**BMC
Systems Biology**

METHODOLOGY ARTICLE

Open Access

CrossMark

Inferring microbial interaction network from microbiome data using RMN algorithm

Kun-Nan Tsai[1,2], Shu-Hsi Lin[1], Wei-Chung Liu[3] and Daryi Wang[1*] ID

Abstract

Background: Microbial interactions are ubiquitous in nature. Recently, many similarity-based approaches have been developed to study the interaction in microbial ecosystems. These approaches can only explain the non-directional interactions yet a more complete view on how microbes regulate each other remains elusive. In addition, the strength of microbial interactions is difficult to be quantified by only using correlation analysis.

Results: In this study, a rule-based microbial network (RMN) algorithm, which integrates regulatory OTU-triplet model with parametric weighting function, is being developed to construct microbial regulatory networks. The RMN algorithm not only can extrapolate the cooperative and competitive relationships between microbes, but also can infer the direction of such interactions. In addition, RMN algorithm can theoretically characterize the regulatory relationship composed of microbial pairs with low correlation coefficient in microbial networks. Our results suggested that *Bifidobacterium*, *Streptococcus*, *Clostridium XI*, and *Bacteroides* are essential for causing abundance changes of *Veillonella* in gut microbiome. Furthermore, we inferred some possible microbial interactions, including the competitive relationship between *Veillonella* and *Bacteroides*, and the cooperative relationship between *Veillonella* and *Clostridium XI*.

Conclusions: The RMN algorithm provides the reconstruction of gut microbe networks, and can shed light on the dynamical interactions of microbes in the infant intestinal tract.

Keywords: Microbiome, Next-generation sequencing, Microbial regulatory network, Cooperative and competitive relationship, OTU-triplet model, RMN algorithm

Background

Microbes are the most abundant and diverse organisms on earth and their interactions are crucial in understanding both the ecology and the evolution of microorganisms. Microbial interactions, including mutualism, competition, parasitism and commensalism, are difficult to quantify as the underlying processes usually cannot be observed directly and are often too complex for laboratory experiments [1]. However, recent advances in high-throughput sequencing technology have made large scale surveys of microbial communities feasible. Metagenomic studies and network-based approaches have yield detailed information on the composition of microbial communities, which in turn pave the way to study the structure of microbial ecosystems and their dynamics [2–4].

Elucidating competitive and cooperative relationship is a challenge in generating a microbial interaction network because of the direction of such interactions [5]. Competition and cooperation are the two most studied microbial interactions in the recent times with the former dominating the latter in various microbial communities [6–9]. Recent studies have also shown that competitive interactions can drive the evolution of cooperation in microbial ecosystems [9]. Therefore, identifying competitive and cooperative relationships between microbes is profound importance; however, the directional nature of such interactions also poses as a difficult challenge in network construction.

There are several approaches in constructing a microbial network. One commonly used method is the similarity-based network construction where the co-occurrences of two species over multiple time-series samples are measured to infer their interaction [2, 3, 10–12]. In such networks, nodes correspond to organisms and an edge between two

* Correspondence: dywang@gate.sinica.edu.tw
[1]Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan
Full list of author information is available at the end of the article

Tsai *et al. BMC Systems Biology* (2015) 9:54

Page 2 of 10

nodes represents the significant relationship of two taxa across a set of time series samples. Although other derived approaches [13–15] can identify the pairwise relationships between microbes by using correlation estimation, they do not identify the nature and the strength of these relationships [1]. In general, these methods cannot capture the complexity of microbial interactions and cannot elucidate how microbes regulate each other.

To model complex relationships (one species influencing multiple others), methods employing mathematical and statistical models have been developed recently. For example, the generalized Lotka–Volterra framework was used to model the dynamics of microbial populations in order to estimate parameters governing species-species interactions [16–18]. Moreover, other studies have used nonparametric regression models to infer the dynamic relationships between three microbial populations [19, 20]. These regression models use smooth functions to describe microbial relationships and estimate their relative effects. Because these regression analyses do not use *a priori* knowledge to build dynamic microbial interactions, the prediction from those models often lack biological explanations, especially when the microbial community under examination is a complex one [21]. Therefore, how to construct a meaningful microbial interaction network remains a challenge in microbial ecology.

Due to the importance of competitive and cooperative relationships between microbes in complex ecological communities [8, 22, 23], an inference model for constructing microbial regulatory network should consider these interactions simultaneously. However, all of the studies mentioned above do not incorporate these features into their construction of regulatory networks. Association rule mining can be a promising way to infer complex relationships. A rule-based method, named smooth response surface (SRS) algorithm, was developed to construct complex regulatory networks [24]. Basing on a piecewise linear-quadratic polynomial, the SRS algorithm uses a three-dimensional model surface (3D-MS) to successfully infer the relationships between the target, activator, and repressor nodes.

In this paper, we adopt a similar rationale as the SRS algorithm to detect the cooperative and competitive relationships between microbes for constructing microbial regulatory networks. Our new methodology, named the rule-based microbial network (RMN) algorithm, is an advanced version of the rule-based approach. We used Tanh functions for modeling nonlinear responses as regulatory effects of microbes. To our understanding, Tanh functions have been used to model nonlinear responses in some microbial studies, such as dose influences of pathogens [25], regulatory effects of microbes [26], influences of microbial compositions (phospholipids fatty acids)

[27]. The RMN algorithm uses a nonlinear regulatory OTU-triplet (NRO) model to capture cooperative and competitive relationships and consequently infers the directional interaction between microbes. We applied our new approach to simulated data and showed that the RMN algorithm was capable to reconstruct microbial regulatory networks. Furthermore, we used Koenig' data (2011) to construct the interaction network for the microbial community in infant gut [28]; and our results suggest that *Bifidobacterium*, *Streptococcus*, *Clostridium XI*, and *Bacteroides* play essential roles in regulating the abundance of *Veillonella* in an infant gut microbiome.

Methods
Sample collection
To construct microbial regulatory networks in infant intestinal tracts, we used the pyrosequencing reads of 16S rRNA genes from 61 infant gut samples, which were treated with four diet states at four time periods [28]. These infant gut samples were collected over 61 time points. The average length of 16S rRNA gene sequences were 237 bps [28], we used only 16S rRNA gene sequences with length longer than 200 bps in this study.

Taxonomic classification and OTU assignment
The Ribosomal Database Project (RDP) classifier [29] was used to classify 16S rRNA pyrosequencing reads and to perform an operational taxonomic unit (OTU) assignment. After accomplishing the OTU assignment with a threshold of 80 % confidence, we carried out a further filtering step by removing operational taxonomic units (OTUs) that appeared in fewer than half of the total time points.

Relative abundance analysis
To estimate the relative difference between individual OTUs at each time point, we calculated the relative abundance of individual OTUs. First, we defined the pyrosequencing read number of the ith OTU at j^{th} time point as Y_{ij}. Second, we defined the sum of the pyrosequencing read numbers at j^{th} time point as Y_j. Finally, we divided Y_{ij} by Y_j to obtain the relative proportion X_{ij} of the ith OTU at j^{th} time point. After all pyrosequencing read numbers were transformed into relative proportions, we removed the lth OTU if both the maximum relative proportion of all X_{lj} was less than 0.1 and the coefficient of variation of all X_{lj} was larger than 3.4. In addition, we filtered out the data at the kth time point if both the maximum of all X_{ik} was less than 0.1 and the total number of OTUs was less than 60 % at the kth time point. Finally, the missing values were filled through the Bayesian principal component analysis (BPCA) with K = 5 [30].

Tsai *et al. BMC Systems Biology* (2015) 9:54

Page 3 of 10

Data standardization

The relative proportion of each OTU at each time point was standardized in order to estimate the difference between OTUs at each time point. The relative proportion X_{ij} of the ith OTU at the jth time point was transformed such that it fell within the interval [0,1]. We defined P_{ij} as $(X_{ij} - X_{min})/(X_{max} - X_{min})$, where P_{ij} is the standardized relative proportion (SRP) of the ith OTU at the jth time point, while X_{min} and X_{max} are the minimum and maximum relative proportions respectively.

Network construction

We developed a rule-based microbial network (RMN) algorithm to construct microbial regulatory networks. The RMN algorithm is based on a nonlinear regulatory OTU-triplet model, named the NRO model for short. Figure 1 shows the flowchart of the RMN algorithm, and the details are as follows:

(i) To investigate how each OTU was involved in microbial regulatory network, we constructed a nonlinear regulatory OTU-triplet model by using triplets of OTUs across all time points (or samples). The nonlinear regulatory OTU-triplet model has three assumptions: First, a cooperation–competition system is assumed to exist in the microbial regulatory network. An OTU-triplet is defined as $\{O_m, O_c, O_t\}$, where O_m, O_c, and O_t represent the mth, c^{th} and t^{th} OTUs respectively. Here, O_m and O_c are assumed to be the cooperator and the competitor of O_t respectively. In other words, the relationship between the mth and the tth OTUs is assumed to be cooperative, while that between the cth and the tth OTUs is assumed to be competitive. Note that O_m, O_c and O_t represent different microbes in genus level. Second, it is assumed that the standardized relative proportion (SRP) of O_t is high when the SRP of O_m is high and the SRP of O_c is low. Conversely, the SRP of O_t is assumed to be low when the SRP of the O_m is low and the SRP of the O_c is high. Third, the relationship between O_m, O_c and O_t can be modeled by using a three-dimensional model surface (3D-MS) (Fig. 2) which is based on a piecewise nonlinear-quadratic polynomial:

$$P(P_m,,P_c) = \begin{cases} 2\tanh(1.1P_m)(1-\tanh(1.1P_c)), \\ \quad if\ 0 \le P_m < 0.5\ and\ 0.5 < P_c \le 1; \\ 1-2(1-\tanh(1.1P_m))\tanh(1.1P_c), \\ \quad if\ 0.5 < P_m \le 1\ and\ 0 \le P_c < 0.5; \\ \tanh(1.1P_m)-\tanh(1.1P_c)+0.5,\ otherwise \end{cases} \quad (1)$$

In (1), P_m and P_c are the standardized relative proportions (SRPs) of O_m and O_c respectively while $P(P_m,P_c)$ is the inferred SRP of O_t.

(ii) To find suitable OTU-triplets close to the 3D-MS for the OTU-triplet regulatory model, we used a lack-of-
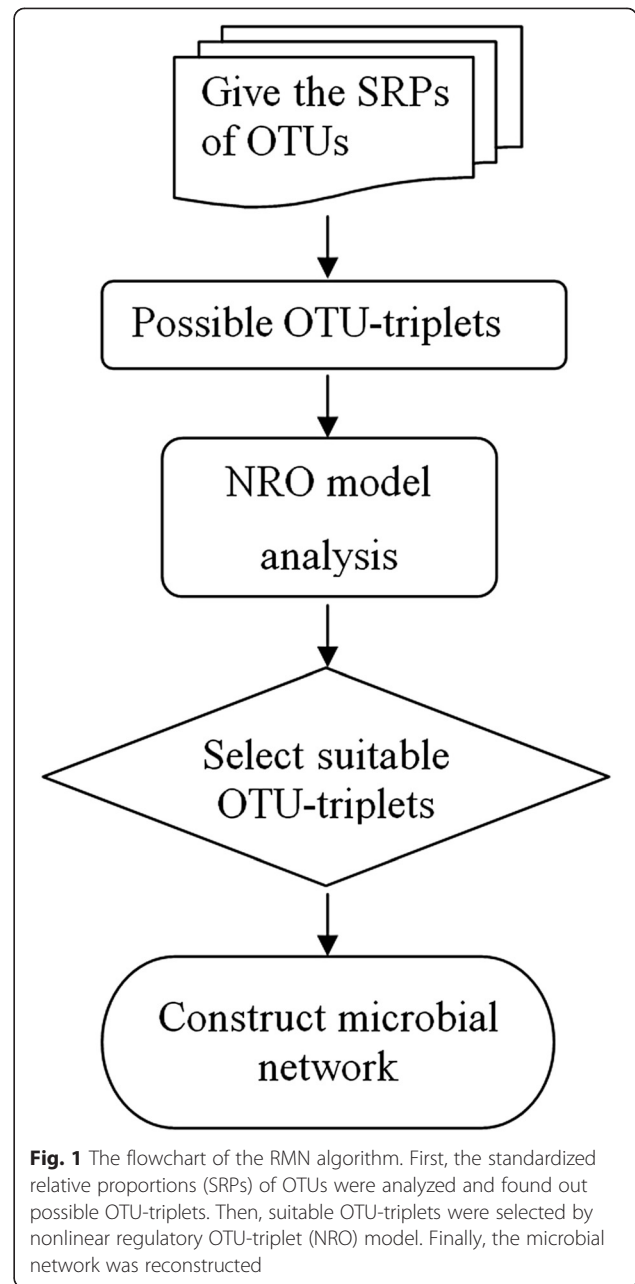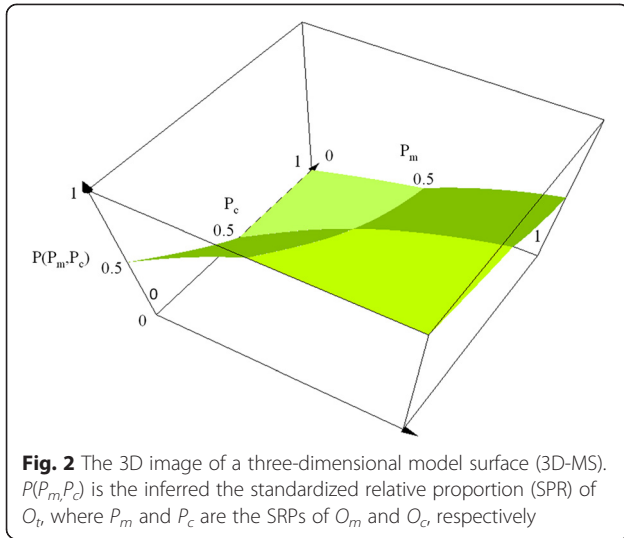

Fig. 1 The flowchart of the RMN algorithm. First, the standardized relative proportions (SRPs) of OTUs were analyzed and found out possible OTU-triplets. Then, suitable OTU-triplets were selected by nonlinear regulatory OTU-triplet (NRO) model. Finally, the microbial network was reconstructed

fit function to analyze all possible OTU-triplets. The lack-of-fit function is defined as:

$$L(O_m,,O_c,,O_t) = \frac{\sum_{n=1}^{k}(\hat{P}_{tn}-P_{tn})^2}{\sum_{n=1}^{k}(P_{tn}-\bar{P}_{tn})^2} \quad (2)$$

where $\hat{P}_{tn}$ is the inferred *SRP* of O_t at the nth time point (e.g. $P(P_mP_c)$ at the nth time point); P_{tn} is the observed SRP of O_t at the nth time point; and $\bar{P}_{tn}$ represents the average of P_{tn} across all k time points. A small $L(O_m, O_c, O_t)$ value indicates a good fit between an OTU-triplet $\{O_m, O_c, O_t\}$ and the 3D-MS. In other

Tsai *et al. BMC Systems Biology* (2015) 9:54

Page 4 of 10

Fig. 2 The 3D image of a three-dimensional model surface (3D-MS). $P(P_m, P_c)$ is the inferred the standardized relative proportion (SPR) of O_t, where P_m and P_c are the SRPs of O_m and O_c, respectively

words, a smaller $L(O_m, O_c, O_t)$ value reveals a stronger relationship among O_m, O_c, and O_t. Similar rationale has been successfully applied for constructing gene regulatory networks [24].

(iii) To check the reliability of OTU-triplets, we used an adjustment function to fix imperfect measurements at one or more time points (or samples). The adjustment function is defined as:

$$D(O_m,,O_c,,O_t) = \sum_{n=1}^{k} \left| \left(\frac{L_{(n)}(O_m,,O_c,,O_t)}{L(O_m,,O_c,,O_t)} \right) \log_{10} \left(\frac{L_{(n)}(O_m,,O_c,,O_t)}{L(O_m,,O_c,,O_t)} \right) \right| \tag{3}$$

where $L_{(n)}(O_m, O_c, O_t)$ is the value of lack-of-fit function when the information from the nth time point is removed. A smaller $D(O_m, O_c, O_t)$ value indicates the information of the OTU-triplet relationship is reliable. Finally, an integrated function, $I(O_m, O_c, O_t)$ is defined as the value of $L(O_m, O_c, O_t)$ multiplied by the value of $(1+ D(O_m, O_c, O_t))$. A smaller $I(O_m, O_c, O_t)$ value indicates a stronger OTU-triplet relationship among O_m, O_c, and O_t.

(iv) Finally, we defined two criteria, $L_d(O_m, O_c, O_t) = L(O_m, O_c, O_t)/(1- L(O_m, O_c, O_t))$ and $D(O_m, O_c, O_t)$, to filter out unreliable OTU-triplets, when $0 \leqq L_d(O_m, O_c, O_t) \leqq L$ and $D(O_m, O_c, O_t) < D$. Note the L and D are constants. We then ranked reliable OTU-triplets basing on the sorted $I(O_m, O_c, O_t)$ values in the ascending order, and selected OTU-triplets with lower $I(O_m, O_c, O_t)$ values as candidates for network construction.

Performance analysis

To evaluate the performance of the RMN algorithm, we simulated a microbial regulatory network consisting of ten gut bacteria, three latent factors, and ten noise factors as follows:

$$O_1(s) = 1.3 \tanh(-0.1F_2(s) + 0.25) + N_1(s)$$
$$O_2(s) = 1.1 \tanh(0.35F_2(s) + 0.1) + N_2(s)$$
$$O_3(s) = \tanh(-0.56O_1(s) + 0.4F_1(s) + 0.5) + N_3(s)$$
$$O_4(s) = \tanh(0.6O_1(s) - 0.15O_2(s) - 0.1O_3(s) + 0.1) + N_4(s)$$
$$O_5(s) = \tanh(0.24O_2(s) - 0.75O_3(s) - 0.05F_1(s) + 0.3) + N_5(s)$$
$$O_6(s) = 1.4 \tanh(0.2O_5(s) - 0.2O_3(s) + 0.15) + N_6(s)$$
$$O_7(s) = 1.3 \tanh(-0.15O_5(s) + 0.2) + N_7(s)$$
$$O_8(s) = \tanh(-0.15F_3(s) + 0.2) + N_8(s)$$
$$O_9(s) = \tanh(-0.41O_6(s) + 0.55O_7(s) + 0.35O_8(s) + 0.18) + N_9(s)$$
$$O_{10}(s) = \tanh(-0.45O_7(s) + 0.7O_8(s) + 0.1F_3(s) + 0.25) + N_{10}(s)$$

$$(4)$$

We used Tanh functions to create simulated a microbial regulatory network because Tanh functions have been used to model regulatory responses of pathogens and microbes [25–27]. In addition, the similar idea has been used to create simulated gene networks for evaluating the performance of GASA method [31]. In detail, we defined this synthetic regulatory model in Fig. 3, where $O_i(s)$ is the abundance proportion of the ith OTU with
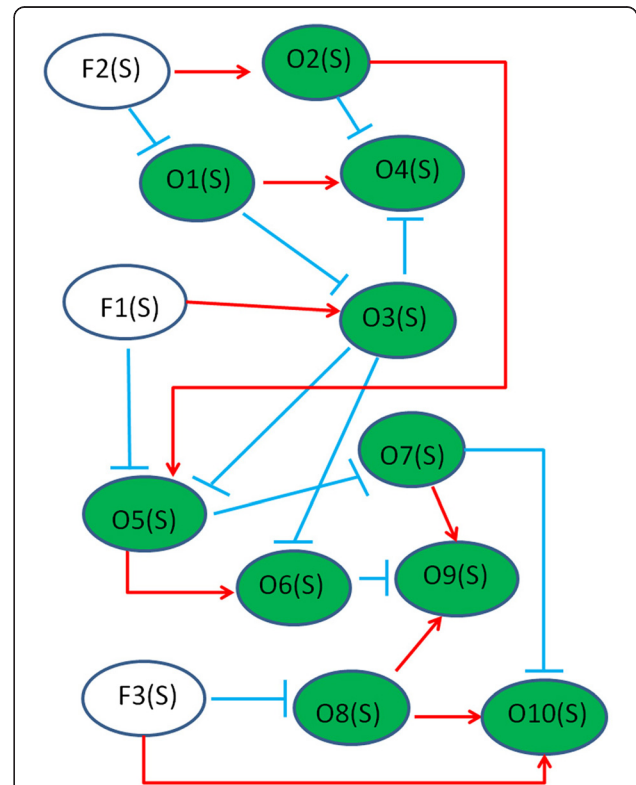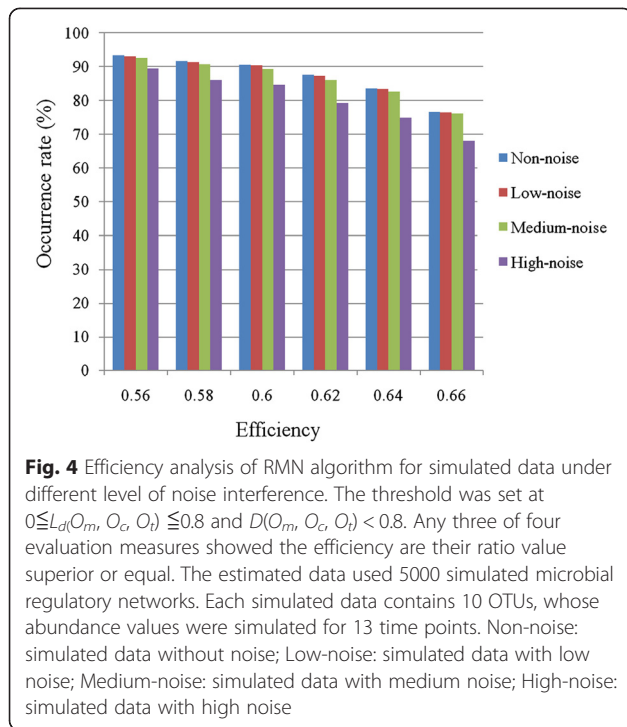

Fig. 3 Simulation of microbial regulatory network. For example, $O_1(s)$ is defined to interact competitively with $O_3(s)$ but interact cooperatively with $O_4(s)$. Here, the competitive and cooperatively relationships are defined as ⊣ and →, individually

Tsai *et al. BMC Systems Biology* (2015) 9:54

Page 5 of 10

Fig. 4 Efficiency analysis of RMN algorithm for simulated data under different level of noise interference. The threshold was set at $0 \leqq L_d(O_m, O_c, O_t) \leqq 0.8$ and $D(O_m, O_c, O_t) < 0.8$. Any three of four evaluation measures showed the efficiency are their ratio value superior or equal. The estimated data used 5000 simulated microbial regulatory networks. Each simulated data contains 10 OTUs, whose abundance values were simulated for 13 time points. Non-noise: simulated data without noise; Low-noise: simulated data with low noise; Medium-noise: simulated data with medium noise; High-noise: simulated data with high noise

i =1,2,...,10; $F_j(s)$ is the abundance proportion of jth latent factor with j = 1, 2, 3; and $N_k(s)$ is the abundance proportion of the kth noise factor with k = 1,2,...,10. Note the value of $O_i(s)$ is between 0 and 1, while $F_j(s)$ is a random variable between 0 and 1. $N_k(s)$ is a random variable representing the noise level in the simulated data (when $N_k(s) = 0$, no noise exists in the simulated data). Three noise levels are assumed in the simulated data: low noise if the value of $N_k(s)$ is equal to the variance of $O_i(s)$ divided by $R_k(s)$; medium noise if $N_k(s)$ is equal to five times the low noise level; and
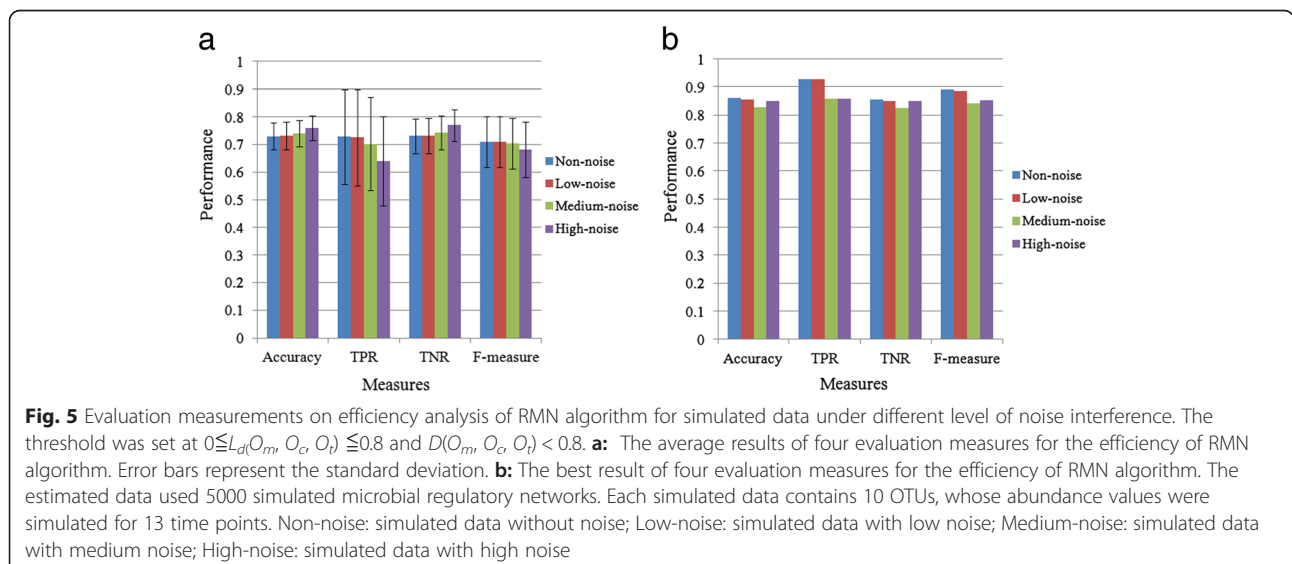
high noise if $N_k(s)$ is equal to ten times the low noise level. Here, $R_k(s)$ represents the signal-noise ratio and its range was between 4 and 12. Note that $R_k(s)$ is an integral and is generated by using time-series data of gut bacteria [19]. In detail, $R_k(s)$ is the average proportion of gut bacteria divided by its standard deviation.

We evaluated the performance of the RMN algorithm by calculating the following measures: true positive rate (TPR), true negative rate (TNR), F-measure, and Accuracy. TPR is defined as TP/(TP + FN), where TP and FN are the numbers of true positive interactions (links) and false negative interactions (links) respectively. TNR is defined as TN/(TN + FP), where TN and FP are the numbers of true negative interactions (links) and false positive interactions (links) respectively. We define F-measure as 2*TPR*TNR/ (TPR + TNR), and accuracy as (TP + TN)/(TP + FN + TN + FP).

Results and discussion

Performance of RMN algorithm using simulation data

We generated simulated data under different levels of noise and evaluated the performance of our RMN algorithm by using the above mentioned four evaluation measures. Then, we analyzed the occurrence rate (%) of 50000 simulated microbial regulatory networks when any three of four evaluation measures showed the efficiency superior or equal to the defined value. (see Fig. 4). All four evaluation measures suggest high performance from our RMN algorithm. Comparing different noise levels, RMN algorithm has shown the best efficiency when there is no noise in the simulated data. Even at medium noise level, over 82 % of the results show efficiency values higher than 0.64. In addition, the assessment from those four evaluation measures shows 0.64 efficiency value on more than 74 % simulated data.

Fig. 5 Evaluation measurements on efficiency analysis of RMN algorithm for simulated data under different level of noise interference. The threshold was set at $0 \leqq L_d(O_m, O_c, O_t) \leqq 0.8$ and $D(O_m, O_c, O_t) < 0.8$. **a:** The average results of four evaluation measures for the efficiency of RMN algorithm. Error bars represent the standard deviation. **b:** The best result of four evaluation measures for the efficiency of RMN algorithm. The estimated data used 5000 simulated microbial regulatory networks. Each simulated data contains 10 OTUs, whose abundance values were simulated for 13 time points. Non-noise: simulated data without noise; Low-noise: simulated data with low noise; Medium-noise: simulated data with medium noise; High-noise: simulated data with high noise

Tsai *et al. BMC Systems Biology* (2015) 9:54

Page 6 of 10

There are no significant differences between the performance results when the noise is varied from zero to medium level (see Fig. 5(a)). Although the performance of our RMN algorithm at the high noise level is worse than those from lower noise levels, its efficiency value is still over 0.68. More interestingly, according to TNR, there were fewer false positive interactions (links) when inferred from the high noise data than that from lower noise data (see Fig. 5(a)). In addition, best efficiency under different noise interferences exceeds 0.82 (see Fig. 5(b)). During developing RMN algorithm, we used two conditions, $L_d(O_m, O_c, O_t)$ and $D(O_m, O_c, O_t)$, to filter out unreliable OTU-triplets, and we tried to select the suitable thresholds for $L_d(O_m, O_c, O_t)$ and $D(O_m, O_c, O_t)$ by analyzing the simulation data (Table 1). The best efficiency for all evaluations reached over 0.82 when the threshold at $L_d(O_m, O_c, O_t)$ and $D(O_m, O_c, O_t)$ were used (see Table 1). Moreover, the efficiency for all evaluations reached over 0.71 when the threshold at $L_d(O_m, O_c, O_t)$ and $D(O_m, O_c, O_t)$. All in all, our efficiency evaluations demonstrate that the performance of our RMN algorithm is robust under three latent factors and different noise levels.

Inferring the regulatory relationship of microbial pairs with low correlation coefficient

Recent study shows that correlations between the abundances of microbes do not imply microbial interactions [18]. In order to investigate whether OTU-triplets obtained by RMN algorithm are significantly correlated based on their correlation coefficient and whether RMN algorithm can find the regulator relationship behind standard correlation-based approaches, we analyzed the optimal resulting triplets with true positive links (filtered by RMN algorithm). Here we used Spearman's rank correlation coefficient for the estimation of nonlinear correlation coefficient. The Pr value represents a probability of RMN algorithm prediction for OTU pairs whose nonlinear correlation coefficient is less than 0.5. In detail, Pr value (%) was defined as (O_mO_t_count + O_cO_t_count) *100 %/Total_count, where O_mO_t_count is the number of resulting triplets with a true positive link between cooperator OTU and target OTU pair whose nonlinear correlation coefficient is less than 0.5; O_cO_t_count is the number of resulting triplets with a true positive link between competitor OTU and target OTU pair whose nonlinear correlation coefficient is greater than –0.5; Total count is the number of resulting triplets with a true positive link. As Table 2 shown, under different levels of noise inference, Pr values for most analyzed results were more than 15 %. That may imply the resulting triplets contain OTU pairs with low nonlinear correlation coefficient. For example, more than 15 % OTU pairs with low nonlinear correlation coefficient (less than

Table 1 The best performance result with different parameters and different level of noise interference. Here L0.8 represents for $0 \leqq L_d(O_m, O_c, O_t) \leqq 0.8$ and D0.8 represents for $D(O_m, O_c, O_t) < 0.8$

Non-noise	Accuracy	TPR	TNR	F-measure
L0.8D0.8	0.861	0.929	0.855	0.890
L1.8D0.8	0.833	0.857	0.831	0.844
L2.8D0.8	0.756	0.857	0.747	0.798
L3.8D0.8	0.722	0.786	0.717	0.750
L0.8D1.8	0.822	0.857	0.819	0.838
L0.8D2.8	0.828	0.857	0.825	0.841
L0.8D3.8	0.828	0.857	0.825	0.841
Low-noise	Accuracy	TPR	TNR	F-measure
L0.8D0.8	0.856	0.929	0.849	0.887
L1.8D0.8	0.767	0.786	0.765	0.775
L2.8D0.8	0.756	0.786	0.753	0.769
L3.8D0.8	0.756	0.714	0.759	0.736
L0.8D1.8	0.828	0.857	0.825	0.841
L0.8D2.8	0.811	0.857	0.807	0.831
L0.8D3.8	0.822	0.857	0.819	0.838
Medium-noise	Accuracy	TPR	TNR	F-measure
L0.8D0.8	0.828	0.857	0.825	0.841
L1.8D0.8	0.811	0.786	0.813	0.799
L2.8D0.8	0.767	0.786	0.765	0.775
L3.8D0.8	0.761	0.786	0.759	0.772
L0.8D1.8	0.828	0.857	0.825	0.841
L0.8D2.8	0.828	0.857	0.825	0.841
L0.8D3.8	0.833	0.929	0.825	0.874
High-noise	Accuracy	TPR	TNR	F-measure
L0.8D0.8	0.850	0.857	0.849	0.853
L1.8D0.8	0.800	0.786	0.801	0.793
L2.8D0.8	0.756	0.786	0.753	0.769
L3.8D0.8	0.756	0.786	0.753	0.769
L0.8D1.8	0.833	0.857	0.831	0.844
L0.8D2.8	0.833	0.857	0.831	0.844
L0.8D3.8	0.844	0.857	0.843	0.850

0.5) could be found when the threshold was set at $0 \leq L_d(O_m, O_c, O_t) \leqq 0.8$ and $D(O_m, O_c, O_t) < 0.8$. Based on the threshold, the performances of RMN algorithm were more than 0.82 (see Table 1). In addition, more than 20 % OTU pairs with low nonlinear correlation coefficient (less than 0.5) could be found when the threshold was set at $0 \leq L_d(O_m, O_c, O_t) \leqq 3.8$ and $D(O_m, O_c, O_t) < 0.8$. Based on the threshold, the performances of RMN algorithm were more than 0.71 (see Table 1). In other words, the RMN algorithm may characterize the regulatory relationship composed of OTU pairs with low correlation coefficient, and has a choice to find the regulator relationship behind standard correlation-based approaches.

Tsai *et al. BMC Systems Biology* (2015) 9:54

Page 7 of 10

Table 2 The nonlinear correlation coefficient of OTU pairs, predicted by RMN algorithm as co-regulated triplets, under different level of noise interference. A triplet predicted by RMN algorithm is composed of cooperator, competitor and target. Here L0.8 represents for $0 \leqq L_d(O_m, O_c, O_t) \leqq 0.8$ and D0.8 represents for $D(O_m, O_c, O_t) < 0.8$

Non-noise	O_mO_t_count	O_cO_t_count	Total_count	Pr value (%)
L0.8D0.8	3	3	26	23.1
L1.8D0.8	4	0	24	16.7
L2.8D0.8	7	5	24	50.0
L3.8D0.8	6	1	22	31.8
L0.8D1.8	2	2	24	16.7
L0.8D2.8	1	2	24	12.5
L0.8D3.8	1	1	24	8.3
Low-noise	O_mO_t_count	O_cO_t_count	Total_count	Pr value (%)
L0.8D0.8	3	1	26	15.4
L1.8D0.8	6	2	22	36.4
L2.8D0.8	5	0	22	22.7
L3.8D0.8	1	4	24	20.8
L0.8D1.8	4	2	24	25.0
L0.8D2.8	1	0	24	4.2
L0.8D3.8	3	1	24	16.7
Medium-noise	O_mO_t_count	O_cO_t_count	Total_count	Pr value (%)
L0.8D0.8	5	2	24	29.2
L1.8D0.8	5	3	22	36.4
L2.8D0.8	5	1	22	27.3
L3.8D0.8	7	1	22	36.4
L0.8D1.8	4	1	24	20.8
L0.8D2.8	2	1	24	12.5
L0.8D3.8	4	2	26	23.1
High-noise	O_mO_t_count	O_cO_t_count	Total_count	Pr value (%)
L0.8D0.8	3	3	24	25.0
L1.8D0.8	6	5	22	50.0
L2.8D0.8	7	4	22	50.0
L3.8D0.8	5	0	20	25.0
L0.8D1.8	3	2	24	20.8
L0.8D2.8	2	1	24	12.5
L0.8D3.8	2	2	24	16.7

Table 3 The efficiency analysis under different level of noise interference by using the simulated data who's best performance showed at a threshold on $0 \leqq L_d(O_m, O_c, O_t) \leqq 0.8$ and $D(O_m, O_c, O_t) < 0.8$

Non-noise	Accuracy	TPR	TNR	F-measure
L0.8D0.8	0.861	0.929	0.855	0.890
L1.8D0.8	0.706	1.000	0.681	0.810
L2.8D0.8	0.617	1.000	0.584	0.738
L3.8D0.8	0.556	1.000	0.518	0.683
L0.8D1.8	0.828	0.929	0.819	0.871
L0.8D2.8	0.828	0.929	0.819	0.871
L0.8D3.8	0.828	0.929	0.819	0.871
Low-noise	Accuracy	TPR	TNR	F-measure
L0.8D0.8	0.856	0.929	0.849	0.887
L1.8D0.8	0.678	0.929	0.657	0.769
L2.8D0.8	0.611	0.929	0.584	0.717
L3.8D0.8	0.550	0.929	0.518	0.665
L0.8D1.8	0.761	0.929	0.747	0.828
L0.8D2.8	0.761	0.929	0.747	0.828
L0.8D3.8	0.761	0.929	0.747	0.828
Medium-noise	Accuracy	TPR	TNR	F-measure
L0.8D0.8	0.828	0.857	0.825	0.841
L1.8D0.8	0.706	0.929	0.687	0.790
L2.8D0.8	0.622	0.929	0.596	0.726
L3.8D0.8	0.533	0.929	0.500	0.650
L0.8D1.8	0.744	0.857	0.735	0.791
L0.8D2.8	0.744	0.857	0.735	0.791
L0.8D3.8	0.744	0.857	0.735	0.791
High-noise	Accuracy	TPR	TNR	F-measure
L0.8D0.8	0.850	0.857	0.849	0.853
L1.8D0.8	0.733	0.929	0.717	0.809
L2.8D0.8	0.633	0.929	0.608	0.735
L3.8D0.8	0.572	0.929	0.542	0.685
L0.8D1.8	0.778	0.857	0.771	0.812
L0.8D2.8	0.778	0.857	0.771	0.812
L0.8D3.8	0.778	0.857	0.771	0.812

The construction of microbial regulatory network from infant gut data

To further demonstrate the capability of our RMN algorithm, we constructed microbial interaction networks from a set of real microbial community time series data. The data used consist of 61 infant gut samples treated with four diet states at four time periods [28].

After performing taxonomic classification and OTU assignment, 101 OTUs were obtained (Additional file 1). However, after the filtering process removing OTUs that appeared in less than half of the total time points, only 14 OTUs met the requirement (Additional file 2). In addition, after performing relative abundance analysis and filling missing values, 10 OTUs with 43 time points were obtained. Finally, according to Fig. 3c from Koenig, et al. [28], only 10 OTUs with 36 time points between four steps were selected for analyzing using our RMN algorithm (Additional file 3). Consistently, among 10 OTUs, 4 OTUs (OTU15, OTU16, OTU34, and OTU94) were found to be involved in competitive and cooperative interactions in Freilich's study [32]. Since microbiome data are usually enormous and inherently noisy, the threshold

Tsai et al. BMC Systems Biology (2015) 9:54

Page 8 of 10

was set at $0 \leq L_d(O_m, O_c, O_t) \leq 3.8$ and $D(O_m, O_c, O_t) < 0.8$ for better prediction sensitivity (see Table 3). Before filtering by this threshold, top 100 possible OTU-triplets from 10 OTUs analysis were listed in Additional file 4. After filtering by this threshold, only 3 reliable OTU-triplets from 5 OTUs (Additional file 5) were used to construct microbial regulatory network. The resulting microbial interaction network among five OTUs is displayed in Fig. 6. Gram-negative anaerobic cocci, *Veillonella* (OTU100), is found to interact cooperatively with three OTUs, namely *Clostridium XI* (OTU30), *Bifidobacterium* (OTU16) and *Streptococcus* (OTU94), but competitively with *Bacteroides* (OTU15). Among them, *Bifidobacterium*, *Streptococcus*, and *Bacteroides* have been identified in competitive and cooperative metabolic interactions [32]. In particular, we observed a positive correlation between *Veillonella* and *Bifidobacterium*, both residing in human intestines and oral mucosa as anaerobic commensal organisms [33]. This directional interaction also suggests the commensal relationship between *Veillonella* and *Bifidobacterium*. In addition, *Veillonella* was also observed to have a positive correlation with *Streptococcus*. They have been found to interact cooperatively during the production and degradation of Lactic acid [34–40]. Recently, positive correlations were also found between the abundances of *Veillonella* and *Streptococcus* [41]. In addition, the constructed network also suggests that competitive relationship may exist between *Veillonella* and *Bacteroides*. *Veillonella* and *Bacteroides* are gut-associated obligate anaerobic genera found in maternal faeces, breast milk and neonatal faeces [42]; and the difference in their efficiency on oligosaccharide consumptions may suggest they occupy different metabolic niches [43]. In addition, our network suggests there might be a cooperative relationship between *Veillonella* and *Clostridium XI Clostridium glycolicum*; and this cooperative relationship might suggest their simultaneous involvement in the ear infections in infants. In fact, those two gut bacteria have been found to be associated with bacterial

infections, such as bacteremia [44, 45]. Finally, microbial interaction network from Fig. 6 were analyzed to check if the RMN algorithm can predict microbial interactions from low correlation data. As shown, it confirmed that OTU pairs with low correlations can be predicted by RMN algorithm (Additional file 6). In general, our constructed network can extrapolate cooperative and competitive relationships between microbes in infant guts.

Conclusions

In this paper we presented RMN algorithm, a rule-based algorithm using the OTU-triplet model with parametric weighting function, to construct microbial regulatory networks. RMN algorithm can theoretically extract directional interactions to delineate the cooperative and competitive relationships between microbes from high-throughput sequencing time series data. The efficiency estimation shows the high performance of RMN algorithm on simulated data with three latent factors and different noise levels. We also applied our framework to identify new relationships in infant gut samples. RMN algorithm is both computationally feasible and capable of detecting biologically significant progresses embedded in a microbial community.

Although simple pairwise relationship has been used to describe complex forms of microbial interactions, advanced techniques are still required to infer complex microbial networks. Recent studies show that the cooperative and competitive relationships between microbes coexist in complicated ecological communities [8, 22, 23]. In addition, triplet relation, like activator-repressor-target pattern, has been adopted to construct gene regulatory network form time course gene expression data [24]. Basing on such a rationale, we developed RMN algorithm to describe the relationship among every triplet of microbe as the cooperator-competitor-target. RMN algorithm is used to search for triplet relations of OTUs where the abundances of the target
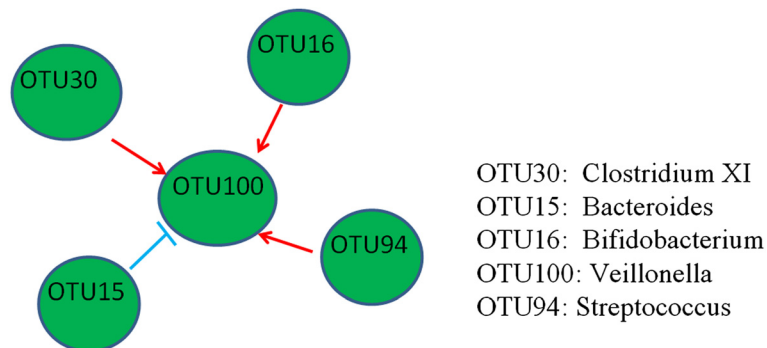

OTU30: Clostridium XI
OTU15: Bacteroides
OTU16: Bifidobacterium
OTU100: Veillonella
OTU94: Streptococcus

Fig. 6 The reconstruction of microbial regulatory networks in infant intestinal tract. Here, the competitive and cooperatively relationships are defined as ⊣ and →, individually

Tsai *et al. BMC Systems Biology* (2015) 9:54

Page 9 of 10

should be high while that for the cooperator is high and that for the competitor is low (conversely, the abundances of the target should be low while that for the cooperator is low and that for the competitor is high). Our evaluation results suggest RMN algorithm has high performance when subjected to simulated data with different levels of noise (see Figs. 4 and 5). As presented in the results, our algorithm has characterized well-known and possible microbial interactions from time series microbiome data of infant gut samples (see Fig. 6).

In summary, networks reconstructed by using simple similarity-based method often fail to capture the complex topology of a real system, and such networks may be full of false positive links. Furthermore, biologically significant relationships are often difficult to be identified from such networks. Our RMN algorithm can produce low false positive links even for highly noisy data (see Fig. 5(a)). Moreover, RMN algorithm is designed to explore complex networks consisted of non-linear, decentralized and directed interactions among microbes. However, having said all these, the drawback of RMN algorithm is that it may not be able to detect simple linear correlation between two OTUs. Another potential concern is that we tested our method using a simulated data with the Tanh functions, the circularity in the model construction and validation can cause bias on our results. Overall, RMN algorithm can theoretically predict microbial interactions from low correlation data (see Tables 1 and 2). Our method should be a promising starting step to identify novel microbial interactions that normally cannot be found by using similarity-based methods. Therefore, by integrating similarity-based methods and our RMN algorithm, one can potentially gain a more accurate picture of microbial interactions towards a better understanding of microbial dynamics.

Additional files

Additional file 1: The table of OTUs for all the samples at the genus level with OTU assignment before filtering process. (XLS 81 kb)

Additional file 2: The table of OTUs for all the samples at the genus level with OTU assignment after filtering process. (XLS 34 kb)

Additional file 3: The table of OTUs for all the samples at the genus level with data standardization. (XLS 26 kb)

Additional file 4: List of top 100 possible OTU-triplets ranking by $I(O_m, O_c, O_t)$ values in the ascending order. These OTU-triplets were selected without using the threshold set at $0 \leqq L_d(O_m, O_c, O_t) \leqq 3.8$ and $D(O_m, O_c, O_t) < 0.8$. (XLS 28 kb)

Additional file 5: List of the reliable OTU-triplets ranking by $I(O_m, O_c, O_t)$ values in the ascending order. These OTU-triplets were selected when the threshold was set at $0 \leqq L_d(O_m, O_c, O_t) \leqq 3.8$ and $D(O_m, O_c, O_t) < 0.8$. (XLS 19 kb)

Additional file 6: The nonlinear correlation coefficient of OTU pairs predicted by RMN algorithm. (XLS 20 kb)

Author details
[1]Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan. [2]Department of Medical Research and Development, Show Chwan Health Care System, Changhua 505, Taiwan. [3]Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan.

References
1. Faust K, Raes J. Microbial interactions: from networks to models. Nat Rev Microbiol. 2012;10(8):538–50. doi:10.1038/nrmicro2832.
2. Zhou J, Deng Y, Luo F, He Z, Yang Y. Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO2. mBio. 2011;2(4). doi:10.1128/mBio.00122-11.
3. Deng Y, Jiang YH, Yang Y, He Z, Luo F, Zhou J. Molecular ecological network analyses. BMC Bioinf. 2012;13:113. doi:10.1186/1471-2105-13-113.
4. Sun Y, Cai Y, Mai V, Farmerie W, Yu F, Li J, et al. Advanced computational algorithms for microbial community analysis using massive 16S rRNA sequence data. Nucleic Acids Res. 2010;38(22), e205. doi:10.1093/nar/gkq872.
5. Freilich S, Kreimer A, Meilijson I, Gophna U, Sharan R, Ruppin E. The large-scale organization of the bacterial network of ecological co-occurrence interactions. Nucleic Acids Res. 2010;38(12):3857–68. doi:10.1093/nar/gkq118.
6. Lo Giudice A, Brilli M, Bruni V, De Domenico M, Fani R, Michaud L. Bacterium-bacterium inhibitory interactions among psychrotrophic bacteria isolated from Antarctic seawater (Terra Nova Bay, Ross Sea). FEMS Microbiol Ecol. 2007;60(3):383–96. doi:10.1111/j.1574-6941.2007.00300.x.
7. Perez-Gutierrez RA, Lopez-Ramirez V, Islas A, Alcaraz LD, Hernandez-Gonzalez I, Olivera BC, et al. Antagonism influences assembly of a Bacillus guild in a local community and is depicted as a food-chain network. ISME J. 2013;7(3):487–97. doi:10.1038/ismej.2012.119.
8. Foster KR, Bell T. Competition, not cooperation, dominates interactions among culturable microbial species. Curr Biol. 2012;22(19):1845–50. doi:10.1016/j.cub.2012.08.005.
9. Celiker H, Gore J. Competition between species can stabilize public-goods cooperation within a species. Mol Syst Biol. 2012;8:621. doi:10.1038/msb.2012.54.
10. Duran-Pinedo AE, Paster B, Teles R, Frias-Lopez J. Correlation network analysis applied to complex biofilm communities. PLoS One. 2011;6(12), e28438. doi:10.1371/journal.pone.0028438.
11. Zhou J, Deng Y, Luo F, He Z, Tu Q, Zhi X. Functional molecular ecological networks. mBio. 2010;1(4). doi:10.1128/mBio.00169-10.
12. Xia LC, Steele JA, Cram JA, Cardon ZG, Simmons SL, Vallino JJ et al. Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. BMC systems biology. 2011;5 Suppl 2:S15. doi:10.1186/1752-0509-5-S2-S15.

Tsai *et al. BMC Systems Biology* (2015) 9:54

Page 10 of 10

13. Fuentes S, van Nood E, Tims S, Heikamp-de Jong I, ter Braak CJ, Keller JJ, et al. Reset of a critically disturbed microbial ecosystem: faecal transplant in recurrent Clostridium difficile infection. ISME J. 2014;8(8):1621–33. doi:10.1038/ismej.2014.13.

14. Xu Z, Hansen MA, Hansen LH, Jacquiod S, Sorensen SJ. Bioinformatic approaches reveal metagenomic characterization of soil microbial community. PLoS One. 2014;9(4), e93445. doi:10.1371/journal.pone.0093445.

15. Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X, et al. Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. ISME J. 2014;8(4):881–93. doi:10.1038/ismej.2013.185.

16. Stein RR, Bucci V, Toussaint NC, Buffie CG, Ratsch G, Pamer EG, et al. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS Comput Biol. 2013;9(12), e1003388. doi:10.1371/journal.pcbi.1003388.

17. 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. doi:10.1073/pnas.1311322111.

18. 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. doi:10.1371/journal.pone.0102451.

19. Trosvik P, Rudi K, Naes T, Kohler A, Chan KS, Jakobsen KS, et al. Characterizing mixed microbial population dynamics using time-series analysis. ISME J. 2008;2(7):707–15. doi:10.1038/ismej.2008.36.

20. Trosvik P, Stenseth NC, Rudi K. Convergent temporal dynamics of the human infant gut microbiota. ISME J. 2010;4(2):151–8. doi:10.1038/ismej.2009.96.

21. Trosvik P, Skanseng B, Jakobsen KS, Stenseth NC, Naes T, Rudi K. Multivariate analysis of complex DNA sequence electropherograms for high-throughput quantitative analysis of mixed microbial populations. Appl Environ Microbiol. 2007;73(15):4975–83. doi:10.1128/AEM.00128-07.

22. Korb J, Foster KR. Ecological competition favours cooperation in termite societies. Ecol Lett. 2010;13(6):754–60. doi:10.1111/j.1461-0248.2010.01471.x.

23. Mitri S, Xavier JB, Foster KR. Social evolution in multispecies biofilms. Proc Natl Acad Sci U S A. 2011;108 Suppl 2:10839–46. doi:10.1073/pnas.1100292108.

24. Xu H, Wu P, Wu CF, Tidwell C, Wang Y. A smooth response surface algorithm for constructing a gene regulatory network. Physiol Genomics. 2002;11(1):11–20. doi:10.1152/physiolgenomics.00060.2001.

25. Yang SX. A neural network model for dose–response of foodborne pathogens. Appl Soft Comput. 2003;3(2):85–96.

26. Mozhayskiy V, Tagkopoulos I. Horizontal gene transfer dynamics and distribution of fitness effects during microbial in silico evolution. BMC Bioinf. 2012;13 Suppl 10:S13.

27. Schryver JC, Brandt CC, Pfiffner SM, Palumbo AV, Peacock AD, White DC, et al. Application of nonlinear analysis methods for identifying relationships between microbial community structure and groundwater geochemistry. Microb Ecol. 2006;51(2):177–88.

28. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A. 2011;108 Suppl 1:4578–85. doi:10.1073/pnas.1000081107.

29. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73(16):5261–7. doi:10.1128/AEM.00062-07.

30. Oba S, Sato MA, Takemasa I, Monden M, Matsubara K, Ishii S. A Bayesian missing value estimation method for gene expression profile data. Bioinformatics. 2003;19(16):2088–96.

31. Chen CM, Lee C, Chuang CL, Wang CC, Shieh GS. Inferring genetic interactions via a nonlinear model and an optimization algorithm. BMC Syst Biol. 2010;4:16.

32. Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, et al. Competitive and cooperative metabolic interactions in bacterial communities. Nat Communicants. 2011;2:589. doi:10.1038/ncomms1597.

33. Verma R, Dhamija R, Ross SC, Batts DH, Loehrke ME. Symbiotic bacteria induced necrotizing pancreatitis. JOP Journal of the pancreas. 2010;11(5):474–6.

34. Chalmers NI, Palmer Jr RJ, Cisar JO, Kolenbrander PE. Characterization of a Streptococcus sp.-Veillonella sp. community micromanipulated from dental plaque. J Bacteriol. 2008;190(24):8145–54. doi:10.1128/JB.00983-08.

35. Palmer RJ, Diaz PI, Kolenbrander PE. Rapid succession within the Veillonella population of a developing human oral biofilm in situ. J Bacteriol. 2006;188(11):4117–24.

36. Kara D, Luppens SB, Cate JM. Differences between single- and dual-species biofilms of Streptococcus mutans and Veillonella parvula in growth, acidogenicity and susceptibility to chlorhexidine. Eur J Oral Sci. 2006;114(1):58–63. doi:10.1111/j.1600-0722.2006.00262.x.

37. Mikx FHM, van der Hoeven JS. Symbiosis of Streptococcus mutans and Veillonella alcalescens in mixed continuous cultures. Archives of Oral Biology. 1975;20(7):407–10. doi:http://dx.doi.org/10.1016/0003-9969(75)90224-1.

38. Mikx FH, van der Hoeven JS, Konig KG, Plasschaert AJ, Guggenheim B. Establishment of defined microbial ecosystems in germ-free rats. I. The effect of the interactions of streptococcus mutans or Streptococcus sanguis with Veillonella alcalescens on plaque formation and caries activity. Caries Res. 1972;6(3):211–23.

39. McBride BC, Van der Hoeven JS. Role of interbacterial adherence in colonization of the oral cavities of gnotobiotic rats infected with Streptococcus mutans and Veillonella alcalescens. Infect Immun. 1981;33(2):467–72.

40. van der Hoeven JS, Toorop AI, Mikx RH. Symbiotic relationship of Veillonella alcalescens and Streptococcus mutans in dental plaque in gnotobiotic rats. Caries Res. 1978;12(3):142–7.

41. Rigsbee L, Agans R, Shankar V, Kenche H, Khamis HJ, Michail S, et al. Quantitative profiling of gut microbiota of children with diarrhea-predominant irritable bowel syndrome. Am J Gastroenterol. 2012;107(11):1740–51. doi:10.1038/ajg.2012.287.

42. Jost T, Lacroix C, Braegger CP, Rochat F, Chassard C. Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. Environ Microbiol. 2014;16(9):2891–904. doi:10.1111/1462-2920.12238.

43. Marcobal A, Barboza M, Froehlich JW, Block DE, German JB, Lebrilla CB, et al. Consumption of human milk oligosaccharides by gut-related microbes. J Agric Food Chem. 2010;58(9):5334–40. doi:10.1021/jf9044205.

44. Fisher RG, Denison MR. Veillonella parvula bacteremia without an underlying source. J Clin Microbiol. 1996;34(12):3235–6.

45. Elsayed S, Zhang K. Clostridium glycolicum bacteremia in a bone marrow transplant patient. J Clin Microbiol. 2007;45(5):1652–4. doi:10.1128/JCM.01845-06.