

SOFTWARE

Open Access



MetaMIS: a metagenomic microbial interaction simulator based on microbial community profiles

Grace Tzun-Wen Shaw, Yueh-Yang Pao and Daryi Wang*

Abstract

Background: The complexity and dynamics of microbial communities are major factors in the ecology of a system. With the NGS technique, metagenomics data provides a new way to explore microbial interactions. Lotka-Volterra models, which have been widely used to infer animal interactions in dynamic systems, have recently been applied to the analysis of metagenomic data.

Results: In this paper, we present the Lotka-Volterra model based tool, the **Metagenomic Microbial Interaction Simulator (MetaMIS)**, which is designed to analyze the time series data of microbial community profiles. MetaMIS first infers underlying microbial interactions from abundance tables for operational taxonomic units (OTUs) and then interprets interaction networks using the Lotka-Volterra model. We also embed a Bray-Curtis dissimilarity method in MetaMIS in order to evaluate the similarity to biological reality. MetaMIS is designed to tolerate a high level of missing data, and can estimate interaction information without the influence of rare microbes. For each interaction network, MetaMIS systematically examines interaction patterns (such as mutualism or competition) and refines the biotic role within microbes. As a case study, we collect a human male fecal microbiome and show that *Micrococcaceae*, a relatively low abundance OTU, is highly connected with 13 dominant OTUs and seems to play a critical role. MetaMIS is able to organize multiple interaction networks into a consensus network for comparative studies; thus we as a case study have also identified a consensus interaction network between female and male fecal microbiomes.

Conclusions: MetaMIS provides an efficient and user-friendly platform that may reveal new insights into metagenomics data. MetaMIS is freely available at: <https://sourceforge.net/projects/metamis/>.

Keywords: Metagenomics, Lotka-Volterra, Network dynamics

Background

Propelled by 16S ribosomal RNA (rRNA) sequencing technologies, there has recently been a growing interest in characterizing the role of complex microbial communities in a diverse ecosystem. As a result, an increasing number of samples from marine, soil [1], animal feces, and mammalian gut microflora [2] has been placed in the public domain. Studies have shown that health status, habitat types, and external perturbations are some of the key factors that can change a microbial community in specific ecosystem niches. For instance, the human gut harbors a vast number of microbial species, and imbalances in the intestinal microbiome have been linked with such chronic diseases as

obesity [3], inflammatory bowel disease [4], and type 2 diabetes [5]. Marine microbes sensitive to changing climates also play an important role in ocean feedback, being associated with such phenomena as surface warming, ice melting, and acidification, as well as climate change [6]. From the human gut to global oceans, metagenomic studies offer new insights into compositional stability. However, a deeper investigation into microbial interactions, including mutualism (+/+), competition (-/-), parasitism or predation (+/-), commensalism (+/0), and amensalism (-/0), as reviewed by Faust and Raes [7], is required to fill gaps in understanding of the relationships between microbial communities and hosts or environments. Fortunately, with recent efforts on bioinformatics, some computational approaches using metagenomic data have suggested that association networking and modeling show promise as tools for characterizing

* Correspondence: dywang@gate.sinica.edu.tw
Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan



multilevel interactions and elucidating the temporal dynamics exhibited by microbial communities.

Discerning the full extent of the web of microbial interactions is a difficult task. The conventional approach is to observe the growth behavior in mixed cultures of only a very few microorganisms [8]. Recently, high-throughput interaction inference approaches, such as Sparse Correlations for Compositional data (SparCC) [9], the Learning Interactions from Microbial Time Series (LIMITS) algorithm [10], co-occurrence networks [11], the SParse Inverse Covariance estimation for Ecological Association Inference (SPIEC-EASI) [12], and the Rule-based Microbial Network (RMN) algorithm [13], have been proposed for modeling microscale dynamics using 16S rRNA marker gene sequences. These approaches may be roughly divided into two categories. Correlation-based methods, including SparCC [9] and co-occurrence networks [11], aim to develop algorithms that combine correlation methods in order to decipher highly dependent temporal microbial communities that have usually proved refractory to classical correlation analysis. Although correlation is straightforward and easy to conduct, it nevertheless does not seem to be a proper measure of species interactions, and is limited to inferring non-directional interactions [11, 12]. Modeling-centered approaches, on the other hand, including the LIMITS [10], SPIEC-EASI [12], and RMN [13] algorithms, rest on special biological assumptions and statistical techniques, and usually employ a combined strategy in order to infer microbial interactions. LIMITS, for instance, combines a sparse linear regression with a bootstrapping strategy in order to incorporate interactive relations iteratively into an interaction network [10]. SPIEC-EASI assumes the underlying ecological association network to be sparse and accordingly relies on sparse inverse covariance selection and a neighborhood selection strategy to reconstruct a non-directional interaction network [12]. The RMN algorithm bypasses the NP-hard problem of finding a network with the optimum number of interactions and proceeds directly to the construction of a triplet subnetwork in which the triplet has a convergent recipient that is repressed by one interaction and simultaneously activated by another [13].

Although much work has been done to date, more study is necessary to ascertain the effects of inferring a direct comprehensive interaction network on a variety of network inference methods. Among the methods mentioned above, the LIMITS and RMN algorithms offer a more sound theoretical basis for inferring a direct interaction network, but cause complications for the comprehensive inference of an interaction network. To that end, early attempts at exploiting a direct comprehensive interaction network from microorganisms have been successfully conducted using the Lotka-Volterra model, as first proposed by Jansen [14] and commonly employed by ecologists, which can describe

systematically a dynamic trophic web of more than two macro-organism populations. When applied to a metagenomic abundance generator, the Lotka-Volterra model can successfully generate a simulated microbial community given a set of known interspecies interactions [10, 11, 15]. When applied to simulating microbial interactions, recent studies on lake ecosystem [16], mouse intestine [17] and cheese-making environments [18] have shown that Lotka-Volterra equations can quantify microbial interactions and successfully predict microbiome temporal dynamics. Moreover, a previous study has demonstrated that the distribution of simulated interaction pairs in an ecological system can be used to predict microbiome stability. For instance, a cooperative network of microbes (i.e., one characterized by mutualism) is often unstable, while a higher proportion of competitive interaction pairs (-/-) helps the host to maintain a stable microbial community [15]. Thus the Lotka-Volterra model, which, as mentioned, is commonly used to illustrate the dynamics of macro-ecological communities, may shed light on the complex world of microbial communities.

Detecting and investigating the structure of interactions in microbial ecosystems is, then, absolutely critical, but the reconstruction of ecosystem-wide association networks using the Lotka-Volterra model is far from straightforward. Here we present a stand-alone tool called MetaMI that aims to facilitate the systematic inference of microbial interactions. The characteristics of MetaMIS are as follows. (i) User-friendly interface: we have constructed an easy-to-use graphic user interface (GUI) for scientists, even those who lack programming skills, to infer microbial interactions. (ii) Network topological visualization: MetaMIS offers two ways to visualize the inferred microbial interactions. If there are N microbes in an interaction network, a general view includes the minimum number of interaction pairs to describe N microbes. A specific view of a single microbe takes into account the interactive behaviors of one microbe in relation to all others. (iii) Maximal detection of rare population: while rare species are usually regarded as noise in most quantitative ecological analysis, MetaMIS provides the opportunity to evaluate the fitness of each rare species in a microbial system by means of an abundance-ranking strategy. (iv) Consensus network: MetaMIS is able to unify multiple interaction networks into a confident network.

To provide a user friendly interface, MetaMIS was designed to accept microbial abundance profiles in regular text format on both Mac and Windows (64-bit) platforms. MetaMIS has been tested using a human male intestinal microbiota dataset composed of 317 time points and 92 microbes at the family level and produced 27 prediction models in around 5 min on a current desktop computer. MetaMIS generates outputs in several formats that can be used with other popular network visualization software, such as Gephi [19] and Cytoscape [20]. The central purpose

of MetaMIS is to provide clues about the interactions among microbes and about specific microbes in a microbial community. To our knowledge, no similar tool is available. MetaMIS is free to the public and can be accessed at a public IP address space without any login requirement: <https://sourceforge.net/projects/metamis/>.

Implementation

MetaMIS: overview

The central organizing metaphor of MetaMIS is the construction of microbial interaction networks, with microbial members, i.e., operational taxonomic units (OTUs). The network is presented with nodes and directed edges, in which nodes are OTUs and directed edges are inferred microbial interactions from source to target. The network is constructed based on Lotka-Volterra dynamics (Eq. (1)), which is a conventional way of investigating fluctuations in the populations of wild animals. MetaMIS is the first tool for inferring metagenomic microbial interactions in manner that is automatic and allows for the direct visualization of microbial interaction networks through a user-friendly interface. Figure 1 outlines the rationale of MetaMIS; Fig. 2 depicts the workflow and key features of MetaMIS using screenshots; and Fig. 3 provides a schematic representation of the interrelationships among these features. The detail operation of MetaMIS is introduced in the supplementary (Additional file 1).

The foundation of MetaMIS was the inference of microbial interactions following an abundance-ranking strategy (Fig. 1) that involves ranking OTUs according to their average abundance levels among samples, generating multiple interaction networks and retaining the maximum number of low abundance OTUs in an interaction network (Fig. 1b). This strategy was derived in a straightforward fashion from an empirical rule that dominant microbes are most likely to be observed and analyzed in experimental microbial abundance profiles, and this approach greatly simplifies the complex problem of finding a conserved interaction sub-network. For each interaction network, there were two possible outcomes (a successful or failed interaction network (Fig. 1a) realized by a generalized form of Lotka-Volterra equation (Eq. (2)). A set of predicted interactions that could successfully regenerate abundance profiles within the prescribed period of time constituted a successful network. Otherwise, failure could be due to inaccurate inference of microbial interactions. The regenerated abundance profiles (successful cases) should be further compared with the original data based on the Bray-Curtis dissimilarity (Eq. (3)). A smaller Bray-Curtis dissimilarity (BCD) would mean that interactions could reproduce microbial abundance similar to the original and were more likely to reveal the underlying interactive relations of a microbial community. These processes are easy to carry out using the user-friendly interface of MetaMIS (Fig. 2).

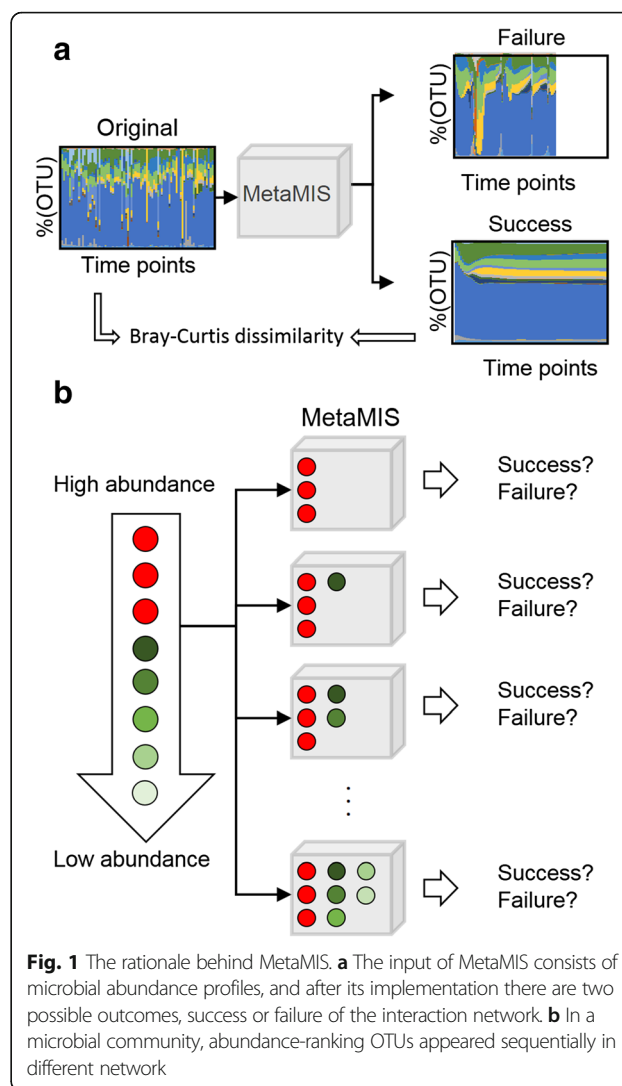


Fig. 1 The rationale behind MetaMIS. **a** The input of MetaMIS consists of microbial abundance profiles, and after its implementation there are two possible outcomes, success or failure of the interaction network. **b** In a microbial community, abundance-ranking OTUs appeared sequentially in different network

Results

Case study: human intestinal microbiome

In the case study, human fecal microbiomes were collected daily from two healthy subjects, one female, for 6 months, and one male, for 15 months [21], which are publicly available at MG-RAST:4457768.3-4459735.3. The male fecal microbiomes containing more time points were used to demonstrate the functionality of MetaMIS. We constructed 27 interaction networks in total over a span of 420 days, the most compact of which was composed of 14 high abundance families. *Micrococcaceae*, the least abundant among the 14 families, influenced the other 13. According our calculations, *Micrococcaceae* repressed *Oxalobacteraceae*, *Bacteroidaceae*, *Porphyromonadaceae*, *Rikenellaceae*, *Eubacteriaceae*, *Lachnospiraceae*, *Ruminococcaceae*, and *Verrucomicrobiaceae*, but activated *Neisseniaceae* and *Prevotellaceae*. Comparative analysis of the male and female fecal microbiomes using MetaMIS revealed a consensus interaction network.

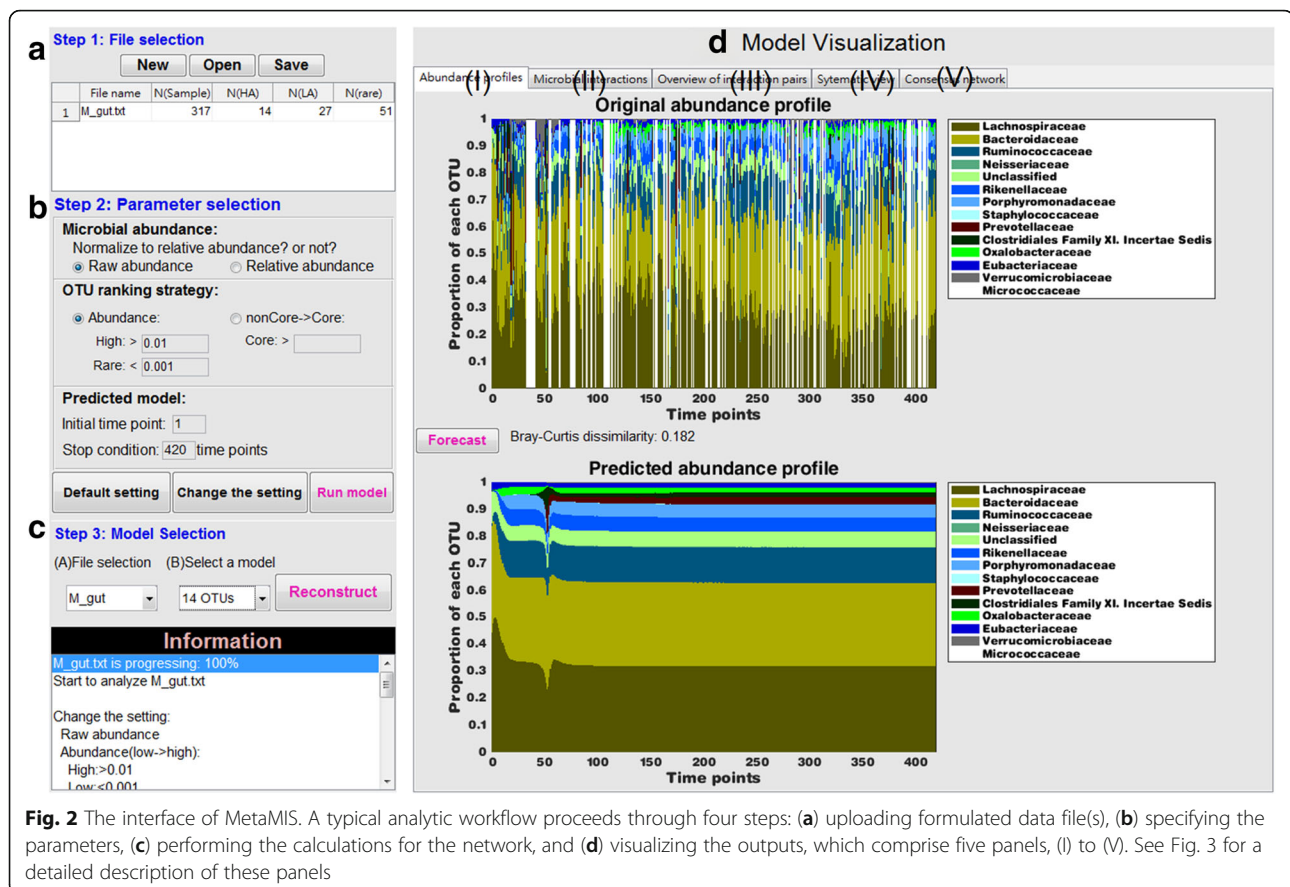


Fig. 2 The interface of MetaMIS. A typical analytic workflow proceeds through four steps: (a) uploading formulated data file(s), (b) specifying the parameters, (c) performing the calculations for the network, and (d) visualizing the outputs, which comprise five panels, (I) to (V). See Fig. 3 for a detailed description of these panels

Functionality of MetaMIS

Using the greengenes taxonomy, the total number of taxa assigned to the family level was 92 over 317 time points for the male fecal microbiome [21]. Using the default settings of MetaMIS, we detected 14 high abundance families, 22 that were low abundance and not rare, and 56 rare families (Fig. 1a), with a total of 27 interaction networks. Results from an interaction network with the 14 most abundant families are schematized in Fig. 3a–g. In general, the original abundance profiles (Fig. 3a), measured by Eq. (1), seem to present more fluctuation than the predicted ones (Fig. 3b), which were generated by Eq. (2). For each interaction outcome, MetaMIS displayed an interaction network containing the minimum number of strongest interactions to cover all families in this network (Fig. 3D-1): Global view). MetaMIS provides a scrolling bar for users to modify more or less interactions according to interactive strength. In brief, the global interaction network showed *Micrococcaceae* was the least abundant among the 14 families (Table 1), but played the most influential role in the system. A specific view served to display the overall interactive relations of *Micrococcaceae* with the other 13 families (Fig. 3D-2: Specific view). *Micrococcaceae* showed strong negative relations with eight bacterial families, *Oxalobacteraceae*, *Bacteroidaceae*,

Porphyromonadaceae, *Rikenellaceae*, *Eubacteriaceae*, *Lachnospiraceae*, *Ruminococcaceae*, and *Verrucomicrobiaceae*, and was positively associated with *Neisseriaceae* and *Prevotellaceae* (Fig. 3D-2). In the specific view, weaker interactions with *Micrococcaceae* could be observed with clarity. *Micrococcaceae* acted as a regulator that strongly influenced the other families but was only slightly influenced by them (Fig. 3D-2). It is worth noting that *Micrococcaceae* tended to repress core microbes but to activate none-core taxa (Table 1).

Furthermore, three approaches were used to visualize the interactive relations between one microbe and the others (Fig. 3e–g). The most frequent interactive relation for *Micrococcaceae*, i.e., ID14, was parasitism or predation (+/-), as shown in Fig. 3e. The interactive strength of each interaction pattern is shown in Fig. 3f. According to the PCA decomposition of the frequency of interaction patterns, ID14 is located in the direction of parasitism or predation (+/-) and amensalism (-/0) (Fig. 3g).

Among 27 successful interaction networks, 18 demonstrated similar predictive power, with BCD (Eq. (3)) ranging from 0.18 to 0.22 (Fig. 3h): 14-OTU, 21-OTU, 37-OTU, ..., and 52-OTU. Other than the 14-OTU and 21-OTU interaction networks, 16 rare families participated sequentially in the remaining successful networks, from 37-OTU to 52-

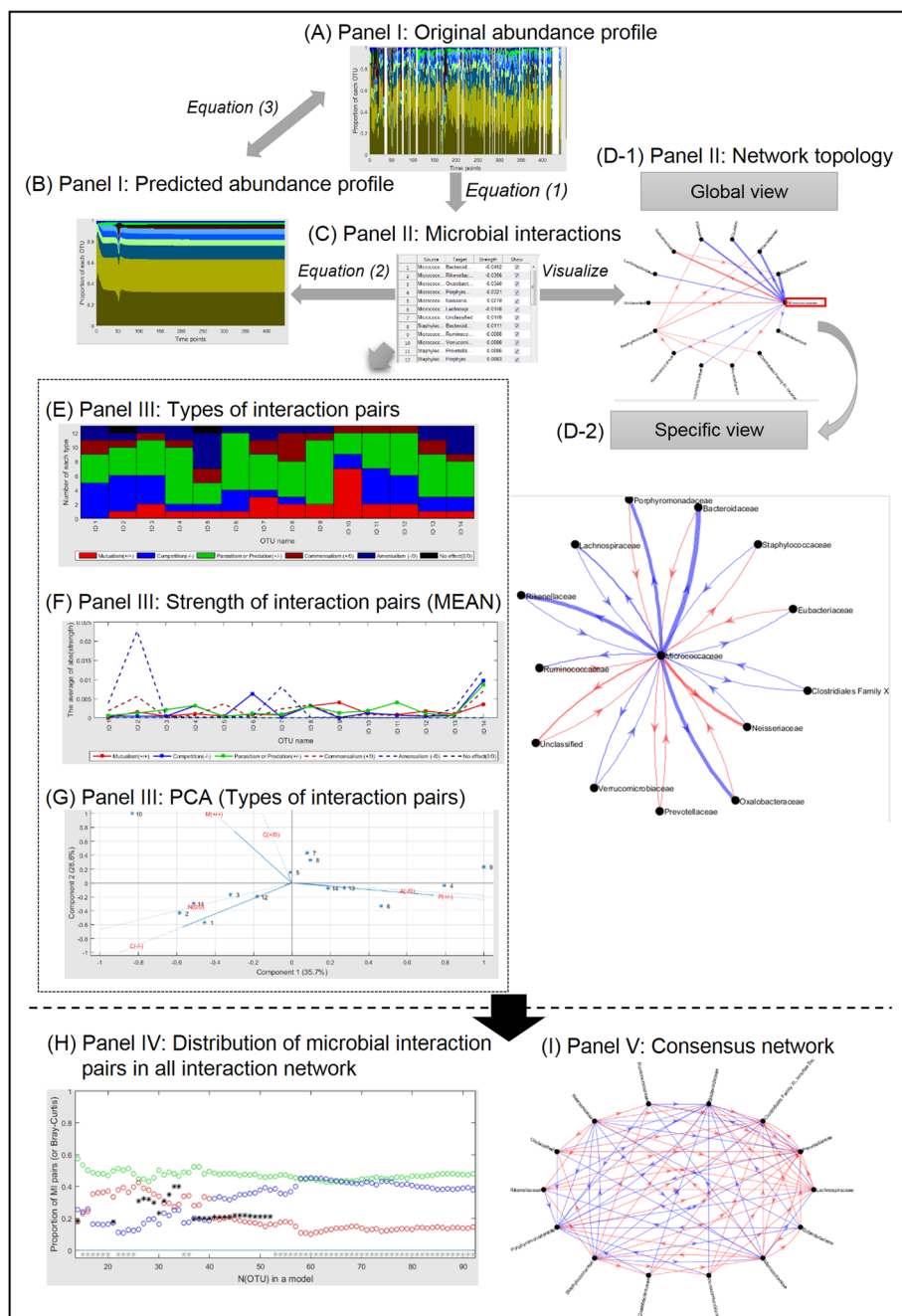


Fig. 3 The analytic schema of MetaMIS. Panel I contains the original (a) and predicted (b) abundance profiles. Inferred microbial interactions are displayed in tabular form (c) and topologically (d), as shown by the global (D-1) and specific views (D-2) in Panel II. Panel III summarizes the distribution of interaction patterns (e) and their interactive strength (f) for each microbe. The PCA plot is intended to help users to identify key microbes (g). Panel IV provides a systematic diagram (h) to monitor and compare the performance from diverse interaction networks. Panel V displays a consensus network (i) in which interactions have more consensus directions among interaction networks

OTU. Among these 16 rare families, *Coriobacteriaceae* (core = 85.8%), *Acidaminococcaceae* (core = 76.3%), and *Clostridiaceae* (core = 98.7%) were frequently present at the 317 time points (Table 1) and showed different abundance profiles with others (measured by Pearson correlation among microbial members (0.06, $p = 0.52$)).

Examining the dependency of interacting pairs

As noted, Lotka-Volterra models have been commonly used to infer animal interactions in ecological studies. For this tool, we applied the Lotka-Volterra model to the investigation of microbial interactions, and further provided a validation calculation by measuring the metabolic complementarity index

Table 1 The male intestinal microbiome was ranked according to the average abundance among 317 time points

MetaMIS ID	Microbial name	Average abundance	Core ratio
ID 1	<i>Lachnospiraceae</i>	24.89%	100.00%
ID 2	<i>Bacteroidaceae</i>	24.61%	100.00%
ID 3	<i>Ruminococcaceae</i>	9.97%	100.00%
ID 4	<i>Neisseriaceae</i>	6.00%	2.20%
ID 5	<i>Unclassified</i>	4.34%	100.00%
ID 6	<i>Rikenellaceae</i>	3.98%	100.00%
ID 7	<i>Porphyromonadaceae</i>	3.60%	99.40%
ID 8	<i>Staphylococcaceae</i>	3.07%	1.30%
ID 9	<i>Prevotellaceae</i>	2.53%	63.40%
ID 10	<i>Clostridiales Family XI. Incertae Sedis</i>	2.23%	54.30%
ID 11	<i>Oxalobacteraceae</i>	1.40%	95.60%
ID 12	<i>Eubacteriaceae</i>	1.30%	99.70%
ID 13	<i>Verrucomicrobiaceae</i>	1.21%	76.00%
ID 14	<i>Micrococcaceae</i>	1.04%	0.90%
ID 15	<i>Actinomycetaceae</i>	0.77%	23.70%
ID 16	<i>Leptotrichiaceae</i>	0.67%	0.90%
ID 17	<i>Enterobacteriaceae</i>	0.56%	82.00%
ID 18	<i>Veillonellaceae</i>	0.54%	55.50%
ID 19	<i>Fusobacteriaceae</i>	0.51%	16.40%
ID 20	<i>Nitrosomonadaceae</i>	0.50%	59.90%
ID 21	<i>Dietziaceae</i>	0.45%	22.40%
ID 22	<i>Campylobacteraceae</i>	0.43%	42.90%
ID 23	<i>Pasteurellaceae</i>	0.39%	18.30%
ID 24	<i>Sporolactobacillaceae</i>	0.36%	0.60%
ID 25	<i>Chromatiaceae</i>	0.33%	94.00%
ID 26	<i>Sphingobacteriaceae</i>	0.28%	98.70%
ID 27	<i>Corynebacteriaceae</i>	0.27%	21.50%
ID 28	<i>Flavobacteriaceae</i>	0.26%	1.30%
ID 29	<i>Streptococcaceae</i>	0.25%	65.90%
ID 30	<i>Desulfovibrionaceae</i>	0.24%	65.00%
ID 31	<i>Erysipelotrichaceae</i>	0.23%	98.70%
ID 32	<i>Dermacoccaceae</i>	0.19%	3.50%
ID 33	<i>Blattabacteriaceae</i>	0.16%	4.70%
ID 34	<i>Ectothiorhodospiraceae</i>	0.13%	0.60%
ID 35	<i>Nocardioideaceae</i>	0.12%	9.50%
ID 36	<i>Aerococcaceae</i>	0.11%	9.10%
ID 37	<i>Coriobacteriaceae</i>	0.10%	85.80%
ID 38	<i>Spirochaetaceae</i>	0.10%	0.90%
ID 39	<i>Acidaminococcaceae</i>	0.09%	76.30%
ID 40	<i>Intrasporangiaceae</i>	0.09%	0.60%
ID 41	<i>Clostridiaceae</i>	0.09%	98.70%
ID 42	<i>Acholeplasmataceae</i>	0.08%	0.30%
ID 43	<i>Rhodocyclaceae</i>	0.07%	0.60%
ID 44	<i>Nocardiaceae</i>	0.07%	1.90%

Table 1 The male intestinal microbiome was ranked according to the average abundance among 317 time points (*Continued*)

ID 45	<i>Comamonadaceae</i>	0.06%	5.40%
ID 46	<i>Bradyrhizobiaceae</i>	0.06%	4.40%
ID 47	<i>Enterococcaceae</i>	0.05%	2.80%
ID 48	<i>Caulobacteraceae</i>	0.05%	1.30%
ID 49	<i>Pasteuriaceae</i>	0.05%	0.30%
ID 50	<i>Gemmatimonadaceae</i>	0.05%	0.90%
ID 51	<i>Burkholderiaceae</i>	0.05%	2.50%
ID 52	<i>Spiroplasmataceae</i>	0.05%	0.30%

The core ratio represents the percentage frequency of one OTU appeared across time-series samples
Only 52 of 92 OTUs are listed hereThe large core ratio represents that this OTU is present in more time-series samples

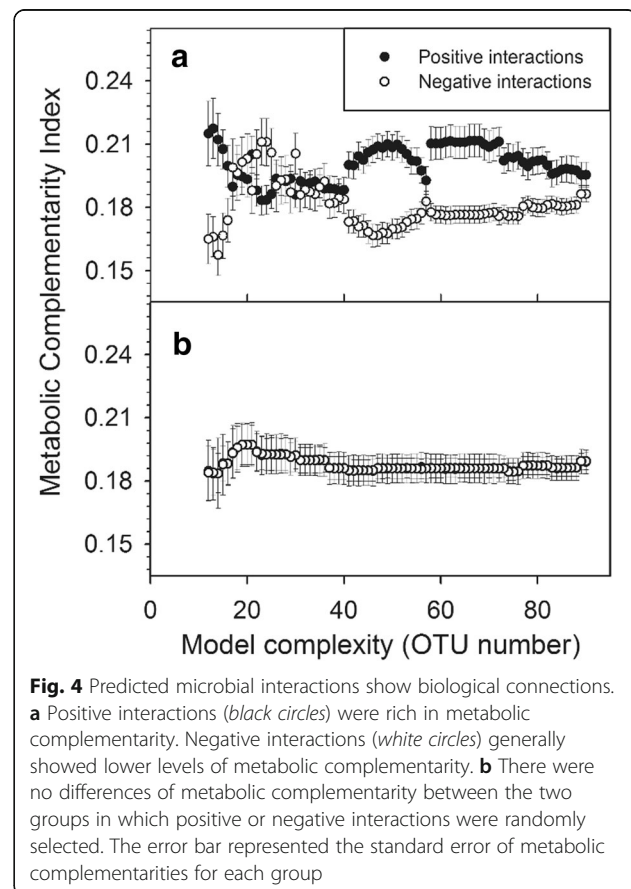
of the datasets. Metabolic complementarity is an index that measures the trophic relations between two microbes based on a metabolic network [22]. The index may reflect the interdependence of each microbe pair, in which the metabolic waste of one microbe is necessary for the other. We observed that positive interactions within the male intestinal microbiome tended to be associated with a larger metabolic complementary index while negative interactions tended to reach a lower level (Fig. 4a). Alternatively, if the interaction of two microbes is set up randomly, the trophic relations will show no significant difference between two groups (Fig. 4b). Thus, the results using male intestinal microbiomes suggested that the inferred interaction was reasonable.

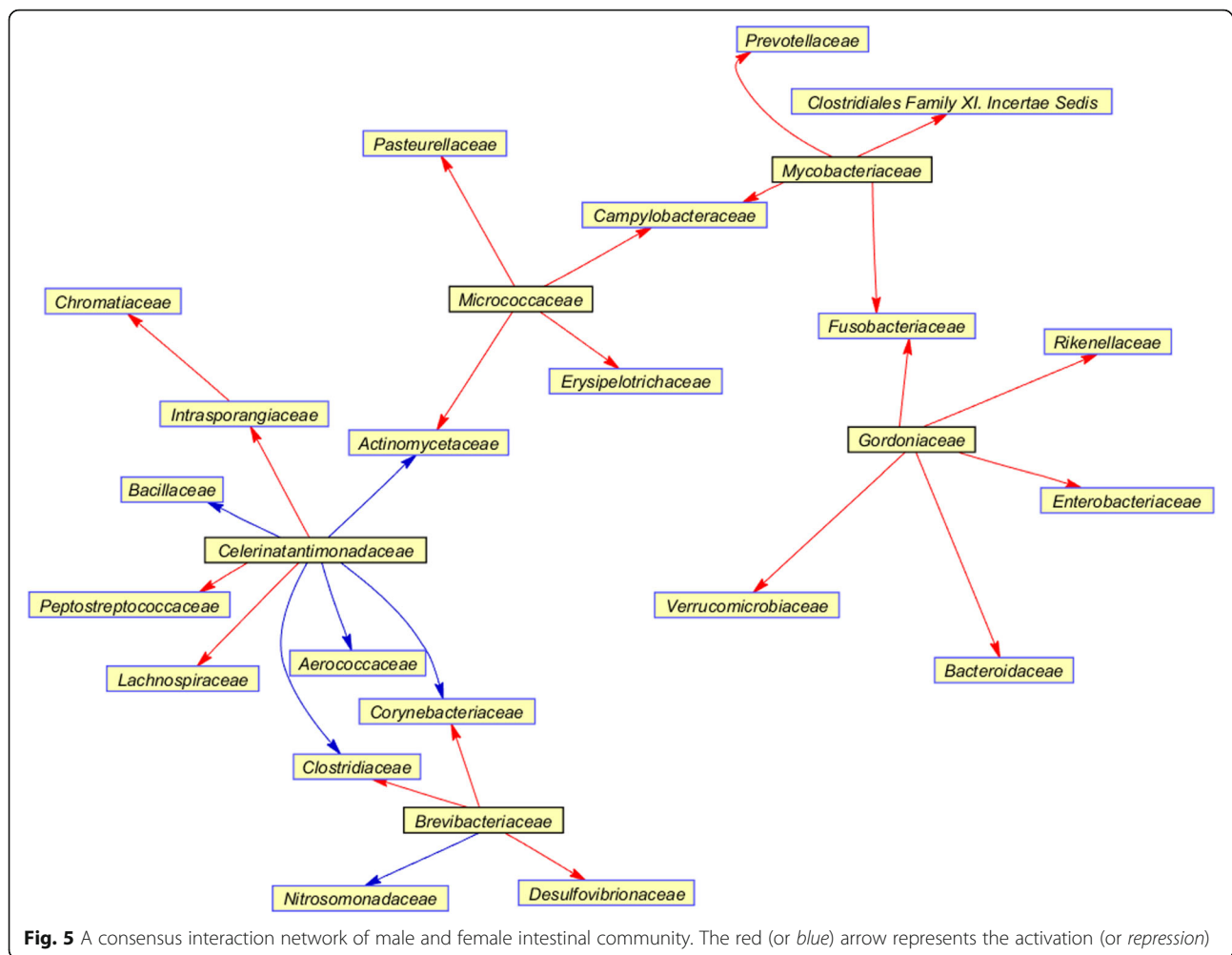
Comparative study

MetaMIS is able to organize multiple interaction networks into a consensus interaction network. In this section, we identify consistent microbial interactions among male and female fecal microbiomes via consensus interaction networks. In the analysis of female fecal microbiome, we focused on the influence of rare or low abundance families on the inference of microbial interactions. The female fecal microbiome contained 9 high, 11 non-rare, and 49 rare families. The latter 60 rare or low abundance families were tested to determine their influence on the high abundance 9-OTU interactive network independently. Our results showed that the female intestinal interactive network (BCD = 0.175) was greatly influenced by rare or low abundance families, 7 out of 60 relatively low abundance OTUs showed significant improved effects in generating the interaction profiles (the median of BCD was 0.167, $p < 0.05$, Student's t test).

For each microbiome (male and female), a consensus interaction network was organized from the comparison of all interaction networks using one sample z -test for proportions, instead of measuring the change of interaction strengths. The female microbiome, containing 69 families over 124 time-series points, in which 63 were overlapped with the male microbiome, generated 1,128 confident positive interactions and 937 negative interactions. The male microbiome produced more interactions in its consensus network, for a total

of 1,618 positive and 2,643 negative interactions. With regard to the absolute interactive strength, 26 stronger interactions among 26 families were coherent between the male and female microbiomes (Fig. 5). The relative abundance or core ratio of 26 families is shown in Table 2. Acting as transmitters, the rare families *Celerinatantimonadaceae*, *Micrococcaceae*, *Brevibacteriaceae*, *Gordoniaceae*, and *Mycobacteriaceae* played key roles to influence others. *Celerinatantimonadaceae* repressed four rare or low abundance non-core families, *Bacillaceae*, *Actinomycetaceae*, *Aerococcaceae*, and *Corynebacteriaceae*, and one rare core families, *Clostridiaceae*. However, *Micrococcaceae* and *Brevibacteriaceae* tended to





activate low level non-core families. *Gordoniaceae* had strong positive association with high abundant core families, *Verrucomicrobiaceae*, *Bacteroidaceae*, *Enterobacteriaceae*, and *Rikenellaceae*. *Mycobacteriaceae* colonized in male intestinal tracts activated two highly abundant non-core families, *Prevotellaceae* and *Clostridiales Family XI. Incertae Sedis*. The community of these highly abundant families, acting as receptors, seemed to be greatly influenced by rare or low abundance microbes. Furthermore, *Micrococcaceae* was also identified as an influential bacterial family, not only in the male 14-OTU interaction network, but also in this consensus interaction network, reflecting its common role in the male and female biomes.

Discussions and conclusions

The Lotka-Volterra equations, which are canonical in mathematical ecology, provide variable ways to illustrate the importance of nonlinear dynamics [23]. Recently Lotka-Volterra models have been applied in the field of metagenomics to investigate microbial interactions because of their usefulness in reverse-engineering multispecies ecosystems

[17, 18]. In this context, these models serve to simulate multi-species microbial communities with known interaction relations [10, 11, 15] that can be adjusted for systematic stability analysis [15]. Recent work, including studies of yeast-bacterium interactions on the surface of cheese [18] and microbial interactions in murine intestinal communities [17], have demonstrated that Lotka-Volterra models can be used to reverse-engineer the interactive behaviors of an ecosystem, even in response to such external perturbations as antibiotic intervention. These studies are important for understanding the application of Lotka-Volterra models to the comprehensive inference of dynamic biological systems in the effort to decipher the interrelationships between species.

In this paper, we have presented a user-friendly, stand-alone GUI tool, MetaMIS, that is designed to provide rapid and accurate predictions of microbial interactions that can help to reveal temporal changes in microbial communities. The integrated diagrammatic presentation can aid in revealing mechanically interactive links between microbes. We offered as examples three interaction networks inferred from a human male, female,

Table 2 OTUs showing consistent microbial interactions between male and female intestines and their taxonomic abundance and core levels

	Female		Male	
	ABUN	CORE	ABUN	CORE
<i>Lachnospiraceae</i>	34.61%	100.00%	24.89%	100.00%
<i>Bacteroidaceae</i>	31.04%	100.00%	24.61%	100.00%
<i>Verrucomicrobiaceae</i>	8.90%	96.77%	1.21%	76.03%
<i>Enterobacteriaceae</i>	1.44%	78.23%	0.56%	82.02%
<i>Nitrosomonadaceae</i>	1.04%	72.58%	0.50%	59.94%
<i>Corynebacteriaceae</i>	0.49%	21.77%	0.27%	21.45%
<i>Erysipelotrichaceae</i>	0.40%	100.00%	0.23%	98.74%
<i>Prevotellaceae</i>	0.37%	83.87%	2.53%	63.41%
<i>Desulfovibrionaceae</i>	0.27%	91.94%	0.24%	64.98%
<i>Chromatiaceae</i>	0.26%	54.03%	0.33%	94.01%
<i>Fusobacteriaceae</i>	0.10%	34.68%	0.50%	16.40%
<i>Clostridiales Family XI. Incertae Sedis</i>	0.09%	59.68%	2.23%	54.26%
<i>Rikenellaceae</i>	0.06%	87.90%	3.98%	100.00%
<i>Pasteurellaceae</i>	0.05%	38.71%	0.39%	18.30%
<i>Peptostreptococcaceae</i>	0.04%	80.65%	0.03%	65.30%
<i>Micrococcaceae</i>	0.04%	3.23%	1.04%	0.95%
<i>Mycobacteriaceae</i>	0.04%	4.03%	0.02%	1.89%
<i>Clostridiaceae</i>	0.03%	90.32%	0.09%	98.74%
<i>Campylobacteraceae</i>	0.03%	30.65%	0.43%	42.90%
<i>Aerococcaceae</i>	0.02%	21.77%	0.11%	9.15%
<i>Bacillaceae</i>	0.02%	49.19%	0.02%	28.08%
<i>Actinomycetaceae</i>	0.02%	29.03%	0.77%	23.66%
<i>Brevibacteriaceae</i>	0.01%	4.84%	0.03%	0.32%
<i>Intrasporangiaceae</i>	0.01%	1.61%	0.09%	0.63%
<i>Gordoniaceae</i>	0.01%	1.61%	0.03%	0.63%
<i>Celerinatantimonadaceae</i>	0.01%	0.81%	0.03%	0.63%

and a mixed-gender fecal microbiome. Those inferred relationships receive some support in the literature. For example, some strains of *Micrococcaceae* have been shown to possess considerable antibacterial activity [24] and antibiotic-resistance ability that counters the inhibitory effect of *Lactobacillus*, *Lact. sake* CL35 [25]. Furthermore, we found that *Micrococcaceae* consistently activated two microbes, *Neisseriaceae* and *Prevotellaceae*, which is consistent with the studies showing that the use of antibiotic agents significantly increases the incidence of members of the *Prevotellaceae* family in the mucosal-associated microbiome [26]. The antimicrobial effect of *Micrococcaceae* [24] and *Neisseriaceae* [27] might therefore balance those dominant microorganisms and thereby help to maintain innate homeostasis and to achieve a more diverse intestinal ecosystem. Overall, these reported microbial functions and characteristics were consistent with the microbial interactions that we inferred.

In the case of consensus network, *Mycobacteriaceae*, which is defined as a rare family in the male microbiome and a non-core family in the female microbiome, and is associated with tuberculosis [28, 29], also exhibited a similar interaction pattern in both genders. On the other hand, several studies have noted that sex hormones and microbes together trigger a gender bias in such autoimmune diseases as type 1 diabetes (T1D) [30] and systematic lupus erythematosus (SLE) [31]. As suggested, the distribution of *Enterobacteriaceae* and *Peptostreptococcaceae* correlated strongly with the concentration of androgen as conditions in which male nonobese diabetic (NOD) mice experienced a lower risk of T1D [30]. However, our data suggest that the role of *Enterobacteriaceae* and *Peptostreptococcaceae* in the male and female samples could be the same considering the interaction patterns with the other microbes. Furthermore, we suggest that, when analyzing metagenomic abundance profiles, considerable care is required in determining the cutoff for low abundance or rare OTUs, informative interactions from low level members may be lost.

Conclusion

In sum, here we have presented an easy-to-follow workflow designed to infer microbial interactions using Lotka-Volterra models for 16S-rRNA microbial abundance profiles. MetaMIS allows researchers to analyze interactive relations conveniently and to visualize network topology directly through an intuitive graphic user interface. The abundance-ranking strategy of MetaMIS produces a variety of interaction networks and allows maximum information to be gathered regarding low-abundance members of the microbial community. Among different interaction networks, users can trace changes in interactive relations or utilize a consensus network that contains a set of OTUs with qualified interactions in order to identify key microbes. The publicly available MetaMIS is expected to undergo continuous development; future plans include: organizing interaction networks across different dataset, establishing topological analyses to extract key OTUs based on their topological nature, plugging in a functional annotated package for microorganisms, and, in the longer term, developing a pathway dependent interaction cascade. We view the current version of MetaMIS as a first step toward facilitating the interpretation of metagenomic studies in the context of the rapidly expanding knowledge of microbial genomes and the growing databases that store that knowledge.

Methods

Implementation

MetaMIS was performed as an off-line GUI coded by a commercial software package (MATLAB R2015b, The MathWorks, Inc., Natick, Massachusetts, United States). It runs properly on Mac and Windows (64-bit) platforms.

Before the execution of MetaMIS, the Matlab runtime should be installed, which is a simple one-click process.

Data preprocessing

Before they use MetaMIS, we recommended that users perform two kinds of data preprocessing for a metagenomic microbial abundance profile. First, 16S rRNA amplicon microbial profiles should be corrected based on 16S rRNA gene copy number (GCN) information, since GCN bias may compromise the accuracy of microbial abundance profiles and significantly influence biological interpretations [32]. Second, microbial abundance profiles should be normalized by transformation to relative abundance, which is done by dividing the minimum number of total reads for all samples, and finally deleting OTUs without abundance values for all samples. The aim of this process is to ascertain which low abundance OTUs are present.

The classification of OTUs according to population size

According to the average abundance across samples in which the zero count was not included in the average calculation, microbial OTUs may be categorized into three groups as follows. The high abundance group is characterized by OTUs with average abundance greater than 1%. Rare species are characterized by an average abundance lower than 0.1%. The remaining organisms are assigned to the low abundance, non-rare group.

The inference of microbial interactions

In a metagenomic microbial abundance profile, there are $i = 1, \dots, L$ microbes or taxonomic labels, i.e. OTUs, and $k = 1, \dots, T$ time points. Time-series samples with total reads smaller than 5,000 are automatically deleted in MetaMIS. Next, a discrete-time Lotka-Volterra model (Eq. (1)) [33] coupled with a partial least square regression (PLSR) is used to infer microbial interactions, from which the number of PLS components containing the minimum estimated mean-squared error is determined. PLSR is a powerful method for handling a highly correlated time series data structure [34].

$$\frac{\ln(x_i(t_{k+1})) - \ln(x_i(t_k))}{t_{k+1} - t_k} = r_i + \sum_{j=1}^L M_{ij} x_j(t_k) \quad (1)$$

where $x_i(t_k)$ represents microbial abundances for any OTU i at the time t_k , r_i is the growth rate of OTU i , and M_{ij} characterizes the interactive effect of OTU j on i . In general, $M_{ij} > 0$ means that OTU j has an activated ability to OTU i , while $M_{ij} < 0$ means that the repressive effect of OTU j on i , and $M_{ij} = 0$ shows no interaction between OTU i and j . Notice that MetaMIS chooses the components as predictors using above method, the result may effect the estimated interaction strengths and signs.

The criteria for a successful interaction network

After microbial interactions have been estimated, they can be placed into a generalized Lotka-Volterra model (Eq. (2)) [33] in order to evaluate the possibility of regenerating microbial profiles over time T . The initial condition can be any time-series sample. The default setting is the first one.

$$\frac{d}{dt} x_i(t_k) = r_i x_i(t_k) + x_i(t_k) \sum_{j=1}^L M_{ij} x_j(t_k) \quad (2)$$

A set of microbial interactions is considered to constitute a successful interaction network when microbial abundance profiles can be successfully regenerated using estimated microbial interactions over time T . If regenerated abundances meet the divergence before the end of the threshold time T , the corresponding microbial interactions represent the failure to form an interaction network.

For each successful interaction network, the concordance between the predicted abundance profiles and the original ones was measured by Bray-Curtis dissimilarity (Eq. (3)) [35].

$$BCD(x_{it_k}, x_{it_k}^*) = \frac{\sum_{i=1}^L |x_{it_k} - x_{it_k}^*|}{\sum_{i=1}^L (x_{it_k} + x_{it_k}^*)} \quad (3)$$

where $x_{it_k}^*$ is the estimated microbial abundance of OTU i at the t_k . This index ranges from 0 and 1. The larger the value, the more dissimilar are the two abundance profiles, and vice versa.

The filtering thresholds for interaction networks

A microbial community with N OTUs can generate $N - N_{HA} + 1$ interaction networks by the default settings of MetaMIS, where $N_{HA} \geq 3$ represents the number of high abundance OTUs. The initial N -dimensional network contained $N(N-1)$ interactions from weakest to strongest in the entire community. Then, an OTU with lowest abundance value was discarded and the remaining $N-1$ OTUs produced $(N-1)(N-2)$ interactions. The strategy of leaving the lowest one out was performed until there were only N_{HA} high abundance OTUs in an interaction network.

For these $N - N_{HA} + 1$ interaction outcomes, one sample z -test for proportions was used to measure the concordance of predicted interactive relations among networks. 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%, we were able to conclude that this interaction relation was concordant among networks and directed positively, and vice versa.

The evaluation of inferred microbial interactions

Microbial interactions predicted by MetaMIS were evaluated using a metabolic complementarity index [22]. The metabolic complementarity ranges from 0 to 1 as a measure of the trophic dependence between two microbes; thus an index of 1 means that all of the nutrients required by one microbe can be synthesized by another microbe from metabolic precursors. In a recent study [22], 23,562 trophic interactions between pairs of microbial species based on genome-scale metabolic network models were transformed to 19,182 microbial interactions on the family level. For each transformation, indices from the common family were averaged.

Mapping the trophic information into predicted interactions from MetaMIS was done in order to explore the underlying biological connections. For each interaction network, network size was a key component of the mapping rate, 9.7 to 62.9% for the male gut and 15.9 to 62.2% for the female gut. A larger network size usually correlated with a lower mapping rate.

Additional file

Additional file 1: User guide of MetsMIS. (PDF 4192 kb)

Abbreviations

BCD: Bray-Curtis dissimilarity; GUI: Graphic user interface; LIMITS algorithm: The Learning Interactions from Microbial Time Series algorithm; MetaMIS: **M**etagenomic **M**icrobial **I**nteraction **S**imulator; NOD mice: Nonobese diabetic mice; OTUs: Operational taxonomic units; RMN algorithm: The Rule-based Microbial Network algorithm; rRNA: ribosomal RNA; SLE: Systemic lupus erythematosus; SparCC: Sparse Correlations for Compositional data; SPIEC-EASI algorithm: The SParse Inverse Covariance estimation for Ecological Association Inference; T1D: Type 1 diabetes

Acknowledgements

We thank Dr. Jason C.H. Shih for his valuable comments.

Funding

This work was supported by the National Science Council of Taiwan [Grant No: MOST 103-2313-B-001-004], and Sustainability Science Research Project, Academia Sinica.

Availability of data and material

All data generated or analysed during this study are included in this published article and its supplementary information files.

Authors' contributions

GTWS developed the tool of MetaMIS and performed the network analysis for the human microbial abundance dataset. YYP assisted the development of MetaMIS. GTWS and DW were major contributors in writing the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Received: 5 July 2016 Accepted: 19 November 2016

Published online: 25 November 2016

References

- Gilbert JA, Meyer F, Jansson J, Gordon J, Pace N, Tiedje J, Ley R, Fierer N, Field D, Kyrpides N, et al. The earth microbiome project: meeting report of the "1 EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6 2010. *Stand Genomic Sci.* 2010;3(3):249–53.
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature.* 2007;449(7164):804–10.
- Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. *Nature.* 2006;444(7122):1022–3.
- Kamada N, Seo SU, Chen GY, Nunez G. Role of the gut microbiota in immunity and inflammatory disease. *Nat Rev Immunol.* 2013;13(5):321–35.
- Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature.* 2012;490(7418):55–60.
- Danovaro R, Corinaldesi C, Dell'anno A, Fuhrman JA, Middelburg JJ, Noble RT, Suttle CA. Marine viruses and global climate change. *FEMS Microbiol Rev.* 2011;35(6):993–1034.
- Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol.* 2012;10(8):538–50.
- Stolyar S, Van Dien S, Hillesland KL, Pineda N, Lie TJ, Leigh JA, Stahl DA. Metabolic modeling of a mutualistic microbial community. *Mol Syst Biol.* 2007;3:92.
- Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLoS Comput Biol.* 2012;8(9):e1002687.
- Fisher CK, Mehta P. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS One.* 2014;9(7):e102451.
- Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front Microbiol.* 2014;5:219.
- Kurtz ZD, Muller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol.* 2015;11(5):e1004226.
- Tsai KN, Lin SH, Liu WC, Wang D. Inferring microbial interaction network from microbiome data using RMN algorithm. *BMC Syst Biol.* 2015;9:54.
- Jansen W. A permanence theorem for replicator and Lotka-Volterra systems. *J Math Biol.* 1987;25(4):411–22.
- Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: networks, competition, and stability. *Science.* 2015;350(6261):663–6.
- Dam P, Fonseca LL, Konstantinidis KT, Voit EO. Dynamic models of the complex microbial metapopulation of lake mendota. *NPJ Syst Biol Appl.* 2016;2:16007.
- Marino S, Baxter NT, Huffnagle GB, Petrosino JF, Schloss PD. Mathematical modeling of primary succession of murine intestinal microbiota. *Proc Natl Acad Sci U S A.* 2014;111(1):439–44.
- Mounier J, Monnet C, Vallaes T, Arditi R, Sarthou AS, Helias A, Irlinger F. Microbial interactions within a cheese microbial community. *Appl Environ Microbiol.* 2008;74(1):172–81.
- Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *ICWSM.* 2009;8:361–2.
- 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(11):2498–504.
- 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(5):R50.
- Levy R, Borenstein E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc Natl Acad Sci U S A.* 2013;110(31):12804–9.
- Edelstein-Keshet L. *Mathematical models in biology*, vol. 46: Siam; 1988.
- Selwyn S, Marsh P, Sethna TN. In vitro and in vivo studies on antibiotics from skin micrococccaceae. Springer: Penicillins and Cephalosporins; 1976. p. 391–6.
- Gonzalez B, Diez V. The effect of nitrite and starter culture on microbiological quality of "chorizo"—a Spanish dry cured sausage. *Meat Sci.* 2002;60(3):295–8.
- Hill DA, Hoffmann C, Abt MC, Du Y, Kobuley D, Kirn TJ, Bushman FD, Artis D. Metagenomic analyses reveal antibiotic-induced temporal and spatial changes in intestinal microbiota with associated alterations in immune cell homeostasis. *Mucosal Immunol.* 2010;3(2):148–58.
- Ogle J, Finck A, Aho E. Production of Antimicrobial Substances by Nonpathogenic Neisseria Species. *FASEB J.* 2015;29(1 Supplement):575–7.
- Nhamoyebonde S, Leslie A. Biological differences between the sexes and susceptibility to tuberculosis. *J Infect Dis.* 2014;209 suppl 3:S100–6.

29. Neyrolles O, Quintana-Murci L. Sexual inequality in tuberculosis. *PLoS Med.* 2009;6(12):e1000199.
30. Yurkovetskiy L, Burrows M, Khan AA, Graham L, Volchkov P, Becker L, Antonopoulos D, Umesaki Y, Chervonsky AV. Gender bias in autoimmunity is influenced by microbiota. *Immunity.* 2013;39(2):400–12.
31. Zhang H, Liao X, Sparks JB, Luo XM. Dynamics of gut microbiota in autoimmune lupus. *Appl Environ Microbiol.* 2014;80(24):7551–60.
32. Angly FE, Dennis PG, Skarshewski 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;2:11.
33. Hofbauer J, Sigmund K. *Evolutionary games and population dynamics.* UK: Cambridge university press; 1998.
34. Wold H. Partial least squares, *Encyclopedia of statistical sciences.* 1985.
35. Lance G, Williams W. Computer programs for hierarchical polythetic classification ("similarity analyses"). *Comput J.* 1966;9(1):60–4.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

