

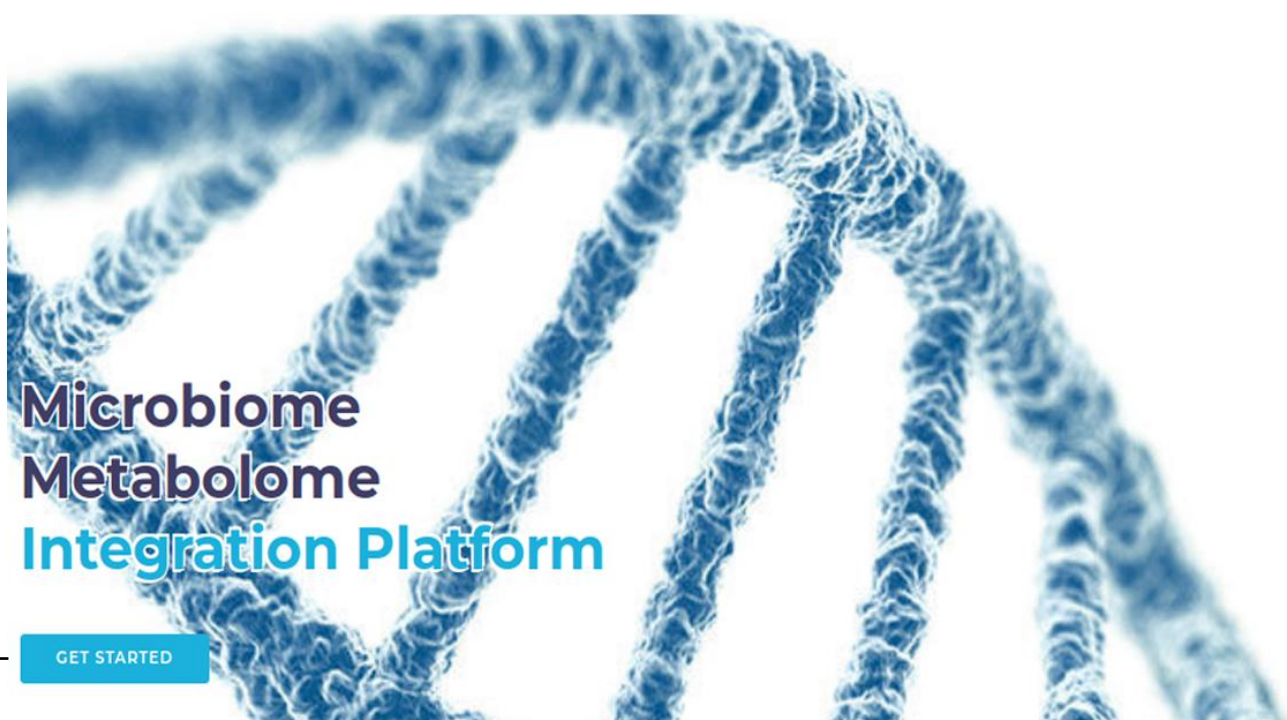


## WORKING MANUAL OF MMIP

### Notes:

1.  Red arrow and number box indicates additional functionality.
2.  Black dotted arrow indicates feature information.
3. All the comparison and the analyses are between two groups (not more).
4. By the term “significant”, statistical significance is referred to.
5. All the analysis diagrams are made using Plotly and will feature the following tools (as shown below) on the top right corner of each diagram, allowing the user to access features like zoom, edit, download and other interesting tools.





# Microbiome Metabolome Integration Platform

GET STARTED

## Overview

Welcome to MMIP Server.

This server is designed for prediction of the metabolites that could be produced by the microbial community during different health condition by using their metagenomics data (16S-rRNA). Furthermore, it also helps you to compare your real-time metabolomic (untargeted) data with predicted metabolite and helps to see what could be the most probable source of the metabolites.

- ✔ Helps in predicting metabolites from the metagenomic data.
- ✔ Helps in finding the most probable source of the metabolite (microbe responsible for the production).
- ✔ Predict important feature at OTU, Compound, Enzyme level by machine learning approaches.
- ✔ Helps in interlinking important feature.

**You can access preprint here:** Anupam Gautam, Debaleena Bhowmik, Sayantani Basu, Abhishake Lahiri, Wenhuan Zeng, Sandip Paul. Microbiome Metabolome Integration Platform (MMIP): a web-based platform for microbiome and metabolome data integration and feature identification. *bioRxiv*, 2023; doi: <https://doi.org/10.1101/2023.04.04.535534>

## Services

Please Select any one of the two module to start your analysis.



### Module-I

Perform general prediction of compounds from metagenomic data (16S-rRNA)



### Module-II

Allow user to perform comparative analysis between real-time metabolic data and one being predicted

Click on either of the tabs to go to the respective "upload" pages.

Click to use the respective services.

Scroll

(Fig. 1)

## UPLOAD PAGES

### Module-I:

Returns to the homepage.

Details of different releases

# MMIP

HOME

MANUAL

RELEASE-NOTES

CONTACT US

Note: please read the [manual](#) before use or look at example datasets (works best with SVM model; dataset2 is taken from qiime2 tutorial) for appropriate input formats. Several new features have been implemented (check [Release-Notes](#)). Please notify us if you encounter any bugs or issues.

#### Name

Enter name

#### Email

Enter Email

#### Institute Name

Enter Institute Name

Download example dataset for Greengenes  Download example dataset for 'Other databases' 

Select OTU Table:  No file chosen

Select Metadata:  No file chosen

Select Column for Grouping:  Select Group:

Note: To select 'Other Database', kindly click "Yes" below and provide the corresponding files.

Select Database:  18may2012  13\_5  13\_8  Other Database

Do you want to provide below mentioned files:  Yes  No

Select representative sequences:  No file chosen

Or Paste Link:

Select taxonomy file:  No file chosen

Or Paste Link:

Enter abundance value to filter biom table (1%=0.01,0.25%=0.0025):

Select Model:  SVM  RandomForest  DecisionTree

Submit Job

Example Result

→ Select machine-learning model

→ Add the taxonomy file (**.txt or.tsv format**)

→ Add rep-set file (**.fasta format**)

→ Add the Metadata file (**.txt format**)

→ Add the OTU/ASV table (**.biom format**)

Add abundance filtering value to filter OTU/ASV table

Select in-case of databases other than Greengenes

(In order to select "Other database", kindly click "Yes" below and provide the corresponding files.)

Select Greengenes database version used

Note: Please maintain the file formats with extensions.

(Fig. 2)

## Module-II:

# MMIP



[HOME](#) [MANUAL](#) [RELEASE-NOTES](#) [CONTACT US](#)

Note: please read the [manual](#) before use or look at example datasets (works best with SVM model; dataset2 is taken from qiime2 tutorial) for appropriate input formats. Several new features have been implemented (check [Release-Notes](#)). Please notify us if you encounter any bugs or issues.

### Name

### Email


### Institute Name

Download example dataset for Greengenes  Download example dataset for 'Other databases' 

Select OTU Table:  No file chosen

Select Metadata:  No file chosen

Select Column for Grouping:  Select Group:

Select Metabolomics Feature table (optional):  No file chosen [Download example file](#) 

Note: To select 'Other Database', kindly click "Yes" below and provide the corresponding files.

Select Database:  18may2012  13\_5  13\_8  Other Database

Do you want to provide below mentioned files:  Yes  No

Select representative sequences:  No file chosen Or Paste Link:

Select taxonomy file:  No file chosen Or Paste Link:

Enter abundance value to filter biom table (1%=0.01, 0.25%=0.0025):

Select Model:  SVM  RandomForest  DecisionTree

→ Add metabolite table from metabolomic processing (.txt format)

Note: Please maintain the file format with extension.

(Fig. 3)

## RESULT PAGES

### USER AND JOB RELATED INFORMATION:

To search keywords or values within the table.

Each of the tabs takes the user to the respective result sections.

MMIP

JOB METABOLITES ALPHA-DIVERSITY TAXONOMY BETA-DIVERSITY PATHWAY FEATURE CONTACT US

Job Summary

Show 10 entries Search:

S.NO	Content	Description
1	Job Id	215
2	User Name	Name of User
3	Biom File Name	otu_table_20000.biom
4	Meta Data File Name	mapping_file.txt
5	Grouping Column	Description
6	Total Number of Sample Group1 (CRC)	20

Showing 1 to 10 of 12 entries Previous 1 2 Next

(Fig. 4)

1 Scroll the table for additional information including number of samples in group 2, Institute name, email id, start date, time and status of the job.

## METABOLITES:

Add range of desired p values.

Click on the compound ids to get to the KEGG page of the respective compound.

Minimum p-value:

Maximum p-value:

Show  entries

Search:

VISUALIZE ALL COMPOUND 2

S.NO	KEGG Compound ID	p-value	Box Plot
1	C00064	0.032828	
2	C00117	0.035297	
3	C00120	0.014059	
4	C00167	0.046764	
5	C00185	0.003461	
6	C00194	0.024378	

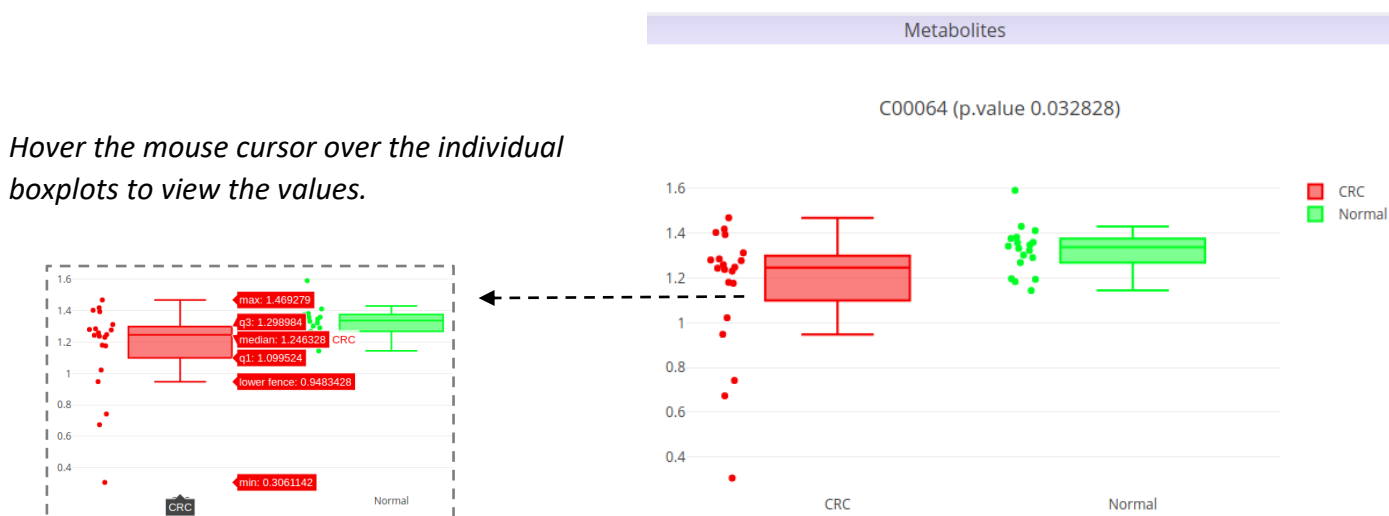
S.NO	File Name	Download
1	Statistically Significant Compound Detail (kruskal-walis test)	
2	Statistically Significant Compound Detail (Wilcox test)	
3	Statistically Significant Compound Detail (Anova-1-way test)	
4	Statistically Significant Compound Detail (t-test)	
5	All Compound CMP Score	

Showing 1 to 5 of 5 entries

3

4

Click to get the boxplot of the compound in the same page (above the table).



(Fig. 5)

2 Click the tab to view all the compounds and more. The following page will open with the significant compounds. The boxplot shows the median values and differences between the two groups.

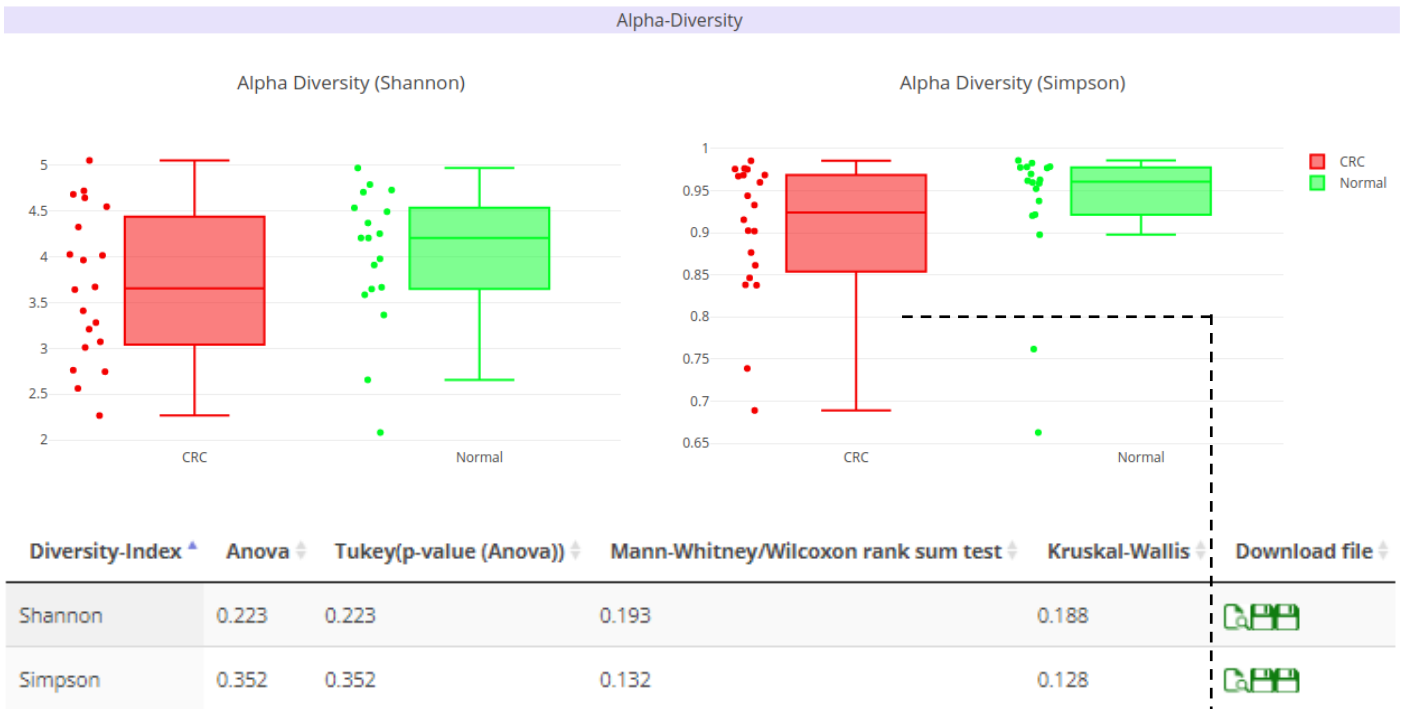


(Fig. 6)

3 Click to download the data table (tsv format).

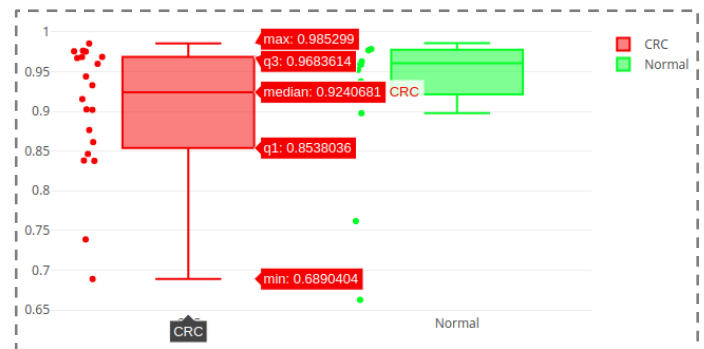
4 Click to open the data table in a new webpage (tsv format).

# ALPHA DIVERSITY:



Showing 1 to 2 of 2 entries

*Hover the mouse cursor over the individual boxplots to view the values.*



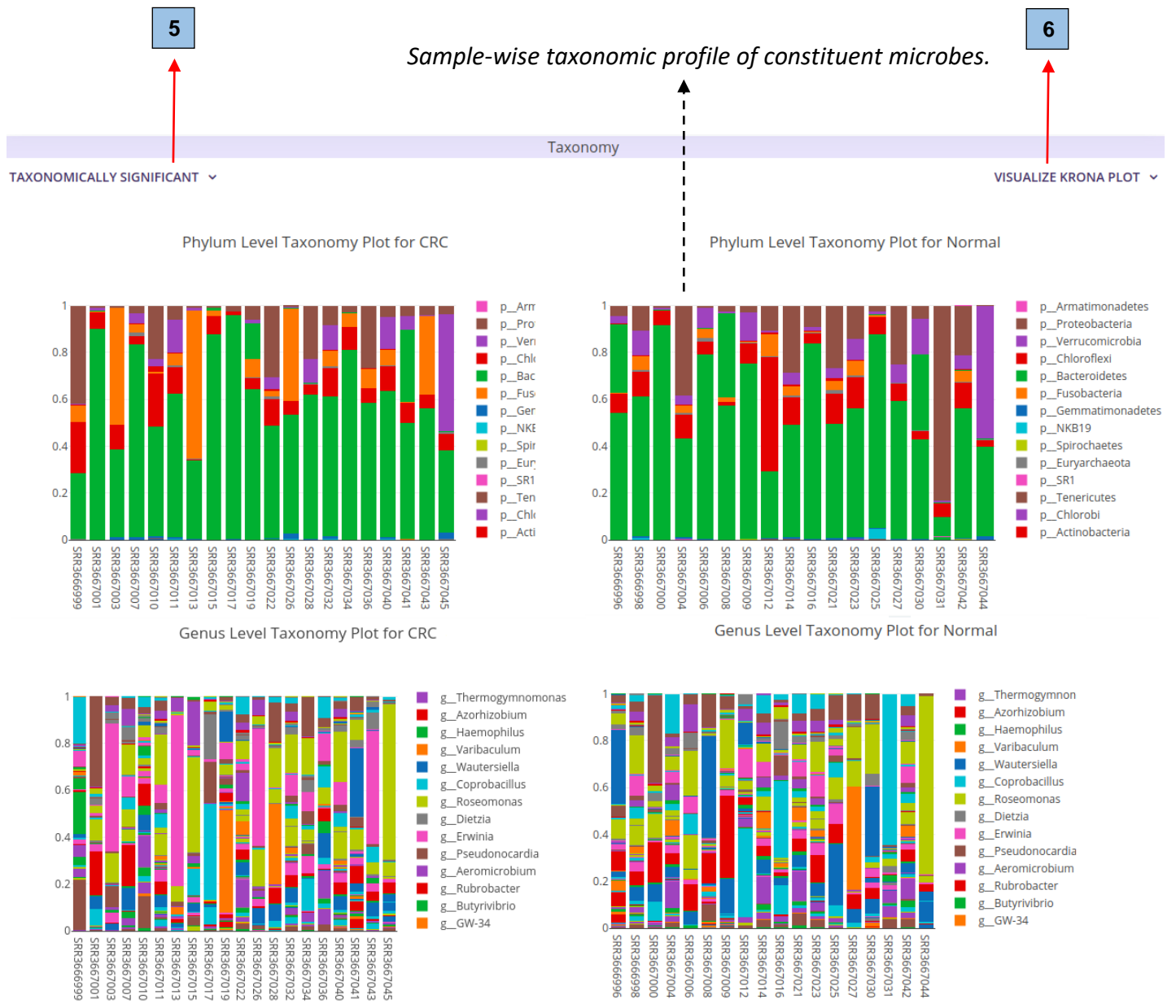
Note: **EFFECTIVE NUMBER** added.  
(Click to go to a new page and view.)



(Fig. 7)



# TAXONOMIC DIVERSITY:

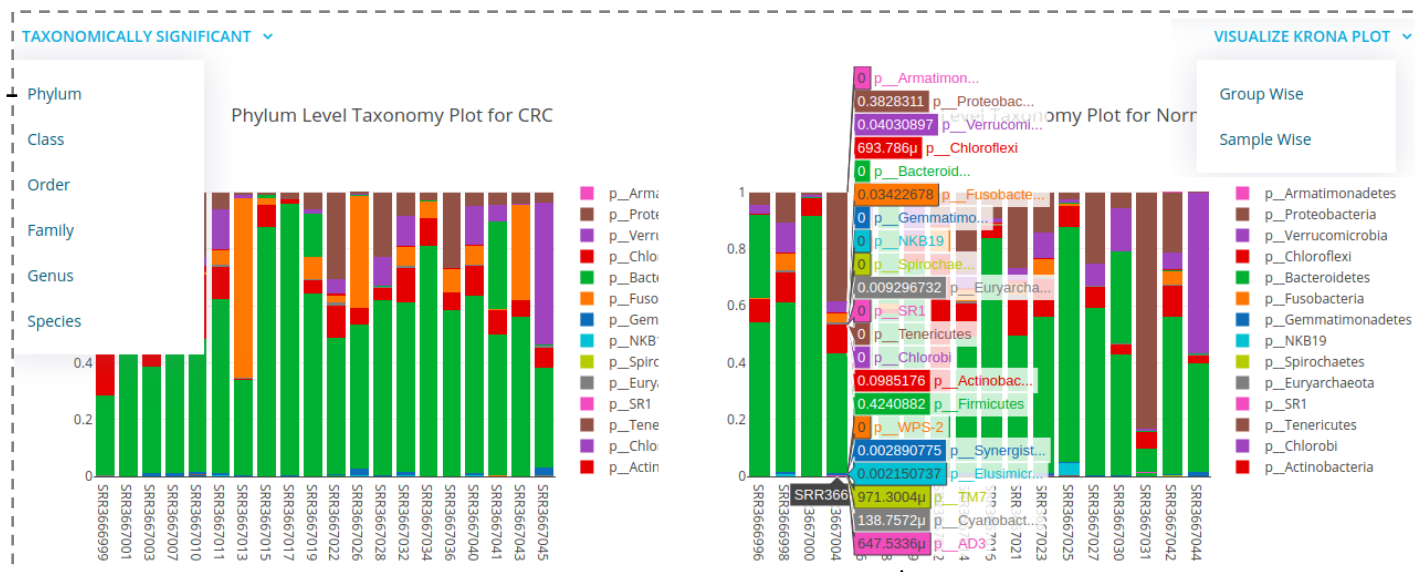


S.No	File Description	Download File
1	Phylum Level Taxonomy Data Group1	
2	Phylum Level Taxonomy Data Group2	
3	Genus Level Taxonomy Data Group1	
4	Genus Level Taxonomy Data Group2	
5	Statistically Significant OTUs Group1	
6	Statistically Significant OTUs Group2	

Showing 1 to 6 of 6 entries

(Fig. 8)

5 Click on this tab to find a drop-down list (shown below), from which user can select the desired taxonomic level to check significant ones and is visualized in a new webpage, also depicted below.



(Fig. 9)

Click on one of the taxonomic level to go to the following page showing the taxa that are significant.

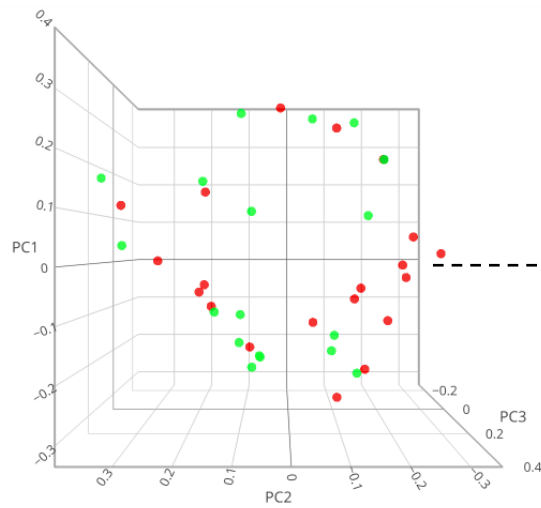
Hover the mouse cursor over the individual samples to view the different taxa and their relative abundance.



(Fig. 10)



**BETA DIVERSITY:**

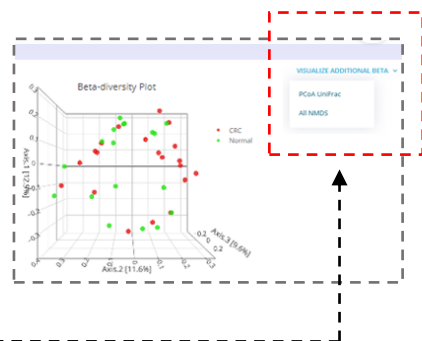
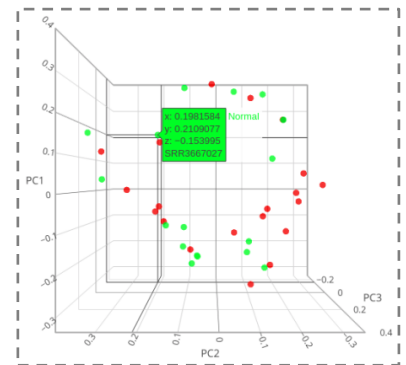


	F.Model	p-value
Bray-Curtis (Adonis test)	0.760695687284777	0.797

Showing 1 to 1 of 1 entries

File Name	Download
Beta-Diversity p-value detail (adonis)	
Beta-Diversity Plot Data & Distance Matix	

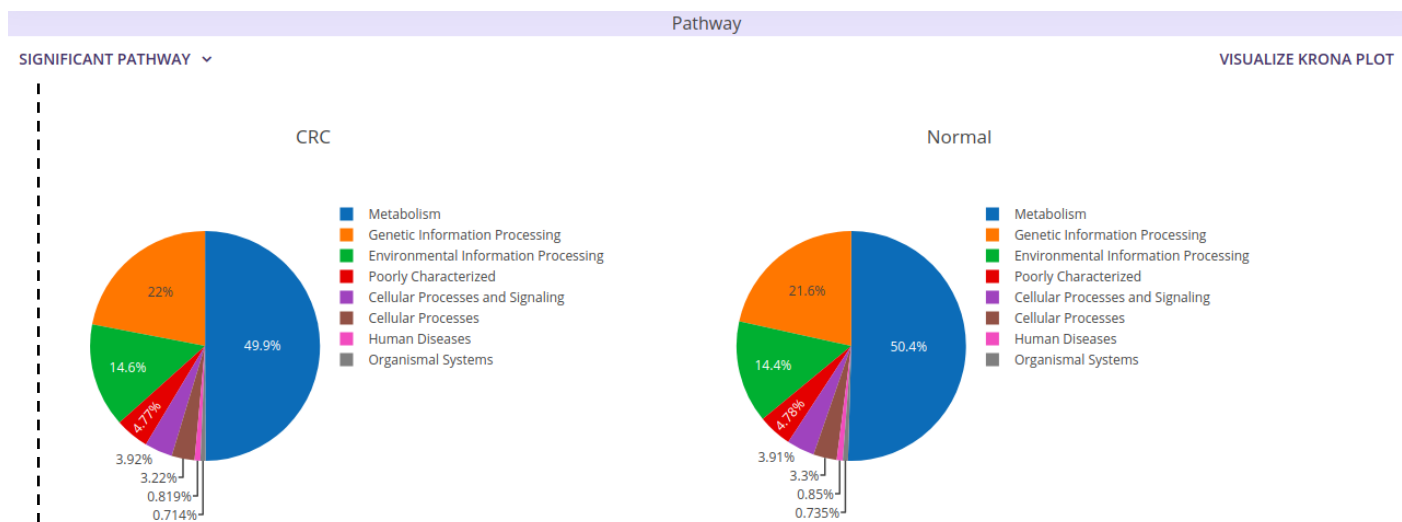
Hover the mouse cursor over the individual dots to view the values and sample number.



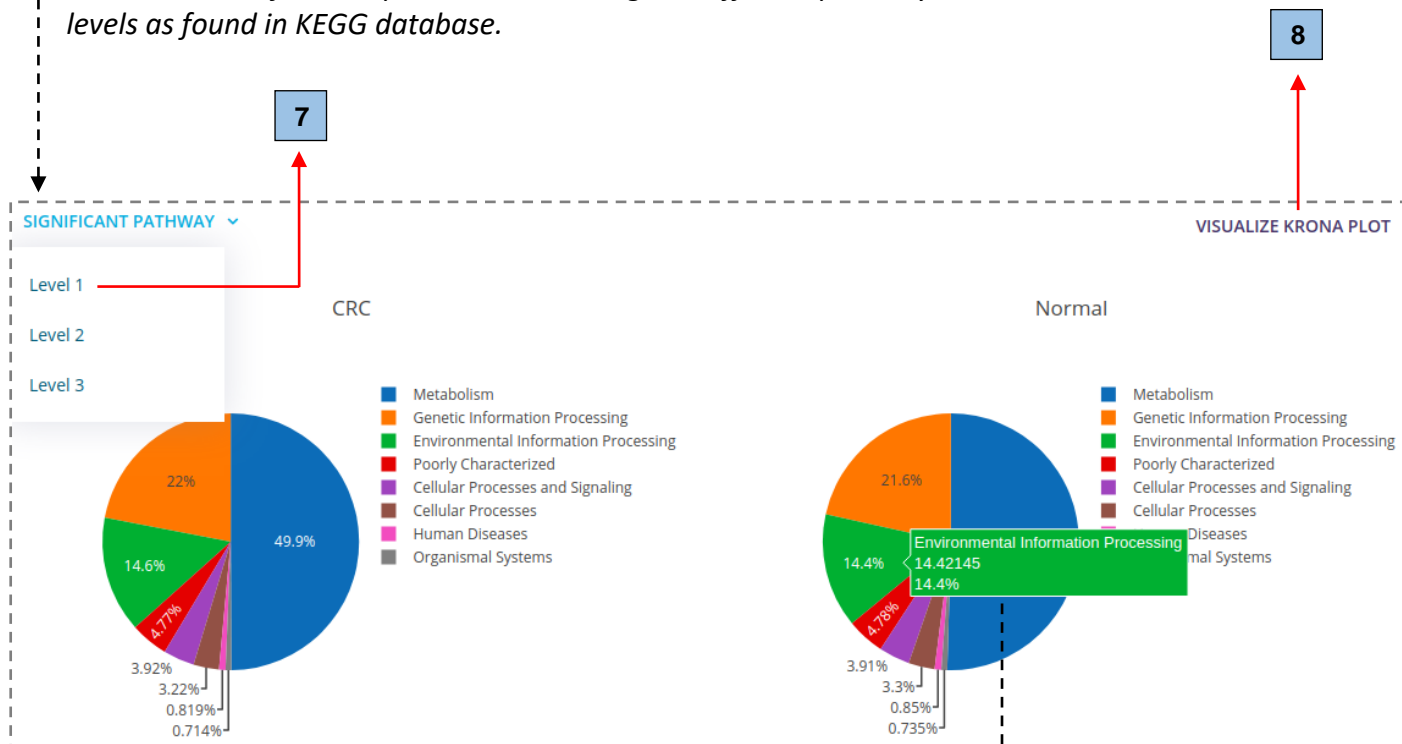
Note: **UNIFRAC (WEIGHTED/UNWEIGHTED)** added.  
 (Click to go to a new page and view.)

(Fig. 12)

# PATHWAY ANALYSIS:



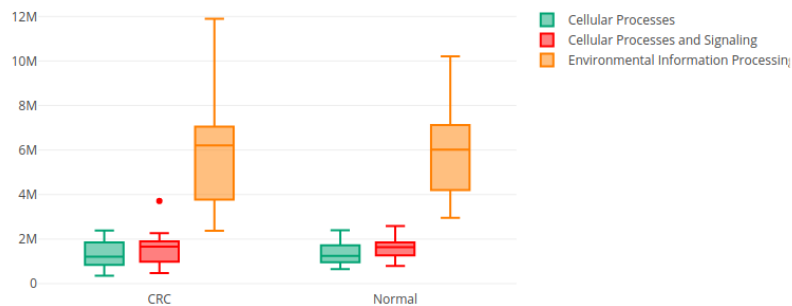
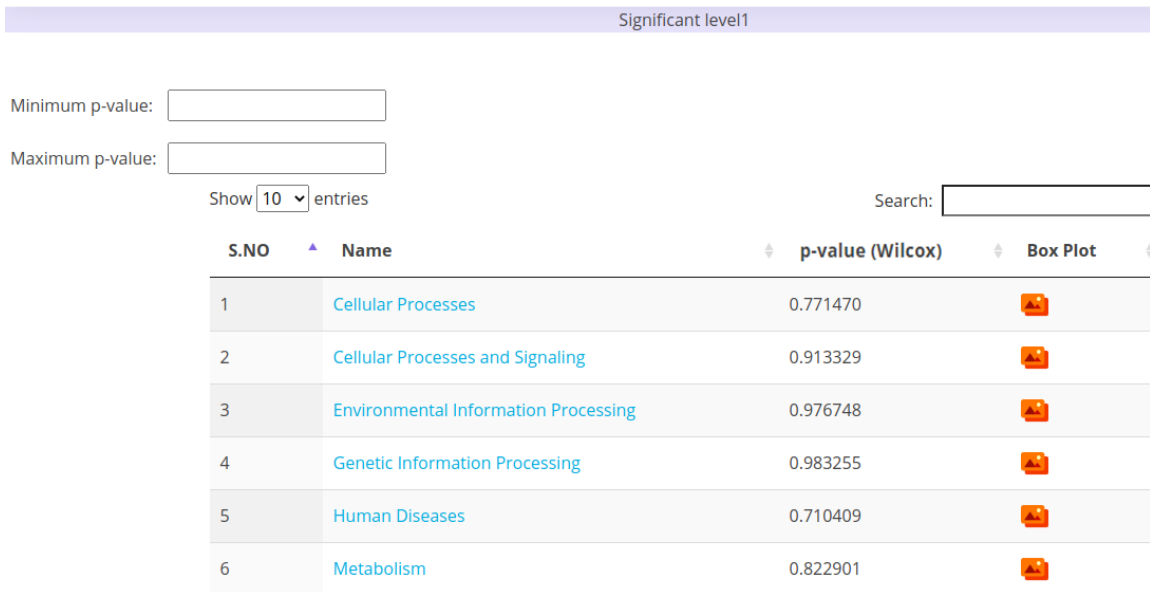
Click on the tab for a drop-down list showing the different pathway levels as found in KEGG database.



Hover the mouse cursor over the different section to view the pathway name and the percentage of its contribution to overall metabolism in that group.

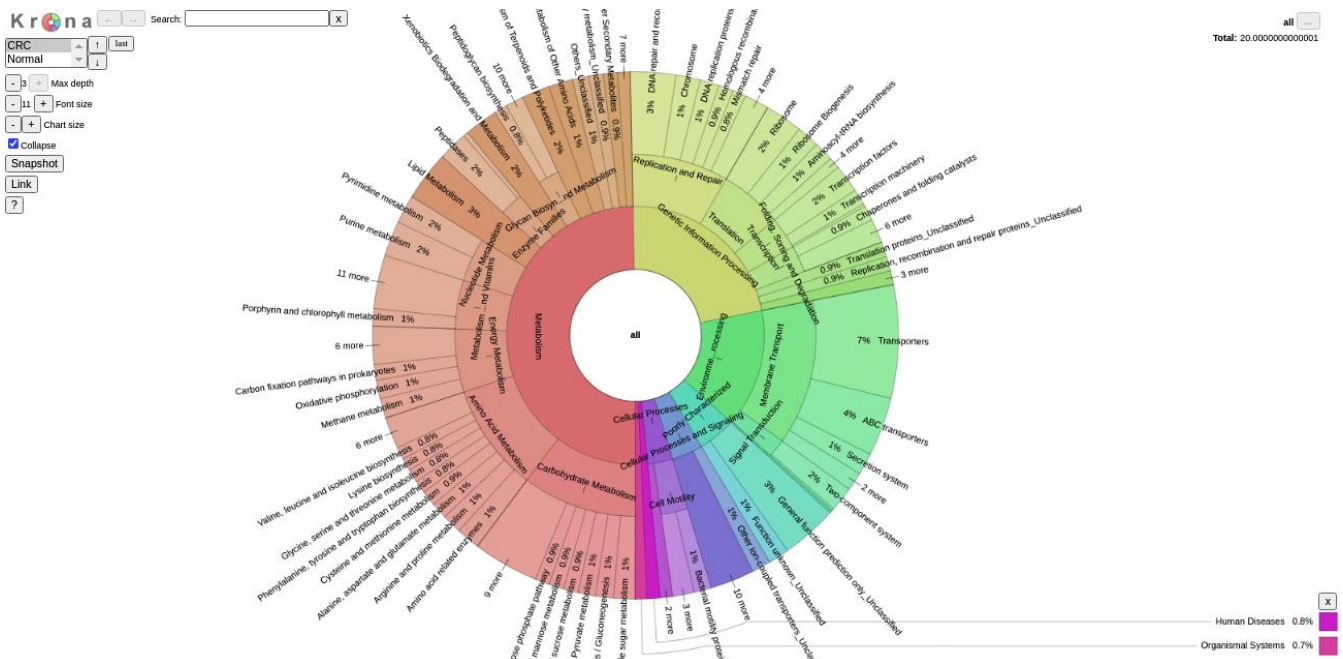
(Fig. 13)

7 Click on one of the levels to check the significant pathways in that level and it opens in a new page.



(Fig. 14)

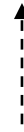
8 Click on this tab to visualize the Krona plot for metabolic pathway abundances between two groups.



(Fig. 15)

# REAL TIME V/S PREDICTED METABOLITES:

*In case of MMIP Module-II this additional tab is observed.*



## MMIP

JOB METABOLITES ALPHA-DIVERSITY TAXONOMY BETA-DIVERSITY PATHWAY **R VS P** FEATURE CONTACT US

### RealTime Vs Predicted Metabolites

Minimum p-value:

Maximum p-value:

Show  entries

Search:

S.NO	KEGG Compound ID	p-value	Mantel Statistic R	Visualize
1	C00021	0.461000	-0.017709	
2	C00022	0.643000	-0.049009	
3	C00025	0.823000	-0.093974	
4	C00026	0.892000	-0.104520	
5	C00031	0.434000	0.004802	
6	C00036	0.339000	0.018716	
7	C00037	0.601000	-0.022063	
8	C00041	0.712000	-0.051428	
9	C00042	0.810000	-0.088605	
10	C00046	0.409000	0.009349	

Showing 1 to 10 of 20 entries

Previous  2 Next

File Name	Statistical-test	Download file
Mantel Correlation Detail	Mantel	

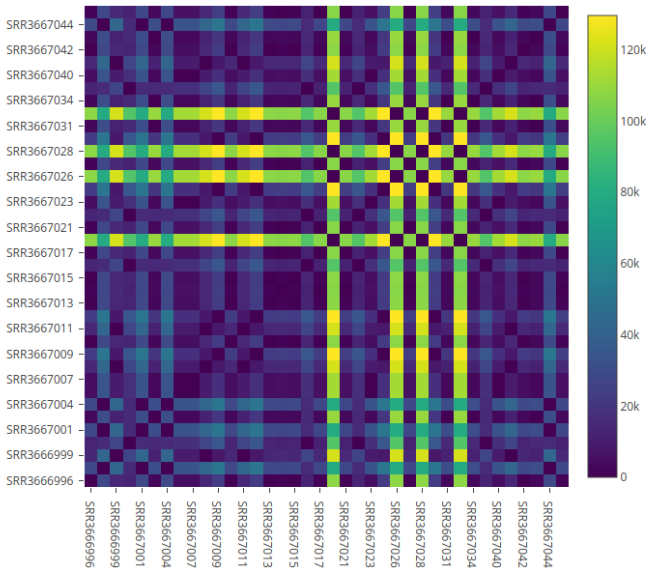
Showing 1 to 1 of 1 entries

*Click to open the correlation results of this metabolite in a new page (as shown in fig 17).*

(Fig. 16)

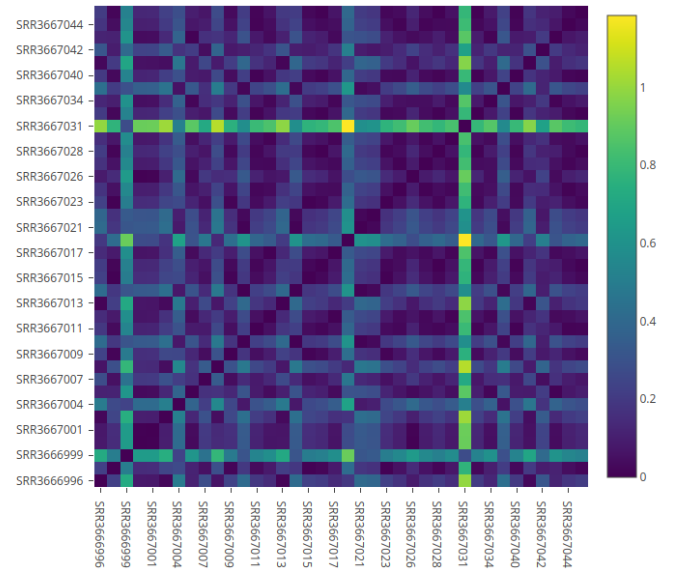
Correlation Result For C00021

Distance Matrix For Real Time Metabolite (C00021 Method=Euclidean)



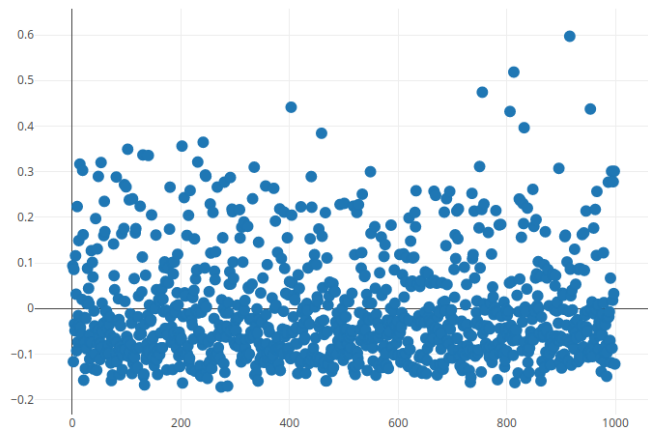
[Download Distance Matrix Data for Above Plot](#)

Distance Matrix For Predicted Time Metabolite (C00021 Method=Euclidean)



[Download Distance Matrix Data for Above Plot](#)

Mantel Test (Pvalue= 0.461, R-Statistics=-0.0177093134075238)



[Download All Permuataion Data For Above Plot](#)

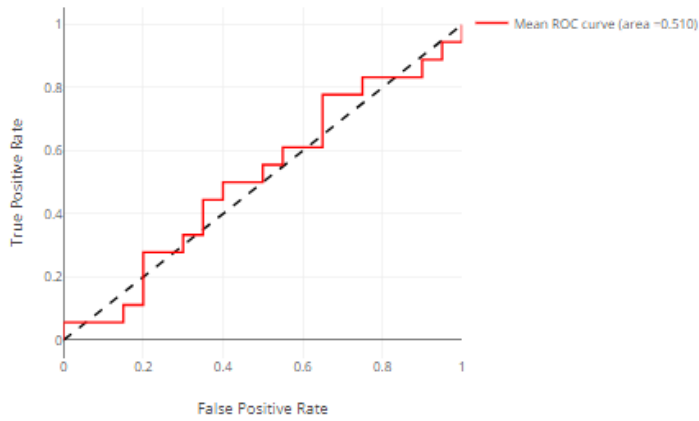
(Fig. 17)



## FEATURE PREDICTION:

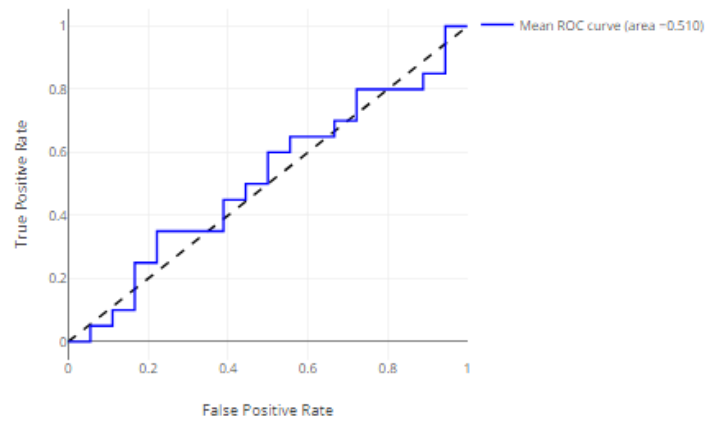
### Feature ROC Curve

ROC Curve For Normal Feature



[Download ROC Curve Data for Normal Feature](#)

ROC Curve For CRC Feature



[Download ROC Curve Data for CRC Feature](#)

Show  entries

Search:

Model Evaluation metric <sup>▲</sup>	F1 Score <sup>⬇</sup>	Accuracy Score <sup>⬇</sup>	Recall Score <sup>⬇</sup>	Precision <sup>⬇</sup>	ModelParameter & ConfusionMatrix <sup>⬇</sup>
Model Name: SVM	0.313	0.421	0.278	0.357	

Showing 1 to 1 of 1 entries

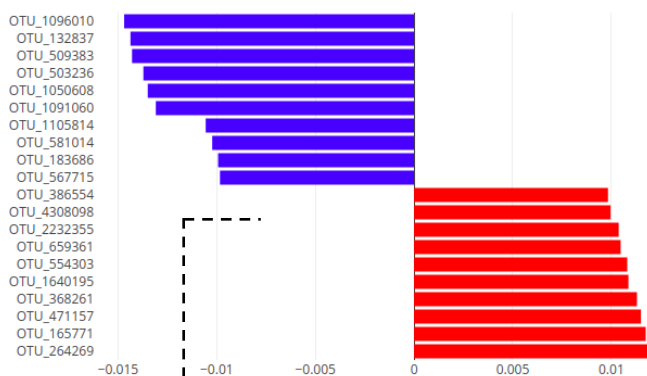
Previous  Next

*ROC curve and evaluation metrics of the selected machine learning model shown.*

(Fig. 18)

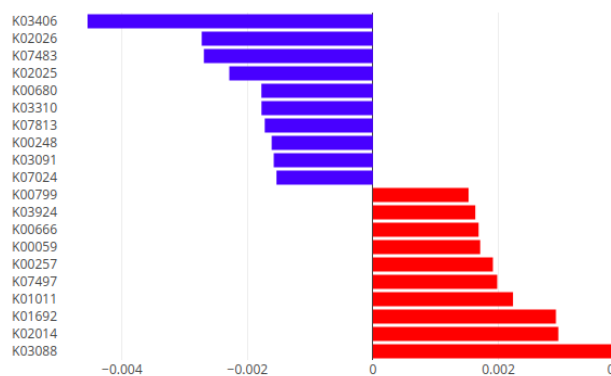
Feature Identified

Community Structure



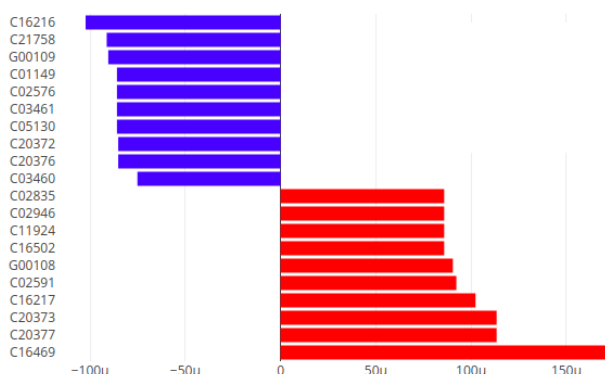
[Download All Feature For Community Structure](#)

KEGG



[Download All Feature For KEGG](#)

CMP



[Download All Feature For CMP](#)

The feature table can be downloaded (.tsv format).

The top 10 features each of positive (in red) and of negative (in blue) are shown here. Also hover mouse over diagram to view values.

Feature Category

Normal Feature

Show  entries

KEGG Compound ID	p-value (Anova)	Visualize cor
<a href="#">C02591</a>	0.56764712027688	<a href="#">LINK</a>
<a href="#">C02835</a>	0.5254858214149772	<a href="#">LINK</a>
<a href="#">C02946</a>	0.5254858214149772	<a href="#">LINK</a>
<a href="#">C11924</a>	0.5254858214149772	<a href="#">LINK</a>
<a href="#">C16217</a>	0.5107272046458358	<a href="#">LINK</a>
<a href="#">C16469</a>	0.6895104179008892	<a href="#">LINK</a>

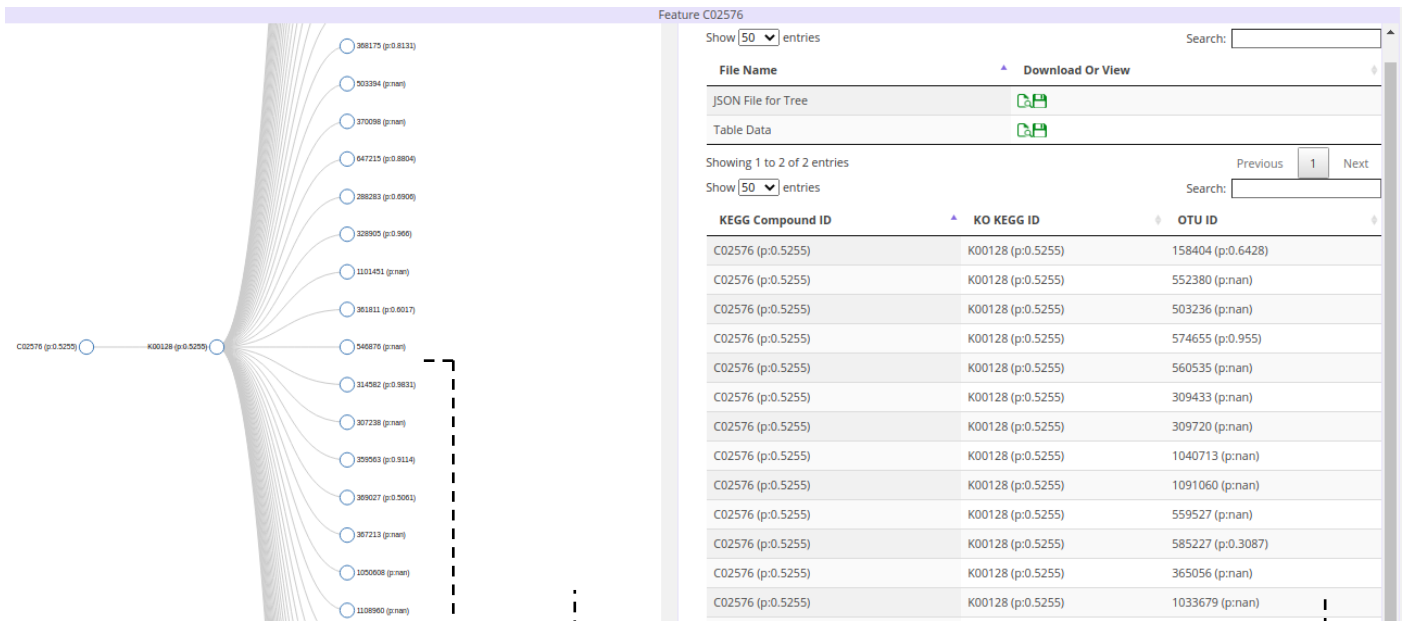
CRC Feature

Show  entries

KEGG Compound ID	p-value (Anova)	Visualize cor
<a href="#">C01149</a>	0.5254858214149752	<a href="#">LINK</a>
<a href="#">C02576</a>	0.5254858214149752	<a href="#">LINK</a>
<a href="#">C03460</a>	0.6294221232575476	<a href="#">LINK</a>
<a href="#">C03461</a>	0.5254858214149752	<a href="#">LINK</a>
<a href="#">C05130</a>	0.5254858214149752	<a href="#">LINK</a>
<a href="#">C16216</a>	0.5107272046458358	<a href="#">LINK</a>

(Fig. 19)

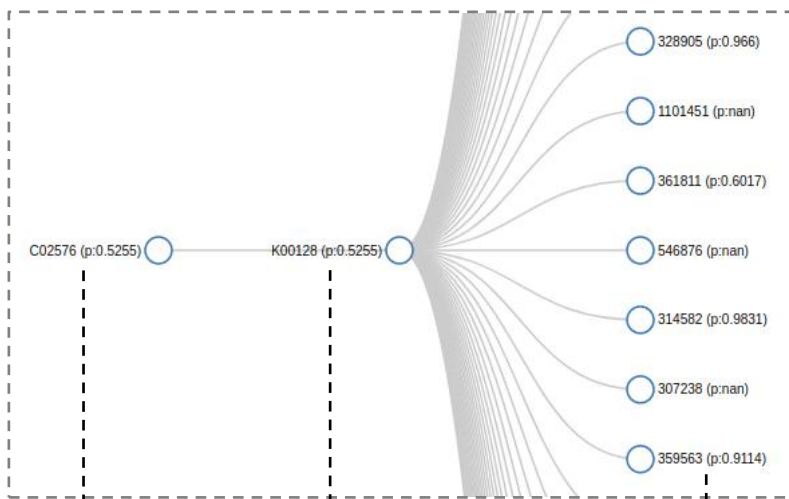
9 Click on the link to visualize the following page which shows the predicted microbial source of metabolites (predicted as feature by our algorithm), in a tree as well as tabular form. [all p values provided in adjacent brackets]



Expanded view of the tree.

Save the tree (found on the top of the diagram).

The adjacent feature links represented in a tabular format and can be downloaded.



Predicted Metabolite

The pathway that the predicted metabolite is associated with.

All the OTU ids (microbes) that could probably be the source of the predicted metabolite.

(Fig. 20)