biology

MDPI

Article

The Self-Organization of Marine Microbial Networks under Evolutionary and Ecological Processes: Observations and Modeling

Zhenghua Liu [1,2], Jianjun Wang [2], Delong Meng [1], Liangzhi Li [1], Xueduan Liu [1], Yabing Gu [1], Qingyun Yan [3], Chengying Jiang [4] and Huaqun Yin [1,*]

[1] School of Minerals Processing and Bioengineering, Key Laboratory of Biometallurgy of Ministry of Education, Central South University, Changsha 410083, China; liuzhenghua2017csu@163.com (Z.L.); delong.meng@gmail.com (D.M.); 205601006@csu.edu.cn (L.L.); xueduanliu@yahoo.com (X.L.); guyabing0207@163.com (Y.G.)
[2] State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China; jjwang@niglas.ac.cn
[3] School of Environmental Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China; yanqy6@mail.sysu.edu.cn
[4] Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; jiangcy@im.ac.cn
* Correspondence: yinhuaqun_cs@csu.edu.cn

Simple Summary: The properties and structure of ecological networks in marine microbial communities determine ecosystem functions and stability; however, the principles of microbial network assemblages are poorly understood. In this study, we revealed the influences of species phylogeny and niches on the self-organization of marine microbial co-occurrence networks and provided a mathematical framework to simulate microbial network assemblages. Our results provide deep insights into network stability from the perspective of network assembly principles and not just network properties, such as complexity and modularity.

Abstract: Evolutionary and ecological processes are primary drivers of ecological network constrictions. However, the ways that these processes underpin self-organization and modularity in networks are poorly understood. Here, we performed network analyses to explore the evolutionary and ecological effects on global marine microbial co-occurrence networks across multiple network levels, including those of nodes, motifs, modules and whole networks. We found that both direct and indirect species interactions were evolutionarily and ecologically constrained across at least four network levels. Compared to ecological processes, evolutionary processes generally showed stronger long-lasting effects on indirect interactions and dominated the network assembly of particle-associated communities in spatially homogeneous environments. Regarding the large network path distance, the contributions of either processes to species interactions generally decrease and almost disappear when network path distance is larger than six. Accordingly, we developed a novel mathematical model based on scale-free networks by considering the joint effects of evolutionary and ecological processes. We simulated the self-organization of microbial co-occurrence networks and found that long-lasting effects increased network stability via decreasing link gain or loss. Overall, these results revealed that evolutionary and ecological processes played key roles in the self-organization and modularization of microbial co-occurrence networks.

Keywords: microbial ecology; microbial community; network; self-organization; marine

check for updates

Citation: Liu, Z.; Wang, J.; Meng, D.; Li, L.; Liu, X.; Gu, Y.; Yan, Q.; Jiang, C.; Yin, H. The Self-Organization of Marine Microbial Networks under Evolutionary and Ecological Processes: Observations and Modeling. *Biology* **2022**, *11*, 592. https://doi.org/10.3390/biology11040592

Academic Editor: Ian Blaby

Received: 27 March 2022
Accepted: 7 April 2022
Published: 13 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The alterations of species interactions in ecological network, whether direct and indirect, are giving rise to cascading extinctions [1,2]. Thus, there is an urgent need to

explore the determinants of species interactions and the ways that they drive assemblage dynamics of ecological networks. Direct species interactions are widely considered to be constrained by evolutionary and ecological processes [3–8]. For examples, phylogeny and optimal niches or functional traits, the proxies of evolutionary and ecological processes, respectively, are key predictors of species interactions [9–13]. However, the question of which process plays the dominant role in constructing species ties is still understudied. Additionally, few studies pay attention to the relative contributions of these processes to indirect species interactions within ecological networks.

Ecological networks are inherently hierarchical [14], and thus the assembly driven by evolutionary and ecological processes can be characterized from various network levels, such as nodes, sub-networks (e.g., motif [15] and module), and the whole network. The metrics at node level, e.g., the degree [16] and within- and among-module connectivity [17], could measure the topological role of species in ecological networks [18–20]. However, less is known on how such topological properties are jointly driven by the phylogeny and niches of these species. At the sub-network level, motifs are common network backbones, and their elements may have similar traits or shared evolutionary histories [15,21,22]. Modules are purely identified by mathematical methods, such as the algorithms of greedy modularity optimization [23] and short random walks [24], but constrained by evolutionary and ecological processes revealed by previous reports on pollination networks, seed-dispersal networks and microbial networks [17,20,25,26]. At the whole network level, both processes are observed as primary drivers that shape microbial networks [27–29], but how these processes shape network global properties, such as self-organization and modularity, is less reported.

Such self-organization and modularization of ecological networks could be essentially revealed by establishing a mathematical model and simulating network assembly. In order to construct ecological networks, there are the three key features that must be re-emphasized from previous studies [30]: (i) the aggregations of nodes are dynamic; the network can gain or lose them, and they can connect or disconnect with each other [31], meaning that species have dynamic connections with other species. (ii) The nodes harbor specific properties, such as phylogeny and niches. (iii) External forces such as environmental disturbance can also act on network structures. Moreover, we expect that evolutionary and ecological processes should further take into account the self-organization of ecological networks. It is critical for us to understand how ecological networks respond to environmental changes and access network stability beyond posteriori knowledge.

Here, our objective is to study the evolutionary and ecological processes that shape marine microbial co-occurrence networks across multiple network levels. Based on global ocean metagenomic data from the TARA project across three depth layers and two size fractions [32], we analyzed the effects of phylogeny and niches on microbial co-occurrence networks across the levels of nodes, motifs, modules and the whole network, and further developed a novel mathematical model for simulating the self-organization of microbial networks based on a scale-free network [33,34]. We focused on three key questions: (1) How are direct and indirect interactions in microbial co-occurrence network constrained by both evolutionary and ecological processes? (2) What is the relative importance of evolutionary and ecological processes in determining co-occurrence network assemblages? (3) How do the characteristics of these two processes underpin co-occurrence network stability under environmental disturbance?

2. Materials and Methods

2.1. Datasets

We obtained meta data from the global ocean data of TARA project [32], and the samples were grouped into six subsets according to the three depth layers (SRF: surface water layer; DCM: deep chlorophyll maximum layer; MES, mesopelagic zone) and two size fractions (free-living: 0.22–3.0 μm; particle-associated: 0.80–5.0 μm) [32,35]. For the consistency of downstream analyses, the read depths of each sample of metagenomic 16S

rRNA tags were rarefied to 10,000. Phylogenetic tree of representative OTUs was extracted from SILVA 128 database (QIIME release) [36] based on 97% similarity.

2.2. Network Construction, Motif and Latent Space Analyses

In each sub-dataset, OTUs observed in over a half of the samples were selected for network constructions. The correlations between OTUs were inferenced via SparCC tool [37–39] and carried out by 'sparcc' function (iter = 100, inner_iter = 100, th = 0.3) [38]. The correlations between OTUs that were no less than 0.65 were kept for network constructions. The memberships of nodes were identified by the fast-greedy modularity optimization method [23] via 'cluster_fast_greedy' function. The frequencies of each node for motif positions in undirected networks were counted using Simmons's method, which was conducted via the 'mcount' function (six_node = TRUE) [40]. For each node in the network with the largest connectivity, the positions within the latent space in two dimensionalities and the probabilities of clustering memberships were calculated by the Variational Bayes Latent Position Cluster Model (VBLPC) [41]. To determine the optimal number of groups, we searched the minimum Bayes information criterion (BIC) with desired group scopes (Figure S1A). The latent space distances between OTUs were calculated as Euclidean distances. These network analyses were conducted in R version 3.6.1 with the packages of 'SpiecEasi' V1.0.7, 'iGraph' V1.2.4.2, 'bmotif' V1.0.0 and 'VBLPCM' V2.4.5 [38,40–42].

2.3. Statistical Analyses

Phylogenetic tree was transferred into pairwise distance matrix via 'cophenetic' function. The optimal habitat value, as a niche for each out, was calculated based on environmental variables [43]. Between-OTU niche differences were calculated as Euclidean distances, and between-OTU differences in topology properties of network motifs and module levels were calculated as Bray distance based on motif position frequencies and the probabilities of clustering memberships. We used multiple logistic regression in the distance matrices with 999 permutations [44] to quantify the effects of phylogeny and niches on network adjacent matrices (binary data). Similarly, we also used multiple linear regression in distance matrices to examine the regression coefficients of pairwise phylogeny and habitat distance against pairwise motif or module differences. These analyses were performed with packages of 'base', 'vegan' V2.5–6 [45] and 'ecodist' V 2.0.1 [46].

To estimate species preference in phylogeny and niches within networks, we proposed the novelty network-based metrics (Figure 1), that is, the mean neighbor phylogeny distance ($MNPD$) and mean neighbor niche distance ($MNND$), respectively. These two metrics were calculated by the following formula:

$$MNPD_i = \frac{1}{n} \times \sum_{j}^{n} PD_{ij} \; i \neq j \tag{1}$$

and:

$$MNND_i = \frac{1}{n} \times \sum_{j}^{n} ND_{ij} \; i \neq j \tag{2}$$

where PD_{ij} and ND_{ij} denote phylogenetic and niche distance between species i and its neighbor j, and n is the number of neighbors belonging to species i. Further, we constructed 1000 random networks corresponding to empirical networks according to Erdos-Renyi model and generated a null distribution of $null.MNPD_i$ and $null.MNND_i$ values. The standardized effect scores of $ses.MNPD_i$ and $ses.MNND_i$ were given by:

$$ses.MNPD_i = \frac{obs.MNPD_i - mean(null.MNPD_i)}{sd(null.MNPD_i)} \tag{3}$$

and:

$$ses.MNND_i = \frac{obs.MNND_i - mean(null.MNND_i)}{sd(null.MNND_i)} \tag{4}$$

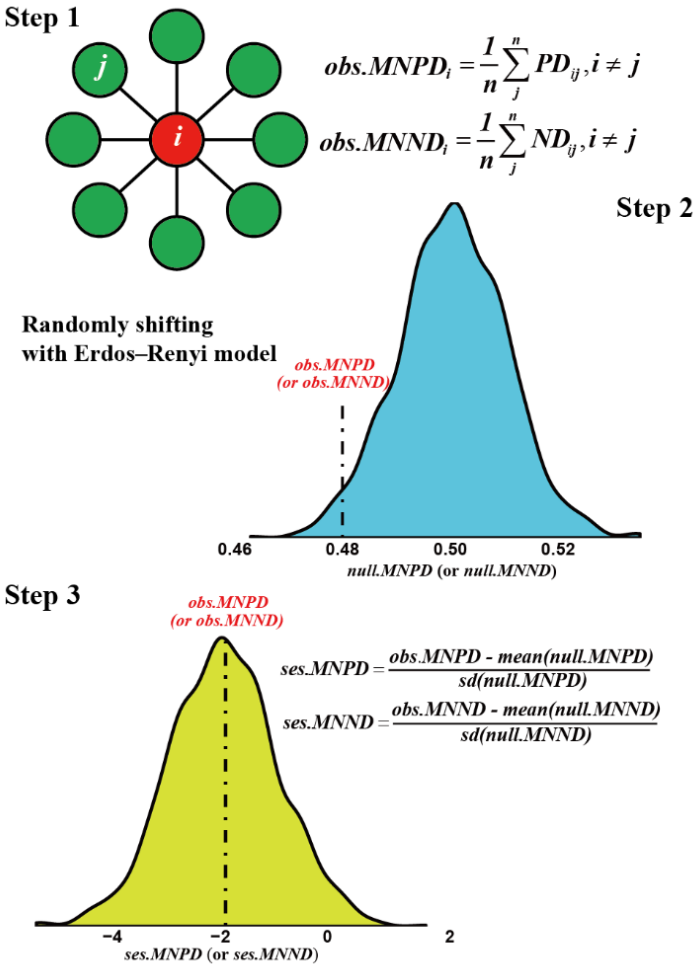

Figure 1. The flowchart of the summarizing procedure for estimating mean neighbor phylogeny and niche distance (*MNPD* and *MNND*) at node level. First, the observed *MNPD* and *MNND* (*obs.MNPD* and *obs.MNND*) were quantified based on empirical networks. Second, the links of empirical networks were randomly shifting with the Erdos–Renyi model, and null distributions of *MNPD* and *MNND* (*null.MNPD* and *null.MNND*) were generated. Third, the standardized effect sizes of *MNPD* and *MNND* (*ses.MNPD* and *ses.MNND*) were calculated based on observed and null values. The absolute values of *ses.MNPD* (*ses.MNND*) larger than two show statistical significance (two sides, $p < 0.05$).

To further explore the design principles of ecological networks at the motif level, we calculated the preferences in phylogeny and niche of motif position j ($j = 1, 2, 3 \ldots, 148$) by:

$$ses.MNPD_j = \sum_i (f_{ij} \times ses.MNPD_i) \tag{5}$$

and:

$$ses.MNND_j = \sum_i (f_{ij} \times ses.MNND_i) \tag{6}$$

where f_{ij} is the frequency of the node i in motif position j ($\sum_i f_{ij} = 1$). As for the module level, the preferences in the phylogeny and niches of module k were given by:

$$ses.MNPD_k = \frac{\sum_i (\lambda_{ij} \times ses.MNPD_i)}{\sum_i \lambda_{ij}} \tag{7}$$

and:

$$ses.MNND_k = \frac{\sum_i (\lambda_{ij} \times ses.MNND_i)}{\sum_i \lambda_{ij}} \quad (8)$$

where λ_{ij} is the probability of node i clustering into module k.

We tested the linear relationships between the phylogenetic or niches distance and network path distance with 'lm' functions. Further, a random forest algorithm [47] was used to classify the network path distance based on species phylogeny and niche distance. Finally, we calculated the Pearson's correlations between phylogeny or niches distance and latent space distance across network path distance. These analyses were performed with packages of 'base' and 'randomForest' V4.6–14.

2.4. Mathematical Mechanism Model

2.4.1. The dynamic of a Scale-Free Network in Ecology Community

Most of ecological networks were generally self-organized and exhibited scale-free property, such as food webs [48,49], species co-occurrence [20,50] and pollination networks [51,52]. In scale-free networks, the probability p of a vertex with certain k links (or degree) followed a power law distribution [33,34,53]:

$$p(k) = k^{-\gamma} \quad (9)$$

where γ is an attraction factor. A greater γ means the weaker ability of a node to attract other nodes [54]. The value of $\gamma < 3.47$ holds for almost all of empirical networks, and the most of these are range from 2 to 3 [53,55].

To explore the assemblage dynamic of microbial co-occurrence networks, we took evolutionary and ecological processes into account and extended the above formula (Figure 2A). The connectivity strength c_{ij} between species i and j is given by:

$$c_{ij} = W_P k_{i,j}^{-\gamma_{P,i,j}} + W_N k_{i,j}^{-\gamma_{N,i,j}} \quad i \neq j;\ i, j \in V \quad (10)$$

where W_P and W_N are the weight of preferences for the similarity of phylogeny and niches to c_{ij}, respectively, and $k_{i,j}$ is the average degree of i and j. The parameters of $\gamma_{P,i,j}$ and $\gamma_{N,i,j}$ are the mutual attraction factors between i and j driven by phylogeny and niche similarity, as described by following formula:

$$\gamma_{P,ij} = -a + \sum_l w_{P,l} \times P_l \quad (11)$$

and:

$$\gamma_{N,ij} = -a + \sum_l w_{N,l} \times N_l \quad (12)$$

where a is a constant representing the original attraction between species, $w_{P,l}$ and $w_{N,l}$ are the weights of the preference of phylogeny and niches similarity among species across network path distance l. If $l = 0$, $w_{P,0}$ and $w_{N,0}$ reflect the direct phylogeny (P_0) and niche (N_0) attraction, respectively, originating from the preferences for phylogeny and niche similarity between i and j. If $l > 0$, $w_{P,l}$ and $w_{N,l}$ reflect the weight of the indirect phylogeny (P_l) and niche (N_l) attraction, respectively; attraction received from neighboring ($l = 1$); and the non-neighboring ($l > 1$) vertexes with a distance of l to i or j.

We assumed that $w_{P,l}$ and $w_{N,l}$ accorded with an exponential attenuation model:

$$w_{P,l} = \frac{A_P}{1 + \exp(B_P \times (l+1))} \quad l = 0,\ 1,\ 2, \ldots, m \quad (13)$$

and:

$$w_{N,l} = \frac{A_N \times (1 + Eh)}{1 + \exp(B_N \times (l+1))} \quad l = 0,\ 1,\ 2, \ldots, m \quad (14)$$

where A_P and A_N reflects the relative importance of phylogeny and niches, respectively. B_P and B_N is the attenuation rate of weight across network path distance l, Eh ($Eh > 0$) is the degree of environment heterogeneity. Then, W_P is given by:

$$W_P = \sum_l w_{P,l} / \sum_l (w_{P,l} + w_{N,l}) \tag{15}$$

and:

$$W_N = 1 - W_P \tag{16}$$

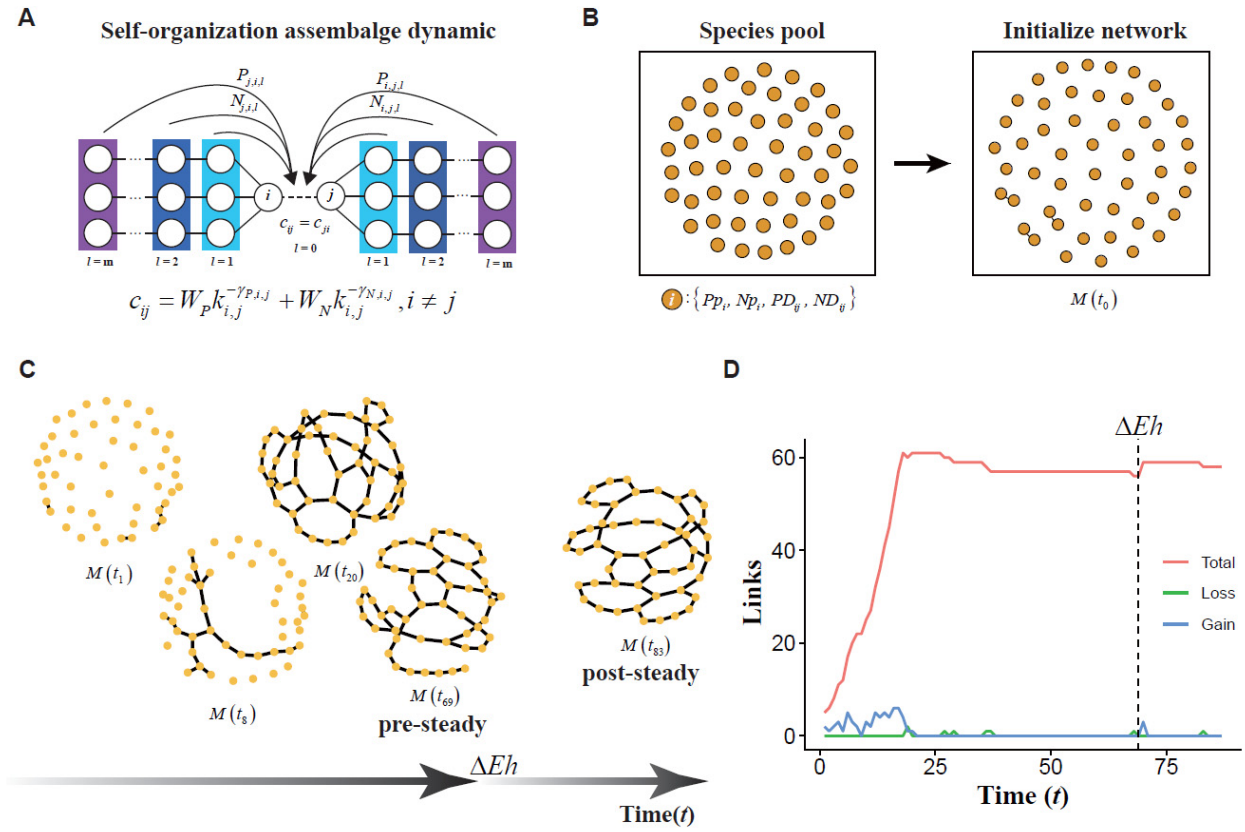

Figure 2. The simulation of self-organized assemblages in microbial networks. (**A**) The self-organized assemblages of microbial networks constrained by species phylogeny and niches. The connectivity c_{ij} between species i and j is constrained by phylogeny (P) and niches (N). These two constraints could be directly affected by this pair of species (l = 0) and their neighbor (l = 1) or non-neighbor species (l > 1). (**B**) Characteristics of species pool and network initialization. In species pool, species i has preferences in phylogeny (Pp_i) or niche similarity (Np_i). The phylogeny and niche distance between species i and j are denoted by PD_{ij} and ND_{ij}, respectively. For network initialization, three pairs of species are randomly selected and linked. The initialized network is denoted by symmetrical adjacent matrix M (t = 0). (**C**) The simulation of self-organized assemblages in microbial networks. Symmetrical adjacent matrix M (t) was repeatedly updated according to network assembly procedures until termination events were triggered (See details in Supplementary Material) and the complete assembly of a pre-steady network. When environment heterogeneity changes (ΔEh), M (t) becomes involved in iterations until iteration termination and achieves a post-steady state. (**D**) The link dynamic in network assembly process. The red line is the number of network links across assembly time. The green and blue lines are the number of loss and gain links in assembly processes.

The mutual attractions between pair nodes driven by phylogeny (P_l) and niches (N_l) in network path distance l are given by following formula:

$$P_l = \frac{1}{2}\left(P_{i,j,l} + P_{j,i,l}\right), \;\; P_{i,j,0} = P_{j,i,0} \tag{17}$$

and:

$$N_l = \frac{1}{2}\left(N_{i,j,l} + N_{j,i,l}\right), \;\; N_{i,j,0} = N_{j,i,0} \tag{18}$$

where $P_{i,j,0}$ and $N_{i,j,0}$ represent the direct mutual attractions of phylogeny and niches between i and j, respectively, given by:

$$P_{i,j,0} = \frac{1}{2} \times \frac{Pp_i + Pp_j}{PD_{ij}} \tag{19}$$

and:

$$N_{i,j,0} = \frac{1}{2} \times \frac{Np_i + Np_j}{ND_{ij}} \tag{20}$$

where Pp_i and Np_i denote the preference of species i in phylogeny and niche similarity, respectively. The phylogenetic and niche distances between i and j are represented by PD_{ij} and ND_{ij}, respectively. Considering the indirect effects on the co-occurrence of i and j, the phylogenetic and niche attractions for i or j from non-neighbor species of j or i with a distance of l are given by Formulas (19) and (20), respectively, as follows:

$$P_{i,j,l} = \frac{1}{2} \times \left(\frac{1}{q} \times \sum_{k=1}^{q} \frac{Pp_{(k,i,l)} + Pp_j}{PD_{(k,i,l)j}} + \frac{1}{p} \times \sum_{o=1}^{p} \frac{Pp_{(o,j,l)} + Pp_i}{PD_{(o,j,l)j}}\right) \tag{21}$$

and:

$$N_{i,j,l} = \frac{1}{2} \times \left(\frac{1}{q} \times \sum_{k=1}^{q} \frac{Np_{(k,i,l)} + Np_j}{ND_{(k,i,l)j}} + \frac{1}{p} \times \sum_{o=1}^{p} \frac{Np_{(o,j,l)} + Np_i}{ND_{(o,j,l)j}}\right) \tag{22}$$

where $k \neq j,\ o \neq i,\ k, o\ \epsilon V.$ (k, i, l) and (o, j, l) is the aggregations of neighbor or not-neighbor nodes belonging to i and j with distance l, respectively. The size of (k, i, l) and (o, j, l) are q and p, respectively.

2.4.2. The Specie Pool Construction

In order to obtain the aggregation of species for ecological network simulations, we generated species pools with regard to phylogeny and niches (Figure 2B). The community phylogeny was constructed by a random phylogenetic tree according to the Paradis's algorithms [56], and the branch length was generated from uniform distribution $U(0, 1)$. The niche distance between species i and j is given by:

$$ND_{ij} = cPD_{i,j} + \varepsilon \quad \text{for } PD_{ij} < PS \tag{23}$$

and:

$$ND_{ij} = U(\beta_1 cPS, \beta_2 cPS) + \varepsilon \quad \text{for } PD_{ij} \geq PS;\ \beta_1 < \beta_2 \tag{24}$$

where PS is phylogeny signal in the optimal niche, which is a threshold of the phylogenetic distance for the linear relationship between phylogenetic and niche distance. The parameters of c, β_1 and β_2, are constant, and ε is noise sampled from the normal distribution $N(0,\ 0.01)$. Finally, pairwise phylogenetic distance PD_{ij} and niche distance ND_{ij} are standardized by Z-score transformation.

2.4.3. Modeling of Network Assembly

At time t_0, the species pool with v species and a symmetrical adjacency matrix $M_{v \times v}(t = 0)$ $(m_{i,j} = m_{j,i},\ m_{i,j} \in 0,\ 1)$ with $e(t = 0)$ links is generated. Additionally,

we then repeated the following five steps until there were no loss and gain links among nodes or the number of iterations reached the maximum value. The brief procedures are shown in Figure 2.

Step 1: At time t, existing species i in networks are likely to interact with new introduced species or other existing species that have no interaction with species i. Regarding this, through Formula (25), we first calculated the pre-connectivity strength $pc_{i,j}$ between connected node i and its non-neighbor node j with a pre-average degree of i and j: $pk_{i,j} = k_{i,j}(t) + 1$:

$$pc_{i,j} = W_p pk_{i,j}^{-\gamma_{p,i,j}} + W_h pk_{i,j}^{-\gamma_{h,i,j}} \tag{25}$$

Step 2: To restrict the number of novel interactions generated at time t, we define a threshold pc_{th} that is a quantile of pc_{ij}, with a given probability p_{cth} and a minimal threshold pc_{min}^{con}. If $pc_{ij} > \min\left(pc_{th},\ pc_{min}^{con}\right)$, node i would prepare to connect node j with probability $\min\left(pc_{ij}, 1\right)$ through a binomial experiment. After that, a pre-connectivity adjacency matrix M_{con} is created.

Step 3: Then, we generate a transitive adjacency matrix M_{tra} through:

$$M_{tra} = M(t) + M_{con} \tag{26}$$

Step 4: After the introduction of novel species or interactions, original network patterns are changed and some existing interactions between species are weakened and may disconnect. Thus, we calculated the connectivity strength matrix C_{tra} based on M_{tra} through Formula (25), and then defined a threshold of connectivity strength pc_{min}^{dis} for disconnection. If $c_{tra,i,j} < pc_{min}^{dis}$, the link between i and j would disconnect with a probability p_{dis} through the binomial experiment. After that, a pre-disconnected matrix M_{dis} was created.

Step5: Finally, we obtained the new adjacency matrix at time $t + 1$ through:

$$M(t+1) = M_{tra} - M_{dis} \tag{27}$$

2.4.4. Sensitive Analysis

We conducted a sensitivity analysis for the network assemblage model by Gaussian generalized linear model (GGLM) and random forest classification. First, the importance of phylogeny was set as twice as niche in the network assemblage, and the other of parameters varied within the given ranges (Table S1). Then, we varied the rest of parameters and generated 100 sets. For each parameter set, we sampled 100 replicates from modeling simulations. Finally, we defined the sensitivity of modeling outcomes to parameters as the regression coefficients of GGLM, where the dependent variables were the contributions of phylogeny and niche distance to network path distance given by random forest classification, and the independent variables were variable parameters.

To simulate the response of networks in the changes of environment heterogeneity, we examined the effects of the indirect interactions on the response of network dynamics in environment heterogeneity. In this model, environment heterogeneity is represented by Eh, which alters the weight of the contributions of phylogeny and niches to the probability of species co-occurrence. After completion of pre-steady network $G(V, E)$ with presetting Eh (default $Eh = 0$), Eh was increased to 2 and adjacent matrix of G was repeatedly updated. When the iterations achieved termination, we obtained a post-steady network $G'(V', E')$. Finally, we calculated link turnover, including the percentages of link gain (G) and loss (L) by following formula:

$$G = \frac{|E_g|}{|E|} \times 100\%,\ E_g \subseteq E',\ E_g \cap\ E = O \tag{28}$$

and:

$$L = \frac{|E_g|}{|E|} \times 100\%,\ E_g \subseteq E,\ E_g \cap\ E' = O \tag{29}$$

where $|E|$ represents the number of links in pre-steady network.

3. Results

Species phylogeny and niche similarity significantly correlated ($p < 0.05$) with, but showed differentiated contributions to, the marine microbial co-occurrence network structure at the motif, module and whole-network level (Figure 3A). Particularly, phylogeny made more contributions to network construction than niches at the motif level (Figure 3A; paired t-test, $p = 0.07$), while it was inconsistent at the module and whole-network level (Figure 3A; both $p > 0.1$). Specifically, at mesopelagic zone (MES), phylogeny, rather than niches, was a primary driver of particle-associated communities across three network levels (Figure 3A; $p = 0.016$), but not for free-living communities (Figure 3A). When the network path distance between species was considered, both phylogeny and niche distance positively correlated with network path distance up to around three, and then showed incongruent trends (Figure 3B). Such a phenomenon was further supported by the fact that phylogeny and niche distances were significantly ($p < 0.05$) positively related to network path distance within the same modules (Figure 3C), and these relationships were stronger than those across modules (Figure 3C). These results implied phylogeny and niche processes had strong constraints on network assemblage within a short range of network path distance, which may be particularly important for network module clustering.

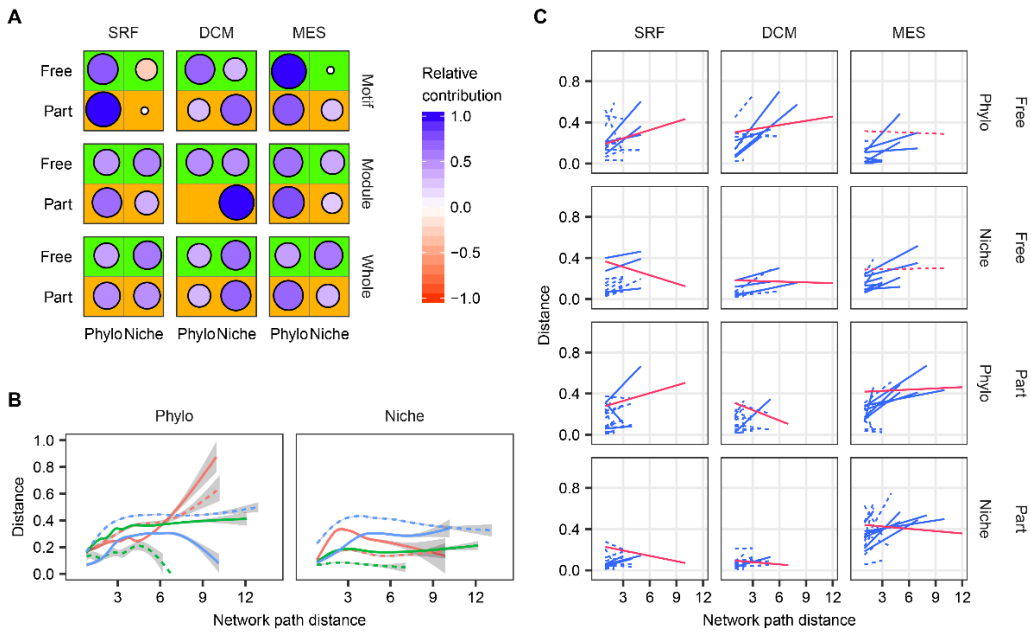

Figure 3. The effects of species phylogeny and niches on marine microbial network constructions. (**A**) The relative contributions of species phylogeny (Phylo) and niche distance to network construction across motif, module and whole-network (whole) levels in free-living (Free) and particle-associated (Part) communities at surface water layer (SRF), deep chlorophyll maximum layer (DCM) and mesopelagic zone (MES), where pattern size and color intensity are the strength of relative contribution, and colors represent the positively (blue) or negatively (red) relative contributions. (**B**) The relationships between network path distance and phylogeny or niche distance at the whole network level, where color represent depth layers (red: SRF; green: DCM blue: MES), and line type represents community fractions (solid: Free; dashed: Part). The curves were fitted by generalized additive models. The shaded region indicates standard errors of the curves. (**C**) The linear relationships between network path distance and phylogeny or niche distance across modules (blue line) or within the same modules (red line), where line type is the statistical significance of linear regression analysis (solid: $p < 0.05$).

To determine whether species prefer to interact with phylogenetically close species or ecologically similar species, we developed novel indicators of mean neighbor phylogeny distance (*MNPD*)- and mean neighbor niche distance (*MNND*)-based ecological networks. Compared with the module level, nodes and motif positions showed the highest and lowest divergence, respectively, in *ses.MNPD-ses.MNND* plots (Figure 4A). Across six groups of microbial communities, over 95% of nodes showed a preference in phylogeny or niche-similar neighbors (*ses.MNPD* < 0 or *ses.MNND* < 0) and more than 65% of them showed a statistical significance (*ses.MNPD* < −2 or *ses.MNND* < −2; Figure 4B). For motif positions, all *ses.MNPD* and *ses.MNND* values were less than 0. This was especially true for the mesopelagic zone as all motif positions showed significant preferences in phylogeny or niche similarity: *ses.MNPD* or *ses.MNND* values < −2 (Figure 4B). These results indicate that phylogeny and niche processes have stronger constraints on network motifs or modules than on nodes. In addition, *ses.MNPD* had more significantly ($p < 0.05$) positive influences on node connectivity in particle-associated communities than in free-living ones, which were supported by the generalized linear model (Table 1).

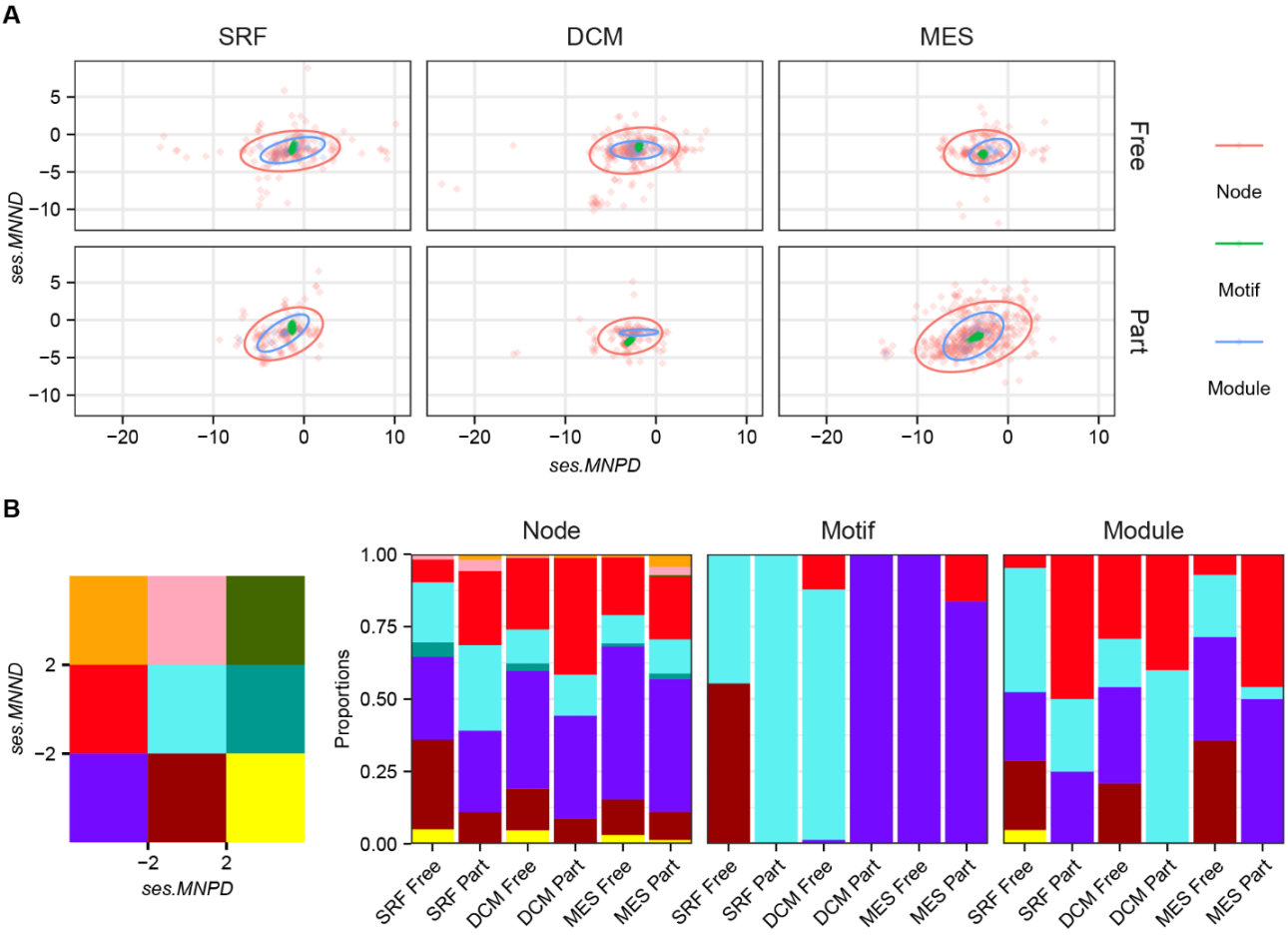

Figure 4. The mean neighbor phylogeny and niche distance of species in microbial networks. (**A**) The *ses.MNPD–ses.MNND* plot for marine microbial communities across node, motif and module levels, where *ses.MNPD* or *ses.MNND* of motifs and modules were calculated according to the proportions of their node members. (**B**) The block proportions in *ses.MNPD–ses.MNND* plots across nodes, motifs and modules, where a *ses.MNPD* (or *ses.MNND*) larger than 2 indicates that one object has a significant preference in a more distantly related phylogeny (or niche) at same network level, while a *ses.MNPD* (or *ses.MNND*) less than −2 indicates one object has significant preference in a more closely related phylogeny (or niche).

Moreover, a random forest classification analysis [47] showed that phylogeny and niche distance affected both direct and indirect species interactions, that is, network path distances = 1 and >1, respectively (Figure 5). Regarding a large network path distance, the contributions of phylogeny and niches for the network path distance between two species generally increased and then decreased with a network path distance of around six. Such patterns showed a peak in the network path distances from two to five (Figure 5A). Notably, the contribution of phylogeny was around two times higher than that of niches in particle-associated communities at MES, and showed stronger long-lasting effects on indirect species interactions (Figure 5A). The highest importance of phylogeny and niches to network path distance classification were 0.286 and 0.247, respectively (Figure 5A).

Table 1. The influences of *ses.MNPD* and *ses.MNND* on node connectivity were analyzed by the generalized linear model. SRF: surface water layer; DCM: deep chlorophyll maximum layer; MES: mesopelagic zone.

Size Fraction	Layer	*ses.MNPD*	*ses.MNND*
Free-living	SRF	0.133	−0.0236
	DCM	0.2302	0.8291 *
	MES	−0.1677	−0.2277
Particle-associated	SRF	0.7122 ***	−0.0927
	DCM	0.3903 **	−0.0811
	MES	0.1315	−0.7282 **

***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$.

To further assess the influence of phylogeny and niches on indirect species interactions, we calculated the latent space distance between species with a variational Bayes method [41] and estimated the Pearson's coefficients between the distance matrices of latent space and phylogeny or niches across the network path distance. We found that species phylogeny or niche distance significantly correlated with the latent space distance ($p < 0.05$; Figure 5B), and these correlations generally decreased with increasing network path distance when ignoring modularity (Figure 5B). Interestingly, when including modularity, the constraints of phylogeny on species latent distance in ecological networks weakened. This is especially true for the particle-associated communities in the deep chlorophyll maximum layer (DCM; Figure S1B) and there were no significant ($p > 0.05$) correlations between species phylogenetic and latent space distances.

To disentangle the ecological significance of the observed indirect species interactions from the perspective of network stability, we established a mathematical model based on scale-free network and simulated the processes of microbial co-occurrence network self-organized assemblages (Figure 2). In a sensitive analysis of model parameters, our results showed that the distance of long-lasting effects l had a greater influence than the other parameters, such as community size S and attraction constant a, regarding the contribution of phylogeny and niche to indirect interactions, particularly in long-distance indirect interactions (Tables S2 and S3). In addition, the strength of the phylogeny signal in the niche optima of species significantly ($p < 0.05$) enhanced the contribution of niches to direct interactions in the community assemblages, primarily driven by phylogeny (Table S3).

We further simulated the response of the ecological network in environment heterogeneity (defined as continuous variable *Eh*), which could decease the weight of phylogeny preference on the probability of species co-occurrence [57]. After completeness of pre-steady network assemblage, *Eh* was shifted from 0 to 20 and the post-steady network and was completed until the iterations of the matrix of connection strength terminated again (Figure 2C). To estimate network stability, we then calculated the link turnover [58], that is, the percentage of the gain and loss links. When species indirect interactions were considered in network construction, the ecological network had a better stability in more heterogeneous environments. Specifically, towards the long distance of long-lasting effects l, the percentage of gain (Figure 6A) and loss (Figure 6B) links reduced. Further, the influ-

ences of long-lasting effects on network stability in reducing link turnover were stronger in large, as opposed to small, communities (ANOVA $p < 0.05$). Moreover, the percentage of gain links showed unimodal patterns along the gradient of phylogenetic signals in optimal niches, with a peak at around 0.08 (Figure 6C). For the percentage of loss links, however, there was a minor variation until the phylogenetic signal increased up to around 0.01, and then it decreased rapidly (Figure 6D).

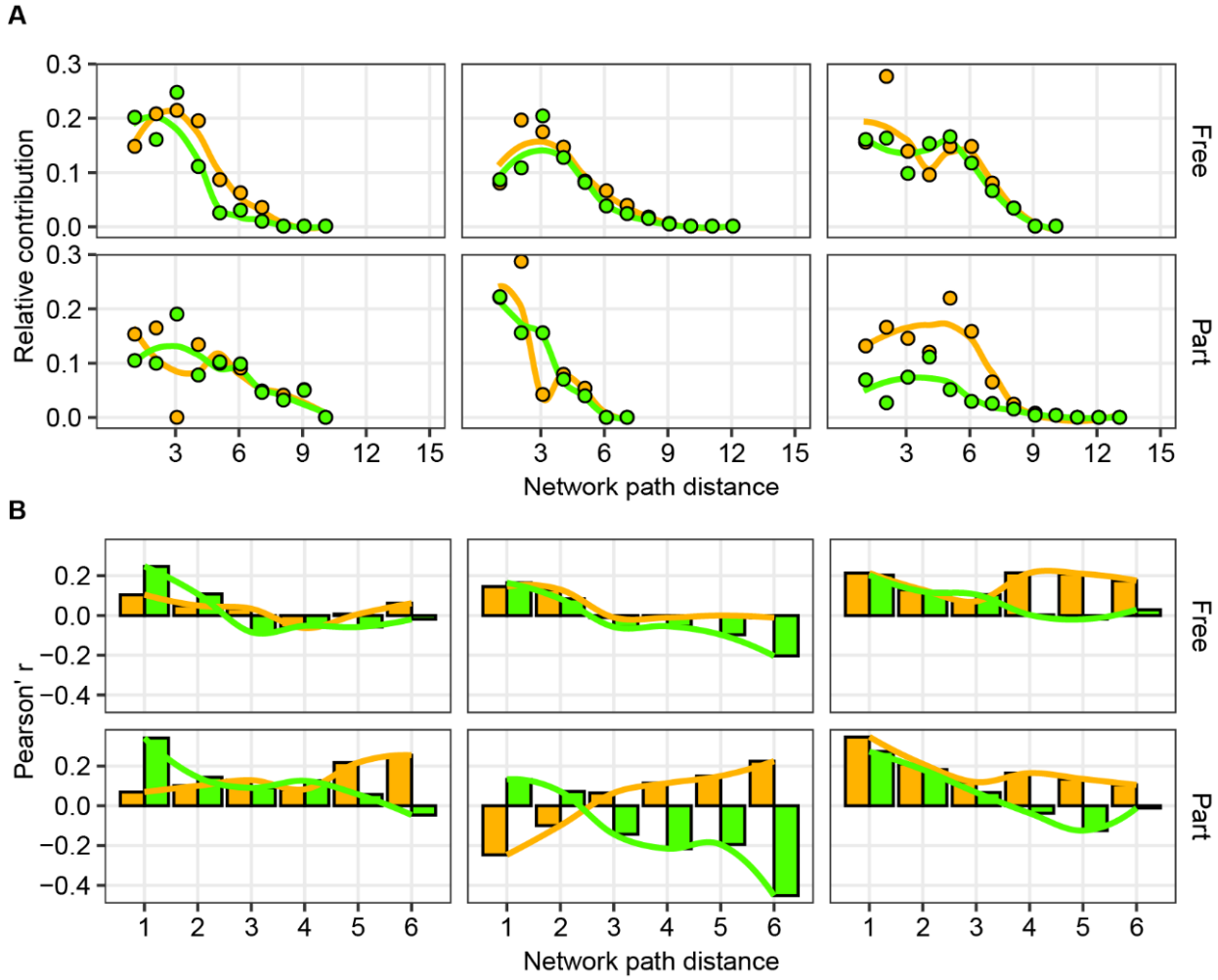

Figure 5. The contribution of species phylogeny and niche distance to their direct and indirect interactions. (**A**) The relative contribution of species phylogeny and niche distance to classifications of network path distance estimated by random forest classification analysis. The colors of points or lines vary with phylogeny (orange) and niche (green). The line is a Loess fit based on the points. (**B**) The Pearson's correlations between species phylogeny or niche distance and network latent space distance ignore the potential for modularity across the network path distance. The colors of bars or lines vary with phylogeny (orange) and niche (green). The lines are a Loess fit based on the values of bars.

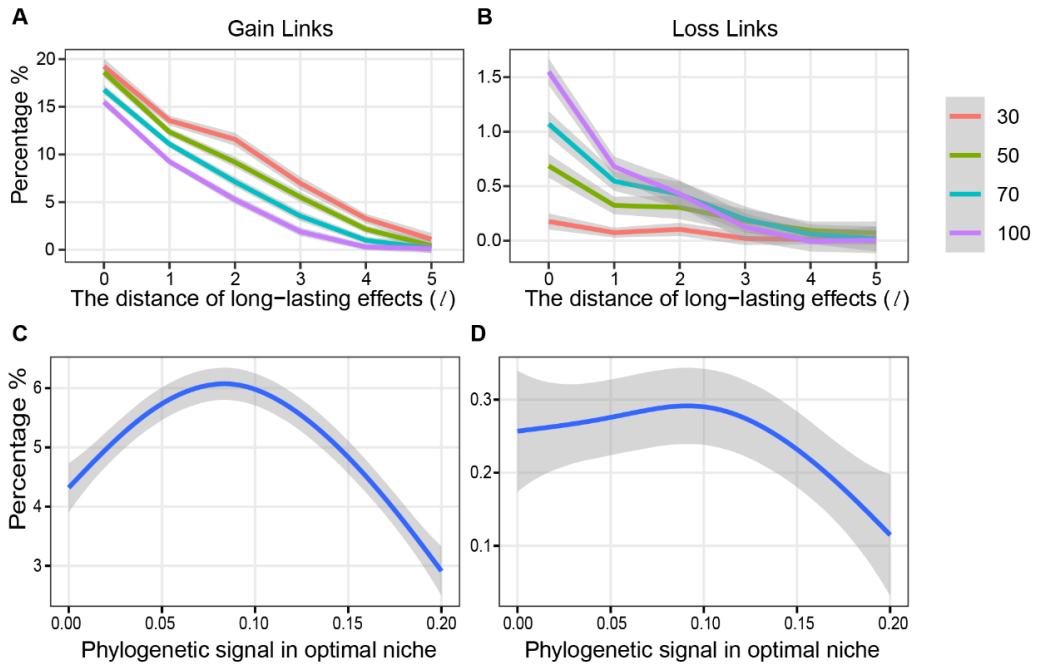

Figure 6. The response of network links in the changes in environmental heterogeneity. The gain links (**A**) and loss links (**B**) of the network across the distance of long-lasting effects (l) and community size (N; labeled with different colors) when environmental heterogeneity changes. The lines are a Loess fit based on 600 iteration points (6 × 100 replicates). The shaded regions indicate standard errors of the Loess curves. The gain links (**C**) and loss links (**D**) of the network across the phylogenetic signal in optimal niches (here, l = 2 and S = 50). The lines are fitted by generalized additive models (GAM) based on 10,000 iteration points (100 × 100 replicates). The shaded regions indicate standard errors of the curves of GAM.

4. Discussion

The evolutionary and ecological processes on direct [5,6,8,59,60] and indirect species interactions [1,61–64] attracted extensive attention; however, the relative importance of these two processes on ecological networks and its assemblage dynamics are less understood. Here, we clearly showed that species interactions were evolutionarily and ecologically constrained in marine microbial co-occurrence networks across the four levels of nodes, motifs, modules and whole networks. We further found that evolutionary processes revealed stronger long-lasting effects on indirect species interactions than ecological processes and dominated the network assemblage in spatially homogeneous but microscopically heterogeneous environments, such as particle-associated communities in MES. Finally, we simulated the assemblage dynamics of microbial co-occurrence networks with mathematically modeling and revealed that, regarding large network path distances, long-lasting effects improved network stability, as indicated by the decreasing link turnover.

We found that evolutionary and ecological processes generally constrained direct and indirect species interactions in marine microbial co-occurrence networks. For direct interactions, the findings were consistent with previous reports on predator–prey networks and competition networks [11,60]. For instance, over 80% of the observed realized and unrealized predator–prey interactions, involving 20 ground beetles and 115 prey species, could be predicted with phylogeny and functional trait information [11]. We further found that node connectivity in ecological networks was also associated with phylogeny and niches, implying that there was a phylogenetical conservatism [65] or niche specialization [66] in species interactions.

For indirect interactions, the findings on the strong effects of the two processes were consistent with previous reports of empirical networks among trophic levels, such as plant–animal-mutualistic [67] or predator–prey networks [1]. For example, with regard to the

evolutionary processes in mutualistic networks, it has stronger constraints on consumer levels than resource levels and maintains network stability via positive within-guild indirect interactions between phylogenetically related species [67,68]. As for ecological processes, species density and function traits mediate indirect species coexistence and coextinction in plant–insect communities [1,62,69]. However, these reports show that indirect species interactions merely happen among three or four trophic levels, while we highlight that the two species indirectly connected within an ecological network could be generally constrained by phylogeny and niches, no matter whether they belong to different trophic levels or not.

Species interactions are fundamental for the assembly of ecological networks, and we expect that the assembly of the co-occurrence network is also driven by evolutionary and ecological processes. However, the ways that co-occurrence network assembly is driven by these two processes is a top-level issue, and could help us to essentially understand the self-organization of the microbial co-occurrence network. With integrations of empirical results, we proposed three reasonable principles of co-occurrence network assembly in marine microbial communities.

First, the phylogeny and niche are the primary drivers used to build the basic network blocks, such as motifs. For species, our null model analyses suggested that over 74% of species preferred to co-occur with phylogeny- or niche-related species, which was strongly supported by previous studies [20,26,28,70]. For motifs, all motif positions displayed the accordant preferences in phylogeny and niche similarity, which revealed that evolutionary and ecological processes had strong influences on building network motifs. It should be noted that, here we did not further consider that the effects of phylogeny and niches may vary regarding the type and strength of species interactions [10,12,71–74].

Second, both drivers had long-lasting effects on indirect species interactions within co-occurrence networks, and the strength of such effects decayed toward large network path distances. The long-lasting effects observed here are that the phylogeny and niche distance of pair species partly contribute to their indirect interactions, despite the large network path distance between them. Such findings agree with previous studies on food webs [12,63,75]. Furthermore, we revealed that the strength of long-lasting effects generally weakened with the increasing network path distance. Interestingly, the long-lasting effects of phylogeny and niches mostly occurred within a network path distance of around six, which may explain the phenomenon that the long-lasting effects were stronger within the same modules than across different modules. Overall, these long-lasting effects are short-range and could be the key mechanisms to enhance network modularity.

Third, phylogeny was more important than niches in determining direct or indirect species interactions when spatial homogeneity was greater. As marine depth increased and spatial homogeneity increased, the relative contributions of phylogeny to network constructions increased, while niches decreased (Figure 3A). For example, in the mesopelagic zone and surface water layer, phylogeny dominated the network assembly of particle-associated communities. However, free-living communities in the mesopelagic zone have no such phenomenon, which reveals that the microscale stability provided by particle may further strengthen the effects of phylogeny on the co-occurrence network assembly.

Based on these three principles, we proposed a mathematical model based on a scale-free network and simulated the self-organization of the co-occurrence network, which was key to exploring the network properties preceding empirical evidence. For example, regarding large network path distance, the long-lasting effects of both processes reduced link turnover and gave the network a strong resistance to environmental disturbances, indicating that the effects of indirect interactions on network stability may be more important than direct interactions [76]. This is particularly true for trophic networks with a low connectivity and high modularity [77] and for mutualistic networks with a nested architecture [78,79]. Moreover, the strong phylogenetic signals in optimal niches stabilized the successions of the co-occurrence network as expected when the determinants of network assembly shifted from evolutionary to ecological processes [80] (Figure 5B). However, for

weaker phylogenetic signals, the percentage of gain links increased with the increasing phylogenetically signal. This might be caused by the increasing level of species connectivity with a higher strength than the minimum threshold of connections. This priori knowledge highlights that network stability should be further explored from design principles, except for network topologies such as size [81], connections [82] and modularity [83], because the latter are post analytical results and the outcomes of the former. Finally, the contributions of phylogeny and niches to direct and indirect species interactions had a great variety despite being from the same set of parameters, and this variety was greater for large network path distances (Figure S2). On the one hand, this supported multi-stability as the mechanism gave rise to different communities under the same environmental conditions [84]. On the other hand, it also indicated that stochastic processes may greatly affect the contributions of phylogeny and niches to network assembly in remotely indirect interactions.

Here, although we show a reasonable picture on how species phylogeny and niches shape microbial co-occurrence networks, three major caveats could be acknowledged regarding the analytical framework. First, microbial co-occurrence networks were inferenced by the correlation detection technique, which would lead to reasonable false positive rates [39], although we did not consider extremely rare species in network constructions. Meanwhile, specific ecological relationships, such as predation, competition, symbiosis and mutualism, could not be further accessed, and these ecological networks may have different and novel findings. Second, there are many correlation detection methods used to construct microbial networks, such as CoNet [85], RMT [86], MIC [87], LSA [88] and FlashWeave [89], which consistently produce different numbers and types of significant species correlation for the same data. Here, we used the SparCC tool for co-occurrence network constructions because it performed well in compositional data and data sparsity [39] and had significant edges with high correlations for analyses, which could greatly avoid noise and better than other tools. Third, some direct species interactions could be driven by indirect species interactions through other intermediary species, which could not be completely detected by the current methods and thus would decrease the robustness of results.

5. Conclusions

We found that direct and indirect species interactions in marine microbial co-occurrence networks were evolutionarily and ecologically constrained across four network levels of node, motif, module and the whole network, resulting in the small-world phenomenon of ecological networks. Compared to ecological processes, evolutionary processes dominated network assemblages in spatially homogeneous environments and showed stronger long-lasting effects on indirect species interactions. The contributions of two processes decreased toward long network path distance, but such long-lasting effects played a key role in stabilizing network link turnovers under environmental disturbances. These results provide a novel insight into the self-organization and modularization of marine microbial co-occurrence networks and highlight the importance of species phylogeny and niches in ecological network assemblages.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biology11040592/s1, Figure S1: The Pearson's correlations between species phylogeny or niche distance and latent space distance that include the potential for modularity; Figure S2: The coefficient of variation for the importance of phylogeny and niche to network connections toward large network path distance; Table S1: Default parameters for sensitive analysis; Table S2: Influences of model parameters on the explanation of phylogenetic distance for network path distance; Table S3 Influences of model parameters on the explanation of niche distance for network path distance.

Author Contributions: Conceptualization, J.W. and H.Y.; methodology, Z.L., J.W., X.L. and Q.Y.; Formal analysis, Z.L., D.M., L.L. and Y.G.; Writing—Original draft, Z.L.; Writing—Review and editing, J.W., C.J. and H.Y.; visualization, Z.L. and J.W.; project administration: H.Y.; funding acquisition, C.J. and H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Nature Science Foundation of China (NSFC; grant No. 41877345 and No. 91851206) and the Key Research Program of Frontier Sciences of the Chinese Academy of Sciences (grant No. QYZDBSSW-DQC043).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original sequences of 16S rRNA tags, OTU table and meta data are available on TARA Ocean website [90]. Codes and relevant data presented in this paper are available on a GitHub repository at https://github.com/zhenghualiu/_phylo_habitat_network (accessed on 19 March 2021).

Acknowledgments: We are grateful to Luo Feng and Wang Yunsheng for their help of methodology improvement and data collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sanders, D.; Sutter, L.; van Veen, F.J.F. The loss of indirect interactions leads to cascading extinctions of carnivores. *Ecol. Lett.* **2013**, *16*, 664–669. [CrossRef] [PubMed]
2. Donohue, I.; Petchey, O.L.; Kéfi, S.; Génin, A.; Jackson, A.L.; Yang, Q.; O'Connor, N.E. Loss of predator species, not intermediate consumers, triggers rapid and dramatic extinction cascades. *Glob. Chang. Biol.* **2017**, *23*, 2962–2972. [CrossRef] [PubMed]
3. Darwin, C. *The Origin of Species by Means of Natural Selection*; Reprinted by The Modern Library; Random House: New York, NY, USA, 1859.
4. Godoy, O.; Kraft, N.J.; Levine, J.M. Phylogenetic relatedness and the determinants of competitive outcomes. *Ecol. Lett.* **2014**, *17*, 836–844. [CrossRef] [PubMed]
5. Thompson, N.J. The Evolution of Species Interactions. *Science* **1999**, *284*, 2116–2118. [CrossRef] [PubMed]
6. Pellissier, L.; Albouy, C.; Bascompte, J.; Farwig, N.; Graham, C.; Loreau, M.; Maglianesi, M.A.; Melian, C.J.; Pitteloud, C.; Roslin, T.; et al. Comparing species interaction networks along environmental gradients. *Biol. Rev. Camb. Philos. Soc.* **2018**, *93*, 785–800. [CrossRef] [PubMed]
7. Queller, D.; Strassmann, J. Kin preference in a social microbe. *Nature* **2006**, *442*, 881882.
8. Gomez, J.M.; Verdu, M.; Perfectti, F. Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature* **2010**, *465*, 918–921. [CrossRef]
9. Best, R.J.; Caulk, N.C.; Stachowicz, J.J. Trait vs. phylogenetic diversity as predictors of competition and community composition in herbivorous marine amphipods. *Ecol. Lett.* **2013**, *16*, 72–80. [CrossRef]
10. Nuismer, S.L.; Harmon, L.J. Predicting rates of interspecific interaction from phylogenetic trees. *Ecol. Lett.* **2015**, *18*, 17–27. [CrossRef]
11. Brousseau, P.-M.; Gravel, D.; Handa, I.T.; Ferrari, M. Trait matching and phylogeny as predictors of predator-prey interactions involving ground beetles. *Funct. Ecol.* **2018**, *32*, 192–202. [CrossRef]
12. Berlow, E.L.; Dunne, J.A.; Martinez, N.D.; Stark, P.B.; Williams, R.J.; Ulrich, B. Simple prediction of interaction strengths in complex food webs. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 187–191. [CrossRef] [PubMed]
13. Cahill, J.F.; Kembel, S.W.; Lamb, E.G.; Keddy, P.A. Does phylogenetic relatedness influence the strength of competition among vascular plants? *Perspect. Plant Ecol. Evol. Syst.* **2008**, *10*, 41–50. [CrossRef]
14. Lau, M.K.; Borrett, S.R.; Baiser, B.; Gotelli, N.J.; Ellison, A.M. Ecological network metrics: Opportunities for synthesis. *Ecosphere* **2017**, *8*, e01900. [CrossRef]
15. Milo, R.; Shen-Orr, S.; Itzkovitz, S.; Kashtan, N.; Chklovskii, D.; Alon, U. Network motifs: Simple building blocks of complex networks. *Science* **2002**, *298*, 824–827. [CrossRef] [PubMed]
16. Wasserman, S. *Advances in Social Network Analysis: Research in the Social and Behavioral Sciences*; Sage: Thousand Oaks, CA, USA, 1994.
17. Olesen, J.M.; Bascompte, J.; Dupont, Y.L.; Jordano, P. The modularity of pollination networks. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19891–19896. [CrossRef]
18. Yin, H.; Niu, J.; Ren, Y.; Cong, J.; Zhang, X.; Fan, F.; Xiao, Y.; Zhang, X.; Deng, J.; Xie, M.; et al. An integrated insight into the response of sedimentary microbial communities to heavy metal contamination. *Sci. Rep.* **2015**, *5*, 14266. [CrossRef]
19. Shi, S.; Nuccio, E.E.; Shi, Z.J.; He, Z.; Zhou, J.; Firestone, M.K. The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. *Ecol. Lett.* **2016**, *19*, 926–936. [CrossRef]
20. Zhou, J.; Deng, Y.; Luo, F.; He, Z.; Yang, Y. Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO_2. *MBio* **2011**, *2*, e00122-11. [CrossRef]
21. Giling, D.P.; Ebeling, A.; Eisenhauer, N.; Meyer, S.T.; Roscher, C.; Rzanny, M.; Voigt, W.; Weisser, W.W.; Hines, J. Plant diversity alters the representation of motifs in food webs. *Nat. Commun.* **2019**, *10*, 1226. [CrossRef]
22. Bramon Mora, B.; Gravel, D.; Gilarranz, L.J.; Poisot, T.; Stouffer, D.B. Identifying a common backbone of interactions underlying food webs from different ecosystems. *Nat. Commun.* **2018**, *9*, 2603. [CrossRef]

23. Clauset, A.; Newman, M.E.; Moore, C. Finding community structure in very large networks. *Phys. Rev. E* **2004**, *70*, 066111. [CrossRef] [PubMed]
24. Pons, P.; Latapy, M. Computing communities in large networks using random walks. *J. Graph. Algorithms Appl.* **2006**, *10*, 191–218. [CrossRef]
25. Schleuning, M.; Ingmann, L.; Strauss, R.; Fritz, S.A.; Dalsgaard, B.; Matthias Dehling, D.; Plein, M.; Saavedra, F.; Sandel, B.; Svenning, J.C.; et al. Ecological, historical and evolutionary determinants of modularity in weighted seed-dispersal networks. *Ecol. Lett.* **2014**, *17*, 454–463. [CrossRef] [PubMed]
26. Banerjee, S.; Kirkby, C.A.; Schmutter, D.; Bissett, A.; Kirkegaard, J.A.; Richardson, A.E. Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil. *Soil Biol. Biochem.* **2016**, *97*, 188–198. [CrossRef]
27. Burke, C.; Steinberg, P.; Rusch, D.; Kjelleberg, S.; Thomas, T. Bacterial community assembly based on functional genes rather than species. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 14288–14293. [CrossRef]
28. Ju, F.; Xia, Y.; Guo, F.; Wang, Z.; Zhang, T. Taxonomic relatedness shapes bacterial assembly in activated sludge of globally distributed wastewater treatment plants. *Environ. Microbiol.* **2014**, *16*, 2421–2432. [CrossRef]
29. Fontaine, C.; Guimarães, P.R.; Kéfi, S.; Loeuille, N.; Memmott, J.; van der Putten, W.H.; van Veen, F.J.F.; Thébault, E. The ecological and evolutionary implications of merging different types of networks. *Ecol. Lett.* **2011**, *14*, 1170–1181. [CrossRef]
30. Proulx, S.R.; Promislow, D.E.; Phillips, P.C. Network thinking in ecology and evolution. *Trends Ecol. Evol.* **2005**, *20*, 345–353. [CrossRef]
31. Jordi, B.; Stouffer, D.B. The assembly and disassembly of ecological networks. *Philos. Trans. Biol. Sci.* **2009**, *364*, 1781–1787.
32. Pesant, S.; Not, F.; Picheral, M.; Kandels-Lewis, S.; Le Bescot, N.; Gorsky, G.; Iudicone, D.; Karsenti, E.; Speich, S.; Troublé, R. Open science resources for the discovery and analysis of Tara Oceans data. *Sci. Data* **2015**, *2*, 150023. [CrossRef]
33. Watts, D.J.; Strogatz, S.H. Collective Dynamics of Small World Networks. *Nature* **1998**, *393*, 440–442. [CrossRef] [PubMed]
34. Barabási, A.-L. Scale-free networks: A decade and beyond. *Science* **2009**, *325*, 412–413. [CrossRef]
35. Sunagawa, S.; Coelho, L.P.; Chaffron, S.; Kultima, J.R.; Labadie, K.; Salazar, G.; Djahanschiri, B.; Zeller, G.; Mende, D.R.; Alberti, A. Structure and function of the global ocean microbiome. *Science* **2015**, *348*, 1261359. [CrossRef] [PubMed]
36. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glockner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. [CrossRef] [PubMed]
37. Friedman, J.; Alm, E.J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **2012**, *8*, e1002687. [CrossRef] [PubMed]
38. Kurtz, Z.D.; Muller, C.L.; Miraldi, E.R.; Littman, D.R.; Blaser, M.J.; Bonneau, R.A. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* **2015**, *11*, e1004226. [CrossRef]
39. Weiss, S.; Van Treuren, W.; Lozupone, C.; Faust, K.; Friedman, J.; Deng, Y.; Xia, L.C.; Xu, Z.Z.; Ursell, L.; Alm, E.J.; et al. Correl. Detect. Strateg. Microb. Data Sets Vary Widely Sensit. Precision. *ISME J.* **2016**, *10*, 1669–1681. [CrossRef]
40. Simmons, B.I.; Sweering, M.J.M.; Schillinger, M.; Dicks, L.V.; Sutherland, W.J.; Di Clemente, R.; Matthiopoulos, J. bmotif: A package for motif analyses of bipartite networks. *Methods Ecol. Evol.* **2019**, *10*, 695–701. [CrossRef]
41. Salter-Townshend, M.; Murphy, T.B. Variational Bayesian inference for the Latent Position Cluster Model for network data. *Comput. Stat. Data Anal.* **2013**, *57*, 661–671. [CrossRef]
42. Csardi, G.; Nepusz, T. The igraph software package for complex network research. *InterJ. Complex Syst.* **2006**, *1695*, 1–9.
43. Stegen, J.C.; Lin, X.; Konopka, A.E.; Fredrickson, J.K. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* **2012**, *6*, 1653–1664. [CrossRef]
44. Lichstein, J.W. Multiple regression on distance matrices: A multivariate spatial analysis tool. *Plant Ecol.* **2006**, *188*, 117–131. [CrossRef]
45. Oksanen, J.; Blanchet, F.G.; Kindt, R.; Legendre, P.; Minchin, P.R.; O'hara, R.; Simpson, G.L.; Solymos, P.; Stevens, M.H.H.; Wagner, H. Package 'vegan'. *Community Ecol. Package Vers.* **2013**, *2*, 1–295.
46. Goslee, S.C.; Urban, D.L. The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* **2007**, *22*, 1–19. [CrossRef]
47. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
48. Dunne, J.A.; Williams, R.J.; Martinez, N.D. Food-web structure and network theory: The role of connectance and size. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12917–12922. [CrossRef] [PubMed]
49. Montoya, J.M.; Sol, R.V. Small world patterns in food webs. *J. Theor. Biol.* **2002**, *214*, 405–412. [CrossRef]
50. Ma, B.; Wang, H.; Dsouza, M.; Lou, J.; He, Y.; Dai, Z.; Brookes, P.C.; Xu, J.; Gilbert, J.A. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J.* **2016**, *10*, 1891–1901. [CrossRef]
51. Olesen, J.M.; Bascompte, J.; Dupont, Y.L.; Jordano, P. The smallest of all worlds: Pollination networks. *J. Theor. Biol.* **2006**, *240*, 270–276. [CrossRef]
52. Riverahutinel, A.; Bustamante, R.O.; Marín, V.H.; Medel, R. Effects of sampling completeness on the structure of plant-pollinator networks. *Ecology* **2012**, *93*, 1593–1603. [CrossRef]
53. Strogatz, S.H. Exploring complex networks. *Nature* **2001**, *410*, 268. [CrossRef] [PubMed]
54. Zhang, W. A random network based, node attraction facilitated network evolution method. *Selforganizology* **2016**, *3*, 1–9.
55. Aiello, W.; Chung, F.; Lu, L. A Random Graph Model for Power Law Graphs. *Exp. Math.* **2001**, *10*, 53–66. [CrossRef]
56. Holland, B.R. Analysis of Phylogenetics and Evolution with R. *Biometrics* **2007**, *63*, 1304–1305. [CrossRef]

57. Kembel, S.W. Disentangling niche and neutral influences on community assembly: Assessing the performance of community phylogenetic structure tests. *Ecol. Lett.* **2009**, *12*, 949–960. [CrossRef]
58. Olesen, J.M.; Stefanescu, C.; Traveset, A. Strong, Long-Term Temporal Dynamics of an Ecological Network. *PLoS ONE* **2011**, *6*, e26455. [CrossRef]
59. Mittelbach, G.G.; Schemske, D.W. Ecological and evolutionary perspectives on community assembly. *Trends Ecol. Evol.* **2015**, *30*, 241–247. [CrossRef]
60. Violle, C.; Nemergut, D.R.; Pu, Z.; Jiang, L. Phylogenetic limiting similarity and competitive exclusion. *Ecol. Lett.* **2011**, *14*, 782–787. [CrossRef]
61. Strauss, S.Y. Indirect effects in community ecology: Their definition, study and importance. *Trends Ecol. Evol.* **1991**, *6*, 206–210. [CrossRef]
62. Holland, P.D.V. Stable Coexistence in Insect Communities Due to Density- and Trait-Mediated Indirect Effects. *Ecology* **2005**, *86*, 3182–3189.
63. Cazelles, K.; Araújo, M.B.; Mouquet, N.; Gravel, D. A theory for species co-occurrence in interaction networks. *Theor. Ecol.* **2015**, *9*, 39–48. [CrossRef]
64. Golubski, A.J.; Westlund, E.E.; Vandermeer, J.; Pascual, M. Ecological Networks over the Edge: Hypergraph Trait-Mediated Indirect Interaction (TMII) Structure. *Trends Ecol. Evol.* **2016**, *31*, 344–354. [CrossRef] [PubMed]
65. Stouffer, D.B.; Sales-Pardo, M.; Sirer, M.I.; Bascompte, J. Evolutionary conservation of species' roles in food webs. *Science* **2012**, *335*, 1489–1492. [CrossRef] [PubMed]
66. Coux, C.; Rader, R.; Bartomeus, I.; Tylianakis, J.M. Linking species functional roles to their network roles. *Ecol. Lett.* **2016**, *19*, 762–770. [CrossRef]
67. Rezende, E.L.; Lavabre, J.E.; Guimarães, P.R.; Jordano, P.; Bascompte, J. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* **2007**, *448*, 925. [CrossRef]
68. Elias, M.; Fontaine, C.; van Veen, F.J. Evolutionary history and ecological processes shape a local multilevel antagonistic network. *Curr. Biol.* **2013**, *23*, 1355–1359. [CrossRef]
69. Griffin, C.A.; Thaler, J.S. Insect predators affect plant resistance via density- and trait-mediated indirect interactions. *Ecol. Lett.* **2006**, *9*, 338–346. [CrossRef]
70. Wu, L.; Yang, Y.; Chen, S.; Zhao, M.; Zhu, Z.; Yang, S.; Qu, Y.; Ma, Q.; He, Z.; Zhou, J.; et al. Long-term successional dynamics of microbial association networks in anaerobic digestion processes. *Water Res.* **2016**, *104*, 1–10. [CrossRef]
71. Kunstler, G.; Lavergne, S.; Courbaud, B.; Thuiller, W.; Vieilledent, G.; Zimmermann, N.E.; Kattge, J.; Coomes, D.A. Competitive interactions between forest trees are driven by species' trait hierarchy, not phylogenetic or functional similarity: Implications for forest community assembly. *Ecol. Lett.* **2012**, *15*, 831–840. [CrossRef]
72. Freilich, S.; Zarecki, R.; Eilam, O.; Segal, E.S.; Henry, C.S.; Kupiec, M.; Gophna, U.; Sharan, R.; Ruppin, E. Competitive and cooperative metabolic interactions in bacterial communities. *Nat. Commun.* **2011**, *2*, 589. [CrossRef]
73. Mehdiabadi, N.J.; Jack, C.N.; Tiffany Talley, F.; Platt, T.G.; Kalla, S.E.; Gad, S.; Queller, D.C.; Strassmann, J.E. Social evolution: Kin preference in a social microbe. *Nature* **2006**, *442*, 881–882. [CrossRef] [PubMed]
74. Naughton, H.R.; Alexandrou, M.A.; Oakley, T.H.; Cardinale, B.J. Phylogenetic distance does not predict competition in green algal communities. *Ecosphere* **2015**, *6*, 1–19. [CrossRef]
75. Morueta-Holme, N.; Blonder, B.; Sandel, B.; McGill, B.J.; Peet, R.K.; Ott, J.E.; Violle, C.; Enquist, B.J.; Jørgensen, P.M.; Svenning, J.-C. A network approach for inferring species associations from co-occurrence data. *Ecography* **2016**, *39*, 1139–1150. [CrossRef]
76. Thébault, E.; Fontaine, C. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* **2010**, *329*, 853–856. [CrossRef] [PubMed]
77. Van Veen, F.F.; Memmott, J.; Godfray, H.C.J. Indirect Effects, Apparent Competition and Biological Control. In *Trophic and Guild in Biological Interactions Control*; Springer: Dordrecht, The Netherlands, 2006; pp. 145–169.
78. Ghazoul, J. Floral diversity and the facilitation of pollination. *J. Ecol.* **2006**, *94*, 295–304. [CrossRef]
79. Bastolla, U.; Fortuna, M.A.; Pascual-García, A.; Ferrera, A.; Luque, B.; Bascompte, J. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* **2009**, *458*, 1018. [CrossRef]
80. Pelletier, F.; Garant, D.; Hendry, A.P. Eco-evolutionary dynamics. *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364*, 1483–1489. [CrossRef]
81. Fung, T.; Farnsworth, K.D.; Reid, D.G.; Rossberg, A.G. Impact of biodiversity loss on production in complex marine food webs mitigated by prey-release. *Nat. Commun.* **2015**, *6*, 6657. [CrossRef]
82. Dunne, J.A.; Williams, R.J.; Martinez, N.D. Network topology and biodiversity loss in food webs: Robustness increases with connectance. *Ecol. Lett.* **2002**, *5*, 558–567. [CrossRef]
83. Garay-Narváez, L.; Flores, J.D.; Arim, M.; Ramos-Jiliberto, R. Food web modularity and biodiversity promote species persistence in polluted environments. *Oikos* **2014**, *123*, 583–588. [CrossRef]
84. Gonze, D.; Lahti, L.; Raes, J.; Faust, K. Multi-stability and the origin of microbial community types. *ISME J.* **2017**, *11*, 2159–2166. [CrossRef] [PubMed]
85. Faust, K.; Sathirapongsasuti, J.F.; Izard, J.; Segata, N.; Gevers, D.; Raes, J.; Huttenhower, C. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* **2012**, *8*, e1002606. [CrossRef] [PubMed]
86. Deng, Y.; Jiang, Y.-H.; Yang, Y.; He, Z.; Luo, F.; Zhou, J. Molecular ecological network analyses. *BMC Bioinform.* **2012**, *13*, 113. [CrossRef] [PubMed]

87. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Turnbaugh, P.J.; Lander, E.S.; Mitzenmacher, M.; Sabeti, P.C. Detecting novel associations in large data sets. *Science* **2011**, *334*, 1518–1524. [CrossRef]
88. Ruan, Q.; Dutta, D.; Schwalbach, M.S.; Steele, J.A.; Fuhrman, J.A.; Sun, F. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* **2006**, *22*, 2532–2538. [CrossRef]
89. Tackmann, J.; Rodrigues, J.F.M.; von Mering, C. Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. *Cell Syst.* **2019**, *9*, 286–296. [CrossRef]
90. Tara Oceans. Available online: http://ocean-microbiome.embl.de/companion.html (accessed on 18 July 2019).