

MetOrigin User Tutorial

Dr. Yan Ni's Research Lab

2022-08 version

MetOrigin is developed by Dr. Yan Ni's research lab. Users can login MetOrigin through <http://metorigin.met-bioinformatics.cn>

MetOrigin includes seven main steps: (1) Load data; (2) Origin Analysis; (3) Function Analysis; (4) Correlation Analysis; (5) Sankey Network; (6) Network Summary; (7) Download Results.

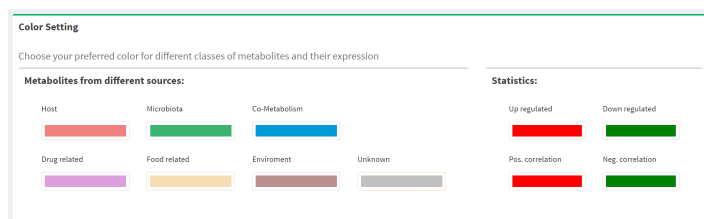
1. Load data

MetOrigin offers two different modes of data analysis: **Simple MetOrigin Analysis (SMOA)**, and **Deep MetOrigin Analysis (DMOA)**. The data formats required for data analysis depends on the type of data analysis mode. In addition, users should choose the correct host information, such as human, mouse, rat, pig etc.



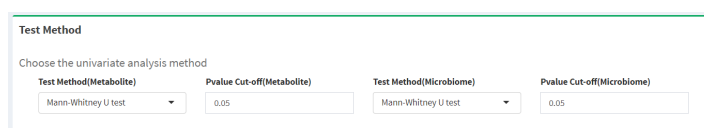
The screenshot shows two adjacent form fields. The left field, titled 'Analysis type', contains two radio buttons: 'Simple MetOrigin Analysis' (unselected) and 'Deep MetOrigin Analysis' (selected). The right field, titled 'Host', is a dropdown menu with 'Homo sapiens (human)' selected.

In the **Color Setting** box, users can set custom colors of metabolites from different sources, the up/down-regulated expressions and positive/negative-correlated relationship in subsequent statistical analysis.



The screenshot shows the 'Color Setting' interface. It has a title 'Color Setting' and a subtitle 'Choose your preferred color for different classes of metabolites and their expression'. The main area is divided into two sections. The left section, 'Metabolites from different sources:', has four columns: 'Host' (red bar), 'Microbiota' (green bar), 'Co-Metabolism' (blue bar), and 'Drug related' (purple bar). Below these are 'Food related' (yellow bar), 'Environment' (brown bar), and 'Unknown' (grey bar). The right section, 'Statistics:', has four columns: 'Up-regulated' (red bar), 'Down-regulated' (green bar), 'Pos. correlation' (red bar), and 'Neg. correlation' (green bar).

In the **Test Method** module, the user can select the statistical method (in DMOA mode) according to the characteristics of the data. According to the different grouping information, two methods, "Student's T Test" and "Mann-Whitney U test", or two methods of "Pearson" and "Spearman" are provided respectively. Users can also choose the "Auto" option, and then the software will automatically selects the statistical method for each variable based on whether the data conform to a normal distribution and homogeneity of variance. At the same time, the user needs to set the threshold for statistical difference here.



The screenshot shows the 'Test Method' interface. It has a title 'Test Method' and a subtitle 'Choose the univariate analysis method'. The main area is divided into two sections. The left section, 'Test Method(Metabolite)', has a dropdown menu with 'Mann-Whitney U test' selected. The right section, 'Test Method(Microbiome)', has a dropdown menu with 'Mann-Whitney U test' selected. Both sections have a 'Pvalue Cut-off' input field with '0.05' entered.

a. Simple MetOrigin Analysis (SMOA):

Click the button **"Browse"** to choose and upload a table of metabolites or click the button **"Load Example Data"** for testing. The metabolite table must contain at least one column of "HMDBID", "KEGGID" or "Name", and a column of 0/1 values indicating statistical significance (1- significant, 0-

nonsignificant). If the "Diff" column is missing, all metabolites will be considered as differential metabolites.

Show entries Search:

	HMDBID	KEGGID	Name	Diff
1	HMDB0000357	C01089	3-Hydroxybutyric acid	1
2	HMDB0000020	C00642	p-Hydroxyphenylacetic acid	1
3	HMDB00006524	C02938	3-Indoleacetonitrile	1
4	HMDB0000673	C01595	Linoleic acid	1
5	HMDB0000510	C00956	Aminoadipic acid	1
6	HMDB0000806	C06424	Myristic acid	1
7	HMDB0000857	C02656	Pimelic acid	1
8	HMDB0000072	C00417	cis-Aconitic acid	0
9	HMDB0000500	C00156	4-Hydroxybenzoic acid	0
10	HMDB0000511	C01571	Capric acid	0

Showing 1 to 10 of 83 entries Previous **1** 2 3 4 5 ... 9 Next

b. Deep MetOrigin Analysis (DMOA)

Click the button "**Browse**" to choose and upload three different data files or click the button "**Load Example Data**" for testing. DMOA takes three individual files as input datasets, including a "Sample Info" table with sample names and groupings information e.g., control versus diseased, a "metabolite" table with compound abundance / concentration, and a "microbiome" table with their annotations and abundance from either 16S ribosomal RNA (16s rRNA) gene sequencing or shotgun metagenomic sequencing. You must upload the "Sample Info" table before the other two files.

The sample information table requires sample IDs of metabolite analysis, sample IDs of microbiome analysis, and sample grouping information, e.g., control versus diseased. The table is editable, such as renaming sample names or groupings, unchecking a sample in the column of "Included". Then, users can click the button "save" to save all the modifications.

Search:

	Included	Metabolite_sampleID	Microbiome_sampleID	Grouping
1	<input checked="" type="checkbox"/>	S1	S1	0
2	<input checked="" type="checkbox"/>	S2	S2	0
3	<input checked="" type="checkbox"/>	S3	S3	0
4	<input checked="" type="checkbox"/>	S4	S4	0
5	<input checked="" type="checkbox"/>	S5	S5	0
6	<input checked="" type="checkbox"/>	S6	S6	0
7	<input checked="" type="checkbox"/>	S7	S7	0
8	<input checked="" type="checkbox"/>	S8	S8	0
9	<input checked="" type="checkbox"/>	S9	S9	0
10	<input checked="" type="checkbox"/>	S10	S10	0

The metabolite table should consist of at least one column of "HMDBID", "KEGGID" or "Name", followed by the quantitative values of each sample. To note, the sample IDs should be consistent with the sample IDs in the sample information table.

Show 10 entries

Search:

	HMDBID	KEGGID	Name	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
1	HMDB0002231	C16526	11Z-Eicosenoic acid	34.82	67.06	439.96	95.93	50.94	44.94	105.31	62.63	31.65	32.8
2	HMDB0000008	C05984	2-Hydroxybutyric acid	59.08	24.58	20.74	15.4	28.59	11.5	6	17.43	14.08	10.71
3	HMDB00031594		2-Methylhexanoic acid	60451.11	6693.29	6412.4	7448.82	2151.39	5686.31	29411.44	99267.84	25613.33	5530.88
4	HMDB0002210		2-Phenylglycine	16.56	12.3	230.92	54.13	11.47	15.18	76.81	132.28	613.71	163.2
5	HMDB0003911	C05145	3-Aminoisobutanoic acid	20.06	18.21	25.94	7.76	21.86	7.01	1.18	270.66	15.37	32.86
6	HMDB0000357	C01089	3-Hydroxybutyric acid	86.26	4607.71	4496.53	2276.15	364.22	13002.69	1476.13	178.92	1906.16	90.75
7	HMDB0000754		3-Hydroxyisovaleric acid	2.24	7.56	11.41	6.17	7.71	12.06	4.13	103.11	6.8	8.4
8	HMDB0006524	C02938	3-Indoleacetonitrile	147.96	61.07	576.33	106.92	31.65	55.56	68.5	313.11	1496.39	369.27
9	HMDB0000466	C08313	3-Methylindole	1.9	28.43	1.09	0.18	0.77	0.85	25.06	24.25	0.19	7.86
10	HMDB0000500	C00156	4-Hydroxybenzoic acid	128.25	63.21	92.87	83.58	48	149.81	71.5	73.35	80.88	82.41

Showing 1 to 10 of 83 entries

Previous12345...9Next

The microbiome table requires at least a column of taxonomy annotation information with the correct column name, e.g., kingdom, phylum, class, order, family, genus, and species. To note, the sample ID should be consistent with the sample IDs of microbiome analysis in the sample information table.

Show 10 entries

Search:

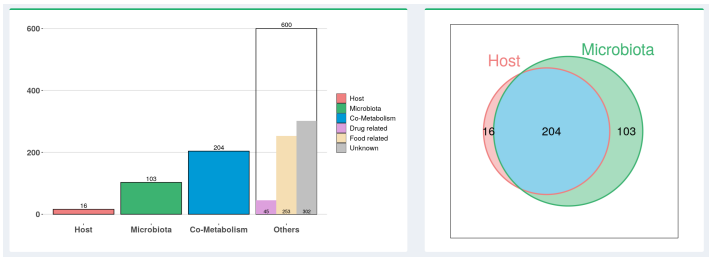
	kingdom	phylum	class	order	family	genus	species	S1	S2	S3	S4	S5	S6
1	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	uncultured bacterium	0	1	0	155	0	9735
2	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	uncultured bacterium	1	0	5113	423	1063	0
3	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ruminococcaceae UCG-004	uncultured bacterium	0	0	3	0	0	0
4	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	Others	23	0	0	1	2	16
5	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes	Others	3	36	13	1	0	27
6	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Lachnospira	uncultured bacterium	0	61	19	5	35	8
7	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Butyrivibrio	uncultured bacterium	4	13	6	1	4	10
8	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Alloprevotella	uncultured bacterium	0	0	0	0	0	0
9	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Coproccoccus 1	uncultured bacterium	0	2	9	5	4	7
10	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Oscillibacter	uncultured bacterium	15	11	30	0	0	2

Showing 1 to 10 of 702 entries

Previous12345...71Next

2. Origin Analysis

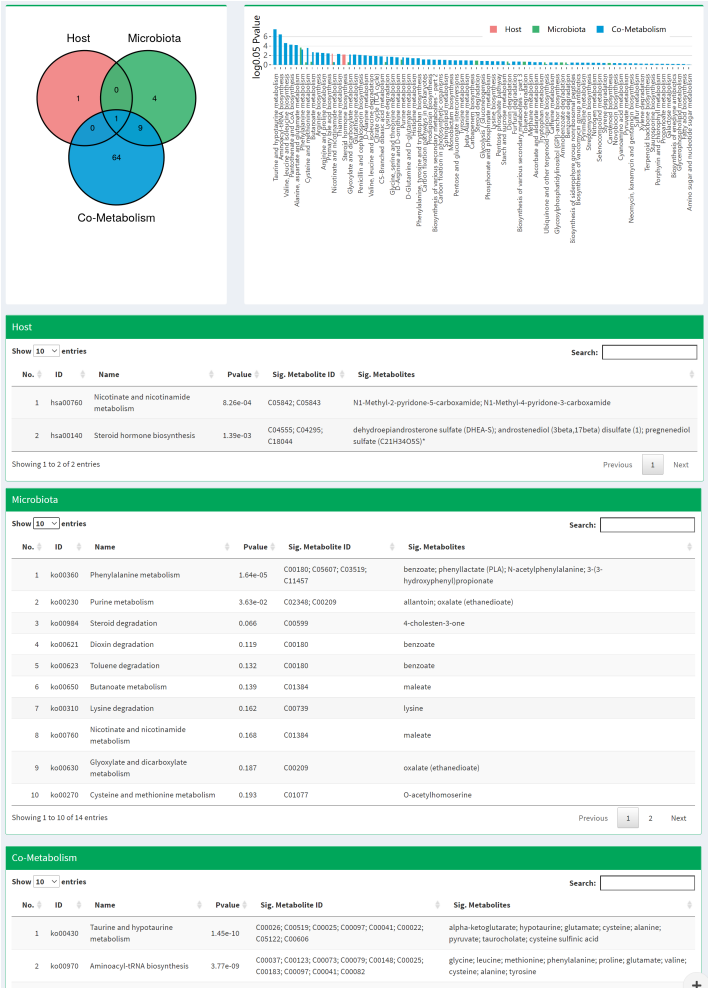
Click the button **"Perform Analysis"** to start the data analysis. As a result, a bar plot and a venn diagram are produced to summarize the total number of metabolites from host, microbiota, co-metabolism, and others.



3. Function Analysis

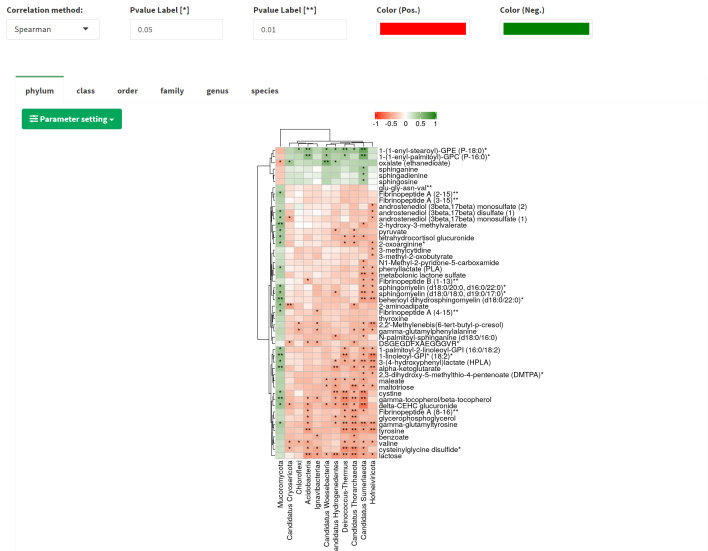
Click the button **"Perform Analysis"** to start the data analysis.

As a result, a bar plot is produced to compare the relative significance of differential metabolic pathways from MPEA analysis according to the differential metabolites from different origins, i.e., bacteria, host, or both. Their corresponding tables with details of pathway enrichment are provided in the same page.



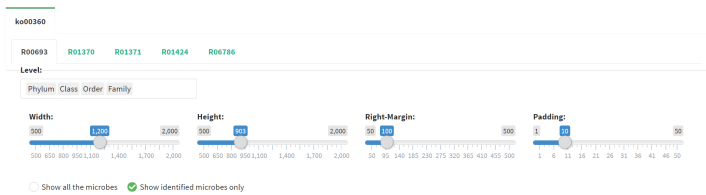
4. Correlation Analysis

MetOrigin provides three classical methods of correlation analysis, including Spearman, Pearson, and Maximal Information Coefficient analysis. In addition, a heatmap of correlation coefficients between differential metabolites and microbes at phylum, class, order, family, genus, species level is produced for visualization. The color block above determines the color of the heatmap. Users can drag the triangle in the bottom right to change the image size, or click the button "Parameter setting" to change the fontsize of the heatmap. It is worth noting that the first p-value will be used as the cut-off for the correlation between metabolites and microbes in the **STA-Sankey** network diagram.



5. Sankey Network

Next, users can click the button "Perform Analysis" to start the data analysis to obtain Sankey network for each reaction of selected metabolic pathways. Users can choose which level of bacteria is viewed using the top left "Level" checkbox and modify the figure size using sliding bars. The figures can be saved and download as *.svg file to the local computer using the camera icon



MetOrigin provides a list of significant metabolic pathways from bacteria or both. To note, the top one metabolic pathway of microbiota and co-metabolism origins is visualized automatically. An interactive table is provided allowing users to remove and add a certain pathway for visualization by simply clicking the corresponding box.

Selected Pathway

Microbiota

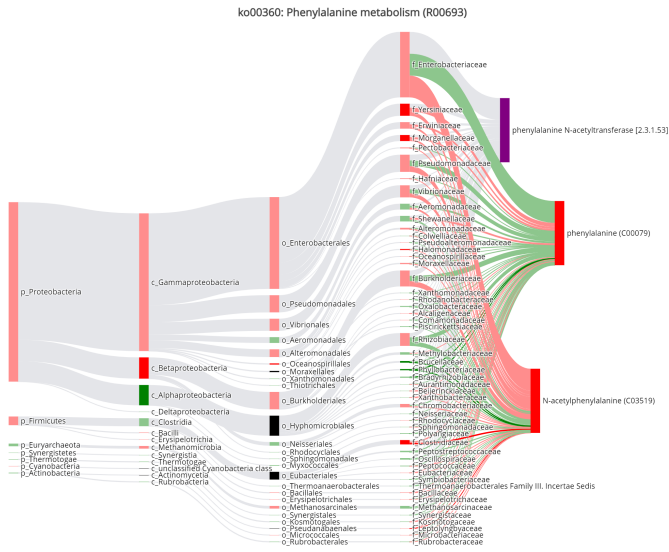
[1] "ko00360"

Co-Metabolism

[1] "ko00430"

ID	Name	Host	Microbiota	Co-Metabolism
ko00430	Taurine and hypotaurine metabolism	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
ko00970	Aminesucyl-tRNA biosynthesis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ko00290	Valine, leucine and isoleucine biosynthesis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ko00770	Pantothenate and CoA biosynthesis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ko00250	Alanine, aspartate and glutamate metabolism	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ko00360	Phenylalanine metabolism	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
ko00270	Cysteine and methionine metabolism	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ko00450	Butyrate metabolism	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ko00220	Arginine biosynthesis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ko00330	Arginine and proline metabolism	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ko00120	Primary bile acid biosynthesis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

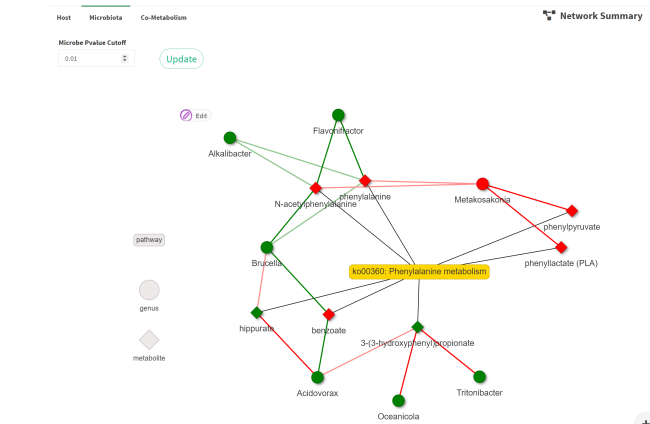
For each metabolic reaction, MetOrigin provides a **BIO-Sankey** and a **STA-Sankey** network. The widths of the bands are linearly proportional to the number of bacteria involved in specific metabolic reaction. The red or green color of bars and bands indicates up or down regulation of bacteria and metabolites, or positive or negative correlations between them. The shades of color (dark or light) indicate the statistical significance of bacteria/metabolite and their correlations, respectively. For details, users can refer to the summary table below.



Colors (default)		Meaning
Simple MetOrigin Analysis (SMOA)		
<div></div>	Dark red bars	Metabolic substrates
<div></div>	Dark green bars	Metabolic products
<div></div>	Purple bars	Metabolic enzymes
Deep MetOrigin Analysis (DMOA)		
<div></div>	Dark red bars	Significantly up regulated microbes or metabolites (FC>1 and p<0.05)
<div></div>	Light red bars	Up regulated microbes or metabolites (FC>1 and p≥0.05)
<div></div>	Dark green bars	Significantly down regulated microbes or metabolites (FC<1 and p<0.05)
<div></div>	Light green bars	Down regulated microbes or metabolites (FC<1 and p≥0.05)
<div></div>	Dark grey bars	Microbes or metabolites with no change (FC=1)
<div></div>	Black bars	Microbes or metabolites in the reference database
<div></div>	Purple bars	Metabolic enzymes
<div></div>	Dark red bands	Significant positive correlation (R>0 and p<0.05)
<div></div>	Light red bands	Positive correlation without statistical significance (R>0 and p≥0.05)
<div></div>	Dark green bands	Significant negative correlation (R<0 and p<0.05)
<div></div>	Light green bands	Negative correlation without statistical significance (R<0 and p≥0.05)
<div></div>	Dark gray bands	No correlation (R=0)
<div></div>	Light gray bands	Reference relationships searched through database

6. Network Summary

Finally, users can click the button "Perform Analysis" to obtain a whole picture of microbiome and metabolome interactions. Three metabolic network and associated microbes are summarized for host, microbiota and co-metabolism, correspondingly. To note, the top one significant metabolic pathway of host, microbiota and co-metabolism origins is visualized automatically. An interactive table is provided allowing users to remove and add a certain pathway for visualization by simply clicking the corresponding box.



7. Download Results

Users can click the button "Download Analysis Results" to download all the figures and tables. And to get the high-resolution PDF figures in previous steps, please download the SVG image files during the analysis and transfer to PDF format by the "SVG Converter" tool.

You make it. Congratulations!

All the analysis results can be downloaded for your further data exploration. If you have any questions or suggestions, please contact us:
bioinformatics_group@aliyun.com

[Download Analysis Result](#)

SVG Converter

If you prefer the high-resolution PDF figures of Sankey plots, please download the SVG image files during the analysis and transfer to PDF format here.

Choose a file

[Browse...](#) No file selected

[Convert](#) [Download PDF](#)

Please cite: Yu, Gang, Cuifang Xu,Danni Zhang, Feng Ju, and Yan Ni. 2022.“MetOrigin: Discriminating the Origins of Microbial Metabolites for

Integrative Analysis of the Gut Microbiome and Metabolome.” iMeta. e10.
<https://doi.org/10.1002/imt2.10>

Contact information:

For any questions or suggestions, please contact our technique support through (bioinformatics_group@aliyun.com), or the principal investigator of this project Dr. Yan Ni(yanni617@zju.edu.cn).