



Developers

Grace Tzun-Wen Shaw (tzunwen@gmail.com)

Yueh-Yang Pao (greanozone@gmail.com)

Daryi Wang (dywang@gate.sinica.edu.tw)

Dr. Daryi Wang's Laboratory

Biodiversity Research Center, Academia Sinica

Contents

1.	What MetaMIS can do?	2
2.	Support platform	3
3.	Installation	3
4.	MetaMIS Input Format	4
4.1	Data format	4
4.2	Test data:	5
5.	Run MetaMIS	6
5.1	Graphical user interface of MetaMIS	6
5.2	Pipeline for data flow and parameter selection	6
5.3	Parameter selection rule for multiple files	11
5.4	Example of Test data	12
6.	MetaMIS Results	14
6.1	Five tab panels for MetaMIS outputs	14
6.2	Export File	26
	Bugs or problems	30
	References	31

1. What MetaMIS can do?

MetaMIS (**M**etagenomic **M**icrobial **I**nteraction **S**imulator) is a Lotka-Volterra model based tool to infer microbial interactions based on microbial community. The functionalities of MetaMIS are listed as follows.

1. Automatically infer microbial interactions from 16S rRNA abundance profiles based on abundance-ranking strategy [1].
2. Visualize the original or predicted abundance profiles.
3. Visualize the interaction networks.
4. MetaMIS can systematically examines interaction patterns, such as mutualism (+/+), competition (-/-), parasitism or predation (+/-), commensalism (+/0), amensalism (-/0), and no effect (0/0).
5. For each interaction network, principal component analysis (PCA) is embedded to refine the biological role inside microbes.
6. The interaction tables generated by MetaMIS can be exported as Gephi or Cytoscape format for advanced topological analysis.
7. Provide a consensus network from multiple interaction networks.
8. The consensus interaction table can be exported as a text file.
9. MetaMIS is publicly free and accessible at
<https://sourceforge.net/projects/metamis/> without login requirement.

2. Support platform

MetaMIS was developed and maintained on Matlab (MATLAB R2016a, The MathWorks, Inc., Natick, Massachusetts, United States). Currently, MetaMIS supports Windows (64 bits) and Mac platform.

3. Installation

<https://sourceforge.net/projects/metamis/>

Please install Matlab runtime before running MetaMIS.

(If your desktop has been installed the Matlab software already, please skip the installation of Matlab runtime program.)

GUI version:

Windows (64bits)

1. Matlab runtime program: MatlabRuntime_Installer.exe
2. Main program: MetaMIS_Win.exe

Mac

1. Matlab runtime program: MatlabRuntime_Installer.app
2. Main program: MetaMIS_Mac.app

4. MetaMIS Input Format

4.1 Data format

MetaMIS accepts a temporal microbial abundance profile as input in which microbial communities should be obtained in a series of temporal samples. A dataset with long time lapses between consecutive data points may have potential problem, the sparse data may cause false results. To simplify the process of data input, we have defined the input file format as follows. First, a temporal microbial abundance profile needs to be processed in advance according to a user-defined taxonomic level. The data must be in a tab separated text format, described as follows (Fig. 1).

1. OTU identifiers are listed in the first column. The first cell can be empty.
Duplicated OTU identifiers are not allowed.
2. The first row is reserved for time-series sample identifiers that must be numeric; and the remaining spaces should be filled with the microbial abundance values, a numeric matrix.
3. Reads or relative abundance tables are both acceptable.
4. The rearrangement of disordered times-series samples is allowed (Fig. 1).
5. Please keep the same unit of time among samples.
6. Time interval can be variable.
7. Decimal time unit is allowed at MetaMIS_v1.02 (Table 1) (**New functionality in MetaMIS_v1.02**).

Table 1 Different precision in the time measurement among different versions.

Time measurements	MetaMIS_v1.00 or v1.01	MetaMIS_v1.02
Integers	Yes	Yes
Decimals	No	Yes (round to four decimal places)
Time interval	variable	variable
Example	0, 1, 2, 5, 8 ...	0, 1.1, 2.12, 3.123, 4.1234...

OTU identifiers	Time-series sample identifiers																		disordered			
	0	9	142	143	144	145	146	147	148	149	150	151	10	152	153	155	156					
Unclassified	139	830	851	803	316	653	1472	690	518	326	1216	641	374	968	622	475	1036					
Acetobacteraceae	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Moraxellaceae	6	0	0	0	2	0	0	0	1	0	1	0	0	0	0	0	1	0				
Neisseriaceae	0	0	0	0	0	0	0	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0
Alcaligenaceae	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0	0	0	0	0
Enterobacteriaceae	1132	2266	3	1363	115	232	170	129	95	33	1	2	174	65	5	1	409					
Pasteurellaceae	0	9	0	1	3	0	2	1	3	114	3	17	1	5	0	30	3					
Bacteroidaceae	5907	27055	30095	22275	15479	17509	24279	23255	17946	15377	25459	15974	21937	10962	21237	28151	18620					
Chromatiaceae	44	3	0	179	0	74	155	2	10	8	0	0	13	5	2	1	199					
Micrococcaceae	49	0	1	1	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
Synechococcaceae	20	68	16	21	20	23	15	7	24	27	66	13	10	12	5	25	19					
Corynebacteriaceae	4808	0	1	1	2	15	1	0	7	12	7	0	0	6	3	0	0					
Mycobacteriaceae	11	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actinomycetaceae	33	0	0	1	10	3	3	4	0	1	3	5	1	0	0	0	19					
Streptomyctaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pseudonocardiaceae	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Acholeplasmataceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bifidobacteriaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Propionibacteriaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Vellonellaceae	1008	6	1	2	93	3	1	50	35	14	62	4	17	7	1	28	8					
Clostridiaceae	16	54	4	10	114	33	24	16	11	18	15	4	4	50	7	223	35					
Rhodobacteraceae	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Methylcytaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kathromonadaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lactobacillaceae	4199	1	0	0	0	0	0	0	0	0	6	5	0	0	1	1	0					
Bradyrhizobiaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sphingomonadaceae	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Flavobacteriaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aberimonadaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ectothiorhodospiraceae	67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Campylobacteraceae	51	0	0	0	25	0	0	5	0	6	14	1	4	2	0	0	2					
Dualobacteraceae	1	0	2192	30	294	588	2677	318	464	530	55	850	0	124	664	95	320					

Figure 1 Data input format.

4.2 Test data:

1. Data description:

A gut metagenomic study containing temporal microbiota from a male and a female was used [2]. Using greengenes taxonomy, the number of total taxa assigned at family level were 92 and 69 for male and female microbiomes. The two datasets were further checked by gene copy number correction program [3].

2. Input file:

(A) M_gut.txt: 92 families over 317 time points

(B) F_gut.txt: 69 families over 124 time points

3. Output files:

(A) M_gut.mat

(B) F_gut.mat

(C) F_gut_EDGE_9.txt

(D) F_gut_NODE_9.txt

(E) MI_CON_Top0to10(P0.90101).txt

5. Run MetaMIS

5.1 Graphical user interface of MetaMIS

There are two main blocks in MetaMIS (Fig. 2).

1. Data loading and preprocess blocks.
2. Visualization of interaction network and analysis.

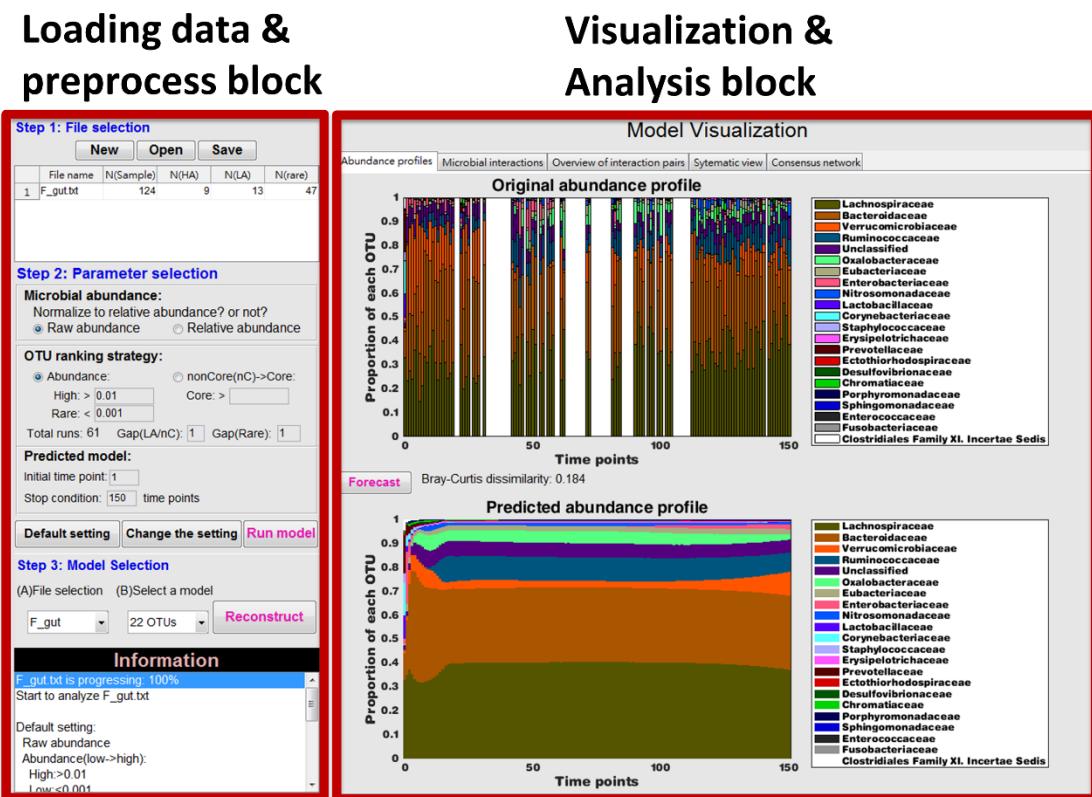


Figure 2 Graphical user interface of MetaMIS.

5.2 Pipeline for data flow and parameter selection

A typical analysis workflow may contain three steps: uploading of formulated data file(s) (Step 1: File selection), specification of the parameters (Step 2: Parameter selection), and performing the model calculations. Input parameters are optional for users or can be set by default.

5.2.1 Step 1: File selection

Select 「New」 to create a new project, e.g. F_gut.txt (Fig. 3).

Select 「Open」 to open a processed dataset, e.g. F_gut.mat.

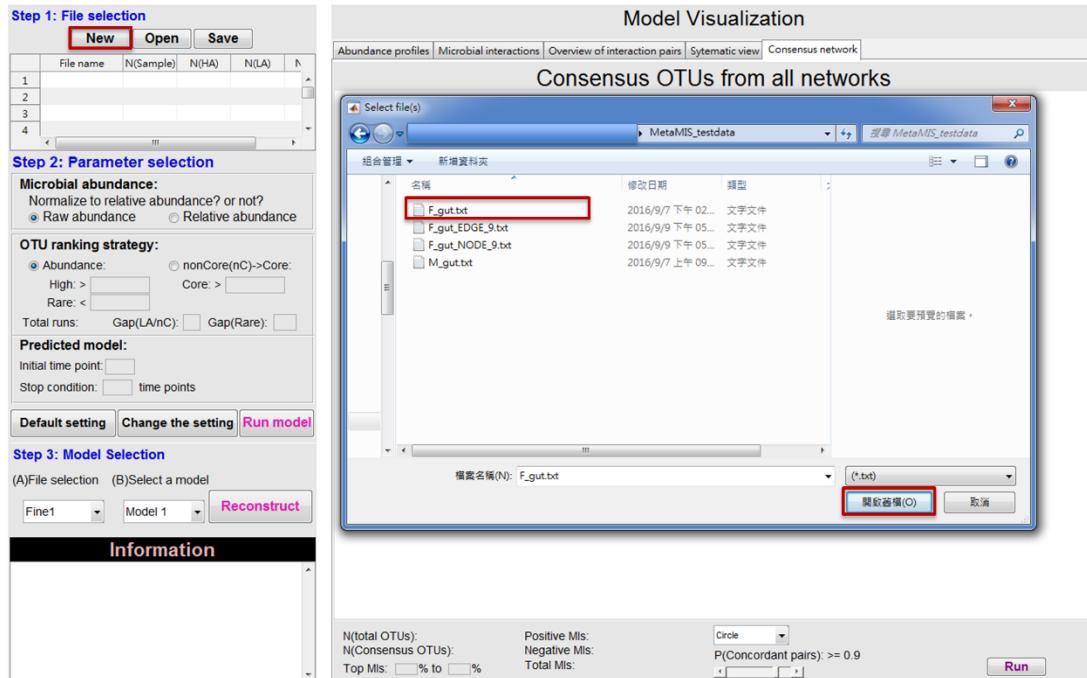


Figure 3 The exemplification of creating a new project.

5.2.2 Step 2: Parameter selection

Parameter properties are introduced under the condition of importing only one data file. When multiple abundance files are imported into MetaMIS, the general rules of parameter selection are described in the section of 5.3 General rules for parameter selection.

① Microbial abundance:

「Raw abundance」: Time-series samples retain the original input reads.

「Relative abundance」: Raw abundances are adjusted to have common total reads in all samples.

④ OTU ranking strategy:

- 「[Abundance](#)」 : OTUs are ranked based on the ratio of average abundance across samples to the total abundance. The default setting of the ratio is >1% for high abundance, <0.1% for rare, and the remaining for low abundance non-rare OTUs. However, the minimum number of high abundance OTUs is set to be three in MetaMIS owing to the model calculations. If the requirement is not fulfilled, the upper bound of the ratio will be automatically changed to satisfy the mentioned condition.
- 「[nonCore->Core](#)」 : Core OTUs are OTUs that consistently present across time-series samples. The default setting of core OTUs is 90%, meaning that those OTUs have detectable reads among over 90% of time-series samples.
- 「[Total runs](#)」 : The total number of running interaction networks. (**New item in MetaMIS_v1.01**)
- 「[Gap\(LA/nC\)](#)」 : Based on abundance ranking strategy (see ref [1] Fig. 1), for each time, only one OTU with highest low abundance is added to sequentially expand the interaction network. The default setting is 1. The larger the value, the less the number of running interaction networks. (**New item in MetaMIS_v1.01**)
- 「[Gap\(Rare\)](#)」 : An interaction network containing rare OTUs means it also contains all high and low abundance OTUs. The functionality of this item is similar to that of [Gap\(LA/nC\)](#). It determines that how many rare OTUs are sequentially added to an interaction network. The default setting is 1. (**New item in MetaMIS_v1.01**)

Note: General rule for 「Total runs」, 「Gap(LA/nC)」 and 「Gap(Rare)」 :

Problem:

If there are hundreds of OTUs in an abundance table, to run all interaction networks is too time-consuming and may have the memory-out problem.

1. Test running:

We used a desktop computer with 32Gb of RAM and Intel(R) Core(TM) i7-4770 CPU @ 3.40 GHz processor to run two larger dataset. The computing and exporting time were shown in Table 2.

- ◆ “Dataset 1”: It took about 20 minutes to infer the 435 interaction networks, and almost 1 hour to save all results to a mat file.
- ◆ “Dataset 2”: It took about 8 hours to run the 1485 interaction networks, but a large amount of outputs caused out-of-memory errors.

Table 2 The computing time and data structure for different dataset.

Dataset	N(samples)	N(OTU)	N(Total runs)	T _{Running}	T _{mat}
F_gut.txt	124	69	61	1.5 mins	12 sec
Dataset 1	26	450	435	~20 mins	~57 mins
Dataset 2	10	1500	1485	~8 hrs	-

2. Modification (Fixed at MetaMIS_v1.01):

For these larger abundance table, we constrained the total runs in order to avoid the memory-out problem. We set a constrained number, N_c, to control the total number of running interaction networks.

At first, all OTUs are ranked from high, low to rare abundance.

- ◆ If the number of total runs is smaller than 200, it follows the original abundance ranking strategy to sequentially add a maximum low abundance or rare OTU.
- ◆ If the number of total runs is greater than 200, the selection strategy of low abundance or rare OTUs is modified as follows (*Equation (1)* and *(2)*).

$$\left. \begin{array}{l} \text{Gap}(LA) = 1 \\ \text{Gap}(Rare) = \frac{N(Rare)}{N_c - N(LA) - 1} \end{array} \right\} \quad \text{if } N(LA) \leq (N_c - N_{Rare}^{Default}) \quad (1)$$

$$\left. \begin{array}{l} \text{Gap}(LA) = \text{Gap}(Rare) \\ = \frac{N(LA) + N(Rare)}{N_c - 1} \end{array} \right\} \quad \text{if } N(LA) > (N_c - N_{Rare}^{Default}) \quad (2)$$

The default setting of constrained number in MetaMIS_v1.01 was shown in Table 3 in which $N_{Rare}^{Default}$ was set to 10.

Table 3 The default setting of N_c and its corresponding Gap(LA) and Gap(Rare).

Constrained Rules	Gap(LA)	Gap(Rare)	Total runs or N_c
$1+N(LA)+N(Rare) < 100$	1	1	$1+N(LA)+N(Rare)$
$100 \leq 1+N(LA)+N(Rare) < 200$	1	1	$1+N(LA)+N(Rare)$
$200 \leq 1+N(LA)+N(Rare) < 300$	1	$\frac{N(Rare)}{200 - N(LA) - 1}$	≤ 200
$300 \leq 1+N(LA)+N(Rare) < 400$	1	$\frac{N(Rare)}{180 - N(LA) - 1}$	≤ 180
$400 \leq 1+N(LA)+N(Rare) < 500$	1	$\frac{N(Rare)}{160 - N(LA) - 1}$	≤ 160
$500 \leq 1+N(LA)+N(Rare) < 600$	1	$\frac{N(Rare)}{150 - N(LA) - 1}$	≤ 150
$600 \leq 1+N(LA)+N(Rare) < 700$	1	$\frac{N(Rare)}{140 - N(LA) - 1}$	≤ 140
$700 \leq 1+N(LA)+N(Rare) < 800$	1	$\frac{N(Rare)}{130 - N(LA) - 1}$	≤ 130
$800 \leq 1+N(LA)+N(Rare)$	1	$\frac{N(Rare)}{120 - N(LA) - 1}$	≤ 120

④ Predicted model:

Using generalized Lotka-Volterra equations [4], the inferred microbial interactions can be used to regenerate abundance profile. In this regeneration process, there are two parameters to be set.

「Initial time point」: According to the input data structure, users can arbitrarily determine any time-series sample as the initial state. Different initial condition may conduct different predicted abundance profiles. The default is the first time-series sample. If there are N_T time-series samples in an abundance table, the maximum initial time point that you can input is N_T which depends on the setup of stop condition (*Equation (3)*). For each time-series sample $i=1, 2, \dots, N_T$, they are sampled at day T_1, T_2, \dots, T_{N_T} .

$$\begin{cases} T_{i-1} < \text{Stop condition} \leq T_i & \rightarrow \quad \{1, 2, \dots, i-1\}_{i=1,2,\dots,N_T} \\ T_{N_T} < \text{Stop condition} & \quad \{1, 2, \dots, N_T\} \end{cases} \quad (3)$$

「Stop condition」: Whether the regenerated profile can be produced till the end of the stop time point is defined as an indicator of success or failure of inferred interactions. The default setting is 80% of time-series sample size.

5.2.3 Step 3: model calculations

「Run model」: Press this button to perform model calculations based on Lotka-Volterra equations to get abundance-ranking interaction networks.

5.3 General rules for parameter selection

④ Importing single OTU table:

See the details in the section of 5.2.2 Step 2: Parameter selection.

④ Importing multiple OTU tables:

MetaMIS can manipulate a batch of multiple OTU tables, e.g. file 1, file 2, ..., and file N_F , simultaneously under the circumstances of sharing common parametric settings. If different OTU tables should be run by different settings of parameters, importing

one abundance file at a time is the solution in the current version of MetaMIS. In general, parameters which would be influenced by multiple importing of files are shown below.

- 「Total runs」, 「Gap(LA/nC)」 and 「Gap(Rare)」 :

For N_F dataset, the one with the most total runs would be selected to setup the three parameters.

- 「Stop condition」 :

The default stop condition is set by the following equation (*Equation (4)*).

$$\max\{\max(T_1^k), \min(T_{N_T}^k)\}_{k=\text{file1}, \text{file2}, \dots, \text{file } N_F} \quad (4)$$

Choosing the smallest terminal time point, $\min(T_{N_T}^k)$, is a loosen parametric strategy to make sure more successful interaction networks produced. However, the given stop condition may be smaller than the first time point in some files. We compare the minimum of the terminal time points, $\min(T_{N_T}^k)$, and the maximum of first time points, $\max(T_1^k)$, to avoid the wrong parametric settings.

5.4 Example of Test data

After F_gut.txt is imported into MetaMIS (Fig. 3), MetaMIS immediately generates the basic data description: there are 124 time-series samples, 9 high abundance, 13 low-abundance non-rare, and 47 rare OTUs in this dataset (Fig. 4(A)).

Case 1:

While pressing the button of 「Default setting」, parameters are automatically determined as follows (Fig. 4(A)).

- 「Raw abundance」 to retain the original data format.
- 「Abundance」 :
 - 「High: > 0.01」 to select high abundance OTUs.
 - 「Rare: < 0.001」 to select rare OTUs.
- 「Total runs」 to convey the total running trials
 - 「Gap(LA/nC)」 is set to 1.
 - 「Gap(Rare)」 is set to 1.
- 「Initial time point」 is set to be the first time point.
- 「Stop condition」 is set to be 80% of terminal time points.

Case 2:

Pressing the button of 「Change the setting」, users can easily modify the parameters. Then, MetaMIS will update the data structure (Fig. 4(B)). The threshold of high abundance OTUs was set to 1 in this test data, and the other parameters adopted default values. MetaMIS will automatically adjust the high abundance threshold to make sure that there are at least three high abundance OTUs in an interaction network. In this case, there would be totally 67 interaction networks running in the next step.

To reduce or increase the total runs, users can adjust two parameters, including Gap(LA/nC) or Gap(Rare). In this case, Gap(Rare) was set to 2, and the total runs changed from 67 to 43.

Once the parameter setting is completed, please press the button 「Run model」 to manipulate the Lotka-Volterra equations for inferring microbial interactions.

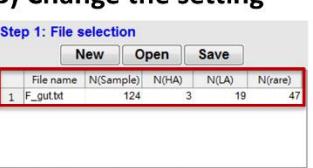
<p>(A) Default setting</p>  <p>Step 2: Parameter selection</p> <p>Microbial abundance: Normalize to relative abundance? or not? <input checked="" type="radio"/> Raw abundance <input type="radio"/> Relative abundance</p> <p>OTU ranking strategy: <input checked="" type="radio"/> Abundance: High: > 0.01 <input type="radio"/> nonCore(nC)->Core: Core: > <input type="text"/> Rare: < 0.001</p> <p>Total runs: 61 Gap(LA/nC): 1 Gap(Rare): 1</p> <p>Predicted model: Initial time point: <input type="text"/> 1 Stop condition: 150 time points</p> <p>Default setting Change the setting Run model</p> <p>Step 3: Model Selection (A)File selection (B)Select a model Reconstruct Fine1 Model 1</p> <p>Information Raw abundance Abundance(low->high): High:>0.01 Low:<0.001 Initial time point: column 1 Stop condition: 150 time points</p>		<p>(B) Change the setting</p>  <p>Step 2: Parameter selection</p> <p>Microbial abundance: Normalize to relative abundance? or not? <input checked="" type="radio"/> Raw abundance <input type="radio"/> Relative abundance</p> <p>OTU ranking strategy: <input checked="" type="radio"/> Abundance: High: > 0.0883 <input type="radio"/> nonCore(nC)->Core: Core: > <input type="text"/> Rare: < 0.001</p> <p>Total runs: 67 Gap(LA/nC): 1 Gap(Rare): 1</p> <p>Predicted model: Initial time point: <input type="text"/> 1 Stop condition: 150 time points</p> <p>Default setting Change the setting Run model</p> <p>Step 3: Model Selection (A)File selection (B)Select a model Reconstruct Fine1 Model 1</p> <p>Information Change the setting: Raw abundance Abundance(low->high): High:>0.0883 Low:<0.001 Initial time point: column 1 Stop condition: 150 time points</p>
<p>(C) Change the total runs</p>  <p>Step 2: Parameter selection</p> <p>Microbial abundance: Normalize to relative abundance? or not? <input checked="" type="radio"/> Raw abundance <input type="radio"/> Relative abundance</p> <p>OTU ranking strategy: <input checked="" type="radio"/> Abundance: High: > 0.0883 <input type="radio"/> nonCore(nC)->Core: Core: > <input type="text"/> Rare: < 0.001</p> <p>Total runs: 43 Gap(LA/nC): 1 Gap(Rare): 2</p> <p>Predicted model: Initial time point: <input type="text"/> 1 Stop condition: 150 time points</p> <p>Default setting Change the setting Run model</p> <p>Step 3: Model Selection (A)File selection (B)Select a model Reconstruct Fine1 Model 1</p> <p>Information Change the setting: Raw abundance Abundance(low->high): High:>0.0883 Low:<0.001 Initial time point: column 1 Stop condition: 150 time points</p>		

Figure 4 Parameter selection. (A) Press the button of 「Default setting」. (B) Press the button of 「Change the setting」. (C) Change the total runs.

6. MetaMIS Results

6.1 Five tab panels for MetaMIS outputs

After the implement of MetaMIS on the female gut communities [2], we got totally 17 interaction networks. As shown in Figure 5, the output page contains five main tab panels to display the selected interaction network. The rightest two panels convey the systematic information from all produced interaction networks. More detailed information is described below from Figure 5 to Figure 16.

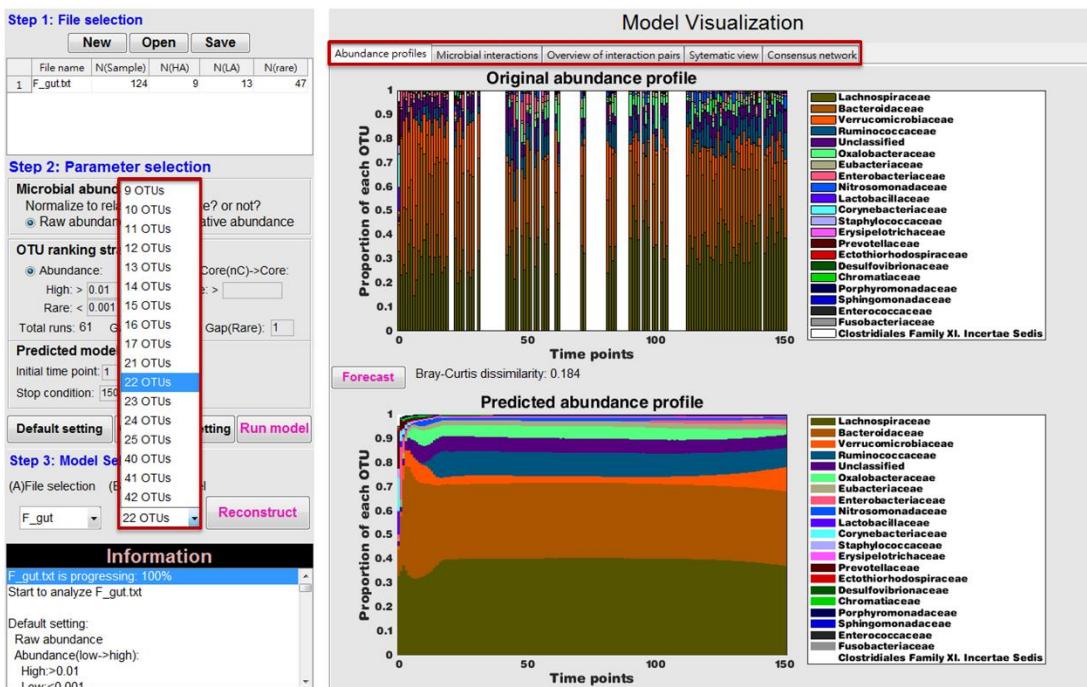


Figure 5 MetaMIS platform showing the original and predicted abundance profiles in the tab panel of “Abundance profiles”. The x-axis is time point and the y-axis is the proportion of each OTU at a time-series sample.

6.1.1 Panel of “Abundance profiles” (Fig. 5)

Original abundance profile directly derived from the input dataset contained 124 time points, which could be discontinuous.

Predicted abundance profile is the data generated (abundance profiles) from the

selected interaction network. If users press the button Forecast, MetaMIS produces another predicted one containing all continuous time points. The stop time point is an option for users to select a preferred terminal time point.

Bray-Curtis dissimilarity, ranged from 0 to 1, measures the difference between the original and predicted abundance profiles. The lower the Bray-Curtis (BC) score, the more similar the two abundance profiles. In the test case, interaction network with 22 OTUs conveyed a BC score 0.184, reflecting the reliability of the generated dataset.

6.1.2 Panel of “Microbial interactions”

Inferred microbial interactions are displayed in two manners. The tabular structure shows the quantitative interactive relationships. The network topology is to visualize the microbial interactions (Fig. 6). Furthermore, the interaction strengths can be optionally chosen for users. There are three strength types (Fig. 6 and 7) in MetaMIS. If different interaction strength type is chosen, the interaction network will be generated based on corresponding strengths.

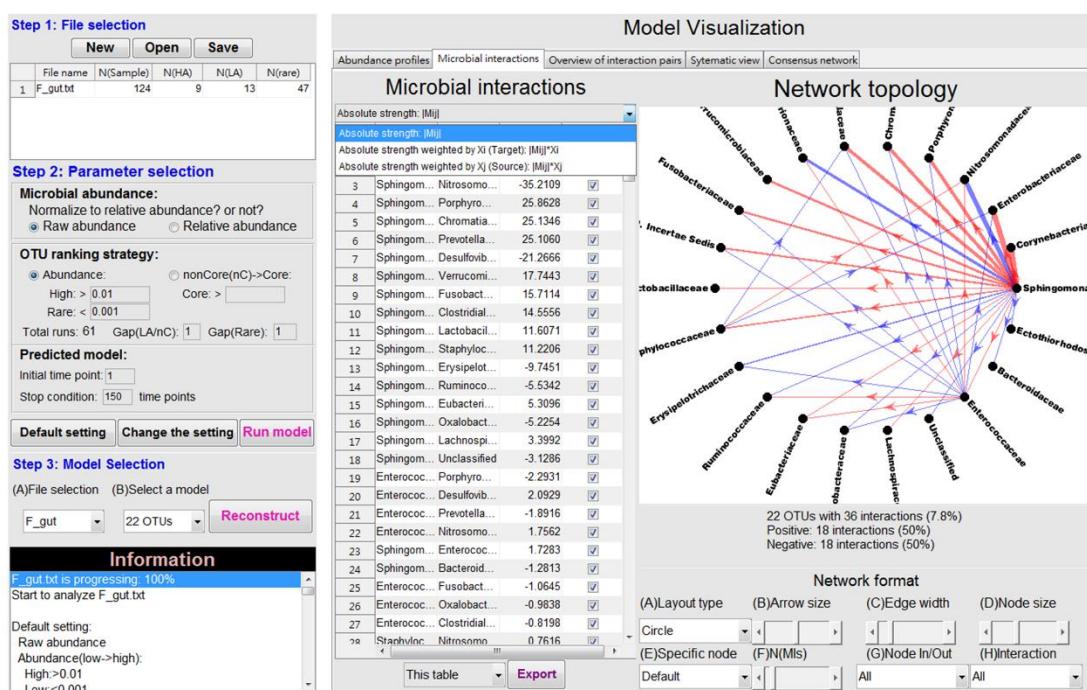


Figure 6 Panel of “Microbial interactions” displays inferred interactions by tabular and network topological view. Interactions were ranked according to the interaction strengths. The blue (or red) line means negative (or positive) interaction between microbes.

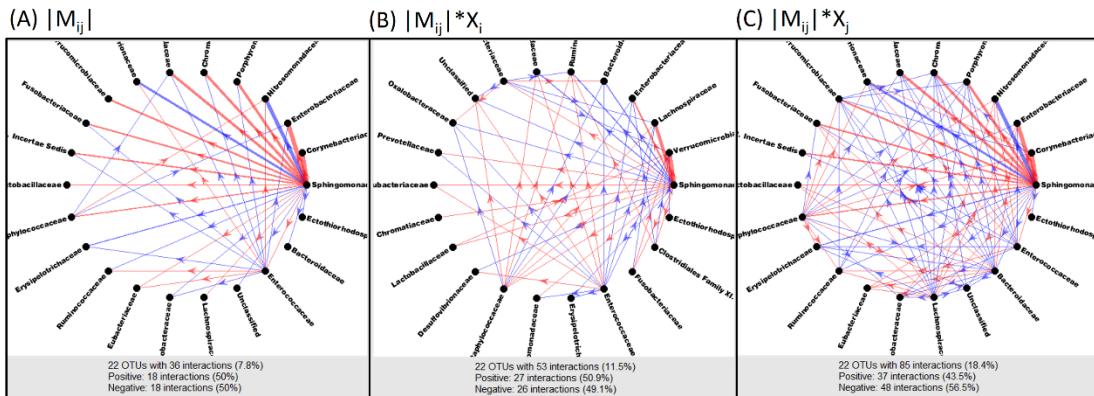


Figure 7 Different types of interaction strengths. (A) $|M_{ij}|$ is absolute interaction strength which can be weighted by target (B) or source abundance (C).

- $|M_{ij}|$: Considering positive or negative strength had similar impacts on interaction networks, interaction strengths are presented by the absolute values.
- $|M_{ij}| * X_i$: Considering abundance values may influence the interaction strengths, all absolute interaction strengths are multiplied by the abundance of target OTUs.
- $|M_{ij}| * X_j$: All absolute interaction strengths are multiplied by abundance of source OTUs.

In the case of test data, there were totally 36 interactions among 22 OTUs. The minimum numbers of interactions covering 8 OTUs are presented in the interactive network (Fig. 6). There are eight optional items in MetaMIS for users to modify the final appearance of the microbial interactive network.

(A) 「Layout type」 :

To keep track of visual-spatial information, we provided four kinds of network topological layouts, such as circle, force, layered, and subspace for optional visualization (Fig. 8).

(B) 「Arrow size」 :

The arrow means the direction of influence from a source OTU to a target one. Users can fine-tune the arrow size to optimize the network visualization.

(C) 「Edge width」 :

The edge width is proportional to the microbial interactive strength. The larger edge width conveys a stronger microbial interactive strength. Users can change

the edge width to highlight or ignore the interactions manually.

(D) 「Node size」 :

Each node denotes an OTU. Dragging the scroll bar can change the node size.

(E) 「Specific node」 :

“Specific node” allowed users to partially observe the connective behaviors of a specific OTU (Fig. 9). The default setting is to show the interactive network of all nodes in a global view.

(F) 「N(MIs)」 :

How many interaction pairs are selected to display in the network (Fig. 9).

(G) 「Node In/Out」 :

Select the option of “Node In” (or “Node Out”), and microbial interactions will converge to (or diverge from) a specific OTU (Fig. 10).

(H) 「Interaction」 :

If the positive or negative interactions are concerned, this is an optional item to change the status (Fig. 11).

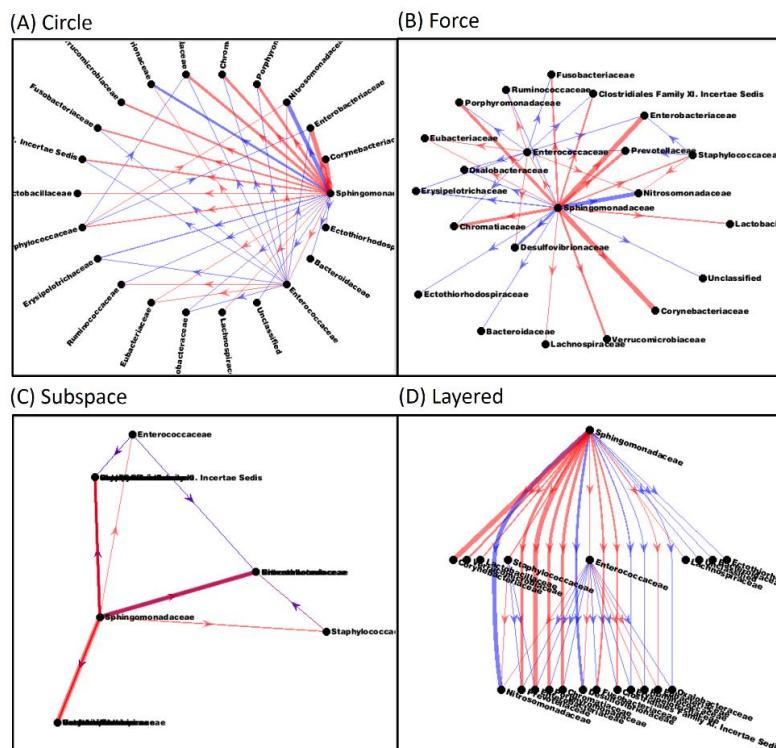


Figure 8 Four kinds of different layout types for network topology.

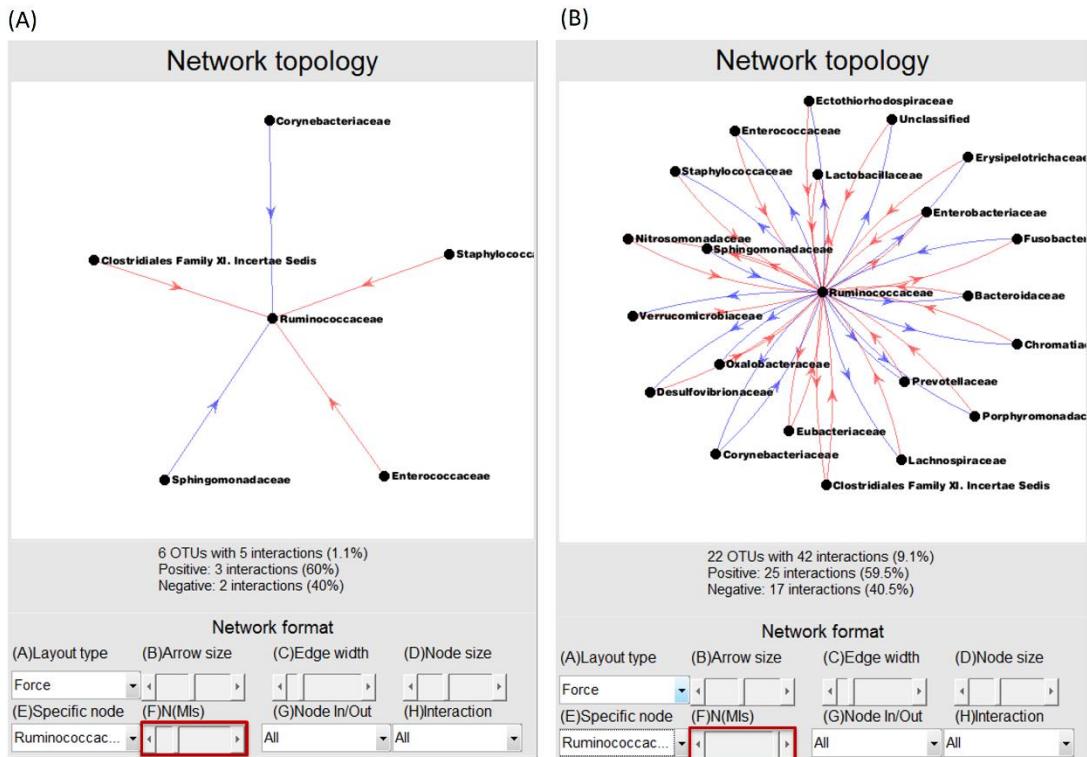


Figure 9 The interactive behaviors of a specific OTU to all others, exemplified by *Ruminococcaceae*. (A) Display fewer interactions with stronger strengths. (B) Display all interactions connected with *Ruminococcaceae*.

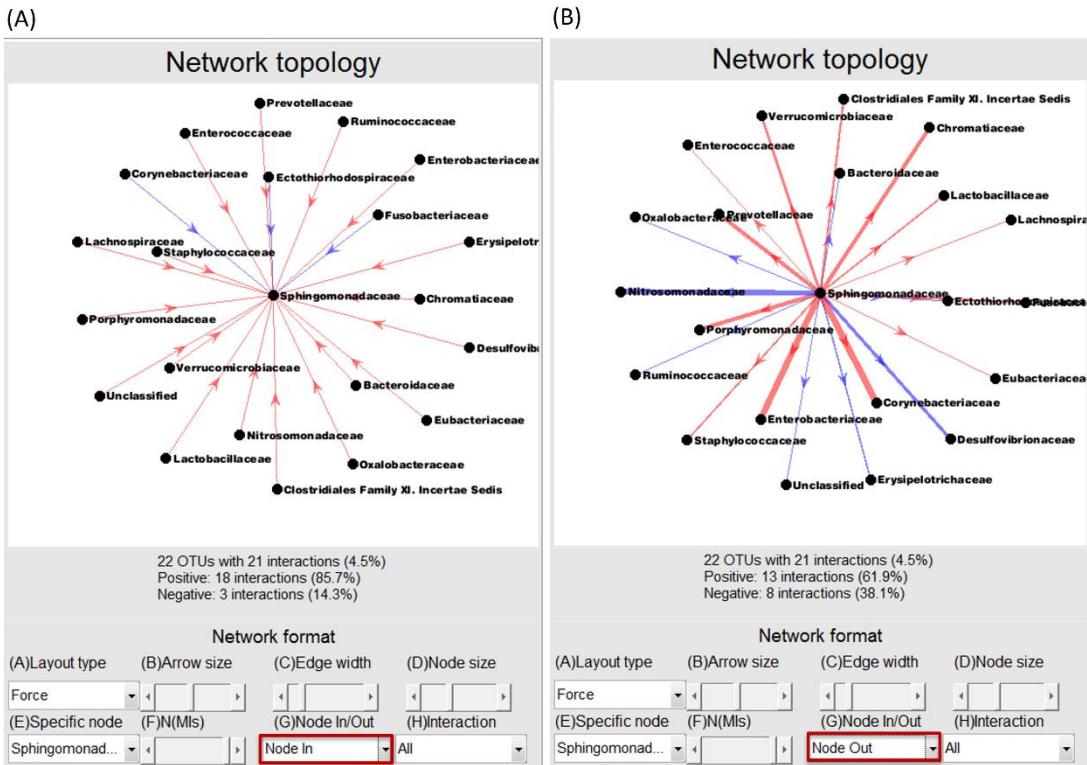


Figure 10 The graphic view of (A) Node In or (B) Node Out. This example was taken by *Sphingomonadaceae* while all interaction relations were selected.

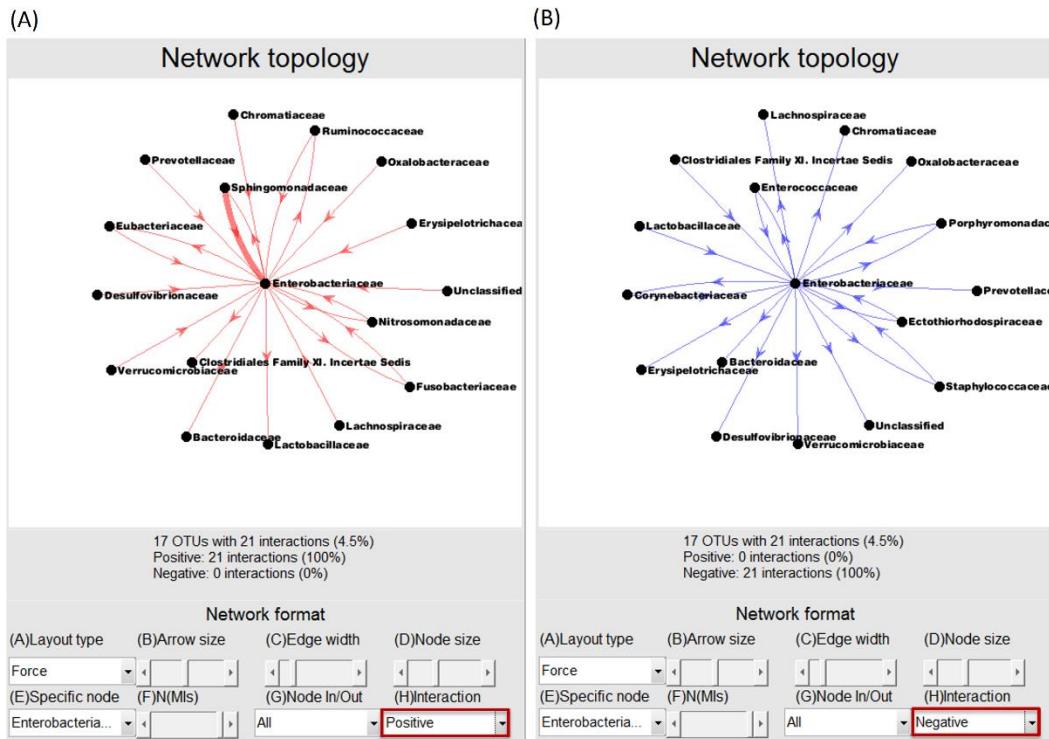


Figure 11 The network topology contains only (A) positive or (B) negative interactions, exemplified by *Enterococcaceae*.

6.1.3. Panel of “Overview of interaction pairs”

The main purpose of this panel is to identify potential key OTUs by observing the interactive behaviors between one OTU and others. If there are N OTUs in an interaction network, each OTU will have at most N-1 interaction-pair relationships. The distribution of six interaction patterns, including mutualism (+/+), competition (-/-), parasitism/predation (+/-), commensalism (+/0), amensalism (-/0), and no effect (0/0), is provided for each OTU (Fig. 12(A)). Users can fine-tune the threshold by turning weaker interactions to zero to reveal the influence of weaker interactions on the total distribution of interaction pairs (Fig. 12(B)). The average or summation of all absolute interaction strength from each interaction pattern is illustrated in Fig. 12(C). The two-dimensional principal component analysis (PCA) plot is supported to identify the major interaction pairs of an OTU (Fig. 12(D)).

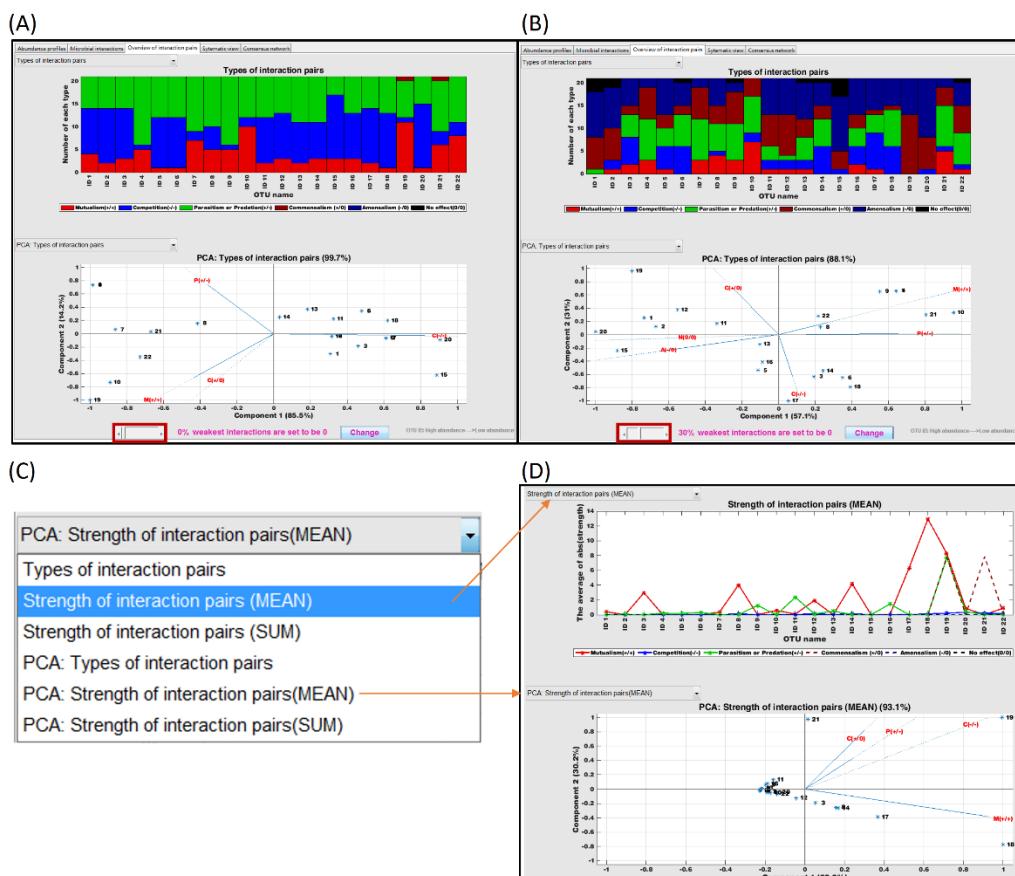


Figure 12 Panel of “Overview of interaction pairs” shows the different types of interaction pairs. PCA plot is supported to reveal the key OTUs among interaction patterns.

6.1.4 Panel of “Systematic view”

To compare outputs from different interaction networks, a systematic perspective is shown in Figure 13. The distribution of three interaction patterns, including mutualism (+/+), competition (-/-), and parasitism/predation (+/-), is shown in Figure 11. The meaning of symbols is illustrated bellow.

○: parasitism or predation (+/-)

○: competition (-/-)

○: mutualism (+/+)

*: Interaction networks with successful outcomes are denoted as black star.

*: Interaction networks with failed outcomes are denoted as gray star.

For any two sequential interaction networks, one OTU may play a critical role in the system. The removal or addition of this lowest abundance OTU may influence the successful or failed outcomes. This column of “Status change” is used to record the status change between two sequential interaction networks (Table 4).

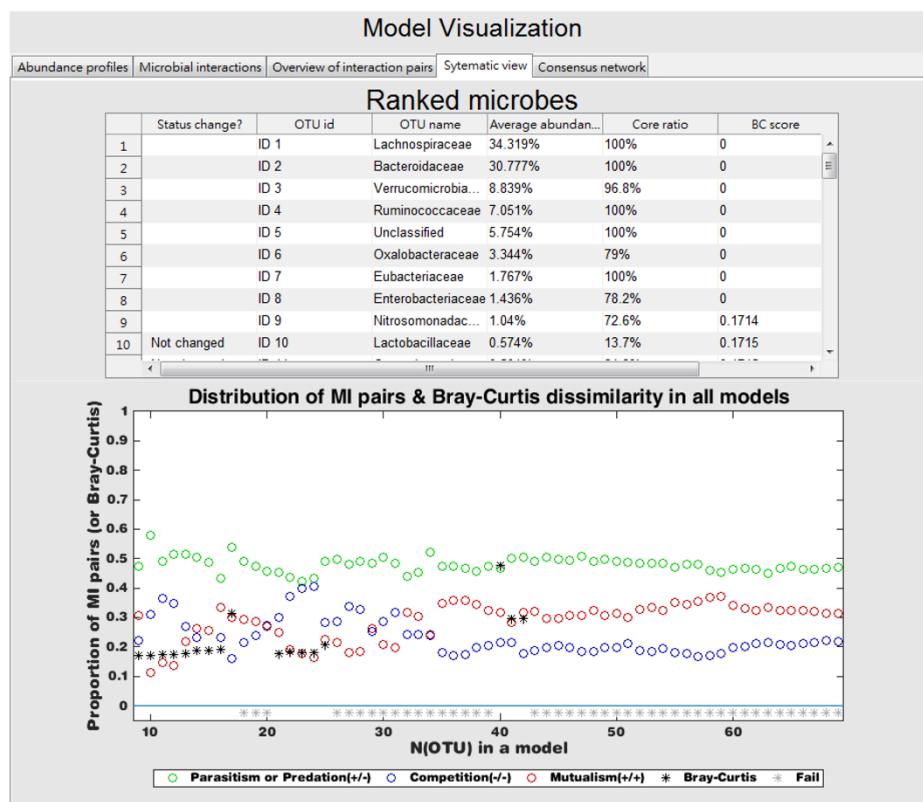


Figure 13 The panel of “systematic view” to examine all interaction networks.

As exemplified in Table 5, *Lactobacillaceae* did not exist in 9-OTU interaction network but in 10-OTU interaction network. The two interaction networks were denoted as success conveying BC scores, 0.1714 and 0.1715 respectively. Consequently, *Lactobacillaceae* would be denoted as “Not changed”. On the other hand, *Porphyromonadaceae* was involved in 18-OTU interaction network but absent in 17-OTU interaction network. The previous one resulted in a successful outcome (BC=0.3141) but the latter was failed (BC=0). Consequently *Porphyromonadaceae* should be annotated as “Status change: failure -> success”.

Table 4 The status change of two sequential interaction networks

Status Change	Network A (with OTU _{lowest abundance})	Network B (without OTU _{lowest abundance})
Not changed	Success	Success
Not changed	Failure	Failure
(Status change: success -> failure)	Success	Failure
(Status change: failure -> success)	Failure	Success

Table 5 The status change of interaction networks from female fecal microbiome.

Status Change?	OTU ID	OTU Name	BC score
	ID 9	<i>Nitrosomonadaceae</i>	0.1714
Not changed	ID 10	<i>Lactobacillaceae</i>	0.1715
Not changed	ID 11	<i>Corynebacteriaceae</i>	0.1745
Not changed	ID 12	<i>Staphylococcaceae</i>	0.1724
Not changed	ID 13	<i>Erysipelotrichaceae</i>	0.1782
Not changed	ID 14	<i>Prevotellaceae</i>	0.1862
Not changed	ID 15	<i>Ectothiorhodospiraceae</i>	0.1863
Not changed	ID 16	<i>Desulfovibrionaceae</i>	0.1907
Not changed	ID 17	<i>Chromatiaceae</i>	0.3141
(Status change: failure -> success)	ID 18	<i>Porphyromonadaceae</i>	0
Not changed	ID 19	<i>Sphingomonadaceae</i>	0
Not changed	ID 20	<i>Enterococcaceae</i>	0
(Status change: success -> failure)	ID 21	<i>Fusobacteriaceae</i>	0.1757

6.1.5 Panel of “Consensus network”

The panel of “Consensus network” is aimed at providing consensus interactions from all interaction networks (Fig. 14). For an interaction pair, M_{ij} , there were n_{ij}^+ and n_{ij}^- interaction networks producing positive and negative outcomes when the interactive direction was fixed. When the ratio of n_{ij}^+ to the summation of n_{ij}^+ and n_{ij}^- was statistically significantly greater than the user-defined threshold for this study, i.e., 90% represented by $P(\text{concordant pairs}) \geq 0.9$ in this panel, we were able to conclude that this interaction relation was concordant among networks and directed positively, and vice versa. One sample z-test for proportions was used to measure the concordance of predicted interactive relations among networks.

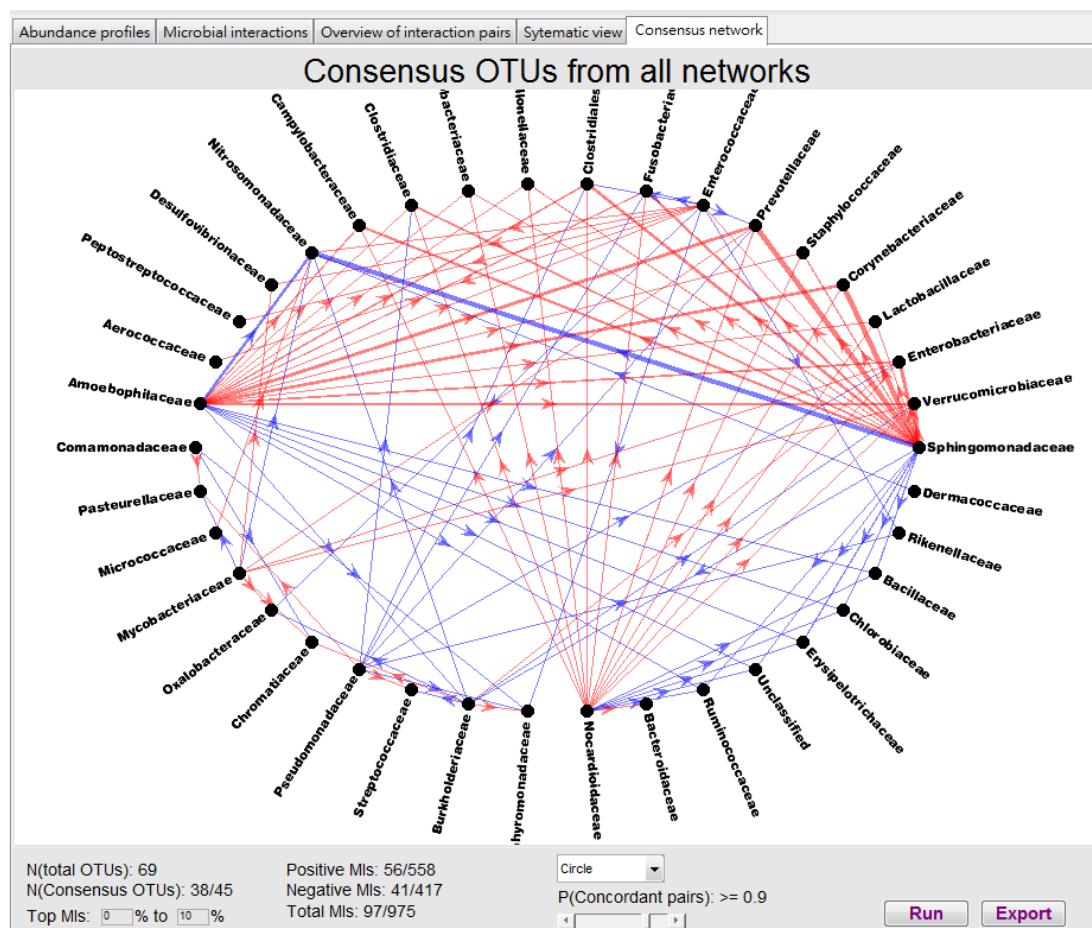


Figure 14 The panel of “Consensus network” can reveal consensus interactions from all interaction networks. The exemplified case was set to show the top 10% of interactions.

There are three optional items to adjust the setting in the panel of “Consensus network” and listed below.

(A) 「Top MIs」 :

This item is used to extract an interaction subnetwork with a specific range of interaction strengths. There are two columns to specify the range of interaction strengths.

For example, if we intended to display the strongest 10% of interactions, the upper and lower bound should be “0” % and “10” % (Fig. 14). On the contrary, the weakest 10% of interactions are extracted by the settings of “90” % and “100” % (Fig. 15).

(B) 「Layout」 | :

Similar to the layout options in the panel of “Microbial interactions”, there are four options, including circle, force, layered, or subspace, in this item.

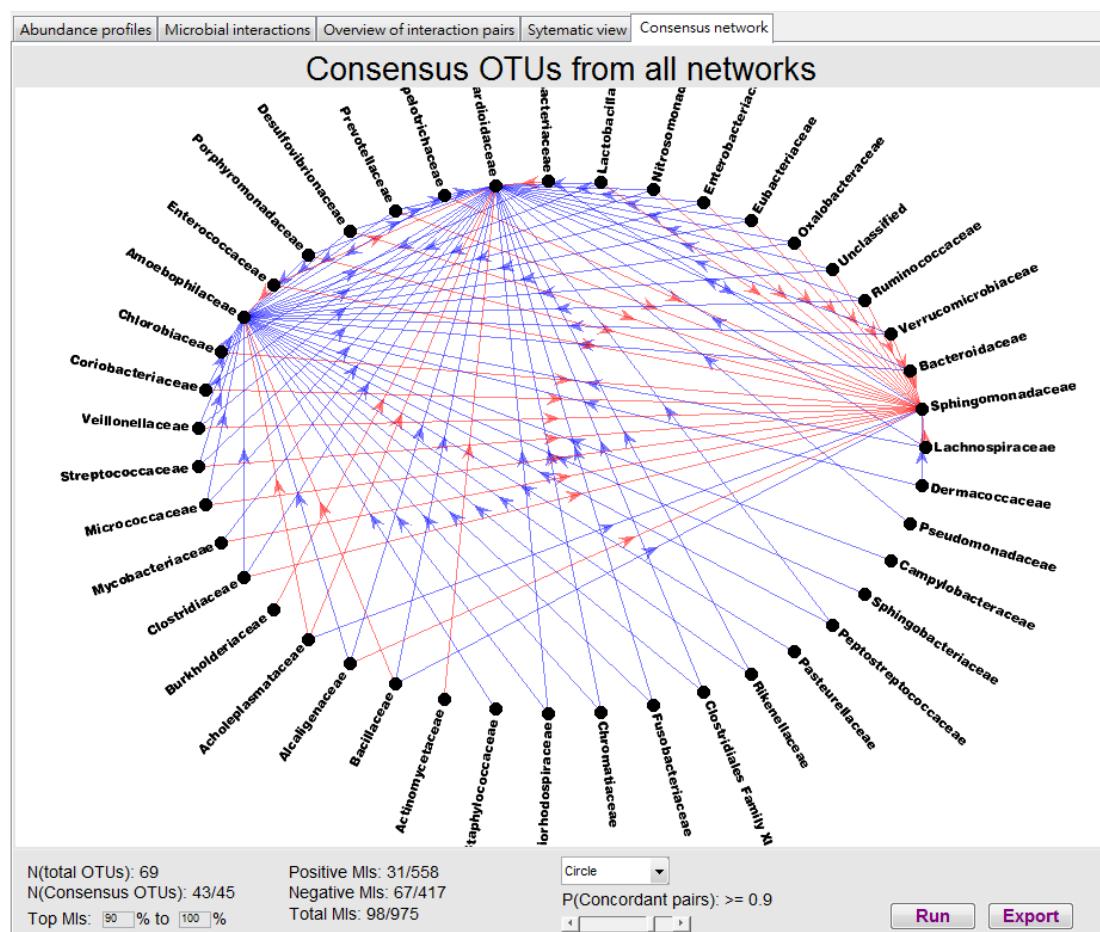


Figure 15 The optional item of “Top MIs” was set to extract the weakest 10% of interaction strengths.

(C) 「P(concordant pairs)」：

P(concordant pairs) is a threshold to determine the interaction consistency from multiple interaction networks. For example, P(concordant pairs) ≥ 0.9 means to collect a set of interaction pairs, usually a source and a target OTUs, with 90% of interactive outcomes being consistent (Fig. 14). If the threshold is raised, fewer interactions are selected (Fig. 16).

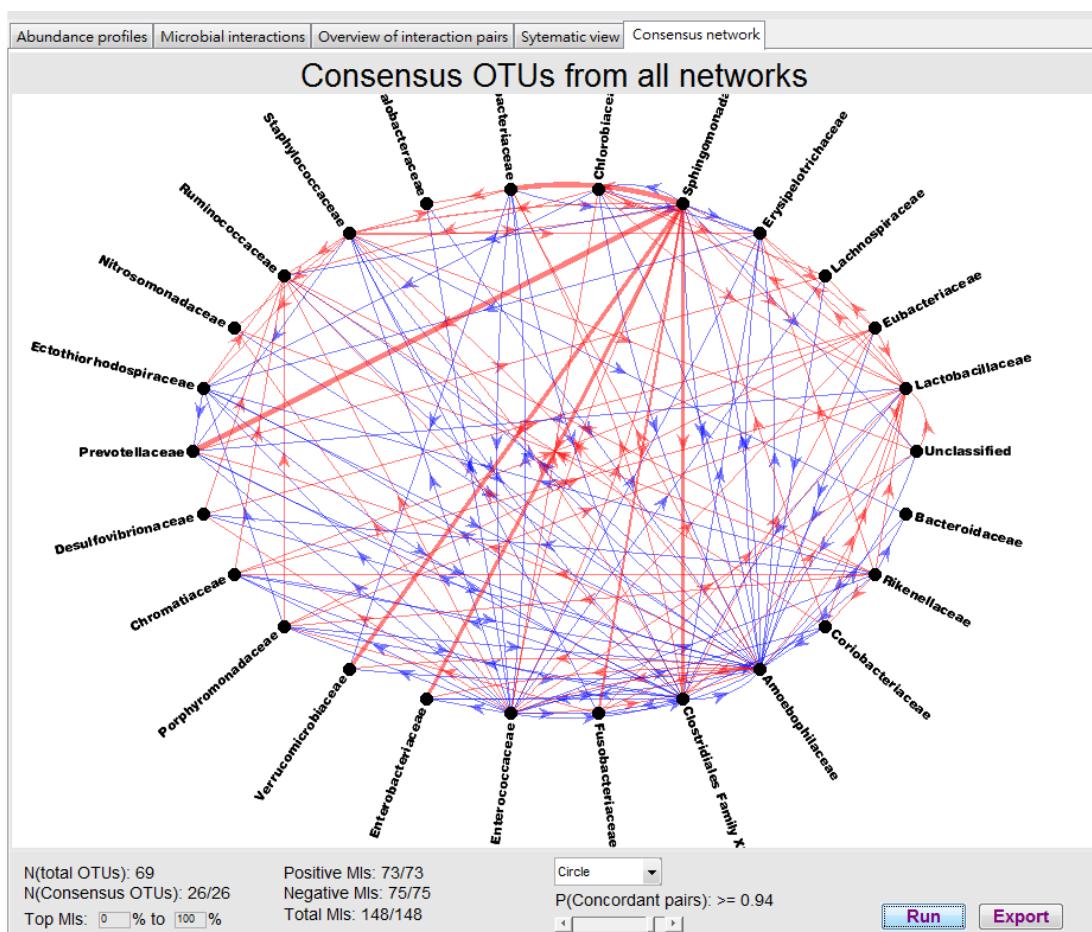


Figure 16 The optional item of “P(concordant pairs)” was set to 0.94 to increase the criteria of selecting consensus interactions.

6.2 Export File

6.2.1 Save to a mat file

After the implement of MetaMIS, all results and parameter settings can be saved into a .mat file by the button of 「Save」 on the left side of MetaMIS (Fig. 2). This .mat file can be loaded directly into the MetaMIS via the button of 「Open」.

6.2.2 Export interaction tables

Exporting the inferred microbial interaction table is an easy step. Choose 「This table」 to export the selected interaction table. Choose 「All tables」 to export all interaction tables (Fig. 6). There are two kinds of tab separated files to be produced. [Filename_EDGE_N\(OTUs\).txt](#) storages the inferred microbial interaction table, e.g. F_gut_EDGE_9.txt. Each OTU identifier has an average abundance value and a core ratio across samples, been reserved in the file of [Filename_NODE_N\(OTUs\).txt](#), e.g. F_gut_NODE_9.txt.

6.2.3 Export consensus interaction tables (**New functionality in MetaMIS_v1.02**)

In the panel of consensus network (section 6.1.5), the button of 「Export」 is to export a consensus microbial interaction table under specific parametric settings, called [MI_CON_TopXtoXX\(P\(concordant pairs\)\).txt](#). The consensus network following the parametric settings of Figure 14 was saved as [MI_CON_Top0to10\(P0.90101\).txt](#).

6.2.4 How to import an interaction table to Gephi [5]?

Step 1: Double click the icon  Gephi

Step 2: Create a new project (Fig. 17)

Step 3: To import the interactions, please follow the indicators in the Fig. 18

Step 4: Users can follow similar steps to import F_gut_NODE_9.txt, which provides OTU information about its average abundance and core ratio among samples. OTU information may help users to understand the topological network.

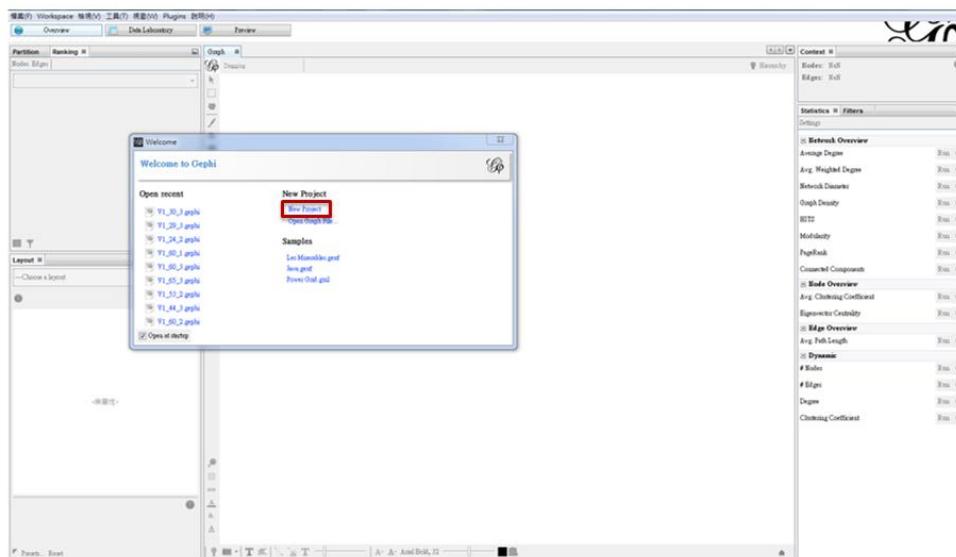


Figure 17 Create a new project in Gephi.

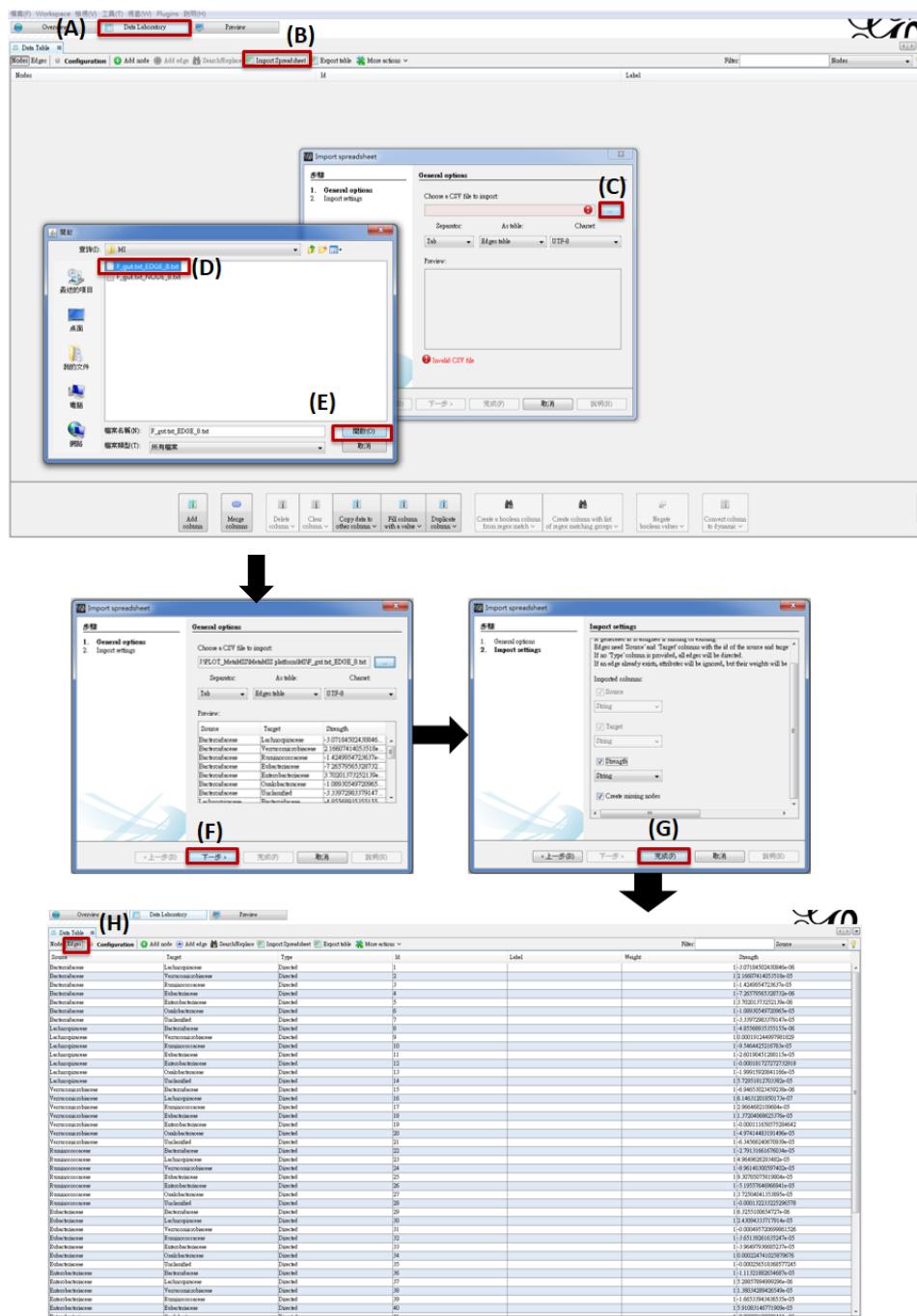
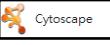


Figure 18 The process of importing an interaction table into Gephi.

6.2.4 How to import an interaction table to Cytoscape [6]?

Step 1: Double click the icon 

Step 2: Create a new project and follow the process in Figure 19.

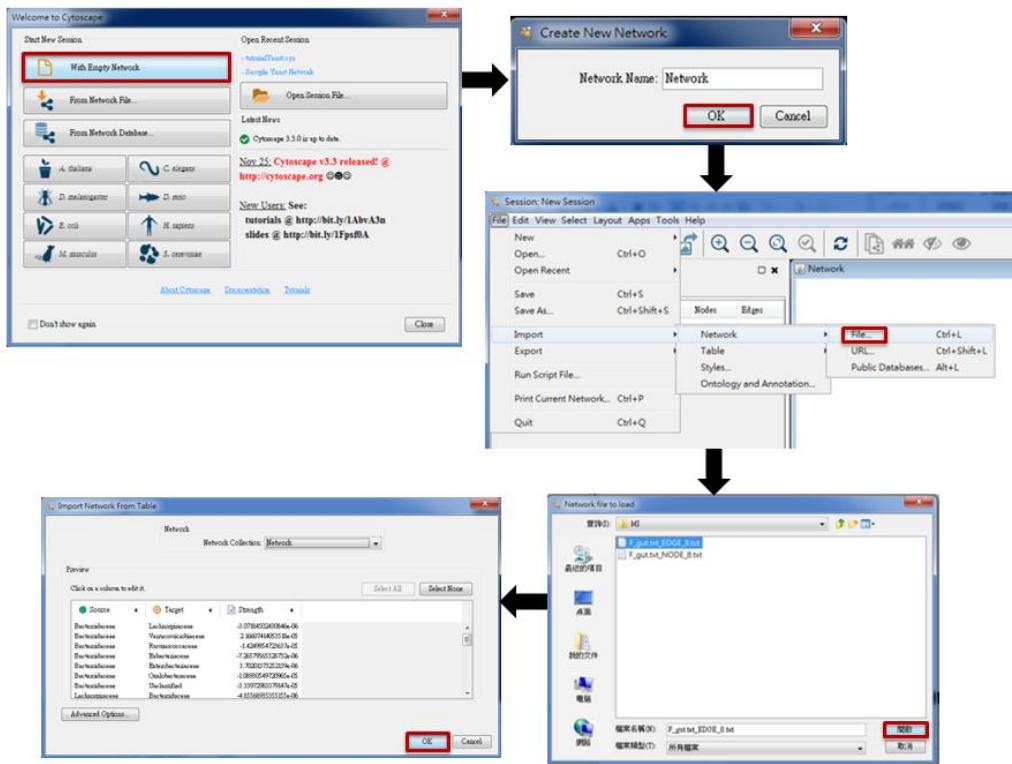


Figure 19 The process of importing an interaction table into Cytoscape.

Bugs or problems

Encounter bugs, problems, or have any suggestions? Please contact Grace Tzun-Wen Shaw (tzunwen@gmail.com)

References

1. Shaw GTW, Pao YY, Wang D: **MetaMIS: a metagenomic microbial interaction simulator based on microbial community profiles.** *BMC Bioinformatics* 2016, **17**:488.
2. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, et al: **Moving pictures of the human microbiome.** *Genome Biol* 2011, **12**:R50.
3. Angly FE, Dennis PG, Skarszewski A, Vanwonterghem I, Hugenholtz P, Tyson GW: **CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction.** *Microbiome* 2014, **2**.
4. Jansen W: **A permanence theorem for replicator and Lotka-Volterra systems.** *Journal of Mathematical Biology* 1987, **25**:411-422.
5. Bastian M, Heymann S, Jacomy M: **Gephi: an open source software for exploring and manipulating networks.** *ICWSM* 2009, **8**:361-362.
6. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: A software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**:2498-2504.