Quantitative Biology 2016, 4(2): 84–91
DOI 10.1007/s40484-016-0071-4

RESEARCH ARTICLE

Exploring the interaction patterns among taxa and environments from marine metagenomic data

Ze-Gang Wei, Shao-Wu Zhang* and Fang Jing

Key Laboratory of Information Fusion Technology of Ministry of Education, College of Automation, Northwestern Polytechnical University, Xi'an 710072, China
* Correspondence: zhangsw@nwpu.edu.cn

Received August 28, 2015; Revised November 26, 2015; Accepted February 7, 2016

The sequencing revolution driven by high-throughput technologies has generated a huge amount of marine microbial sequences which hide the interaction patterns among microbial species and environment factors. Exploring these patterns is helpful for exploiting the marine resources. In this paper, we use the complex network approach to mine and analyze the interaction patterns of marine taxa and environments in spring, summer, fall and winter seasons. With the 16S rRNA pyrosequencing data of 76 time point taken monthly over 6 years, we first use our MtHc clustering algorithm to generate the operational taxonomic units (OTUs). Then, employ the *k*-means method to divide 76 time point samples into four seasonal groups, and utilize mutual information (MI) to construct the four correlation networks among microbial species and environment factors. Finally, we adopt the symmetrical non-negative matrix factorization method to detect the interaction patterns, and analysis the relationship between marine species and environment factors. The results show that the four seasonal microbial interaction networks have the characters of complex networks, and interaction patterns are related with the seasonal variability; the same environmental factor influences different species in the four seasons; the four environmental factors of day length, photosynthetically active radiation, NO_2+ NO_3 and silicate may have stronger influences on microbes than other environment factors.

Keywords: marine microbe; operational taxonomic unit; interaction pattern; network; clustering

INTRODUCTION

Marine microbes account for most of the oceanic activities. They are responsible for 98% primary ocean production and mediate all biogeochemical processes like the flow of nitrogen, carbon, and energy in the ocean [1]. However, most ecological functions and roles among microbial communities and environmental factors are poorly understood, owing to the dilute, microscopic nature of the planktonic microbial community [2]. With the development of high-throughput DNA sequencing technologies that yield a mass of reads of small-subunit rRNA gene (16S rRNA/18S rRNA) and DNA, it is possible for us to describe the compositions of microbial communities, their diversity and how communities may change across space, time or experimental treatments based on these sequence data [3]. However, most of the current analytical approaches of describing and comparing the structure of communities often focus on the total numbers of taxa, the relative abundances of individual taxa and the extent of phylogenetic or taxonomic overlap between communities or community categories [1,3–5]. Although some researchers used the network analysis to explore co-occurrence pattern in soil and ocean [2,6–9], they just used the over-fitting clustering method of operational taxonomic units (OTUs), and adopted the linear correlation approaches (e.g., Pearson or Spearman) to construct the correlation networks for showing the co-occurrence pattern of microbes. They did not mine the communities of networks for further showing the structures of co-occurrence patterns. Clustering the rRNA sequences into OTUs with high accuracy is an essential requirement for downstream analysis, such as obtaining true taxonomic diversity profile of an environmental sample, constructing the correlation networks of microbes and environmental factors. Exploring the interaction patterns among microbes and environment factors can offer new insight into the structure of complex microbial communities, reveal the niche spaces shared by the community members, identify habitat affinities or shared physiologies, and find how the environment influences the microbes, that could guide more experimental settings. In this paper, we will construct the spring, summer, fall and winter correlation networks of microbes and environments by using *k*-means clustering method, mutual information (MI) correlation computing approach and our effective MtHc OTUs clustering algorithm [10], then, introduce a symmetrical non-negative matrix factorization (s-NMF) method to detect the interaction patterns among marine microbes and environments. The aim is to understand the relationship among microbes, environments and seasonal variability and try to determine the microbial interaction pattern difference among seasons and find which environmental factors are closely related with marine microbes.

RESULTS AND DISCUSSION

For exploiting the correlation and co-occurrence patterns of microbes and environments from rRNA read data, we first need an effective clustering method and a reference database to assign the reads to known microbial taxa for obtaining the abundance of species, then, quantify the similarity of two species distributions with a similarity measure (for environment traits, treating them as additional "species"), in the end, we select all significant pairwise relationships to construct four seasonal microbial correlation networks, and detect the communities with s-NMF. Due to the superior performance of MtHc [10] than other four state-of-the-art clustering methods (MSClust [11], ESPRIT-Tree [12], CROP [13] and BEBaC [14]), we select MtHc method to generate the OTUs in this paper.

Topology analysis of four seasonal microbial correlation networks

In order to analyze the microbial diversity and the relationship among OTUs and environmental factors in spring, summer, fall, and winter seasons, we need to construct the correlation networks. In general, mutual information (MI) is a natural generalization of the correlation since it can measure the nonlinear dependency and topology sparseness between variables [15]. Here, we use MI to measure the similarity between two species and obtain the significant pairwise relationships with permutation test. The four seasonal microbial correlation networks are shown in Figure 1. We also compute their topological parameters including the average degree, average clustering coefficient, average power law degree, and modularity and compare them with their corresponding random networks. The comparison results of topological parameters of four seasonal networks and their random networks are summarized in Table 1.

From Table 1, we can see that there is little difference in the topological parameters among spring, summer, fall networks, but there is bigger difference between winter network and other three seasonal networks. These results indicate that the interaction patterns among microbes and environments in winter are significantly different from spring, summer and fall seasons. Compared with random networks, four seasonal microbial correlation networks have bigger average clustering coefficient, average power law degree, and modularity, indicating that the four seasonal microbial associate networks have some characters of complex network.

Interaction patterns detected by s-NMF in seasonal microbial networks

We first used some annotation strategies, such as, BLAST against Greengenes [16], SIVA [17] and RDP [18], to get the annotation information of OTUs at taxonomic level, then adopted s-NMF to mine the four seasonal microbial networks. The structures of interaction patterns detected by s-NMF are shown in Figure 2, from which we can see that the community (or interaction pattern) numbers detected with s-NMF are 4, 3, 3 and 6 in spring, summer, fall and winter networks, respectively. The community number of winter network is more than that of other three seasonal networks, which indicates that the seasonal variability might have the greatest influence on the marine microbe diversity. We also find that some environmental factors are strongly correlative with some special microbes. For instance, in spring microbial network (M1), the environmental factor E12 ($NO_2 + NO_3$) is correlative with OTU57 (*SAR11*), OTU65 (*SAR11*), OTU73 (*SAR11*) and OTU177 (*Roseovarius*). In summer microbial network (M1), E12 is correlative with OTU40 (*Pelagibacter*) and OTU52 (*Pelagibacter*). In fall microbial network (M1), E12 is correlative with OTU40 (*Pelagibacter*), OTU33 (*Pelagibacter*), OTU149 (*Pelagibacter*), OTU183 (*Roseovarius*), OTU67 (*Pelagibacter*), OTU135 (*Pelagibacter*), OTU256 (*Cyanobacteria*), OTU44 (*Pelagibacter*), OTU101 (*SAR11*). In winter microbial network, E12 is correlative with OTU132 (*Pelagibacter*) and OTU228 (*Roseovarius*).

We also analyzed in detail the composition of some communities which include more environmental factors in the four seasonal networks. The community M1 in spring network is composed of 6 environmental factors

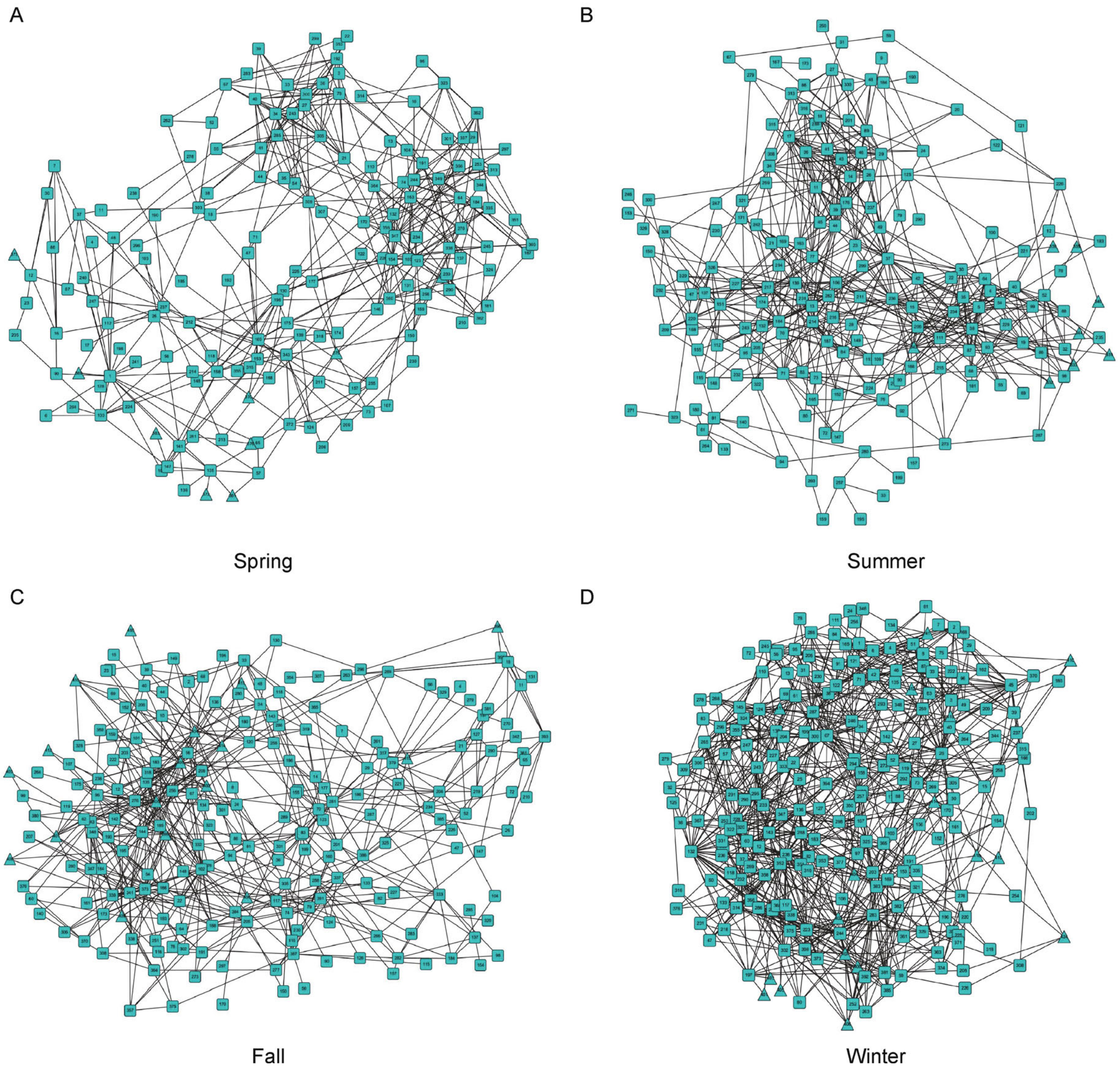

Figure 1. Four correlation networks in spring, summer, fall and winter seasons with MI(□—OTU, Δ—environmental factor).

Table 1. The topological parameters of four seasonal correlation networks and their corresponding random networks.

	Seasonal networks				Random networks			
	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter
Node Number	161	175	208	216	161	175	208	216
Edge Number	413	647	572	980	413	647	572	980
Avg. degree	5.367	5.932	5.123	10.752	5.367	5.932	5.123	10.752
Avg. power law degree	1.241	1.267	1.387	0.812	0.643	0.423	0.671	0.016
Avg. clustering coefficient	0.231	0.271	0.228	0.389	0.012	0.021	0.023	0.038
Modularity	0.565	0.553	0.512	0.371	0.381	0.337	0.412	0.221

(E1, E4, E5, E6, E12, and E14) and 40 OTUs in which the 30 OTUs come from *Bacteria*, 8 from *organelle*, and 2 OTUs have not been annotated. For the 30 *Bacteria* OTUs, 21 OTUs are identified in class level as *Alphaproteobacteria*, 9 OTUs as *Gammaproteobacteria*. For 8 *organelle* OTUs, 6 OTUs come from *Chloroplast* and 2 OTUs from *Mitochondria*. The community M1 in summer network is composed of 8 environmental factors

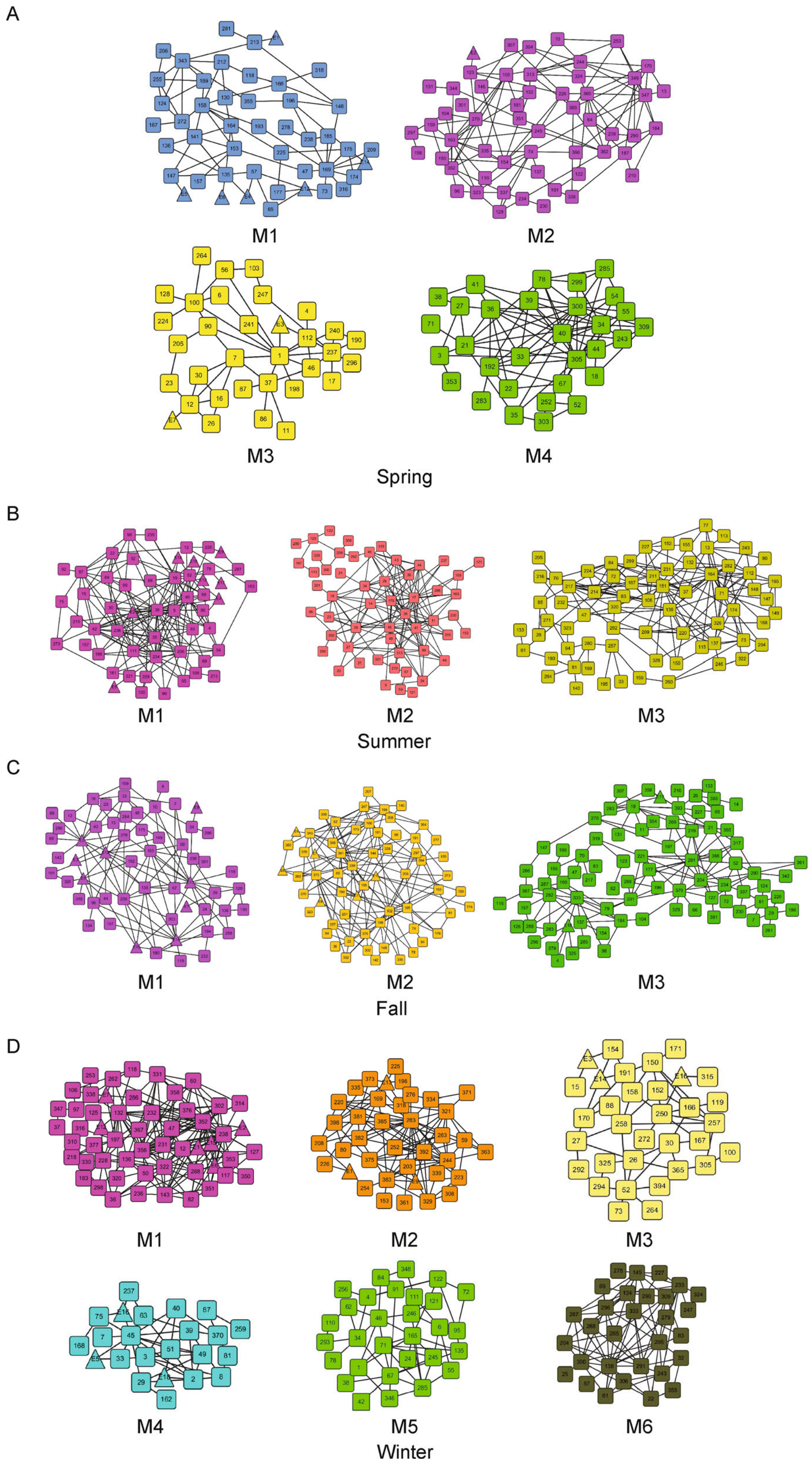

Figure 2. The communities (or interaction patterns) among micorbies and environment factors detected by s-NMF in four seasonal networks (□ — OTU, Δ — environmental factor).

(E2, E4, E5, E9, E12, E14, E17 and E18) and 46 OTUs which belong to *Bacteria*. 24 OTUs are identified in class level as *Alphaproteobacteria*, 19 OTUs as *Gammaproteobacteria*, 2 OTUs as *Verrucomicrobiae*, 1 OTU as *Sphingobacteria*. The community M1 in fall network contains 9 environmental factors (E1, E2, E3, E5, E9, E10, E12, E13 and E14) and 48 OTUs in which 43 OTUs come from *Bacteria*, 5 OTUs from *organelle*. For 43 *Bacteria* OTUs, 36 OTUs are identified in class level as *Alphaproteobacteria*, 3 OTUs as *Betaproteobacteria*, 3 OTUs as *Gammaproteobacteria*, 1 OTU as *Actinobacteria*. The community M1 in winter network consists of 6 environmental factors (E2, E6, E7, E11, E12 and E15) and 49 OTUs in which 41 OTUs come from *Bacteria* and 8 OTUs from *Chloroplast*. For 41 *Bacteria* OTUs, 32 OTUs are identified in class level as *Alphaproteobacteria*, 2 OTUs as *Betaproteobacteria*, 4 OTUs as *Gammaproteobacteria*, 1 OTU as *Flavobacteria*, 1 OTU as *Cyanobacteria*, 1 OTU as *Verrucomicrobiae*. M3 in winter network consists of 3 environmental factors (E3, E14 and E16) and 30 OTUs in which the 29 OTUs come from *Bacteria* and 1 OTU has been not annotated. For 29 *Bacteria* OTUs, 13 OTUs are identified in class level as *Alphaproteobacteria*, 12 OTUs as *Betaproteobacteria*, 3 OTUs as *Actinobacteria*, 1 OTU as *Cyanobacteria*. M4 in winter network includes 3 environmental factors (E5, E16 and E18) and 20 OTUs in which the 18 OTUs come from *Bacteria*, 1 OTU is *Chloroplast* and 1 OTU is unknown. For 18 *Bacteria* OTUs, 13 OTUs are identified in class level as *Alphaproteobacteria*, 2 OTUs as *Betaproteobacteria*, 3 OTUs as *Gammaproteobacteria*.

From Figure 1, we can find that four seasonal networks jointly includes the environment factors of E2 (day length), E5 (photosynthetically active radiation), E12 (NO_2+NO_3) and E14 (silicate), which indicates that the four environmental factors may strongly influence the microbes than other environment factors.

From Figure 2, we can see that the structure of communities in four seasons is significantly different, for example, two communities of fall seasonal network contains more than 7 environment factors, and one community of summer seasonal network contains 8 environment factors, while the communities of spring and winter seasonal networks just contain less than 6 environment factors, meaning that same environmental factors influence different species in four seasons, and more environment factors jointly influence the microbes in fall and summer seasons.

The community structural analysis of four seasonal microbial networks shows that a large fraction microbial interaction in class level occurs among *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria*. The community dense in spring, summer and fall networks is bigger than that of winter network. The correlative relationships between OTUs (taxa) are stronger than that of OTU and environmental factor, which may indicate that biological rather than physical factors can be more important in defining the fine-grain community structure.

Community alignment among four seasons

In order to study the evolution of microbes among four seasons, we align the communities between two seasons with MAGNA network alignment method [19]. MAGNA can optimize any measure of alignment quality, topological or biological and of node or edge conservation. The aligning results of communities in four seasons are shown in Figure 3, from which we can see that several communities of one season evaluate into one communities of another season, and one community of one season evaluates several communities of another season.

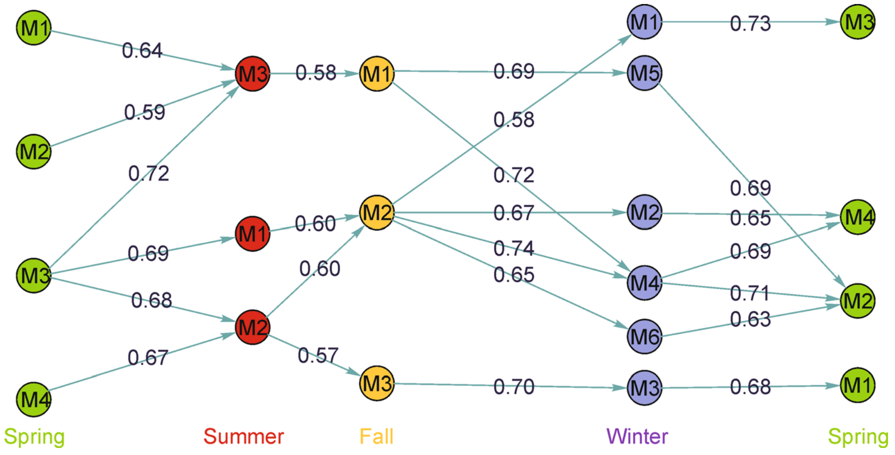

Figure 3. Evolutionary relationship of microbial communities among four seasons. The score on the line is the similarity value between two communities.

For example, M1, M2 and part of M3 in spring season evaluate into M3 in summer, and M2 in fall season evaluates into M1, M2, M4 and M6 in winter.

CONCLUSIONS

Microbial interaction networks provide further investigation angle for microbial community structure and ecological mechanism. As more and more environmental microbiomics data is available, developing the novel methods to explore the potential correlation/interaction patterns among microbial taxa and environmental factors will play a key role in the field of environmental and ecological system biology research. Considering the urgent requirement that needs novel network analytical approaches to move beyond the basic description of compositions and diversity of microbial communities, we use our MtHc OTU clustering method and MI correlation computing approach to construct four seasonal microbial networks from marine 16S rRNA sequences, and employ s-NMF algorithm to detect the potential interaction patterns among microbes and environments. The results show that the four seasonal marine microbial correlation networks have characters of complex networks; the marine microbial interaction patterns are related with the seasonal variability, e.g., the community dense in spring, summer and fall networks is bigger than that of winter network; the interaction between microbe and environmental factor in four seasonal networks is significantly different, that is, the same environmental factors influence different species; the environmental factors of day length, photosynthetically active radiation, NO_2+NO_3 and silicate may strongly influence the microbes than other environment factors. Although these interesting analyze results do not demonstrate that we have a comprehensive view of interactions within marine microbial interaction patterns, our analysis method is more feasible and interesting for exploring the unseen patterns emerged in the complex dataset.

MATERIALS AND METHODS

Datasets

The 16S rRNA sequence data and environmental factor data used in this paper were downloaded from http://vamps.mbl.edu/index.php, which include 969,400 sequences and 18 environmental factors from 76 time point seawater samples taken monthly over 6 years at a temperate marine coastal site in the West English Channel [7]. The 18 environmental factors are serial day (E1), day length (E2), DX1 (E3), DX2 (E4), photosynthetically active radiation (E5), North Atlantic Oscillation data (E6), primary productivity (E7), daily primary productivity (E8), mixed layer depth (E9), the concentrations of ammonia (E10), chlorophyll(E11), NO_2+NO_3 (E12), salinity (E13), silicate (E14), SRP (E15), temperature (E16), total organic carbon (E17) and total organic nitrogen (E18).

Due to the marine climatic changes, it is not fit for partition the four seasons according to the months [20]. Here, according to the environmental data of 6 years, we first use *k*-means method to cluster the 76 samples into four groups which correspond to the spring, summer, fall and winter seasons. This seasonal partition way can better reflect the local seasonal changes. As a result, 15, 24, 17 and 20 of the 76 samples are arranged into winter, spring, summer and fall seasons respectively. The 16S rRNA sequence numbers of winter, spring, summer and fall seasons are 174,885, 256,596, 244,046 and 293,873 respectively. In order to establish the seasonal correlation networks of microbes and environmental factors at the taxonomic level (e.g., species, genus), the 16S rRNA sequences are grouped into species-level operational taxonomic units (OTUs) with our MtHc algorithm [10], which resulted in 8,299 OTUs. MtHc method can accurately estimate the number of species and achieve better cluster quality.

MtHc algorithm

Most heuristic clustering methods are sensitive to the selected seeds that represent the clusters, and a change in the order of the input sequences may alter the clustering results significantly. Hierarchical clustering methods (either based on average-linage or complete-linkage) consider all the sequences in a cluster when forming clusters, which can add the computational burden. Model-based clustering methods and network modularity-based method often have a higher computational complexity, and do not easily deal with the massive 16S rRNA data. To address these problems, we proposed MtHc method in previous work [10] to cluster massive 16S rRNA sequences into OTUs. Comparing with the existing OTU clustering methods, MtHc can achieve higher cluster quality and lower time complexity for millions of 16S rRNA sequences, and also bypass the selection of hard distance threshold. In view of the better clustering performance of MtHc, and dataset containing 969,400 sequences, we select our MtHc method to generate the OTUs in this paper.

MtHc consists of three main phrases: searching motifs, generating crude clusters and merging these crude clusters into OTUs. Suppose all the 16S rRNA sequences can construct a complete weighted network, where sequences are viewed as nodes, each pair of sequences is connected by an imaginary edge, and the distance of a pair of sequences represents the weight of the edge. The process of MtHc method [10] can be simply described as: step 1, heuristically search the motif that is defined as n-node sub-graph (in the present study, $n = 3, 4, 5$), in which the grammar-distance between any two nodes is less than a threshold. Step 2, use the motif as a seed to form candidate cluster by computing the distances of other sequences with the motif. Step 3, hierarchically merge the candidate clusters to generate the OTUs by only calculating the distances of motifs between two clusters. By selecting a threshold θ, a series of sequence motifs are searched from the imaginary complete weighted network constructed with all the 16S rRNA sequences. Based on these motifs, MtHc forms a series of crude clusters, then merging them into OTUs by defining another threshold τ. We have discussed in detail how to select these two parameters in our previous work [10].

Correlation networks modeling

In order to explore the interaction patterns among marine microbes and environments and find the marine microbial seasonal variability, we should construct the four seasonal microbial correlation networks. Suppose vector X_μ and Y_ν represent OTU and environmental factor respectively.

$$X_\mu = [x_{\mu 1}, x_{\mu 2}, \ldots, x_{\mu s}, \ldots, x_{\mu S}],\ (\mu = 1, \ldots, 8299) \tag{1}$$

$$Y_\nu = [y_{\nu 1}, y_{\nu 2}, \ldots, y_{\nu s}, \ldots, y_{\nu S}],\ (\nu = 1, \ldots, 18) \tag{2}$$

where $x_{\mu s}$ is the μ-th OTU abundance value in the s-th sampling, that is, $x_{\mu s}$ equals the ratio of the sequence number $N_{\mu s}$ contained in the μ-th OTU and the total sequence number N_s contained in the s-th sampling, $y_{\nu s}$ is the ν-th environmental factor value in the s-th sampling. To reduce the sequencing effort bias, $x_{\mu s}$ is set to zero if $N_{\mu s}$<5. For reducing the false higher correlation between vectors, we also remove these OTU vectors which contain less than 3 non-zero elements. After these processing, we obtain 702 OTUs vectors, in which spring season contains 144, summer 161, fall 194 and winter 203 OTUs, respectively. Then, four microbial abundance matrixes and four environment factor matrixes of spring, summer, fall and winter seasons are produced by normalizing every OTU and environment factor vector with zero-mean normalization method.

Because mutual information (MI) can capture non-linear dependencies and topology sparseness between variables [15], we use MI to compute the correlations between variables.

$$MI(X,Y) = -\sum_{x \in X, y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{3}$$

where $p(x, y)$ is the joint probability distribution of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distributions of X and Y respectively.

The permutation test is used to calculate the statistical significance. We considered that there are robust correlations between OTU-OTU and OTU-environmental factor vector if P-value$\leqslant 0.01$, and there is a robust correlations between environmental factor vectors if P-value$\leqslant 0.05$. In the end, the four marine microbial association networks (Figure 1) of spring, summer, fall and winter seasons are constructed. These four seasonal networks are weighted and undirected.

Symmetrical non-negative matrix factorization (s-NMF) algorithm

A weighted network contained n nodes can be described by an adjacency matrix $\boldsymbol{A} = [\boldsymbol{A}_{ij}]_{n\times n}$, where $\boldsymbol{A}_{ij} \geqslant 0$. The feature matrix $\boldsymbol{O}$ of the network can be calculated from A, which represents the node-node similarity.

Suppose that n nodes can be grouped into r overlapping cliques (or communities). Then, a clique-node similarity non-negative matrix $\boldsymbol{W} = [w_{ki}]_{r\times n}$ is introduced to represent the similarity degree between node and clique, where w_{ki} indicates the closeness degree between node i and clique k. Because $\sum_{k=1}^{r} W_{ki}W_{kj}$ is an approximation of similarity between node i and node j, and O represents the node-node similarity, then, we can use O_{ij} to estimate $\sum_{k=1}^{r} W_{ki}W_{kj}$. Thus, our task is that minimize the function F_G.

$$\begin{aligned}\min_{W \geqslant 0} F_G(\boldsymbol{O},\boldsymbol{W}) &= \|\boldsymbol{O} - \boldsymbol{W}^T\boldsymbol{W}\|_F^2 \\ &= \frac{1}{2}\sum_{ij}[(\boldsymbol{O} - \boldsymbol{W}^T\boldsymbol{W}) \circ (\boldsymbol{O} - \boldsymbol{W}^T\boldsymbol{W})]_{ij}\end{aligned} \tag{4}$$

where $\boldsymbol{A} \circ \boldsymbol{B}$ is the Hadamard product (or element-by-element product) of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. This optimization problem can be solved by a symmetrical non-negative matrix factorization (s-NMF) method which is an improved method of non-negative matrix factorization (NMF) [21]. The iteratively updated rule of s-NMF can be described as follows:

$$\boldsymbol{W}_{k+1} = \boldsymbol{W}_k \circ \frac{[\boldsymbol{W}_k\boldsymbol{O}]}{[\boldsymbol{W}_k\boldsymbol{W}_k^T\boldsymbol{W}_k]} \tag{5}$$

where $\frac{[A]}{[B]}$ is the Hadamard division (or element-by-element division) of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. The stable points of Equation (5) can only fall into the set of NMF's stationary points, hence guaranteeing the convergence of s-NMF. NMF has been proved that it converges to a stationary point in many cases.

By normalizing the column of $\boldsymbol{W}$, we get the fuzzy membership degree matrix $\boldsymbol{U}$. The clique of corresponding to the largest element of each column in $\boldsymbol{U}$ is determined as the final membership clique of each node. That is, if $\boldsymbol{U}_{ki}$ is the maximum in the column i, the node i is classified to the clique k.

In order to determine the optimal number of community r, we iteratively increase r and choose the one which results in the highest modularity Q [22].

$$Q=\frac{1}{2m}\sum_{ij}\left(w_{ij}-\frac{k_ik_j}{2m}\right)\delta(C_i,C_j)$$

where m is the sum of weighted edges, w_{ij} is the weight of edge connecting nodes i and j, k_i is the degree of node i. If node i and j are grouped to the same cluster, $\delta(C_i,C_j)=1$, and otherwise $\delta(C_i,C_j)=0$.

ACKNOWLEDGEMENTS

This paper was supported by the National Natural Science Foundation of China (Nos. 91430111, 61473232 and 61170134).

COMPLIANCE WITH ETHICS GUIDELINES

The authors Ze-Gang Wei, Shao-Wu Zhang and Fang Jing declare they have no conflict of interests.

This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., Arrieta, J. M. and Herndl, G. J. (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. Proc. Natl. Acad. Sci. USA, 103, 12115–12120
2. Steele, J. A., Countway, P. D., Xia, L., Vigil, P. D., Beman, J. M., Kim, D. Y., Chow, C. E. T., Sachdeva, R., Jones, A. C., Schwalbach, M. S., *et al.* (2011) Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J., 5, 1414–1425
3. Gilbert, J. A., Field, D., Swift, P., Thomas, S., Cummings, D., Temperton, B., Weynberg, K., Huse, S., Hughes, M., Joint, I., *et al.* (2010) The taxonomic and functional diversity of microbes at a temperate coastal site: a “multi-omic” study of seasonal and diel temporal variation. PLoS One, 5, e15545
4. Kirchman, D. L., Cottrell, M. T. and Lovejoy, C. (2010) The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. Environ. Microbiol., 12, 1132–1143
5. Jiang, X., Hua, X., Xu, W. and Park, E.K. (2015) Predicting microbial interactions using vector autoregressive model with graph regularization. IEEE ACM T COMPUT. BI., 12, 254–261
6. Zhou, J., Deng, Y., Luo, F., He, Z. and Yang, Y. (2011) Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO_2. MBio, 2, e00122–e11
7. Gilbert, J. A., Steele, J. A., Caporaso, J. G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., McHardy, A. C., Knight, R., Joint, I., *et al.* (2012) Defining seasonal marine microbial community dynamics. ISME J., 6, 298–308
8. Eiler, A., Heinrich, F. and Bertilsson, S. (2012) Coherent dynamics and association networks among lake bacterioplankton taxa. ISME J., 6, 330–342
9. Faust, K. and Raes, J. (2012) Microbial interactions: from networks to models. Nat. Rev. Microbiol., 10, 538–550
10. Wei, Z. G. and Zhang, S. W. (2015) MtHc: a motif-based hierarchical method for clustering massive 16S rRNA sequences into OTUs. Mol. Biosyst., 11, 1907–1913
11. Chen, W., Cheng, Y.M., Zhang, C., Zhang, S.W., Zhao, H. (2013) MSClust: A multi-seeds based clustering algorithm for microbiome profiling using 16S rRNA sequence. J. Micro. methods, 94, 347–355
12. Cai, Y. and Sun, Y. (2011) ESPRIT-Tree: hierarchical clustering analysis of millions of 16S rRNA pyrosequences in quasilinear computational time. Nucleic Acids Res., 39, e95
13. Hao, X., Jiang, R. and Chen, T. (2011) Clustering 16S rRNA for OTU prediction: a method of unsupervised Bayesian clustering. Bioinformatics, 27, 611–618
14. Cheng, L., Walker, A. W. and Corander, J. (2012) Bayesian estimation of bacterial community composition from 454 sequencing data. Nucleic Acids Res., 40, 5240–5249
15. Wang, J., Chen, B., Wang, Y., Wang, N., Garbey, M., Tran-Son-Tay, R., Berceli, S. A. and Wu, R. (2013) Reconstructing regulatory networks from the dynamic plasticity of gene expression by mutual information. Nucleic Acids Res., 41, e97
16. DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G. L. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol., 72, 5069–5072
17. Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J. and Glöckner, F. O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res., 35, 7188–7196
18. Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam, S. A., McGarrell, D. M., Garrity, G. M. and Tiedje, J. M. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res., 33, D294–D296
19. Saraph, V. and Milenković, T. (2014) MAGNA, maximizing accuracy in global network alignment. Bioinformatics, 30, 2931–2940
20. Liu, F., Zhang, S. W., Wei, Z. G., Chen, W. and Zhou, C. (2014) Mining seasonal marine microbial pattern with greedy heuristic clustering and symmetrical nonnegative matrix factorization. Biomed. Res. Int., 2014, 189590
21. Lee, D. D. and Seung, H. S. (1999) Learning the parts of objects by non-negative matrix factorization. Nature, 401, 788–791
22. Nepusz, T., Petróczi, A., Négyessy, L. and Bazsó, F. (2008) Fuzzy communities and the concept of bridgeness in complex networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys., 77, 016107