user can drag nodes horizontally and vertically using ***mouse***
- **Treemap:** user can ***zoom in*** to lower hierarchy levels by click on any block and ***zoom back out*** using the top horizontal bar (one level of the hierarchy is displayed at once)
- **Sunburst:** user can click on any arc to ***zoom in*** and click on the center circle to ***zoom out***

Also, kindly use feature filtration option for large datasets (with too many features) in order to enhance the chart layout and representation (avoid overlap)



Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition

Hierarchical Composition Profiling

View options:

Chart type: Sankey / Treemap / Sunburst

View type: All samples / By Metadata (Species, Group: Pig) / Per sample (100117001)

Submit

Calculate feature count based on: sum / mean

Show abundance value as: Absolute count / Relative proportion (%)

Filter features with counts < 10 based on sum / median

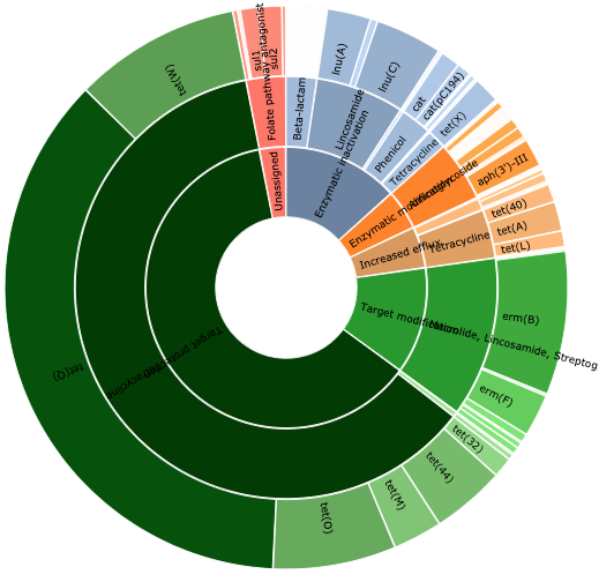
Downloads

Each of the charts have certain interactive functionality associated with it so that user can explore, represent and visualize their data in more better and detailed manner. For instan

- **Sankey Diagram:** user can drag nodes horizontally and vertically using ***mouse***
- **Treemap:** user can ***zoom in*** to lower hierarchy levels by click on any block and ***zoom back out*** using the top horizontal bar (one level of the hierarchy is displayed at once)
- **Sunburst:** user can click on any arc to ***zoom in*** and click on the center circle to ***zoom out***

Also, kindly use feature filtration option for large datasets (with too many features) in order to enhance the chart layout and representation (avoid overlap)

Remember, you can always play around with different subsets of data or groupings. Also, if there is too much information at once, you can filter data according to your interest.



Now you see a 'Sunburst' graph. You can click on any arc to zoom in and click on the center circle to zoom out.

Hierarchical Composition Profiling

View options:

Chart type: Sankey / Treemap / Sunburst

View type: All samples / By Metadata (Species, Group: Pig) / Per sample (100117001)

Submit

Calculate feature count based on: sum / mean

Show abundance value as: Absolute count / Relative proportion (%)

Filter features with counts < 10 based on sum / median

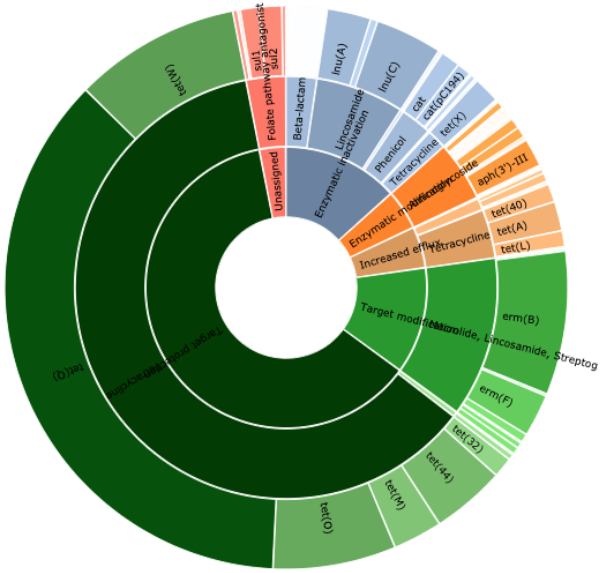
Downloads

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

Each of the charts have certain interactive functionality associated with it so that user can explore, represent and visualize their data in more better and detailed manner. For instance,

- **Sankey Diagram:** user can drag nodes horizontally and vertically using ***mouse***
- **Treemap:** user can ***zoom in*** to lower hierarchy levels by click on any block and ***zoom back out*** using the top horizontal bar (one level of the hierarchy is displayed at once)
- **Sunburst:** user can click on any arc to ***zoom in*** and click on the center circle to ***zoom out***

Also, kindly use feature filtration option for large datasets (with too many features) in order to enhance the chart layout and representation (avoid overlap)



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

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition

Hierarchical Composition Profiling

View options:

Chart type: Sankey / Treemap / Sunburst

View type: All samples / By Metadata (Species, Group: Pig) / Per sample (100117001)

Submit

Calculate feature count based on: sum / mean

Show abundance value as: Absolute count / Relative proportion (%)

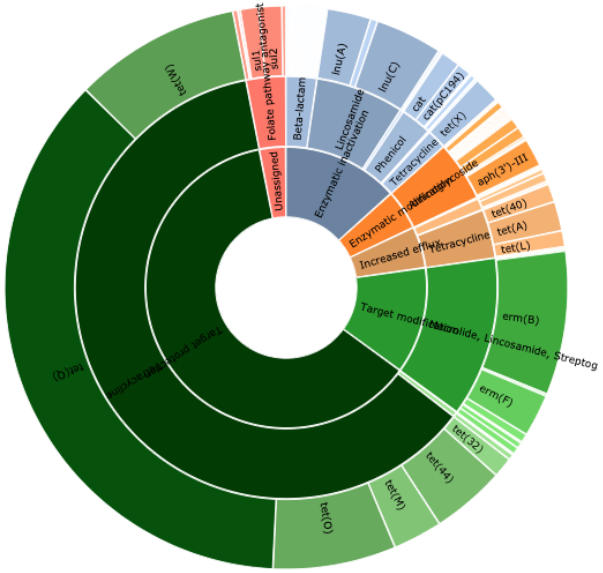
Filter features with counts < 10 based on sum / median

Downloads

Each of the charts have certain interactive functionality associated with it so that user can explore, represent and visualize their data in more better and detailed manner. For instance,

- **Sankey Diagram:** user can drag nodes horizontally and vertically using ***mouse***
- **Treemap:** user can ***zoom in*** to lower hierarchy levels by click on any block and ***zoom back out*** using the top horizontal bar (one level of the hierarchy is displayed at once)
- **Sunburst:** user can click on any arc to ***zoom in*** and click on the center circle to ***zoom out***

Also, kindly use feature filtration option for large datasets (with too many features) in order to enhance the chart layout and representation (avoid overlap)



For alpha diversity – click here.

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

RNA-seq methods

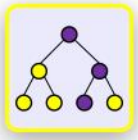
metagenomeSeq

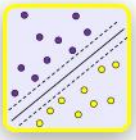
LEfSe

ALDEx2

ANCOM

4. Machine learning (Biomarker prediction)


Random Forest


SVM

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity

Alpha diversity is used to calculate the diversity of features (ARGs) present within a sample. The two most commonly used alpha diversity measurements are richness (numbers) and evenness (distribution).

Alpha diversity analysis & significance testing

General options:

Profile level: Feature (Rownames)

Diversity measure: Chao1

Experimental factor: Species

Statistical method: T-test / ANOVA

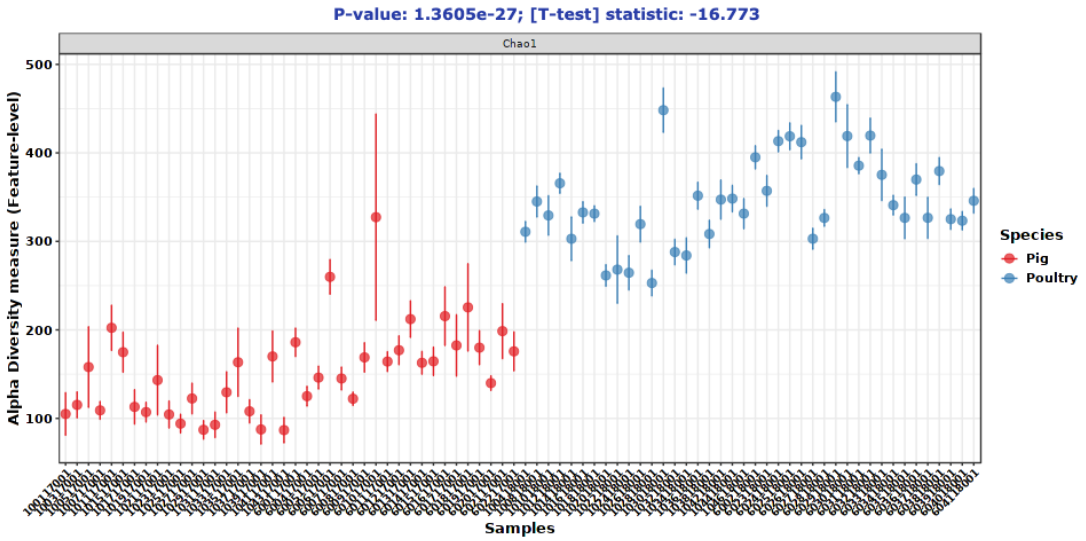
Submit

Downloads

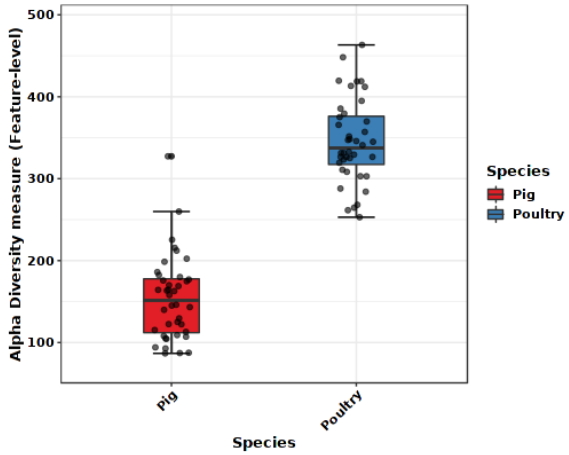
View options:

Color Palette: Set1

Statistical analysis



Individual samples



Groups

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

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity

Alpha diversity analysis & significance testing

General options:

Profile level: Feature (Rownames)

Diversity measure: Chao1

Experimental factor: Species

Statistical method: T-test / ANOVA

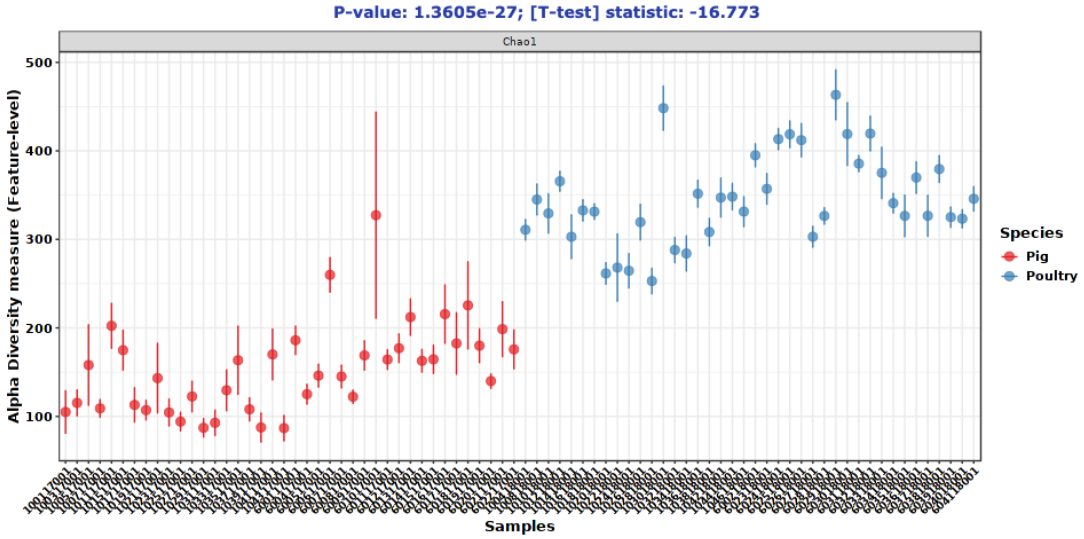
Submit

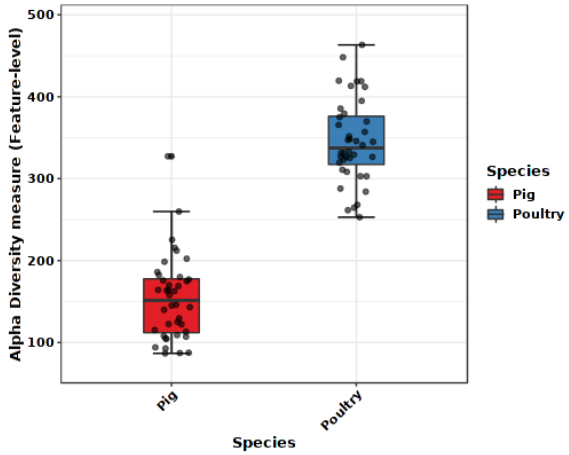
Downloads

View options:

Color Palette: Set1

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





What are the different measures of calculating alpha-diversity in ResistoXplorer?

Alpha diversity (α-diversity) is the mean feature diversity in samples or habitats at a local scale. Richness and Evenness are often used to measure alpha diversity:

Richness – takes into account the number of unique features in the samples, but does not discriminate frequencies.

Evenness – addresses how even the frequencies between unique features are. Typically, this is done with Shannon and Simpson diversity indexes.

In addition, ResistoXplorer provides many metrics to calculate diversity within samples. Most commonly used ones are listed below:

Observed: It estimates the amount of unique features found in each samples (richness);

ACE and Chao1: These metrics estimate diversity by adding a scaling factor to observed richness of features (ARGs) to account for rare observed or unobserved features (richness);

Pielou's evenness & Smith and Wilson's Evar index: These metrics account for how even the features (ARGs) are distributed among all different features present in a sample (evenness);

Shannon, Simpson and Fisher: These metrics account for both richness and evenness.

What are the differences between using features (resistance genes) or assigned taxa for diversity analysis?

ResistoXplorer allows users to compute either diversity based on original features (resistance genes) or on collapsing the data at different functional levels. Note, in the latter case, features (resistance genes) without taxa designation will be collapsed into an “unassigned” category, which could be an arbitrary mix of features (resistance genes) from across different levels. In some cases, features (resistance genes) without genus/species information are frequently both more abundant and more representative of total diversity than are features (resistance genes) with genus/species names.

Because of these issues, to understand the real diversity, it is recommended to first perform diversity analysis at the lowest level before collapsing data by functional assignment. When features (resistance genes) are well annotated or the selected functional level includes the majority of the features (resistance genes), it is biologically useful to perform diversity analysis at higher taxa levels for both data reductions and hypothesis generations.

Analysis Panel

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves

Let's look at rarefaction curves – click here.

Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

RNA-seq methods

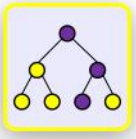
metagenomeSeq

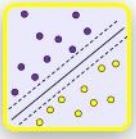
LEfSe

ALDEx2

ANCOM

4. Machine learning (Biomarker prediction)

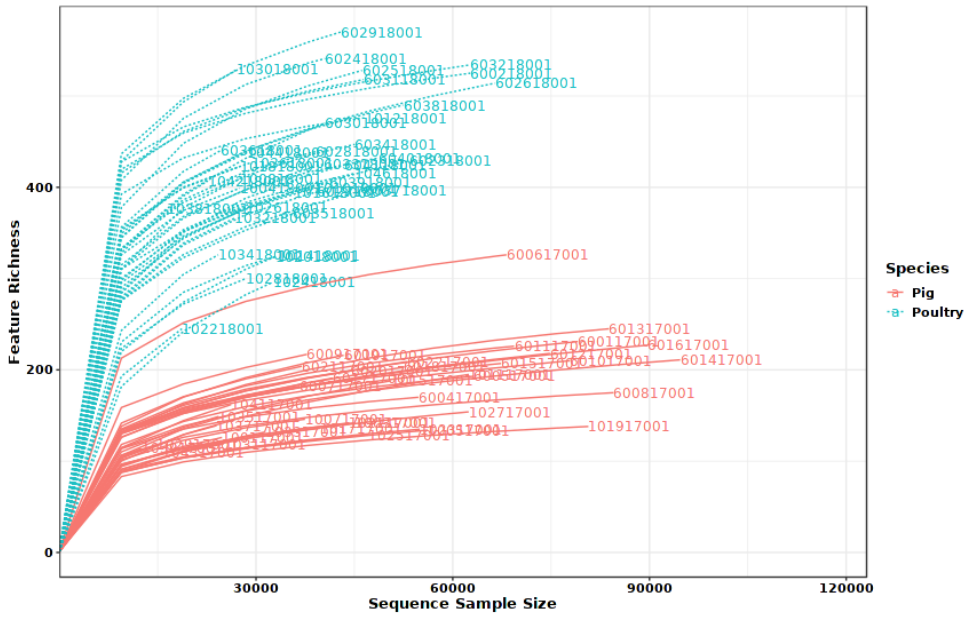

Random Forest


SVM

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity ▸ Rarefaction Analysis

Rarefaction curves

View options: Line color based on: Species | Line type based on: Species | Steps: 10 | Submit | Downloads



Rarefaction analysis will show you the richness of the resistance genes by the sequence sample size. You can customize how you would like to visualize the graph with the options above. If the resistome curves are getting flattened, it is a good indication that you have a good representation of the resistome.

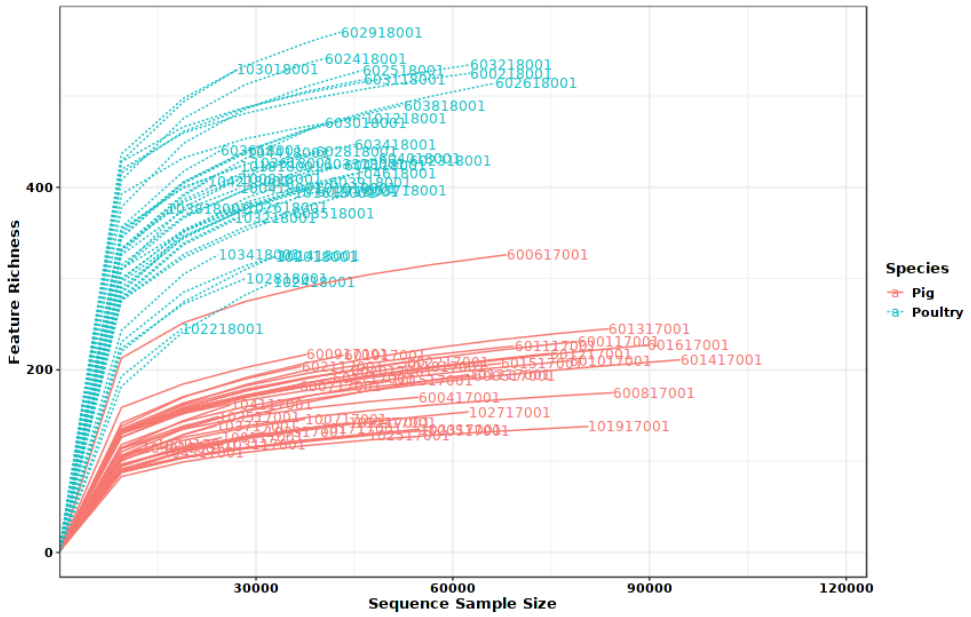
Once you are finished, you can click on 'Analysis panel' to go back to the previous page.

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity ▸ Rarefaction Analysis

Rarefaction curves

View options: Line color based on: Species Line type based on: Species Steps: 10 Submit Downloads

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



Analysis Panel

For ordination analysis – click here.

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

RNA-seq methods

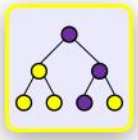
metagenomeSeq

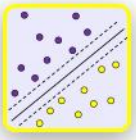
LEfSe

ALDEx2

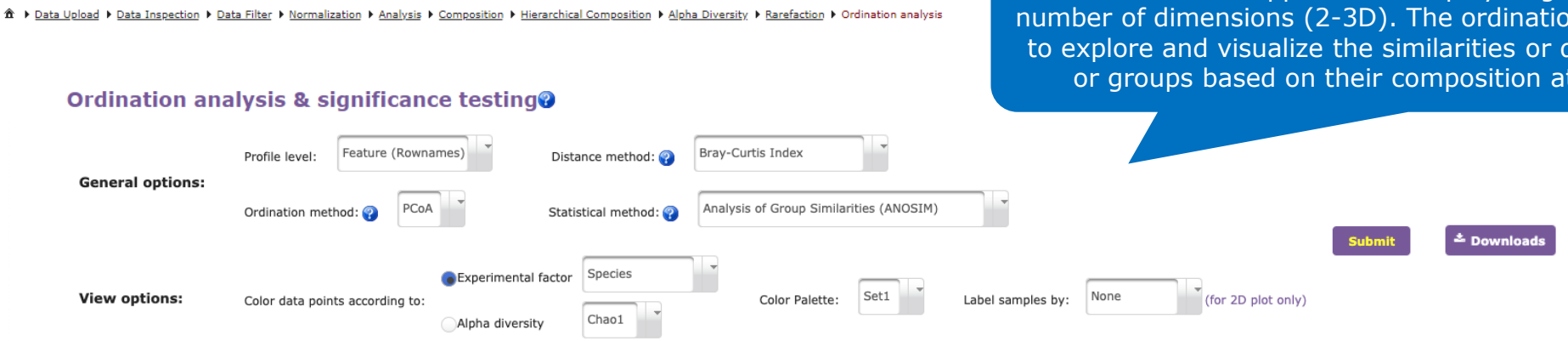
ANCOM

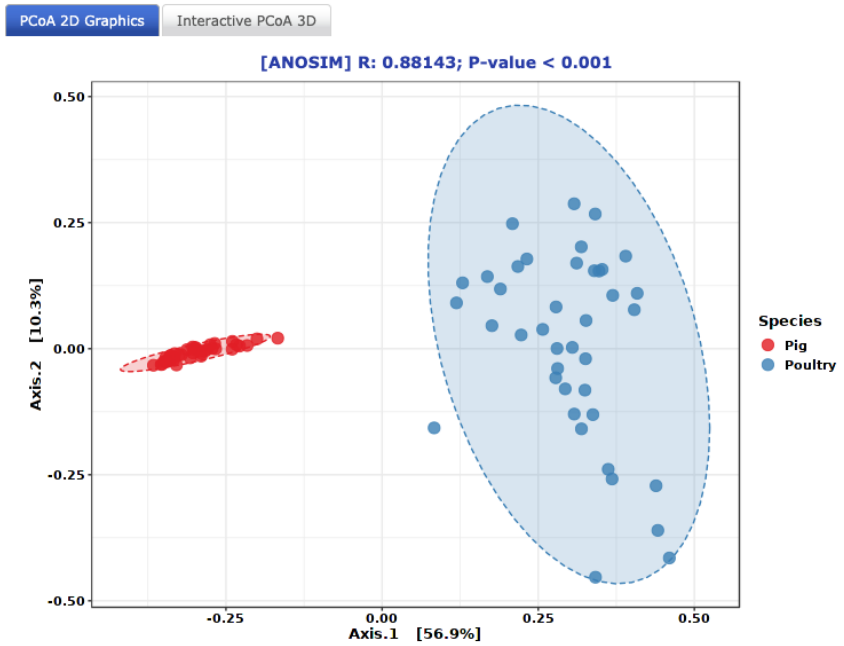
4. Machine learning (Biomarker prediction)


Random Forest


SVM

Ordination is an approach to display “high dimensional” data into lower number of dimensions (2-3D). The ordination analysis function allows users to explore and visualize the similarities or dissimilarities between samples or groups based on their composition at different functional levels.





You can clearly observe the differences in resistome of pig and poultry when clustered in different groups.

Ordination analysis & significance testing

General options:

Profile level: Feature (Rownames)

Distance method: Bray-Curtis Index

Ordination method: PCoA

Statistical method: Analysis of Group Similarities (ANOSIM)

Submit

Downloads

View options:

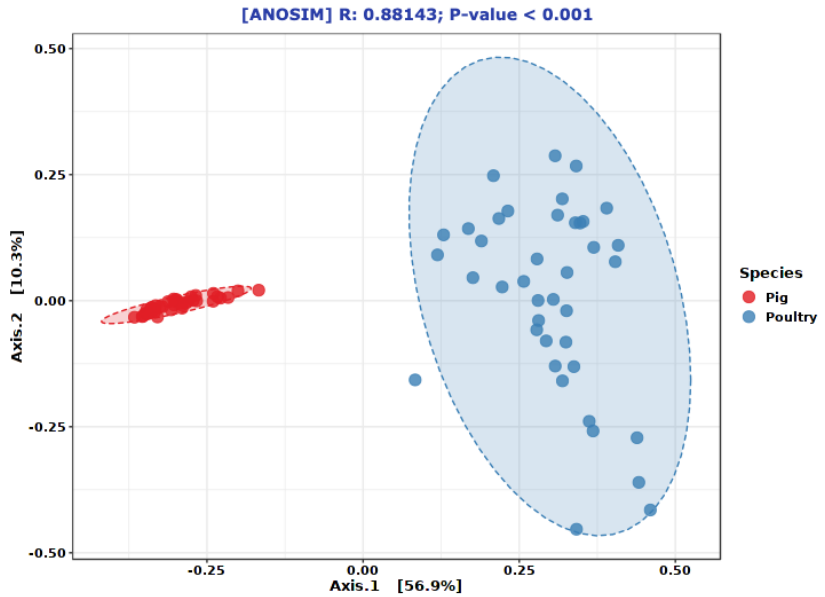
Color data points according to:

Palette: Set1

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

Currently, ResistoXplorer supports three of the most commonly used methods based on ordination: Principal Coordinates Analysis (PCoA), Nonmetric Multidimensional Scaling (NMDS), and Principal Component Analysis (PCA).

PCoA 2D Graphics | Interactive PCoA 3D



Ordination analysis & significance testing

General options:

Profile level: Feature (Rownames)

Distance method: Bray-Curtis Index

Ordination method: PCoA

Statistical method: Analysis of Group Similarities (ANOSIM)

Submit | Downloads

View options:

Color data points according to: Experimental factor — Species

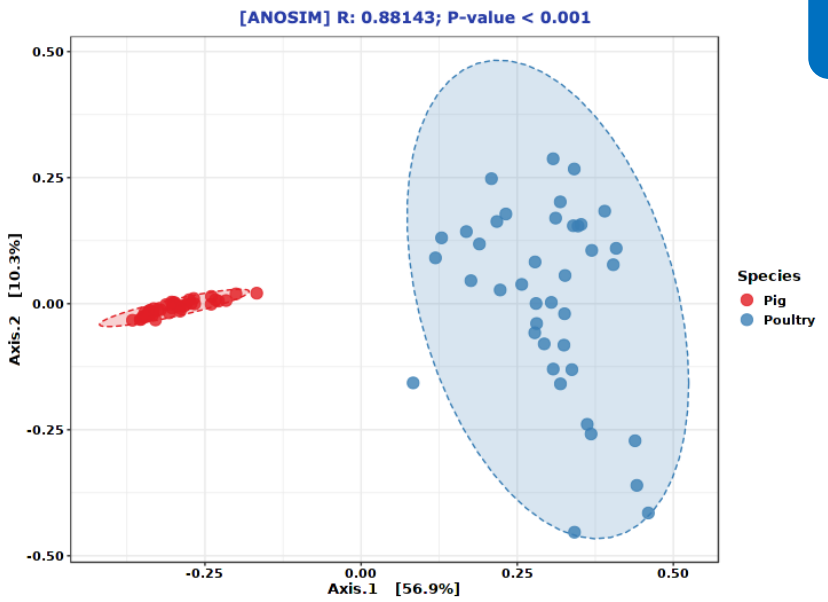
Alpha diversity — Chao1

Set1

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

The statistical methods measure the strength and statistical significance of sample groupings based on distance matrix. You can choose from the options listed in this item.

PCoA 2D Graphics | Interactive PCoA 3D



Ordination analysis & significance testing

General options:

Profile level: Feature (Rownames)

Distance method: Bray-Curtis Index

Ordination method: PCoA

Statistical method: Analysis of Group Similarities (ANOSIM)

Submit

Downloads

View options:

Color data points according to: Experimental factor — Species

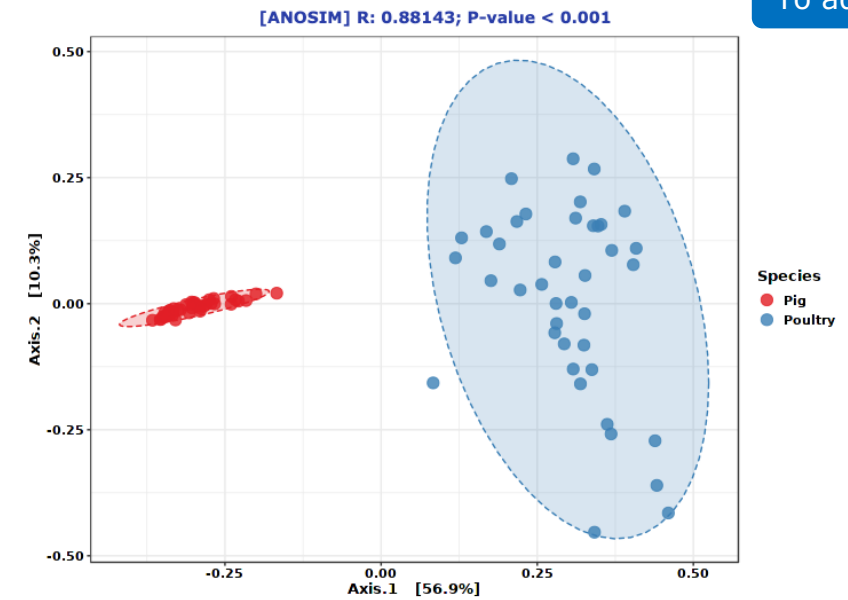
Alpha diversity — Chao1

Color Palette: Set1

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

PCoA 2D Graphics | Interactive PCoA 3D

To add a 3D PCoA analysis, click the button on the left.



Ordination analysis & significance testing

General options:

Profile level: Feature (Rownames)

Distance method: Bray-Curtis Index

Ordination method: PCoA

Statistical method: Analysis of Group Similarities (ANOSIM)

Submit

Downloads

View options:

Color data points according to: Experimental factor — Species

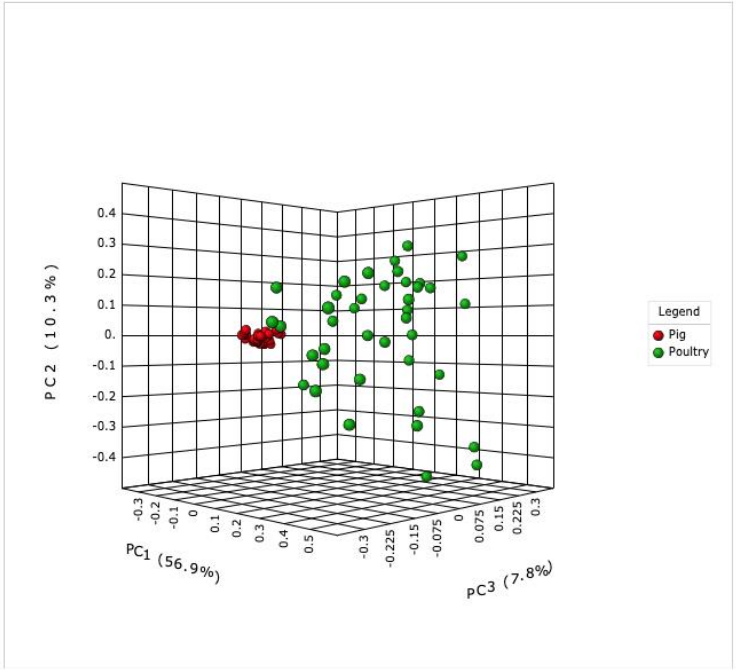
Alpha diversity — Chao1

Color Palette: Set1

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

PCoA 2D Graphics | Interactive PCoA 3D

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



You can click and drag the graph to move around with different angles. You can scroll to zoom and hover over points to get more details about each sample.

To go back, you can click on 'Analysis'.

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity ▸ Rarefaction ▸ Ordination analysis

Ordination analysis & significance testing

General options:

Profile level: Feature (Rownames)

Distance method: Bray-Curtis Index

Ordination method: PCoA

Statistical method: Analysis of Group Similarities (ANOSIM)

Submit

Downloads

View options:

Color data points according to: Experimental factor — Species

Alpha diversity — Chao1

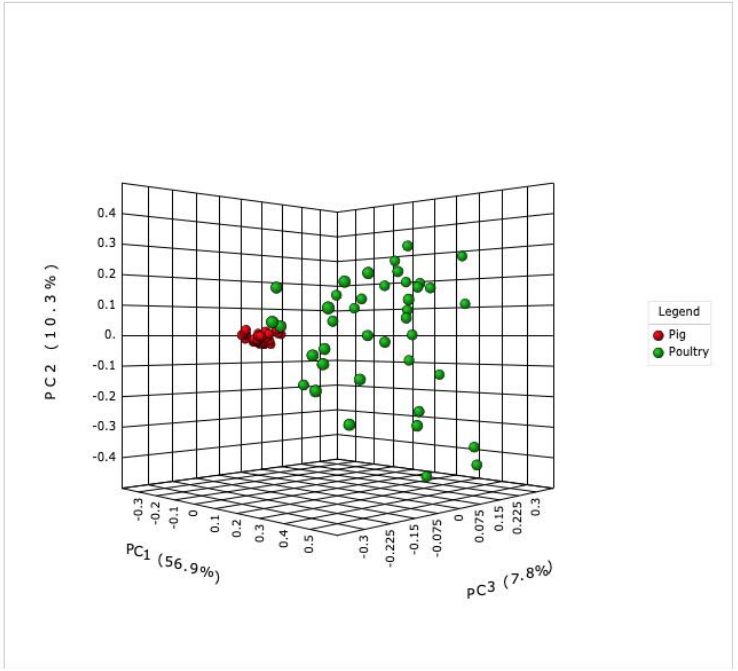
Color Palette: Set1

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.

PCoA 2D Graphics | Interactive PCoA 3D

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



How to choose between different measures to perform ordination analysis?

In contrast with alpha diversity, which compared the diversity within a sample, ordination analysis compares the diversity between-samples. As such, the distance or dissimilarity between each sample pair can be plotted into a graph after ordination.

Metrics supported by ResistoXplorer include:

Distance: Bray-Curtis Index, Jensen-Shannon divergence, Jaccard Index, and Manhattan.

Ordination: Principal Coordinates Analysis (PCoA), Nonmetric Multidimensional Scaling (NMDS), and Principal Component Analysis (PCA).

Statistical method: Analysis of Group Similarities (ANOSIM), Permutational MANOVA (PERMANOVA), and Homogeneity of Group Dispersions (PERMDISP).

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis

Analysis Panel

We will now look at the heatmap – click here.

1. Composition profiling


Visual profiling

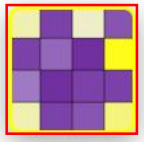

Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

RNA-seq methods

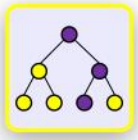
metagenomeSeq

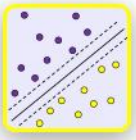
LEfSe

ALDEx2

ANCOM

4. Machine learning (Biomarker prediction)


Random Forest


SVM

Hierarchical clustering & Heatmap visualization

General options:

Profile level: Feature (Rownames)

Clustering Distance: Euclidean

Clustering algorithm: Ward

Cluster samples by: Experimental factor — Species

Current clustering algorithms

Submit

View options:

Color contrast: Viridis

View mode: Overview

Detail View (< 1500 features)

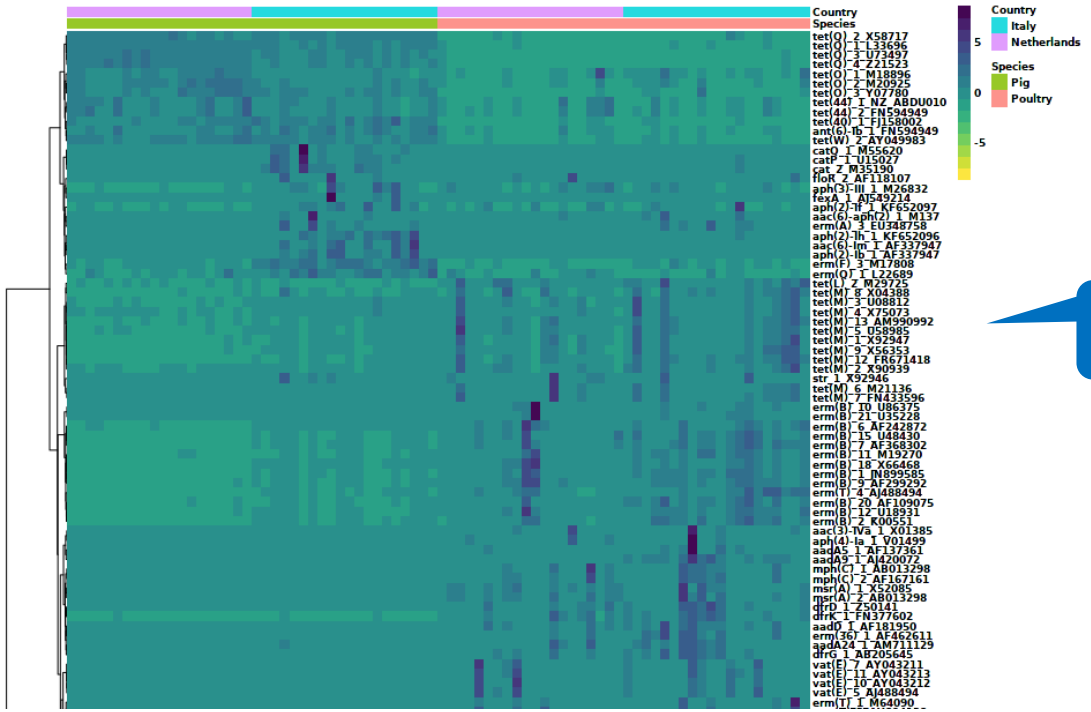
Scale rows

Show feature names

The heatmap visualization allows you analyze the gene abundance across different samples. In the dropdown menu to the left, it is possible to select several customizations such as profile level and color contrast. In addition, you can cluster the different levels of the heatmap together, as observed in the dendrogram to the left side of the graph.

You can see the categorization between Countries and Species here.

Please note: For larger datasets, the heatmap will show upto top 1500 most abundant features.



Here you see the abundance for every feature..

Hierarchical clustering & Heatmap visualization

General options:

Profile level: Feature (Rownames)

Clustering Distance: Euclidean

Clustering algorithm: Ward

Cluster samples by: Experimental factor — Species

Current clustering algorithms

View options:

Color contrast: Viridis

View mode: Overview

Detail View (< 1500 features)

Scale rows

Show feature names

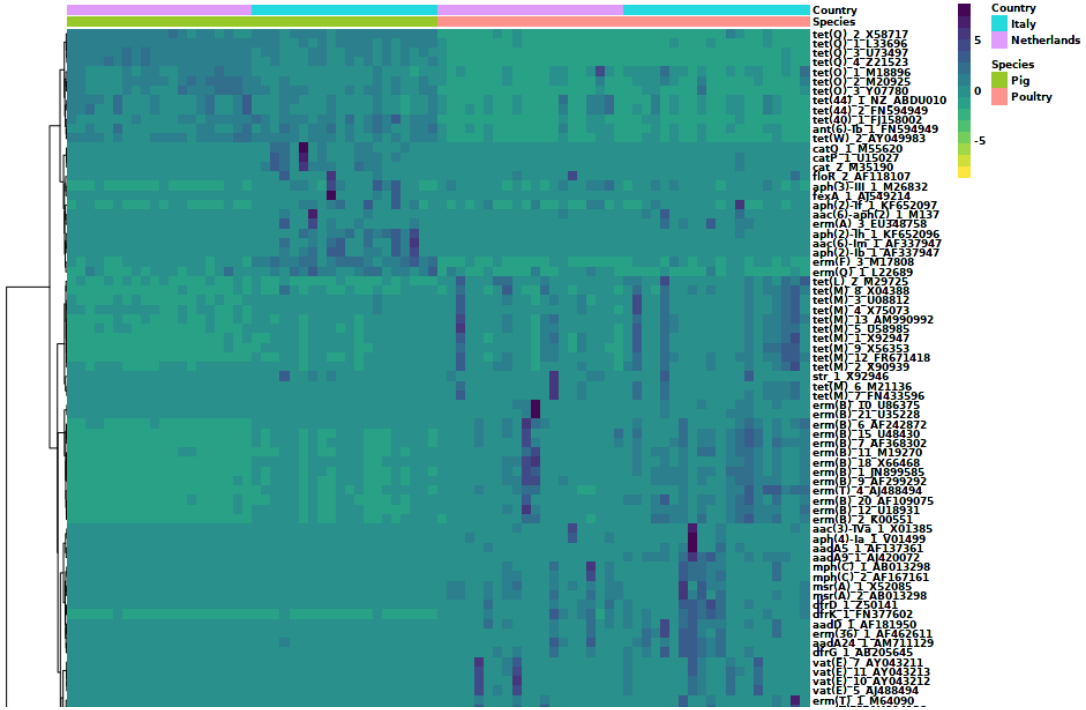
Let us try to look at 'Class' on 'Profile level' and let's change the color to 'Spectral' on the selection panel above.

Submit

Once changes are selected, click on 'Submit'.

Since the graph will be smaller, let's click on detail view as well.

Please note: For larger datasets, the heatmap will show upto top 1500 most abundant features.



Hierarchical clustering & Heatmap visualization

General options:

Profile level: Class

Clustering Distance: Euclidean

Clustering algorithm: Ward

Cluster samples by: Experimental factor — Species

Current clustering algorithms

Submit

Downloads

View options:

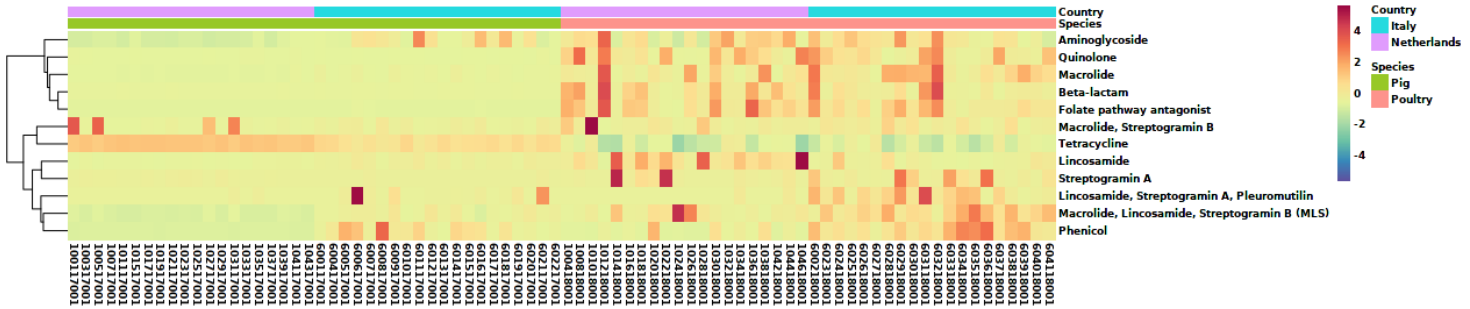
Color contrast: Spectral

View mode: Overview / Detail View (< 1500 features)

Scale rows

Show feature names

Please note: For larger datasets, the heatmap will show upto top 1500 most abundant features.



This the graph produced. We can see, for instance, that the abundance for tetracycline is higher in pig samples as compared to poultry. ResistoXplorer offers several methods of clustering the data, so try them out.

To go back, you can click on 'Analysis panel'.

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity ▸ Rarefaction Analysis ▸ Beta diversity ▸ Heatmap

Hierarchical clustering & Heatmap visualization

General options:

Profile level: Class

Clustering Distance: Euclidean

Clustering algorithm: Ward

Cluster samples by: Experimental factor — Species

Current clustering algorithms

Submit

Downloads

View options:

Color contrast: Spectral

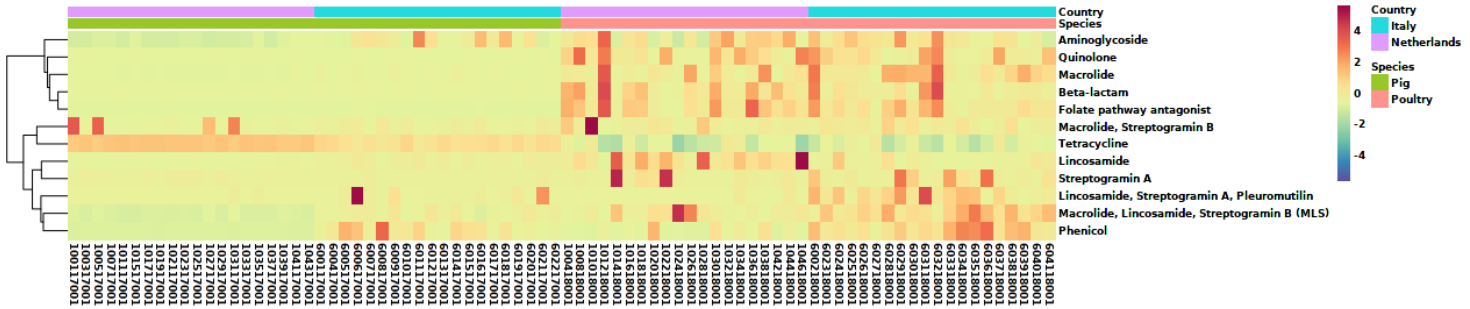
View mode: Overview / Detail View (< 1500 features)

Scale rows

Show feature names

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

Please note: For larger datasets, the heatmap will show upto top 1500 most abundant features.



Analysis Panel

1. Composition profiling


Visual profiling


Rarefaction curves


Ordination analysis

Let's move forward with dendrogram analysis.

2. Clustering ana


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

RNA-seq methods

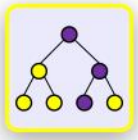
metagenomeSeq

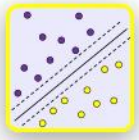
LEfSe

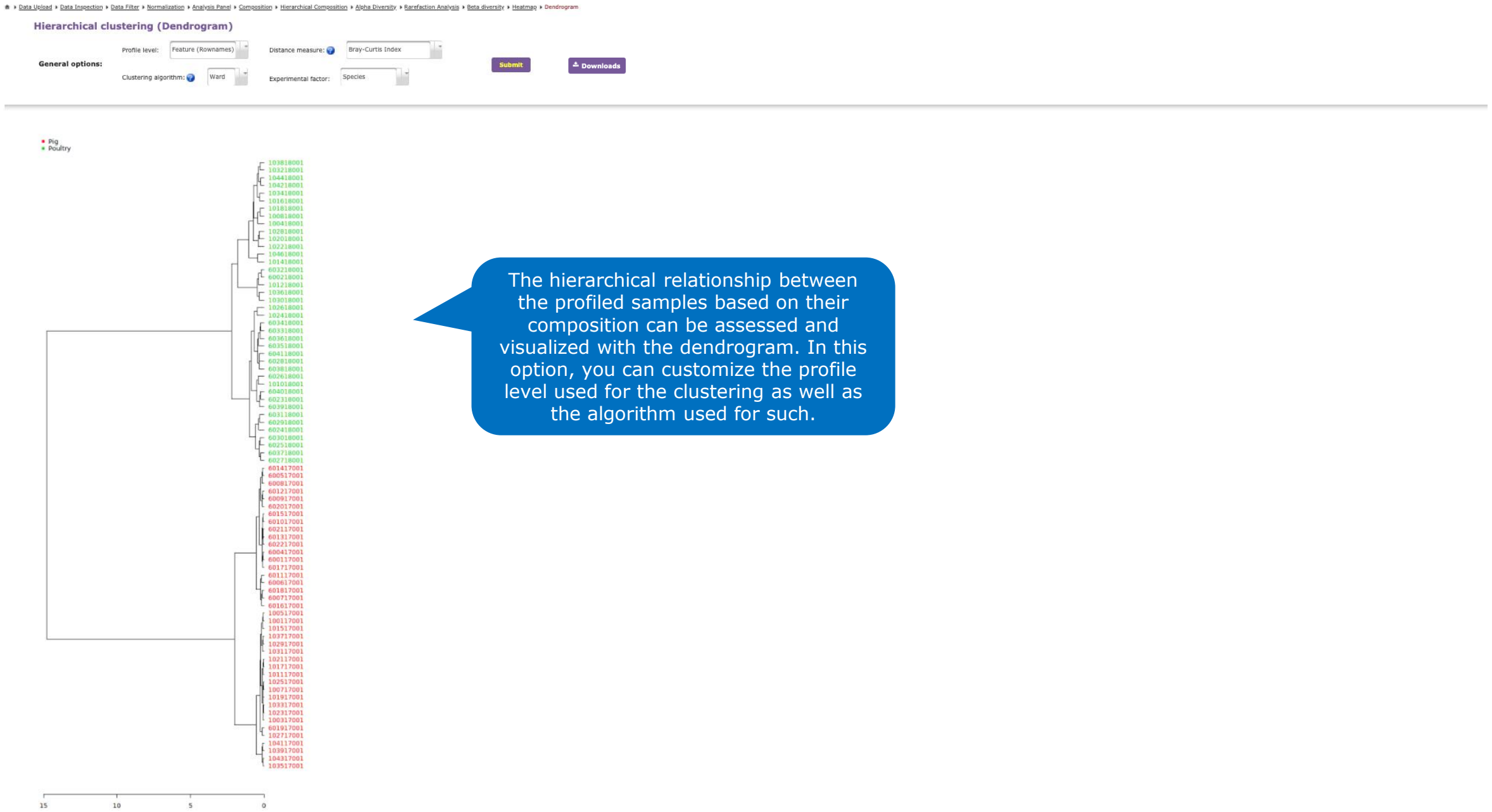
ALDEx2

ANCOM

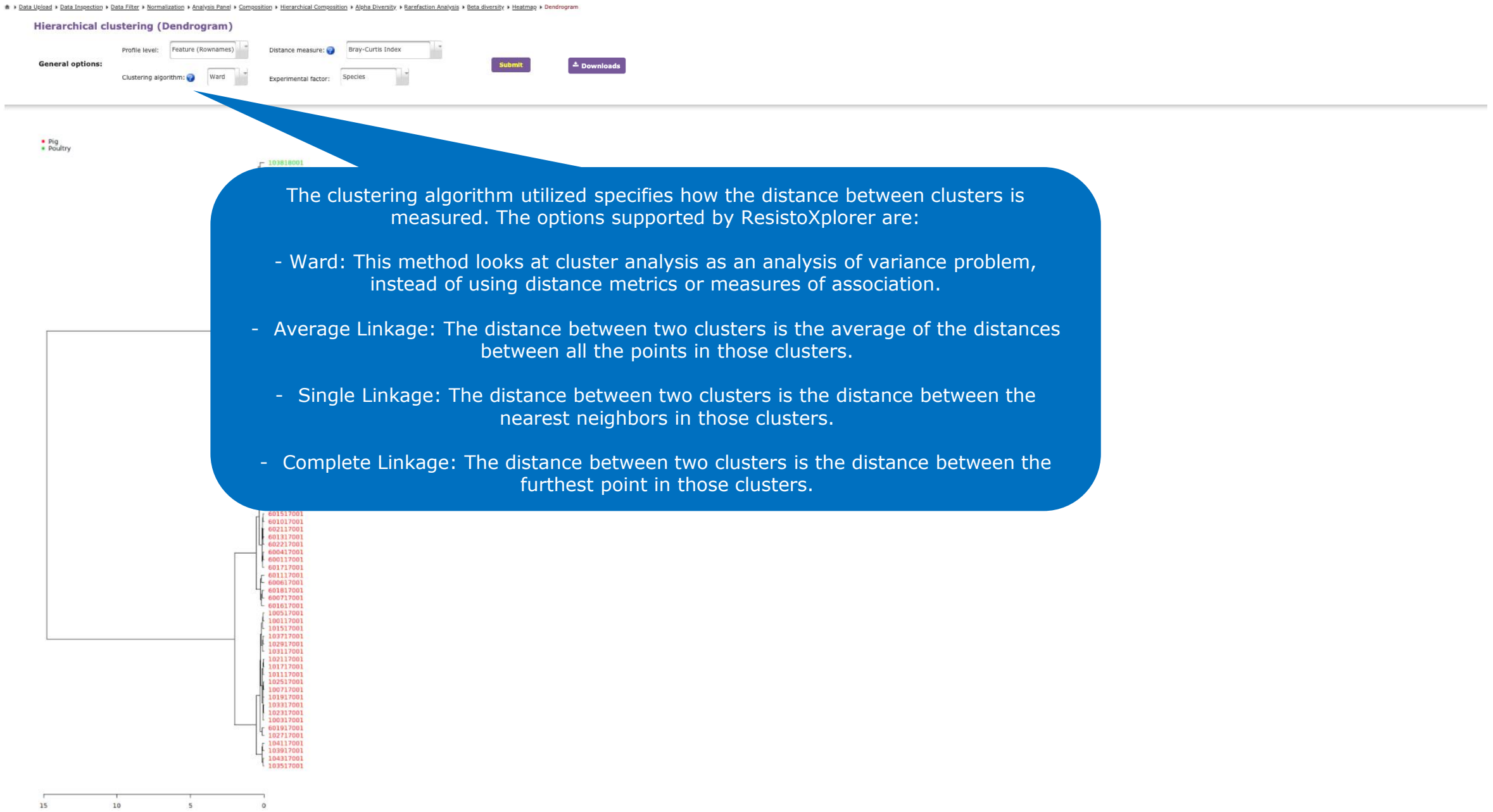
4. Machine learning (Biomarker prediction)


Random Forest


SVM


Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity ▸ Rarefaction Analysis ▸ Beta diversity ▸ Heatmap ▸ Dendrogram
Hierarchical clustering (Dendrogram)
General options:
Profile level:
Feature (Rownames)
Distance measure:
Bray-Curtis Index
Clustering algorithm:
Ward
Experimental factor:
Species
Submit
Downloads
Pig
Poultry
15
10
5
0
The hierarchical relationship between the profiled samples based on their composition can be assessed and visualized with the dendrogram. In this option, you can customize the profile level used for the clustering as well as the algorithm used for such.


Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity ▸ Rarefaction Analysis ▸ Beta diversity ▸ Heatmap ▸ Dendrogram
Hierarchical clustering (Dendrogram)
General options:
Profile level:
Feature (Rownames)
Distance measure:
Bray-Curtis Index
Clustering algorithm:
Ward
Experimental factor:
Submit
Downloads
There are various distance metrics available to quantify the dissimilarity between the samples. The ones supported by ResistoXplorer are: Bray-Curtis Index, Jaccard index, Jensen-Shannon Divergence, Manhattan, Euclidean, and Chao.
Pig
Poultry
15
10
5
0



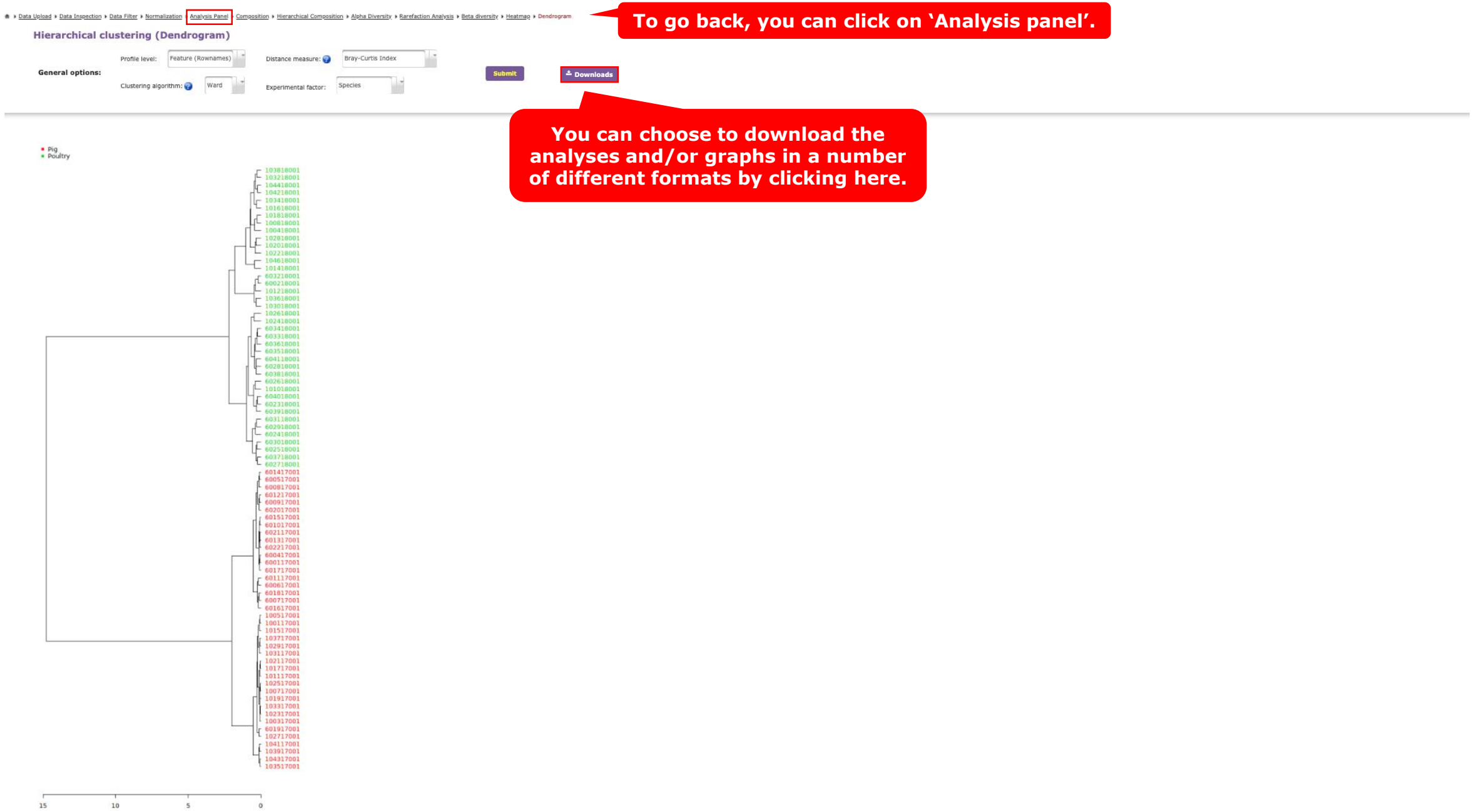
The clustering algorithm utilized specifies how the distance between clusters is measured. The options supported by ResistoXplorer are:

- Ward: This method looks at cluster analysis as an analysis of variance problem, instead of using distance metrics or measures of association.

- Average Linkage: The distance between two clusters is the average of the distances between all the points in those clusters.

- Single Linkage: The distance between two clusters is the distance between the nearest neighbors in those clusters.

- Complete Linkage: The distance between two clusters is the distance between the furthest point in those clusters.


Data Upload
Data Inspection
Data Filter
Normalization
Analysis Panel
Composition
Hierarchical Composition
Alpha Diversity
Rarefaction Analysis
Beta diversity
Heatmap
Dendrogram
Hierarchical clustering (Dendrogram)
General options:
Profile level:
Feature (Rownames)
Distance measure:
Bray-Curtis Index
Clustering algorithm:
Ward
Experimental factor:
Species
Submit
Downloads
Pig
Poultry
15
10
5
0
To go back, you can click on 'Analysis panel'.
You can choose to download the analyses and/or graphs in a number of different formats by clicking here.

Analysis Panel

1. Composition profiling


Visual profiling


Hierarchical


Ordination analysis

Let's look at the core resistome. Click here.

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

RNA-seq methods

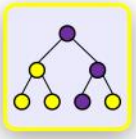
metagenomeSeq

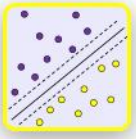
LEfSe

ALDEx2

ANCOM

4. Machine learning (Biomarker prediction)


Random Forest


SVM

Core Resistome

General options:

Profile level: Feature (Rownames)

Analysis type: All samples / By Experimental factor — Species — Group: Pig

Sample prevelance (%): 50

Relative abundance (%): 0.2

Submit

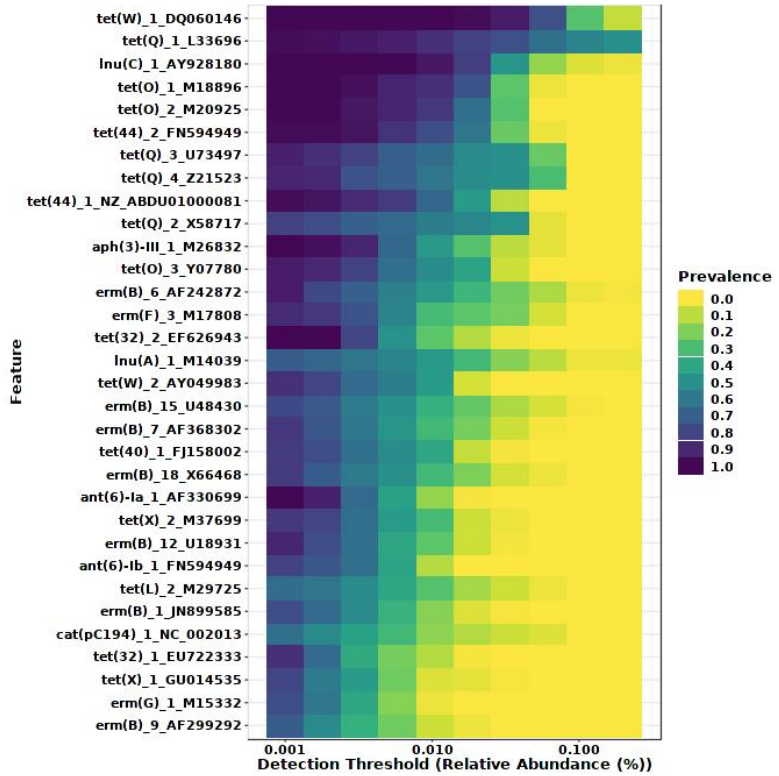
Downloads

View options:

Color contrast: Viridis

View mode: Overview / Detail View (< 1500 features)

Please note: The heatmap will only show upto 1500 core features identified based on sample prevalence and abundance level. The complete results can be downloaded as a table.



The core resistome option allows you to analyze which genes are more prevalent in your sample, and compare between different groups. You can determine the threshold for prevalence in the menu above.

Core Resistome

General options:

Profile level: Feature (Rownames)

Analysis type: All samples / By Experimental factor — Species — Group: Pig

Sample prevelance (%): 50

Relative abundance (%): 0.2

Submit

Downloads

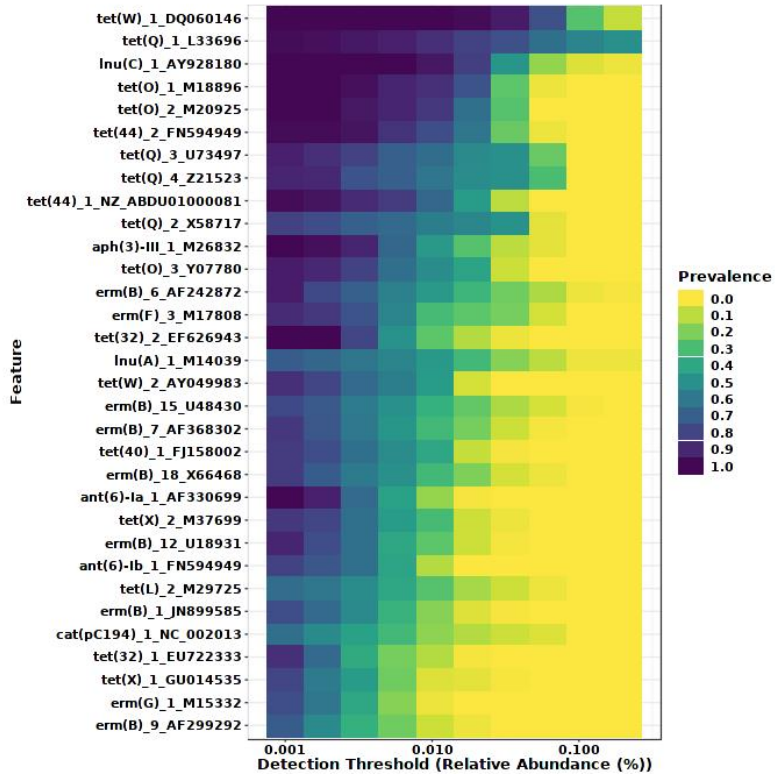
View options:

Color contrast: Viridis

View mode: Overview / Detail View (< 1500 features)

This analysis can be done for all samples or for a particular sample group. This option can be chosen here.

Please note: The heatmap will only show upto 1500 core features identified based on sample prevalence and abundance level. The complete re



To go back, you can click on 'Analysis panel'.

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity ▸ Rarefaction Analysis ▸ Beta diversity ▸ Heatmap ▸ Dendrogram ▸ Core Resistome

Core Resistome

General options:

Profile level: Feature (Rownames)

Analysis type: All samples / By Experimental factor — Species — Group: Pig

Sample prevelance (%): 50

Relative abundance (%): 0.2

Submit

Downloads

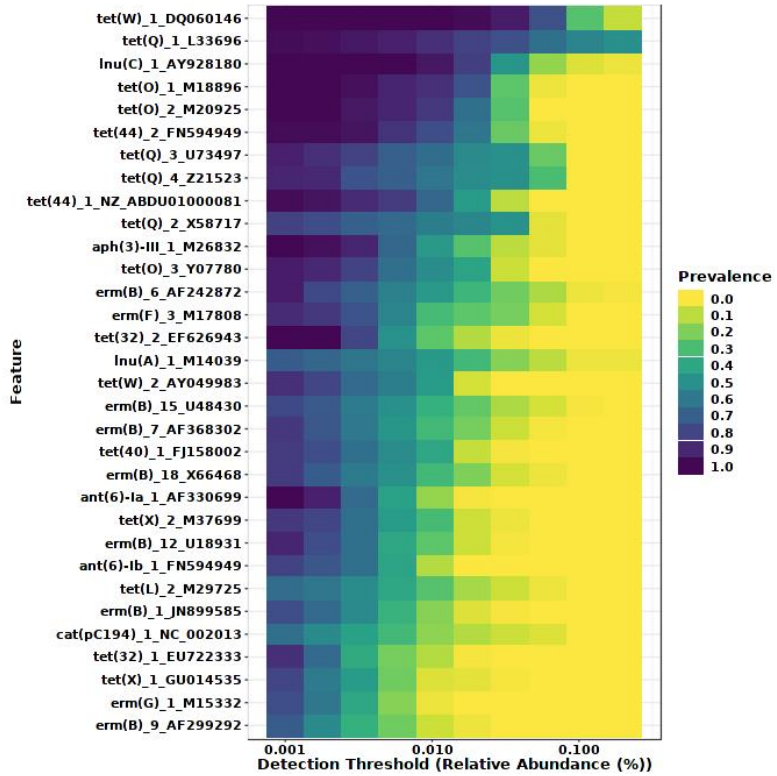
View options:

Color contrast: Viridis

View mode: Overview / Detail View (< 1500 features)

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

Please note: The heatmap will only show upto 1500 core features identified based on sample prevalence and abundance level. The complete results can be downloaded as a table.



Analysis Panel

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

Let's follow to 'Correlation'. Click here.

3. Differential abundance testing

RNA-seq methods

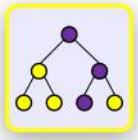
metagenomeSeq

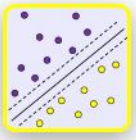
LEfSe

ALDEx2

ANCOM

4. Machine learning (Biomarker prediction)


Random Forest


SVM

Correlation analysis

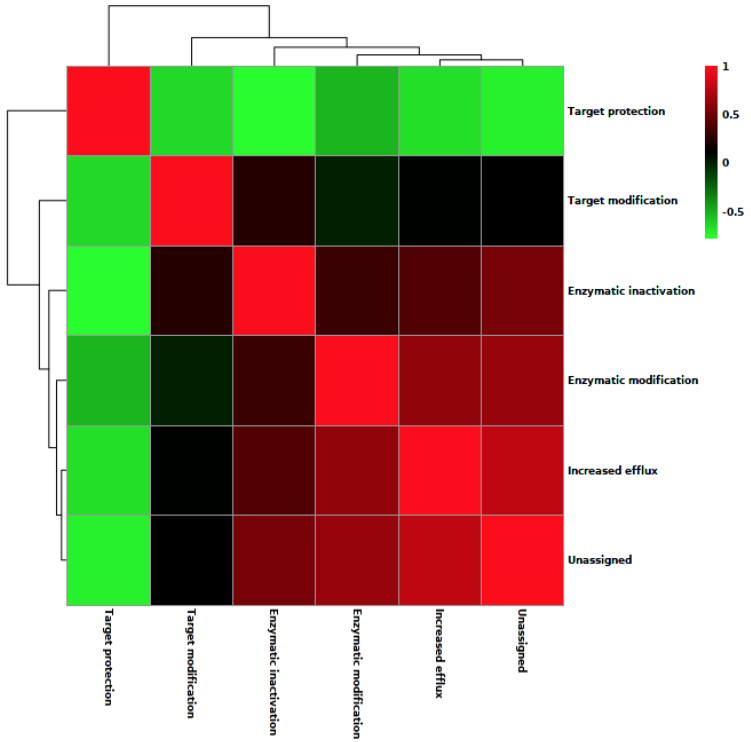
General options: Profile level: Mechanism Method: Pearson r

View options: Color contrast: Red / Green View mode: Overview / Detail View

Submit Downloads

Heatmap view

Please note: For larger datasets, only top 1500 features will be selected based on their ***interquantile range (IQR)***. The heatmap will only show correlations for a maximum of 1500 features.



Correlation analysis shows you how different features are correlated. You can custom the view to see the subdivision of samples you want. ResistoXplorer currently supports 'Pearson r', 'Spearman' and 'Kendall' rank correlations.

To go back, you can click on 'Analysis panel'.

Correlation analysis

General options: Profile level: Mechanism Method: Pearson r

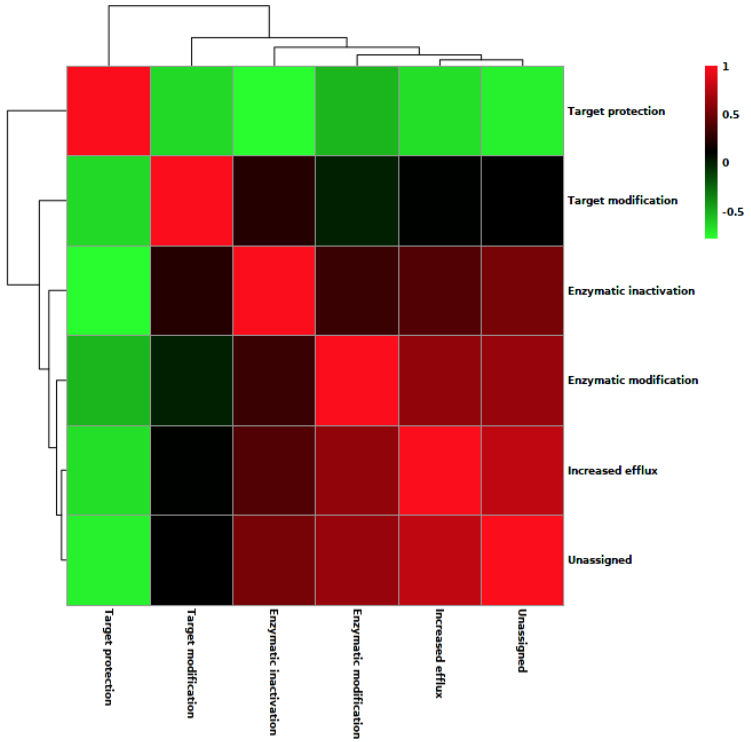
View options: Color contrast: Red / Green View mode: Overview / Detail View

Submit Downloads

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

Heatmap view

Please note: For larger datasets, only top 1500 features will be selected based on their ***interquantile range (IQR)***. The heatmap will only show correlations for a maximum of 1500 features.



What are the different methods for performing differential testing are available in ResistoXplorer?

ResistoXplorer supports both classical and standard as well as more recent compositional data analysis (CoDA) based univariate analysis approaches such as:

edgeR
DESeq2
metagenomeSeq
LEfSe
ALDEx2
ANCOM

We will now go through some details into each of them.

Analysis Panel

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

For RNAseq-based differential analysis, click here.

RNA-seq methods

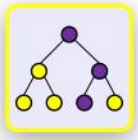
metagenomeSeq

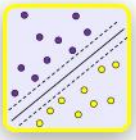
LEfSe

ALDEx2

ANCOM

4. Machine learning (Biomarker prediction)


Random Forest


SVM

RNASeq-based statistical analysis

General options:

Profile level: Feature (Rownames)

Algorithm: DESeq2 / EdgeR

Experimental factor: Species

Adjusted P value cutoff: 0.05

Submit

Two algorithms for RNASeq-based differential analysis are available: DESeq2 and edgeR. Choose your preferred settings for profile level, experimental and p-value cutoff. An example of DESeq2 is shown below.

The table below shows at most 500 features ranked by their P values, with significant ones highlighted in orange

Name	log2FC	lfcSE	Pvalues	FDR	View
tet(40)_1_FJ158002	-5.2795	0.24428	1.3805E-103	3.3407E-101	Details
tet(Q)_4_Z21523	-6.1573	0.31522	5.734E-85	6.9382E-83	Details
tet(Q)_3_U73497	-5.879	0.31033	4.8986E-80	3.9516E-78	Details
tet(W)_2_AY049983	-4.8997	0.25998	3.1316E-79	1.8946E-77	Details
tet(O)_3_Y07780	-4.8759	0.2633	1.4678E-76	7.1041E-75	Details
tet(Q)_1_L33696	-5.5959	0.31488	1.1765E-70	4.7454E-69	Details
tet(O)_2_M20925	-3.9015	0.22905	4.6643E-65	1.6125E-63	Details
ant(6)-Ib_1_FN594949	-4.4491	0.26631	1.1724E-62	3.5465E-61	Details
erm(Q)_1_L22689	-4.1121	0.28656	1.0642E-46	2.8615E-45	Details
erm(F)_3_M17808	-4.8891	0.3506	3.3784E-44	8.1758E-43	Details
tet(A)_4_AJ517790	4.5455	0.32809	1.1974E-43	2.6344E-42	Details
tet(A)_3_AY196695	4.4622	0.32355	2.8745E-43	5.7968E-42	Details
tet(O)_1_M18896	-3.3208	0.24396	3.3804E-42	6.2927E-41	Details
tet(44)_1_NZ_ABDU01000081	-3.6405	0.27116	4.2738E-41	7.3876E-40	Details
tet(A)_5_AJ419171	4.6985	0.3564	1.0985E-39	1.7722E-38	Details
tet(Q)_2_X58717	-5.155	0.39854	2.8713E-38	4.3429E-37	Details
tet(44)_2_FN594949	-3.4075	0.27443	2.1278E-35	3.029E-34	Details
sul1_3_EU855787	4.3451	0.37918	2.1124E-30	2.84E-29	Details
tet(A)_2_X00006	4.6106	0.40447	4.2174E-30	5.3716E-29	Details
dfrK_1_FN377602	5.1222	0.46403	2.4949E-28	3.0188E-27	Details

RNASeq-based statistical analysis

General options:

Profile level: Feature (Rownames)

Algorithm: DESeq2 / EdgeR

Experimental factor: Species

Adjusted P value cutoff: 0.05

Submit

Let's try to change the 'Profile level' to 'Class', the 'Algorithm' to 'edgeR', and the 'P-value cutoff' to '0.001'. When selected, click 'Submit'.

The table below shows at most 500 features ranked by their P values, with significant ones highlighted in orange

Name	log2FC	lfcSE	Pvalues	FDR	View
tet(40)_1_FJ158002	-5.2795	0.24428	1.3805E-103	3.3407E-101	Details
tet(Q)_4_Z21523	-6.1573	0.31522	5.734E-85	6.9382E-83	Details
tet(Q)_3_U73497	-5.879	0.31033	4.8986E-80	3.9516E-78	Details
tet(W)_2_AY049983	-4.8997	0.25998	3.1316E-79	1.8946E-77	Details
tet(O)_3_Y07780	-4.8759	0.2633	1.4678E-76	7.1041E-75	Details
tet(Q)_1_L33696	-5.5959	0.31488	1.1765E-70	4.7454E-69	Details
tet(O)_2_M20925	-3.9015	0.22905	4.6643E-65	1.6125E-63	Details
ant(6)-Ib_1_FN594949	-4.4491	0.26631	1.1724E-62	3.5465E-61	Details
erm(Q)_1_L22689	-4.1121	0.28656	1.0642E-46	2.8615E-45	Details
erm(F)_3_M17808	-4.8891	0.3506	3.3784E-44	8.1758E-43	Details
tet(A)_4_AJ517790	4.5455	0.32809	1.1974E-43	2.6344E-42	Details
tet(A)_3_AY196695	4.4622	0.32355	2.8745E-43	5.7968E-42	Details
tet(O)_1_M18896	-3.3208	0.24396	3.3804E-42	6.2927E-41	Details
tet(44)_1_NZ_ABDU01000081	-3.6405	0.27116	4.2738E-41	7.3876E-40	Details
tet(A)_5_AJ419171	4.6985	0.3564	1.0985E-39	1.7722E-38	Details
tet(Q)_2_X58717	-5.155	0.39854	2.8713E-38	4.3429E-37	Details
tet(44)_2_FN594949	-3.4075	0.27443	2.1278E-35	3.029E-34	Details
sul1_3_EU855787	4.3451	0.37918	2.1124E-30	2.84E-29	Details
tet(A)_2_X00006	4.6106	0.40447	4.2174E-30	5.3716E-29	Details
dfrK_1_FN377602	5.1222	0.46403	2.4949E-28	3.0188E-27	Details

RNASeq-based statistical analysis

You can also change the 'Experimental factor'.

General options:

Profile level: Class

Algorithm: DESeq2 / EdgeR

Submit | Result Table

Experimental factor: Species

Adjusted P value cutoff: 0.00

The table below shows at most 500 features ranked by their P values, with significant ones highlighted in orange

Name	log2FC	logCPM	Pvalues	FDR	View
Tetracycline	-3.3829	20.199	8.0777E-67	9.6933E-66	Details
Folate pathway antagonist	4.7712	14.175	7.9041E-54	4.7425E-53	Details
Beta-lactam	4.2469	13.838	1.4433E-31	5.773E-31	Details
Macrolide	1.7648	11.007	2.5196E-10	7.5588E-10	Details
Aminoglycoside	-1.1916	15.526	8.122E-8	1.9493E-7	Details
Macrolide, Streptogramin B	-1.966	12.538	2.8507E-7	5.7015E-7	Details
Quinolone	1.7815	8.3991	4.6905E-7	8.0409E-7	Details
Lincosamide, Streptogramin A, Pleuromutilin	-1.0386	7.6452	9.9504E-5	1.4926E-4	Details
Streptogramin A	1.2592	9.9407	0.0061635	0.008218	Details
Phenicol	-1.0684	14.527	0.0076633	0.009196	Details
Lincosamide	-0.42468	16.257	0.090756	0.099006	Details
Macrolide, Lincosamide, Streptogramin B (MLS)	0.18334	16.679	0.4288	0.4288	Details

1

To go back, you can click on 'Analysis panel'.

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity ▸ Rarefaction Analysis ▸ Beta diversity ▸ Heatmap ▸ Dendrogram ▸ Core Resistome ▸ Correlation ▸ RNAseq Methods

RNASeq-based statistical analysis

General options:

Profile level: Class

Experimental factor: Species

Algorithm: DESeq2 / EdgeR

Adjusted P value cutoff: 0.00

Submit

Result Table

You can choose to download the result table by clicking here.

The table below shows at most 500 features ranked by their P values, with significant ones highlighted in orange

Name	log2FC	logCPM	Pvalues	FDR	View
Tetracycline	-3.3829	20.199	8.0777E-67	9.6933E-66	Details
Folate pathway antagonist	4.7712	14.175	7.9041E-54	4.7425E-53	Details
Beta-lactam	4.2469	13.838	1.4433E-31	5.773E-31	Details
Macrolide	1.7648	11.007	2.5196E-10	7.5588E-10	Details
Aminoglycoside	-1.1916	15.526	8.122E-8	1.9493E-7	Details
Macrolide, Streptogramin B	-1.966	12.538	2.8507E-7	5.7015E-7	Details
Quinolone	1.7815	8.3991	4.6905E-7	8.0409E-7	Details
Lincosamide, Streptogramin A, Pleuromutilin	-1.0386	7.6452	9.9504E-5	1.4926E-4	Details
Streptogramin A	1.2592	9.9407	0.0061635	0.008218	Details
Phenicol	-1.0684	14.527	0.0076633	0.009196	Details
Lincosamide	-0.42468	16.257	0.090756	0.099006	Details
Macrolide, Lincosamide, Streptogramin B (MLS)	0.18334	16.679	0.4288	0.4288	Details

1

Which RNASeq method should I use for my data (edgeR vs DESeq2)?

Both methods are robust and well established. In addition, they are applicable and useful to metagenomics data as well. The differences between them rely mostly on their normalization method and the algorithms used for estimation of dispersion. While edgeR moderates the dispersion estimate for each gene toward a common estimate across all genes using a weighted conditional likelihood, DESeq2 detects and corrects dispersion estimates that are too low through modeling of the dependence of the dispersion on the average expression strength over all samples. In general, DESeq2 is more robust in estimating differential expression features and usually yields a low false positive rate, while edgeR is more powerful but it can also lead to higher rates of false detection. It is suggested that users utilize multiple methods when running their analyses through ResistoXplorer, specially in terms of differential testing.

For more details about their implementation, please refer to the DESeq2 and edgeR papers.

Analysis Panel

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heat

Core Resistome


Correlation

Let's move to metagenomeSeq. Click here.

3. Diffe

RNA-seq methods

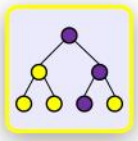
metagenomeSeq

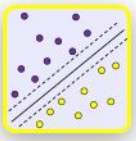
LEfSe

ALDEx2

ANCOM

4. Machine learning (Biomarker prediction)


Random Forest


SVM

metagenomeSeq: statistical analysis for sparse high-throughput sequencing data

This will show the differential testing with metagenomeSeq. All significant findings are highlighted in orange.

General options:

Profile level: Feature (Rownames)

Experimental factor: Species

Statistical model: zero-inflated Gaussian fit (fitZig)

Adjusted P value cutoff: 0.05

Submit

Result Table

The table below shows at most 500 features ranked by their P values, with significant ones highlighted in orange

Name	Pvalues	FDR	View
tet(W)_2_AY049983	2.9249E-20	7.0782E-18	Details
tet(40)_1_FJ158002	1.9021E-19	2.3016E-17	Details
tet(Q)_4_Z21523	4.1365E-19	3.3368E-17	Details
tet(O)_3_Y07780	5.6384E-19	3.4112E-17	Details
tet(Q)_3_U73497	3.1023E-18	1.5015E-16	Details
tet(A)_3_AY196695	4.1152E-18	1.6598E-16	Details
tet(A)_4_AJ517790	3.3007E-17	1.1411E-15	Details
tet(Q)_1_L33696	3.3689E-16	1.0191E-14	Details
tet(A)_5_AJ419171	6.3212E-16	1.6997E-14	Details
tet(Q)_2_X58717	5.0118E-15	1.2129E-13	Details
msr(A)_1_X52085	7.3573E-14	1.4978E-12	Details
dfrK_1_FN377602	7.4273E-14	1.4978E-12	Details
ant(6)-Ib_1_FN594949	8.4707E-14	1.5768E-12	Details
aadA17_1_FJ460181	1.1999E-13	2.0741E-12	Details
tet(O)_2_M20925	2.4222E-13	3.6939E-12	Details
blaTEM-209_1_KF240808	2.4422E-13	3.6939E-12	Details
mph(C)_2_AF167161	5.943E-13	8.4601E-12	Details
sul1_3_EU855787	6.8525E-13	9.2128E-12	Details
blaTEM-208_1_KC865667	1.0103E-12	1.2868E-11	Details
blaTEM-70_1_AF188199	1.1533E-12	1.3955E-11	Details

To go back, you can click on 'Analysis panel'.

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity ▸ Rarefaction Analysis ▸ Beta diversity ▸ Heatmap ▸ Dendrogram ▸ Core Resistome ▸ Correlation ▸ RNAseq Methods ▸ metagenomeSeq

metagenomeSeq: statistical analysis for sparse high-throughput sequencing data

General options:

Profile level: Feature (Rownames)

Experimental factor: Species

Statistical model: zero-inflated Gaussian fit (fitZig)

Adjusted P value cutoff: 0.05

Submit

Result Table

You can choose to download the result table by clicking here.

The table below shows at most 500 features ranked by their P values, with significant ones highlighted in orange

Name	Pvalues	FDR	View
tet(W)_2_AY049983	2.9249E-20	7.0782E-18	Details
tet(40)_1_FJ158002	1.9021E-19	2.3016E-17	Details
tet(Q)_4_Z21523	4.1365E-19	3.3368E-17	Details
tet(O)_3_Y07780	5.6384E-19	3.4112E-17	Details
tet(Q)_3_U73497	3.1023E-18	1.5015E-16	Details
tet(A)_3_AY196695	4.1152E-18	1.6598E-16	Details
tet(A)_4_AJ517790	3.3007E-17	1.1411E-15	Details
tet(Q)_1_L33696	3.3689E-16	1.0191E-14	Details
tet(A)_5_AJ419171	6.3212E-16	1.6997E-14	Details
tet(Q)_2_X58717	5.0118E-15	1.2129E-13	Details
msr(A)_1_X52085	7.3573E-14	1.4978E-12	Details
dfrK_1_FN377602	7.4273E-14	1.4978E-12	Details
ant(6)-Ib_1_FN594949	8.4707E-14	1.5768E-12	Details
aadA17_1_FJ460181	1.1999E-13	2.0741E-12	Details
tet(O)_2_M20925	2.4222E-13	3.6939E-12	Details
blaTEM-209_1_KF240808	2.4422E-13	3.6939E-12	Details
mph(C)_2_AF167161	5.943E-13	8.4601E-12	Details
sul1_3_EU855787	6.8525E-13	9.2128E-12	Details
blaTEM-208_1_KC865667	1.0103E-12	1.2868E-11	Details
blaTEM-70_1_AF188199	1.1533E-12	1.3955E-11	Details

1 2 3 4 5 6 7 8 9 10

How does metagenomeSeq work?

MetagenomeSeq targets to determine variables that are differentially abundant between two or more groups of multiple samples. MetagenomeSeq is designed to address the effects of both normalization and under-sampling of microbial communities on disease association, detection, and the testing of feature correlations.

For more details, please refer to the original paper here metagenomeSeq.

How many statistical models are there in metagenomeSeq and what is the difference between them?

There are two statistical models implemented in metagenomeSeq to model the data:

- fitZig: It is based on a zero-inflated Gaussian mixture model. It can be used when multiple groups are present for differential abundance testing.

- fitFeature: It is based on a zero-inflated Log-Normal mixture model. This approach is recommended by the author. This model currently only supports two-group comparisons.

For more details, please refer to the original paper here metagenomeSeq.

Analysis Panel

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram


Correlation

3. Differential abundance testin

For LEfSe analysis, click here.

RNA-seq methods

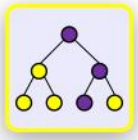
metagenomeSeq

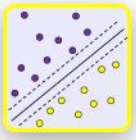
LEfSe

ALDEx2

ANCOM

4. Machine learning (Biomarker prediction)


Random Forest


SVM

Linear Discriminant Analysis (LDA) Effect Size (LEfSe)

General options:

Profile level: Feature (Rownames)

Experimental factor: Species

Adjusted p-value cutoff: 0.05

Log LDA score cutoff: 2.0

Submit

Download

Here you will have the differences when using Linear Discriminant Analysis (LDA) Effect Size (LEfSE)

Result Table | Graphical Summary

The table below shows at most 500 features, with significant ones highlighted in orange

Name	Pvalues	FDR	Pig	Poultry	LDAscore	View
aadA8_1_AF326210	4.0691E-16	2.1611E-14	2.94847	1190.71	2.77	Details
blaTEM-207_1_KC818234	4.8132E-16	2.1611E-14	1.71831	1468.87	2.87	Details
blaTEM-77_1_AF190695	5.532E-16	2.1611E-14	2.08735	918.889	2.66	Details
aadA3_1_AF047479	6.0123E-16	2.1611E-14	2.98593	1745.78	2.94	Details
blaTEM-132_1_AY491682	7.1498E-16	2.1611E-14	0.0	809.713	2.61	Details
blaTEM-29_1_DQ269440	7.1498E-16	2.1611E-14	0.0	806.51	2.61	Details
blaTEM-143_1_DQ075245	7.1498E-16	2.1611E-14	0.0	1126.89	2.75	Details
sul1_9_AY963803	7.3626E-16	2.1611E-14	3.84895	1183.38	2.77	Details
mph(C)_1_AB013298	1.0716E-15	2.1611E-14	0.574171	915.613	2.66	Details
blaTEM-208_1_KC865667	1.0716E-15	2.1611E-14	0.896443	1230.59	2.79	Details
blaTEM-104_1_AF516719	1.0716E-15	2.1611E-14	0.598329	1413.89	2.85	Details
blaTEM-198_1_AB700703	1.0716E-15	2.1611E-14	0.597029	1467.72	2.87	Details
blaTEM-106_1_AY101578	1.1652E-15	2.169E-14	0.968279	812.092	2.61	Details
blaTEM-176_1_GU550123	1.5652E-15	2.5253E-14	1.09762	1300.11	2.81	Details
blaTEM-30_1_AJ437107	1.5652E-15	2.5253E-14	1.12128	1307.73	2.82	Details
blaTEM-76_1_AF190694	1.7003E-15	2.5717E-14	2.36446	1184.51	2.77	Details
blaTEM-122_1_AY307100	1.8468E-15	2.629E-14	1.79657	1114.72	2.75	Details
sul1_3_EU855787	2.1562E-15	2.7006E-14	16.4512	3376.64	3.23	Details
blaTEM-164_1_EU274580	2.2319E-15	2.7006E-14	1.88441	1047.52	2.72	Details
msr(A)_2_AB013298	2.2319E-15	2.7006E-14	1.76562	1109.89	2.74	Details

1 2 3 4 5 6 7 8 9 10

Linear Discriminant Analysis (LDA) Effect Size (LEfSe)

General options:

Profile level: Feature (Rownames)

Experimental factor: Species

Adjusted p-value cutoff: 0.05

Log LDA score cutoff: 2.0

Submit

Downloads

If you click on details, you will be able to see the box plot for each feature.

Result Table | Graphical Summary

The table below shows at most 500 features, with significant ones highlighted in orange

Name	Pvalues	FDR	Pig	Poultry	LDAscore	View
aadA8_1_AF326210	4.0691E-16	2.1611E-14	2.94847	1190.71	2.77	Details
blaTEM-207_1_KC818234	4.8132E-16	2.1611E-14	1.71831	1468.87	2.87	Details
blaTEM-77_1_AF190695	5.532E-16	2.1611E-14	2.08735	918.889	2.66	Details
aadA3_1_AF047479	6.0123E-16	2.1611E-14	2.98593	1745.78	2.94	Details
blaTEM-132_1_AY491682	7.1498E-16	2.1611E-14	0.0	809.713	2.61	Details
blaTEM-29_1_DQ269440	7.1498E-16	2.1611E-14	0.0	806.51	2.61	Details
blaTEM-143_1_DQ075245	7.1498E-16	2.1611E-14	0.0	1126.89	2.75	Details
sul1_9_AY963803	7.3626E-16	2.1611E-14	3.84895	1183.38	2.77	Details
mph(C)_1_AB013298	1.0716E-15	2.1611E-14	0.574171	915.613	2.66	Details
blaTEM-208_1_KC865667	1.0716E-15	2.1611E-14	0.896443	1230.59	2.79	Details
blaTEM-104_1_AF516719	1.0716E-15	2.1611E-14	0.598329	1413.89	2.85	Details
blaTEM-198_1_AB700703	1.0716E-15	2.1611E-14	0.597029	1467.72	2.87	Details
blaTEM-106_1_AY101578	1.1652E-15	2.169E-14	0.968279	812.092	2.61	Details
blaTEM-176_1_GU550123	1.5652E-15	2.5253E-14	1.09762	1300.11	2.81	Details
blaTEM-30_1_AJ437107	1.5652E-15	2.5253E-14	1.12128	1307.73	2.82	Details
blaTEM-76_1_AF190694	1.7003E-15	2.5717E-14	2.36446	1184.51	2.77	Details
blaTEM-122_1_AY307100	1.8468E-15	2.629E-14	1.79657	1114.72	2.75	Details
sul1_3_EU855787	2.1562E-15	2.7006E-14	16.4512	3376.64	3.23	Details
blaTEM-164_1_EU274580	2.2319E-15	2.7006E-14	1.88441	1047.52	2.72	Details
msr(A)_2_AB013298	2.2319E-15	2.7006E-14	1.76562	1109.89	2.74	Details

1 2 3 4 5 6 7 8 9 10

Linear Discriminant Analysis (LDA) Effect Size (LEfSe)

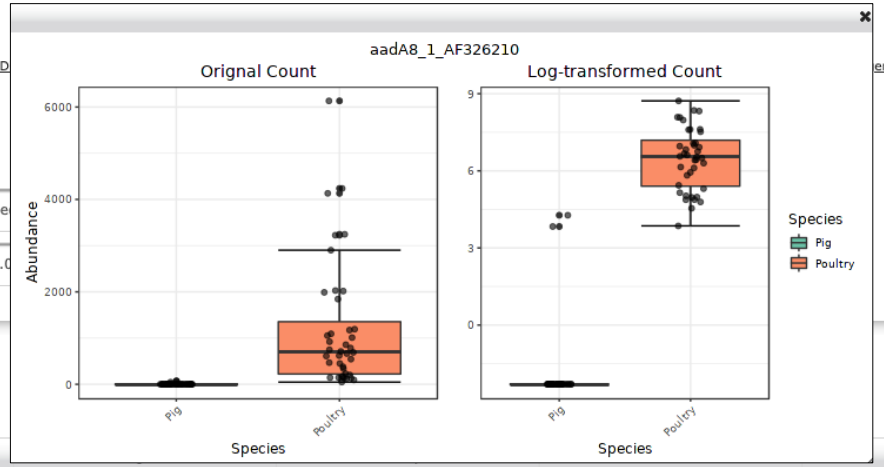
General options:

Profile level: Feature (Rownames)

Experimental factor: Spe

Adjusted p-value cutoff: 0.05

Log LDA score cutoff: 2.0



Result Table | Graphical Summary

The table below shows at most 500 features, with significant ones highlighted in orange

Name	Pvalues	FDR				View
aadA8_1_AF326210	4.0691E-16	2.1611E-14	2.94847	1190.71	2.77	Details
blaTEM-207_1_KC818234	4.8132E-16	2.1611E-14	1.71831	1468.87	2.87	Details
blaTEM-77_1_AF190695	5.532E-16	2.1611E-14	2.08735	918.889	2.66	Details
aadA3_1_AF047479	6.0123E-16	2.1611E-14	2.98593	1745.78	2.94	Details
blaTEM-132_1_AY491682	7.1498E-16	2.1611E-14	0.0	809.713	2.61	Details
blaTEM-29_1_DQ269440	7.1498E-16	2.1611E-14	0.0	806.51	2.61	Details
blaTEM-143_1_DQ075245	7.1498E-16	2.1611E-14	0.0	1126.89	2.75	Details
sul1_9_AY963803	7.3626E-16	2.1611E-14	3.84895	1183.38	2.77	Details
mph(C)_1_AB013298	1.0716E-15	2.1611E-14	0.574171	915.613	2.66	Details
blaTEM-208_1_KC865667	1.0716E-15	2.1611E-14	0.896443	1230.59	2.79	Details
blaTEM-104_1_AF516719	1.0716E-15	2.1611E-14	0.598329	1413.89	2.85	Details
blaTEM-198_1_AB700703	1.0716E-15	2.1611E-14	0.597029	1467.72	2.87	Details
blaTEM-106_1_AY101578	1.1652E-15	2.169E-14	0.968279	812.092	2.61	Details
blaTEM-176_1_GU550123	1.5652E-15	2.5253E-14	1.09762	1300.11	2.81	Details
blaTEM-30_1_AJ437107	1.5652E-15	2.5253E-14	1.12128	1307.73	2.82	Details
blaTEM-76_1_AF190694	1.7003E-15	2.5717E-14	2.36446	1184.51	2.77	Details
blaTEM-122_1_AY307100	1.8468E-15	2.629E-14	1.79657	1114.72	2.75	Details
sul1_3_EU855787	2.1562E-15	2.7006E-14	16.4512	3376.64	3.23	Details
blaTEM-164_1_EU274580	2.2319E-15	2.7006E-14	1.88441	1047.52	2.72	Details
msr(A)_2_AB013298	2.2319E-15	2.7006E-14	1.76562	1109.89	2.74	Details

Linear Discriminant Analysis (LDA) Effect Size (LEfSe)

General options:

Profile level: Feature (Rownames)

Experimental factor: Species

Adjusted p-value cutoff: 0.05

Log LDA score cutoff: 2.0

Submit

Downloads

If you click on 'Graphical Summary', you will find potential biomarkers identified by LEfSE.

Result Table | Graphical Summary

The table below shows at most 500 features, with significant ones highlighted in orange

Name	Pvalues	FDR	Pig	Poultry	LDAscore	View
aadA8_1_AF326210	4.0691E-16	2.1611E-14	2.94847	1190.71	2.77	Details
blaTEM-207_1_KC818234	4.8132E-16	2.1611E-14	1.71831	1468.87	2.87	Details
blaTEM-77_1_AF190695	5.532E-16	2.1611E-14	2.08735	918.889	2.66	Details
aadA3_1_AF047479	6.0123E-16	2.1611E-14	2.98593	1745.78	2.94	Details
blaTEM-132_1_AY491682	7.1498E-16	2.1611E-14	0.0	809.713	2.61	Details
blaTEM-29_1_DQ269440	7.1498E-16	2.1611E-14	0.0	806.51	2.61	Details
blaTEM-143_1_DQ075245	7.1498E-16	2.1611E-14	0.0	1126.89	2.75	Details
sul1_9_AY963803	7.3626E-16	2.1611E-14	3.84895	1183.38	2.77	Details
mph(C)_1_AB013298	1.0716E-15	2.1611E-14	0.574171	915.613	2.66	Details
blaTEM-208_1_KC865667	1.0716E-15	2.1611E-14	0.896443	1230.59	2.79	Details
blaTEM-104_1_AF516719	1.0716E-15	2.1611E-14	0.598329	1413.89	2.85	Details
blaTEM-198_1_AB700703	1.0716E-15	2.1611E-14	0.597029	1467.72	2.87	Details
blaTEM-106_1_AY101578	1.1652E-15	2.169E-14	0.968279	812.092	2.61	Details
blaTEM-176_1_GU550123	1.5652E-15	2.5253E-14	1.09762	1300.11	2.81	Details
blaTEM-30_1_AJ437107	1.5652E-15	2.5253E-14	1.12128	1307.73	2.82	Details
blaTEM-76_1_AF190694	1.7003E-15	2.5717E-14	2.36446	1184.51	2.77	Details
blaTEM-122_1_AY307100	1.8468E-15	2.629E-14	1.79657	1114.72	2.75	Details
sul1_3_EU855787	2.1562E-15	2.7006E-14	16.4512	3376.64	3.23	Details
blaTEM-164_1_EU274580	2.2319E-15	2.7006E-14	1.88441	1047.52	2.72	Details
msr(A)_2_AB013298	2.2319E-15	2.7006E-14	1.76562	1109.89	2.74	Details

To go back, you can click on 'Analysis panel'.

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity ▸ Rarefaction Analysis ▸ Beta diversity ▸ Heatmap ▸ Dendrogram ▸ Core Resistome ▸ Correlation ▸ RNAseq Methods ▸ metagenomeSeq ▸ LEfSe

Linear Discriminant Analysis (LDA) Effect Size (LEfSe)

General options:

Profile level: Feature (Rownames)

Experimental factor: Species

Adjusted p-value cutoff: 0.05

Log LDA score cutoff: 2.0

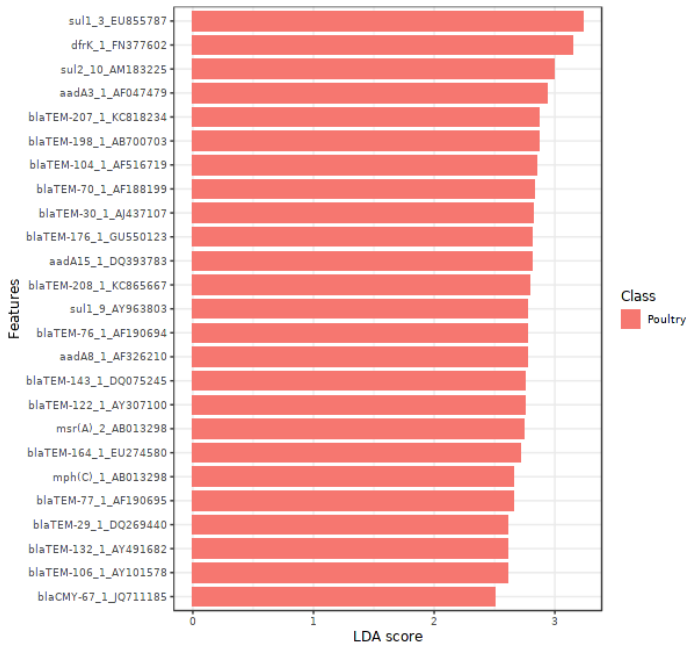
Submit

Downloads

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

Result Table | Graphical Summary

The bar graph showing the LDA scores of (at most) top 25 significant features



How does LDA Effect Size (LEfSe) algorithm work?

The linear discriminant analysis (LDA) effect size (LEfSe) is a method to support high-dimensional class comparisons with a particular focus on metagenomic analyses. LEfSe determines the features (organisms, clades, operational taxonomic units, genes, or functions) most likely to explain differences between classes by coupling standard tests for statistical significance with additional tests encoding biological consistency and effect relevance. Firstly, it uses the non-parametric factorial Kruskal-Wallis (KW) sum-rank test to detect features with significant differential abundance with respect to the class of interest; biological consistency is subsequently investigated using a set of pairwise tests among subclasses using the (unpaired) Wilcoxon rank-sum test. As a last step, LEfSe uses LDA to estimate the effect size of each differentially abundant feature and, if desired by the investigator, to perform dimension reduction.

For more details, please refer to the original paper here: LEfSe.

Analysis Panel

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

RNA-seq methods

metagenomeSeq

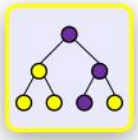
LEfSe

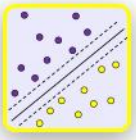
ALDEx2

ANCOM

Let’s move to ALDEx2.

4. Machine learning (Biomarker prediction)


Random Forest


SVM

ALDEx2: ANOVA-Like differential expression tool for high throughput sequencing data

General options:

Profile level: Feature (Rownames)

Experimental factor: Species

No. of Monte Carlo samp

Adjusted P value cutoff: 0.05

Show significant features based on: Welch's t test

ALDEx2 aims on identifying differentially abundant features in compositional (relative) high-throughput sequencing data. This method is a recommended compositional data analysis (CoDA) approach and it estimates per-feature technical variation within each sample using Monte-Carlo instances drawn from Dirichlet distribution.

The table below shows at most 500 features ranked by their P values, with significant ones highlighted in orange

Name	Pvalues	adjPvalues	View
ant(6)-Ib_1_FN594949	1.86034E-23	3.45279E-22	Details
tet(40)_1_FJ158002	1.86034E-23	3.45279E-22	Details
tet(44)_1_NZ_ABDU01000081	1.86034E-23	3.45279E-22	Details
tet(44)_2_FN594949	1.86034E-23	3.45279E-22	Details
tet(O)_1_M18896	1.86034E-23	3.45279E-22	Details
tet(O)_2_M20925	1.86034E-23	3.45279E-22	Details
tet(O)_3_Y07780	1.86034E-23	3.45279E-22	Details
tet(Q)_1_L33696	1.86034E-23	3.45279E-22	Details
tet(Q)_2_X58717	1.86034E-23	3.45279E-22	Details
tet(Q)_3_U73497	1.86034E-23	3.45279E-22	Details
tet(Q)_4_Z21523	1.86034E-23	3.45279E-22	Details
tet(W)_2_AY049983	1.86034E-23	3.45279E-22	Details
erm(F)_3_M17808	2.90678E-23	5.16965E-22	Details
erm(Q)_1_L22689	3.2149E-22	4.63751E-21	Details
tetA(P)_1_AB054980	1.69087E-21	2.38559E-20	Details
ant(6)-Ia_1_AF330699	2.38094E-21	3.17111E-20	Details
aph(3)-III_1_M26832	3.33844E-21	4.54894E-20	Details
tet(W)_1_DQ060146	6.25539E-21	8.27671E-20	Details
tet(X)_2_M37699	2.77781E-20	3.47532E-19	Details
erm(G)_1_M15332	1.71264E-18	1.69588E-17	Details

ALDEx2: ANOVA-Like differential expression tool for high throughput sequencing data

General options:

Profile level: Feature (Rownames) | Experimental factor: Species | No. of Monte Carlo samples: 64

Adjusted P value cutoff: 0.05 | Show significant features based on: Welch's t test

Submit | Result Table

ALDEx2 identifies differential features based on either Wilcoxon Rank Sum test and Welch's t-test (2 groups) or Kruskal-Wallis test and a generalized linear model (multiple groups).

The table below shows at most 500 features ranked by their P values, with significant ones high

Name			
ant(6)-Ib_1_FN594949	1.86034E-23		
tet(40)_1_FJ158002	1.86034E-23		
tet(44)_1_NZ_ABDU01000081	1.86034E-23	3.45279E-22	Details
tet(44)_2_FN594949	1.86034E-23	3.45279E-22	Details
tet(O)_1_M18896	1.86034E-23	3.45279E-22	Details
tet(O)_2_M20925	1.86034E-23	3.45279E-22	Details
tet(O)_3_Y07780	1.86034E-23	3.45279E-22	Details
tet(Q)_1_L33696	1.86034E-23	3.45279E-22	Details
tet(Q)_2_X58717	1.86034E-23	3.45279E-22	Details
tet(Q)_3_U73497	1.86034E-23	3.45279E-22	Details
tet(Q)_4_Z21523	1.86034E-23	3.45279E-22	Details
tet(W)_2_AY049983	1.86034E-23	3.45279E-22	Details
erm(F)_3_M17808	2.90678E-23	5.16965E-22	Details
erm(Q)_1_L22689	3.2149E-22	4.63751E-21	Details
tetA(P)_1_AB054980	1.69087E-21	2.38559E-20	Details
ant(6)-Ia_1_AF330699	2.38094E-21	3.17111E-20	Details
aph(3)-III_1_M26832	3.33844E-21	4.54894E-20	Details
tet(W)_1_DQ060146	6.25539E-21	8.27671E-20	Details
tet(X)_2_M37699	2.77781E-20	3.47532E-19	Details
erm(G)_1_M15332	1.71264E-18	1.69588E-17	Details

To go back, you can click on 'Analysis panel'.

ALDEx2: ANOVA-Like differential expression tool for high throughput sequencing data

General options:

Profile level: Feature (Rownames)

Experimental factor: Species

No. of Monte Carlo samples: 64

Adjusted P value cutoff: 0.05

Show significant features based on: Welch's t test

Submit

Result Table

You can choose to download the result table by clicking here.

The table below shows at most 500 features ranked by their P values, with significant ones highlighted in orange

Name	Pvalues	adjPvalues	View
ant(6)-Ib_1_FN594949	1.86034E-23	3.45279E-22	Details
tet(40)_1_FJ158002	1.86034E-23	3.45279E-22	Details
tet(44)_1_NZ_ABDU01000081	1.86034E-23	3.45279E-22	Details
tet(44)_2_FN594949	1.86034E-23	3.45279E-22	Details
tet(O)_1_M18896	1.86034E-23	3.45279E-22	Details
tet(O)_2_M20925	1.86034E-23	3.45279E-22	Details
tet(O)_3_Y07780	1.86034E-23	3.45279E-22	Details
tet(Q)_1_L33696	1.86034E-23	3.45279E-22	Details
tet(Q)_2_X58717	1.86034E-23	3.45279E-22	Details
tet(Q)_3_U73497	1.86034E-23	3.45279E-22	Details
tet(Q)_4_Z21523	1.86034E-23	3.45279E-22	Details
tet(W)_2_AY049983	1.86034E-23	3.45279E-22	Details
erm(F)_3_M17808	2.90678E-23	5.16965E-22	Details
erm(Q)_1_L22689	3.2149E-22	4.63751E-21	Details
tetA(P)_1_AB054980	1.69087E-21	2.38559E-20	Details
ant(6)-Ia_1_AF330699	2.38094E-21	3.17111E-20	Details
aph(3)-III_1_M26832	3.33844E-21	4.54894E-20	Details
tet(W)_1_DQ060146	6.25539E-21	8.27671E-20	Details
tet(X)_2_M37699	2.77781E-20	3.47532E-19	Details
erm(G)_1_M15332	1.71264E-18	1.69588E-17	Details

How does ALDEx2 work?

ALDEx2 is a package to work with differential abundance analysis for the comparison of two or more conditions. It utilizes a Dirichlet-multinomial model that uses counts to infer abundance. This is optimized for three or more experimental replicates.

The method infers biological and sampling variation to calculate the expected false discovery rate, given the variation, based on a Wilcoxon Rank Sum test and Welch's t-test (for two groups), and a Kruskal-Wallis test, a generalized linear model, or a correlation test (for more than two groups).

All tests report p-values and Benjamini-Hochberg corrected p-values.

For more details, please refer to the original paper here ALDEx2.

Analysis Panel

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

RNA-seq methods

metagenomeSeq

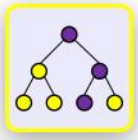
LEfSe

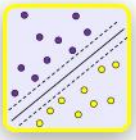
ALDEx2

ANCOM

For ANCOM analysis, click here.

4. Machine learning (Biomarker prediction)


Random Forest


SVM

Analysis of Compositions of Resistomes using ANCOM

General options: Profile level: Feature (Rownames) Experimental factor: Sp

ANCOM is a compositional data analysis approach to compare the composition of microbes in two or more groups. Here, we are using ANCOM approach to identify differentially abundant features in metagenomic (i.e., resistome) data. ANCOM is a compositional data analysis (CoDA) recommended approach.

The table below shows at most 500 features, with significant ones highlighted in orange

Name	W	View
tet(Q)_4_Z21523	240.0	Details
tet(Q)_1_L33696	239.0	Details
tet(Q)_2_X58717	239.0	Details
tet(A)_4_AJ517790	238.0	Details
tet(Q)_3_U73497	238.0	Details
tet(40)_1_FJ158002	237.0	Details
tet(W)_2_AY049983	237.0	Details
erm(F)_3_M17808	235.0	Details
tet(A)_3_AY196695	235.0	Details
tet(A)_5_AJ419171	235.0	Details
tet(O)_3_Y07780	235.0	Details
ant(6)-Ib_1_FN594949	234.0	Details
sul1_3_EU855787	232.0	Details
ant(6)-Ia_1_AF330699	231.0	Details
catQ_1_M55620	230.0	Details
dfrK_1_FN377602	230.0	Details
cmx_1_U85507	229.0	Details
tet(O)_2_M20925	229.0	Details
dfrD_1_Z50141	228.0	Details
erm(Q)_1_L22689	228.0	Details

To go back, you can click on 'Analysis panel'.

Analysis of Compositions of Resistomes using ANCOM

General options: Profile level: Feature (Rownames) Experimental factor: Species Adjusted P value cutoff: 0.05 Submit Result Table

You can choose to download the result table by clicking here.

The table below shows at most 500 features, with significant ones highlighted in orange

Name	W	View
tet(Q)_4_Z21523	240.0	Details
tet(Q)_1_L33696	239.0	Details
tet(Q)_2_X58717	239.0	Details
tet(A)_4_AJ517790	238.0	Details
tet(Q)_3_U73497	238.0	Details
tet(40)_1_FJ158002	237.0	Details
tet(W)_2_AY049983	237.0	Details
erm(F)_3_M17808	235.0	Details
tet(A)_3_AY196695	235.0	Details
tet(A)_5_AJ419171	235.0	Details
tet(O)_3_Y07780	235.0	Details
ant(6)-Ib_1_FN594949	234.0	Details
sul1_3_EU855787	232.0	Details
ant(6)-Ia_1_AF330699	231.0	Details
catQ_1_M55620	230.0	Details
dfrK_1_FN377602	230.0	Details
cmx_1_U85507	229.0	Details
tet(O)_2_M20925	229.0	Details
dfrD_1_Z50141	228.0	Details
erm(Q)_1_L22689	228.0	Details

How does ANCOM work?

ANCOM (analysis of composition of microbiomes) is used for detecting differences in microbial mean taxa abundance. In ResistoXplorer, it has been adapted to analyze the composition of resistomes. The methodology tests the log-ratio abundance of all pairs of features for differences in means using nonparametric statistical tests. The number of significant results involving each feature is used to calculate its significance.

For more details, please refer to the original paper here ANCOM.

Analysis Panel

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

RNA-seq methods

metagenomeSeq

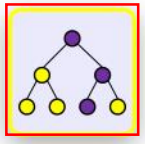
LEfSe

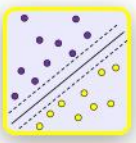
ALDEx2

ANCOM

4. Machine learning (Biomarker prediction)

Let's move to Random Forest. Click here.


Random Forest


SVM

Random Forest

General options:

Profile level: Feature (Rownames)

Experimental factor: Species

Number of trees to grow: 500

Number of predictors to try: 7

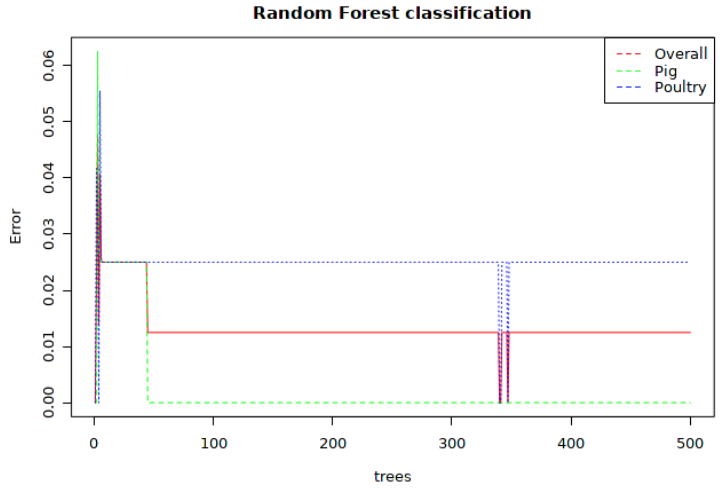
Randomness setting: Use a constant (123456)

Submit

Downloads

Classification Performance | Important Features

Click on 'Important features'.



Random Forests is a powerful machine learning algorithm for classification and identification of predictive features (biomarkers). It operates by constructing a multitude of decision trees (forests) at training time and predicting the class as the majority vote of the individual trees.

Random Forest

General options:

Profile level: Feature (Rownames)

Experimental factor: Species

Number of trees to grow: 500

Number of predictors to try: 7

Randomness setting: Use a constant (123456)

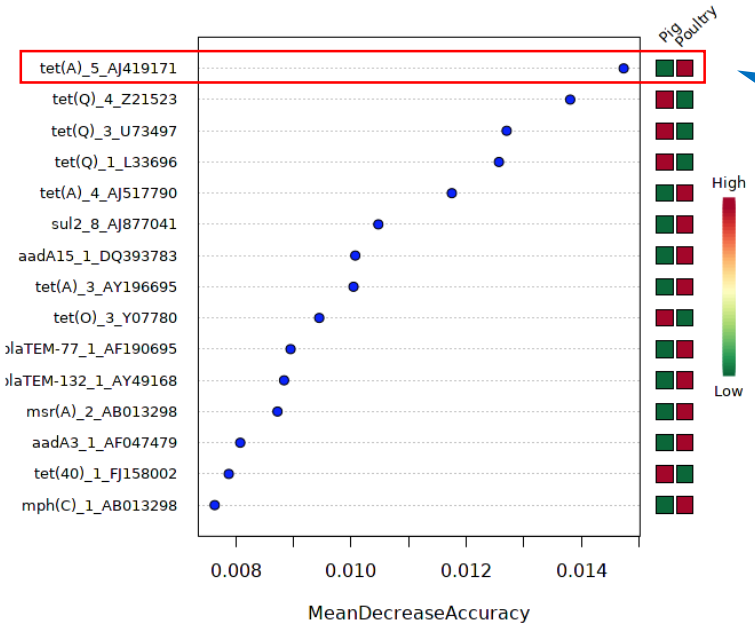
Submit

Downloads

Importantly, Random Forest generates random trees. So if you want to reproduce the same tree, you have to select to 'use a constant' here in this option.

Classification Performance | Important Features

Please Note: Features are ranked by their contributions to classification accuracy (Mean Decrease Accuracy) (only top 15 will be shown)



Now you can observe which antibiotic resistance genes presented a strong predictive value for a specific group. In this case, the gene *tetA* was predictive of the poultry samples as opposed to pig.

To go back, you can click on 'Analysis panel'.

Data Upload ▸ Data Inspection ▸ Data Filter ▸ Normalization ▸ Analysis Panel ▸ Composition ▸ Hierarchical Composition ▸ Alpha Diversity ▸ Rarefaction Analysis ▸ Beta diversity ▸ Heatmap ▸ Dendrogram ▸ Core Resistome ▸ Correlation ▸ RNAseq Methods ▸ metagenomeSeq ▸ LEfSe ▸ ALDEx2 ▸ ANCOM ▸ Random Forest

Random Forest

General options:

Profile level: Feature (Rownames)

Experimental factor: Species

Number of trees to grow: 500

Number of predictors to try: 7

Randomness setting: Use a constant (123456)

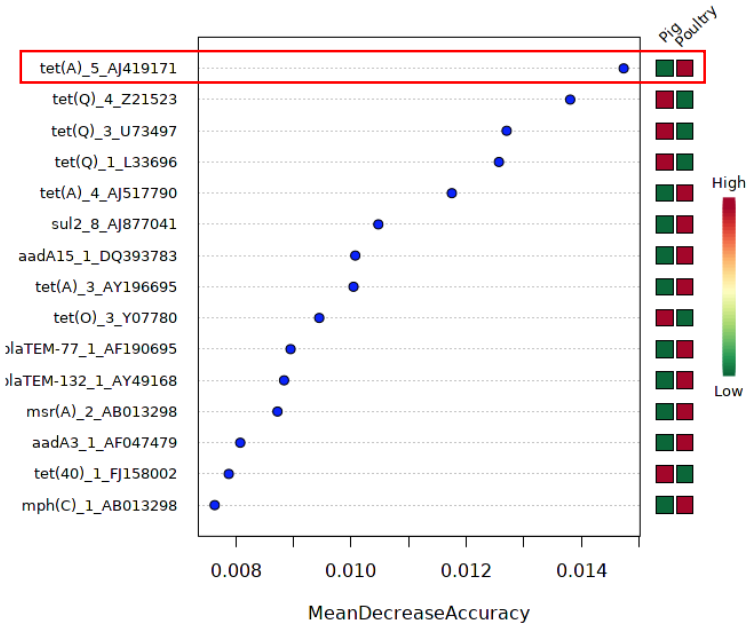
Submit

Downloads

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

Classification Performance | Important Features

Please Note: Features are ranked by their contributions to classification accuracy (Mean Decrease Accuracy) (only top 15 will be shown)



How do Random Forests work?

The Random Forest algorithm uses an ensemble of classification trees (forest) where each tree is grown based on a random subset of features from a bootstrap sample at each branch. The final class prediction is based on the majority vote of the ensemble.

The unbiased estimate of classification errors is obtained by aggregating cross-validation results using bootstrapped samples while the forest is being constructed.

Random forest also measure the importance of each feature based on the increase of the error when it is randomly re-shuffled. It can indicate which groups are easier to predict based on errors.

Analysis Panel

1. Composition profiling


Visual profiling


Hierarchical

Alpha diversity


Rarefaction curves


Ordination analysis

2. Clustering analysis


Heatmap


Dendrogram

Core Resistome


Correlation

3. Differential abundance testing

RNA-seq methods

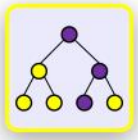
metagenomeSeq

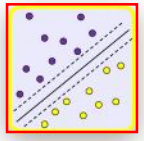
LEfSe

ALDEx2

ANCOM

4. Machine learning (Biomarker prediction)


Random Forest


SVM

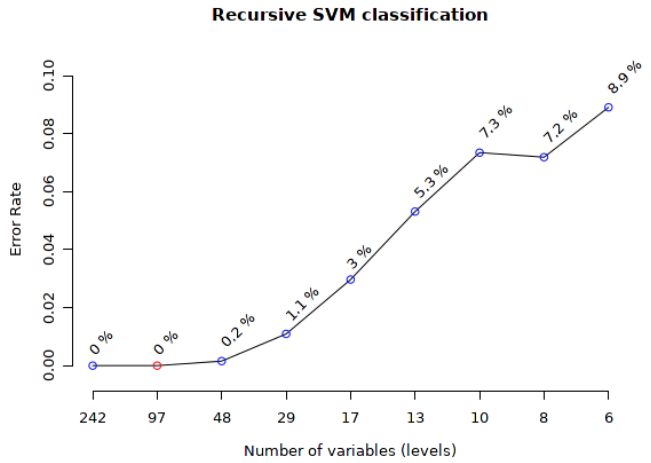
For SVM, click here.

Support vector machine (SVM)

General options: Profile level: Feature (Rownames) Experimental factor: Species Validation method: 10-fold CV Submit Downloads

Classification Performance | Important Features

Click on 'Important features'.



Supper vector machine performs classification recursively using different feature subsets. Features are selected based on their relative contribution in the classification using cross validation error rates. The least important features are eliminated in the subsequent steps. This process creates a series of SVM models (levels). The features used by the best model are then plotted.

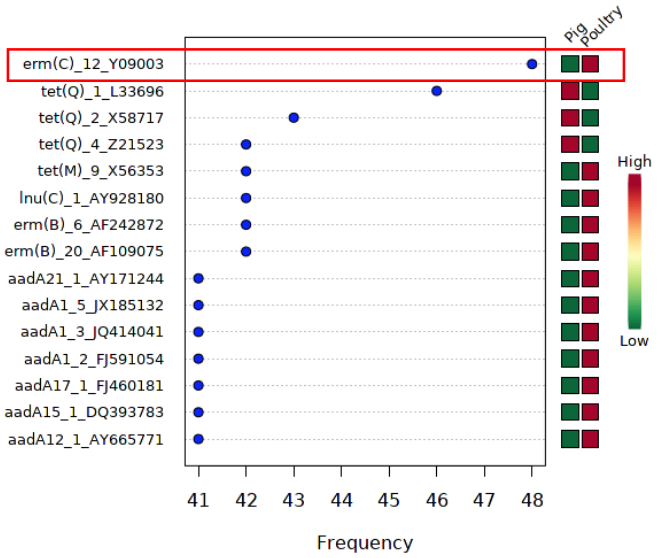
Support vector machine (SVM)

General options: Profile level: Feature (Rownames) Experimental factor: Species Validation method: 10-fold CV Submit Downloads

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

Classification Performance | Important Features

Please Note: Features are ranked by their frequencies being selected in the best classifiers (only top 15 will be shown)



Here, we observe that gene *ermC* was predictive of poultry samples as opposed to pig.

THIS MANUAL IS FINISHED.

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

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.