The ISME Journal (2011), 1–12

www.nature.com/ismej

ORIGINAL ARTICLE

Marine bacterial, archaeal and protistan association networks reveal ecological linkages

Joshua A Steele, Peter D Countway, Li Xia, Patrick D Vigil, J Michael Beman, Diane Y Kim, Cheryl-Emiliane T Chow, Rohan Sachdeva, Adriane C Jones, Michael S Schwalbach, Julie M Rose, Ian Hewson, Anand Patel, Fengzhu Sun, David A Caron and Jed A Fuhrman
Department of Biological Sciences and Wrigley Institute for Environmental Studies, University of Southern California, Los Angeles, CA, USA

Microbes have central roles in ocean food webs and global biogeochemical processes, yet specific ecological relationships among these taxa are largely unknown. This is in part due to the dilute, microscopic nature of the planktonic microbial community, which prevents direct observation of their interactions. Here, we use a holistic (that is, microbial system-wide) approach to investigate time-dependent variations among taxa from all three domains of life in a marine microbial community. We investigated the community composition of bacteria, archaea and protists through cultivation-independent methods, along with total bacterial and viral abundance, and physico-chemical observations. Samples and observations were collected monthly over 3 years at a well-described ocean time-series site of southern California. To find associations among these organisms, we calculated time-dependent rank correlations (that is, local similarity correlations) among relative abundances of bacteria, archaea, protists, total abundance of bacteria and viruses and physico-chemical parameters. We used a network generated from these statistical correlations to visualize and identify time-dependent associations among ecologically important taxa, for example, the SAR11 cluster, stramenopiles, alveolates, cyanobacteria and ammonia-oxidizing archaea. Negative correlations, perhaps suggesting competition or predation, were also common. The analysis revealed a progression of microbial communities through time, and also a group of unknown eukaryotes that were highly correlated with dinoflagellates, indicating possible symbioses or parasitism. Possible 'keystone' species were evident. The network has statistical features similar to previously described ecological networks, and in network parlance has non-random, small world properties (that is, highly interconnected nodes). This approach provides new insights into the natural history of microbes.
The ISME Journal advance online publication, 24 March 2011; doi:10.1038/ismej.2011.24
Subject Category: microbial population and community ecology
Keywords: co-occurrence patterns; stramenopiles; dinoflagellates; SAR11; cyanobacteria; time series

Introduction

The past two decades have seen a revolution in characterizing microbial communities that perform central roles in ocean food webs (Sherr and Sherr, 2008; Caron, 2009; Fuhrman, 2009) and biogeochemical processes (Ducklow, 2000) while making up the majority of global biomass (Pace, 1997; Diez *et al.*, 2001; Fuhrman, 2009). Translating that information into understanding of the actions and interactions of these microbes in complex systems has been difficult. Cultivation (Giovannoni and Stingl, 2007), gene chips (DeSantis *et al.*, 2007) and metagenomic or transcriptomic efforts (Rusch *et al.*, 2007; DeLong, 2009) have provided remarkable information on the potential ecological roles of microorganisms, but these fundamentally reductionist approaches do not readily describe the interactions among microbes within a community or with their environment. Unlike the situation with animals and plants, only rarely can we observe interactions among microbes, such as grazing by protists (Sherr and Sherr, 2008) or localized syntrophy (Orphan *et al.*, 2001); characterizing whole communities of microbes poses major challenges. We recently began investigating the use of network analysis based upon natural environmental co-occurrence patterns to examine the complex interactions among microbes and their environment, initially looking at bacteria and environmental parameters only (Ruan *et al.*, 2006; Fuhrman and Steele, 2008; Fuhrman, 2009). Other recent reports have also started to apply similar approaches, such as a meta-analysis of public 16S rRNA databases to demonstrate co-occurrence networks of microbes,

Correspondence: JA Steele, Division of Geological and Planetary Sciences, California Institute of Technology, Mail Code 100-23, Pasadena, CA 91125, USA.
E-mail: josh.a.steele@gmail.com
Received 2 November 2010; revised 4 February 2011; accepted 4 February 2011

including possible symbiotic associations and connections with environmental metadata (Chaffron *et al.*, 2010). A related approach has been used to show patterns from changes in functional gene expression in soil bacteria during a CO_2 addition experiment (Zhou *et al.*, 2010). These studies published to date have included only bacteria and some basic environmental parameters, but little information on the rest of the microbial community (for example, archaea and protists) that are themselves inherently of great interest and also no doubt influence the bacteria. To fully examine the factors that control microbial ecological processes, we have in this study expanded our co-occurrence network analysis to include a broad variety of bacteria, archaea and protists, as well as environmental parameters.

We investigated linkages within a microbial plankton community in response to changing environmental conditions by sampling a biological feature, the subsurface chlorophyll maximum layer, off the southern California coast monthly from August 2000 to March 2004. Co-occurrence patterns of all variables over time, with and without time lags, were calculated and the results were used to create visual networks of parameters. These show which microbes co-occur or do not co-occur, and the environmental conditions that correlate positively or negatively with these relationships. These networks reveal elements of the natural history of various microbes without isolation, enrichment or manipulation, and our application of this approach discovered numerous interesting features of this system, including possible 'keystone' species.

Materials and methods

Site and sample collection

Monthly samples were collected from the deep chlorophyll maximum depth in the University of Southern California Microbial Observatory at the San Pedro Ocean Time Series site at 33° 33′ N, 118° 24′ W offshore from Los Angeles, CA, USA. This site and the bacterial, eukaryotic and archaeal communities have been previously characterized (Fuhrman *et al.*, 2006; Countway and Caron, 2006; Beman *et al.*, 2010; Countway *et al.*, 2010).

Characterization of the microbial community

To identify and estimate changes in the relative abundance of hundreds of different types of bacteria, eukarya and archaea, we employed molecular fingerprinting techniques and quantitative PCR coupled with rRNA clone libraries. Briefly, the bacterial community composition was estimated by automated ribosomal intergenic spacer analysis (ARISA) using 16S-internal transcribed spacer-23S rRNA gene clone libraries to identify the ARISA peaks, as described previously (Fuhrman *et al.*, 2006; Fisher and Triplett, 1999; Brown *et al.*, 2005). The protistan community composition was estimated by terminal restriction fragment length polymorphism (TRFLP) using 18S rRNA gene clone libraries to identify fragments as described previously (Countway *et al.*, 2005, 2010; Caron *et al.*, 2009; Vigil *et al.*, 2009). The archaeal populations were estimated by quantitative PCR of 16S rRNA gene and amoA genes as described in Beman *et al.* (2010). For the ARISA and TRFLP analyses, fragment lengths were used as genetic signatures that distinguish specific organisms, although not all fragments can be linked to specific microbial taxa at this time. DNA quantities were standardized before amplification with the ARISA and TRFLP primers and standardized again before being analyzed on the ABI-377XL sequencer (Applied Biosystems, Foster City, CA, USA; ARISA) and Beckman CEQ8000 sequencer (Beckman Coulter, Fullerton, CA, USA; TRFLP). Internal size standards were included with each sample and the ARISA and TRFLP fluorescent peaks were conservatively binned in the manner described in Fuhrman *et al.* (2006). Changes in relative abundance were examined by ranking only within taxa, not between them, to obviate PCR quantitation artifacts (Fuhrman and Steele, 2008; Fuhrman, 2009). We measured temperature, salinity, density, sample depth, nutrients (NO_2, NO_3, SiO_3 and PO_4), dissolved oxygen, pigments (chlorophyll *a* and phaeophytin), total abundance of bacteria and viruses, and estimated the bacterial heterotrophic growth rates using ^{3}H-leucine and ^{3}H-thymidine incorporation following previous studies (for example, Fuhrman *et al.*, 2006; Countway *et al.*, 2010).

Statistical and network analyses

To examine associations between the microbial populations and their environment, we analyzed the correlations, of the microbial operational taxonomic units (OTUs) with each other and with biotic and abiotic parameters over time using local similarity analysis (Ruan *et al.*, 2006). Local similarity analysis calculates contemporaneous and time-lagged correlations based on normalized ranked data and produces correlation coefficients, referred to in the text as local similarity correlations (ls), that are analogous to a Spearman's ranked correlation (Ruan *et al.*, 2006). Any parameters (OTUs or physico-chemical parameters) which occurred in <6 months were excluded from the analysis. We included 212 variables in the analysis: 96 bacterial OTUs, 97 eukaryotic OTUs and 4 archaeal groups defined by quantitative PCR (Total Archaea, Crenarchaea, Euryarchaea and amoA-containing archaea), 9 physico-chemical measurements (temperature, salinity, density, depth, nutrients and dissolved oxygen) and 6 biotic measurements (bacterial abundance, viral abundance, heterotrophic bacterial growth rates and pigments). To

sort through, condense and visualize the correlations generated by the analysis, we used Cytoscape (Shannon *et al.*, 2003) to create networks showing the 212 variables with 1005 significant local similarity correlations ($P \leqslant 0.01$) with a false discovery rate (that is, q-value) <0.062 (Storey, 2002). False discovery rates are calculated from the observed P-value distribution and vary with the P-value (for example, for the connections with $P = 0.01$, 0.005, 0.001 and $q = 0.062$, 0.043, 0.014, respectively). These correlation networks include nodes that consisted of OTUs—as a proxy for 'species' (Brown *et al.*, 2005; Countway *et al.*, 2005; Vigil *et al.*, 2009; Beman *et al.*, 2010)—defined by molecular fingerprinting, and the biotic and physico-chemical environmental parameters; the local similarity scores (rank correlations that may include a 1-month time lag) among the taxa and the environmental parameters constituted the edges, that is, connections between nodes.

Three network topology characteristics for our network and for a random network of identical size were calculated using Network Analyzer in Cytoscape (Assenov *et al.*, 2008): Clustering coefficient (Cl), the average fraction of pairs of species one link away from a species that are also linked to one another (Watts and Strogatz, 1998; Albert and Barabási, 2002); characteristic path length (L), the average shortest path between all pairs of species (Watts and Strogatz, 1998); and degree distribution, the nodes which have k edges and the probability $P(k)$ that a node will have k edges (Watts and Strogatz, 1998; Dunne *et al.*, 2002; Oleson *et al.*, 2006). We treated edges as undirected when calculating path length and clustering, because the relationships are correlative and could imply interactions in either direction. Best fit estimates for the degree distributions were performed using MATLAB (7.10.0, Mathworks, Natick, MA, USA). Random network characteristics were calculated from networks, which had the same number of nodes and edges as the empirical networks randomly assembled using the Erdös–Réyni model networks using the default settings on the Random Networks routines in Cytoscape (Erdös and Rényi, 1960; Shannon *et al.*, 2003). As a measure of standard effect size to compare topological characteristics from ecological, biological and social networks with similarly sized random networks, we calculated the log response ratio according to the procedures in Gurevitch and Hedges (2001). The log response ratio was used for such comparisons as unknown variances in studies chosen for comparison does not invalidate the measure (Gurevitch and Hedges, 1999; Hedges *et al.*, 1999).

Results and discussion

A fundamental unknown in microbial ecology (especially among morphologically non-descript microorganisms) is whether a particular OTU definition allows us to discern distinct niches (West and Scanlan, 1999; Achtman and Wagner, 2008; Fuhrman and Steele, 2008; Koeppel *et al.*, 2008). Networks that reveal connections among particular OTUs are one way to approach that question (Fuhrman and Steele 2008; Fuhrman, 2009). Even among phylogenetically related microbes which serve as 'hubs' in the subnetworks (that is, the OTU groups around which the subnetworks were built, such as SAR11 OTUs or stramenopiles; Figure 1), there are unique combinations that tend to co-occur (Figures 1–4). Niche separation based on physico-chemical conditions has previously been shown for *Prochlorococcus* (West and Scanlan, 1999; Rocap *et al.*, 2002), SAR11 (Carlson *et al.*, 2009) and the eukaryotic alga *Ostreococcus* (Rodriguez *et al.*, 2005), but we extend the niche parameters to include co-occurring organisms, and we have been able to distinguish many distinct SAR11 OTUs (Figure 1a; Table 1).

It is appropriate to consider limitations and possible biases in the molecular fingerprinting techniques that generate the relative abundance data used to describe the bacterial and protistan communities in this study. As PCR-based techniques, molecular fingerprints are subject to all of the biases inherent to that technique including uneven and non-specific amplification efficiencies (for example, Suzuki and Giovannoni, 1996; Polz and Cavanaugh, 1998; Suzuki *et al.*, 1998) and variation in DNA or in rRNA gene copy number per cell (Fogel *et al.*, 1999; Crosby and Criddle, 2003). However, it has been suggested that within an environment and a study, these PCR bias effects will have a small influence between samples as they should apply equally to all samples (Yannarell and Triplett, 2005; Schütte *et al.*, 2008). In this study, careful standardization of the amount of template DNA used in PCRs and fingerprint analysis should further minimize these variations and allow for comparisons of relative abundance within these communities. Previous work at this microbial ocean time series demonstrated a strong correlation ($r^2 = 0.86$) between the relative fluorescence of *Prochlorococcus* OTUs identified by ARISA and independent estimates of abundance by flow cytometry (Brown *et al.*, 2005). Another issue is the detection limit of these fingerprinting techniques. We have calculated our detection limit with ARISA and TRFLP to be $\sim$0.1% of the bacteria and protistan community in this system; and extensive 16S and 18S clone libraries from these communities have boosted our confidence both in identifying OTUs and in our level of sampling of the community (Brown and Fuhrman, 2005; Brown *et al.*, 2005; Countway *et al.*, 2005, 2010; Hewson and Fuhrman, 2006; Vigil *et al.*, 2009). At this level, we are sampling the more dominant members of the community and are well aware that there are ecological roles being performed by rare or sometimes rare members as has been

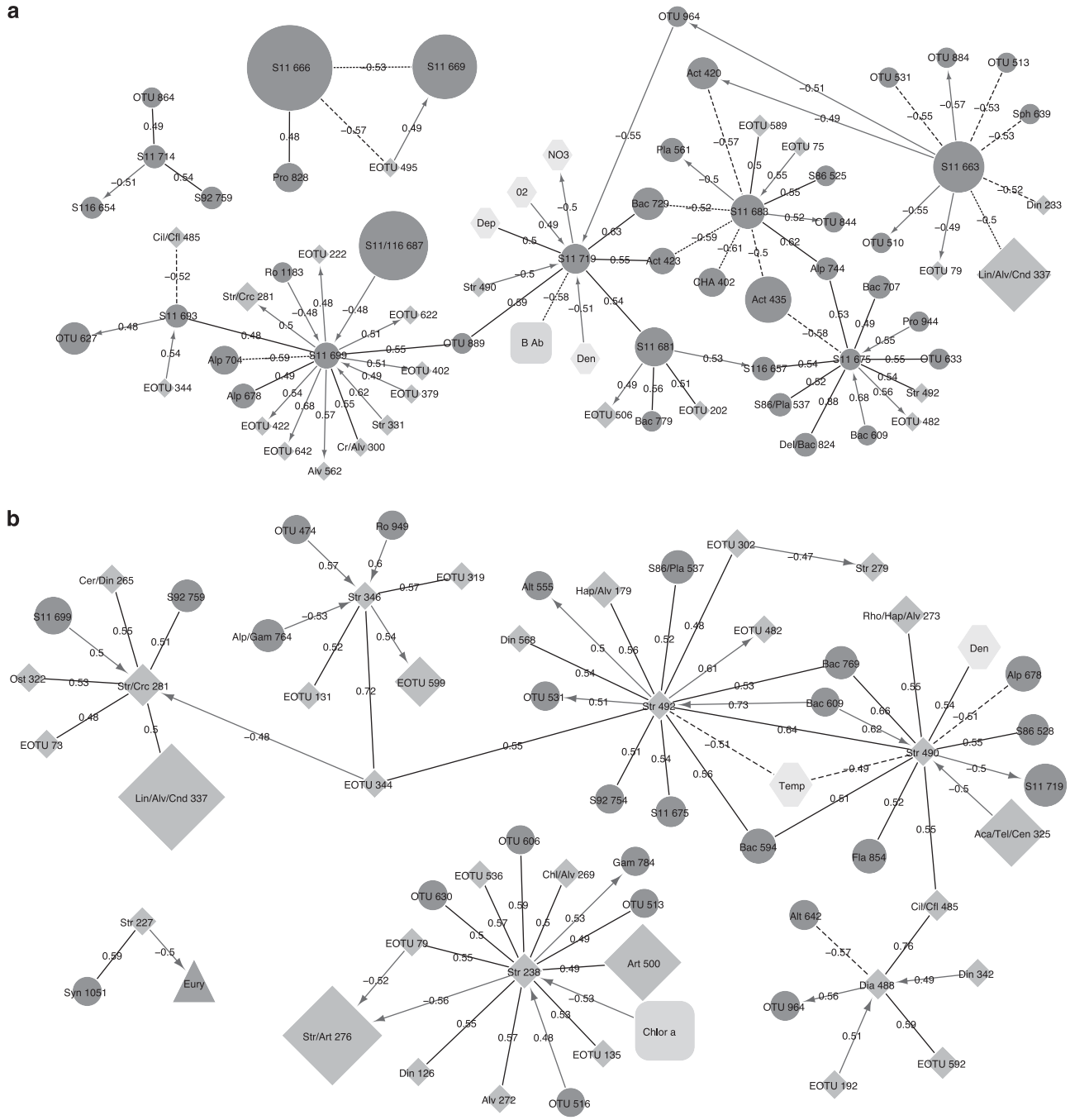

Figure 1 Subnetworks organized around 10 SAR11 OTUs (abbreviated S11, **a**) and 7 stramenopile OTUs (abbreviated Dia: Diatom, Str: stramenopile, **b**). Abbreviated names are followed by OTU fragment size for bacteria and eukaryotes. Circles are bacteria, diamonds are eukaryotes, triangles are archaea, squares are biotic environmental variables and hexagons are abiotic environmental variables. Sizes of the bacterial and eukaryotic nodes indicate the average abundance of the OTUs as measured by ARISA and TRFLP, respectively. Solid lines show a positive correlation, dashed lines show a negative correlation, arrows indicate a 1-month shift in the correlation. Abbreviations for other nodes are translated in Table 1.

shown with deep sequencing projects (for example, Huber *et al.*, 2007). The detection limit must also be kept in mind, as there are OTUs on the edge of this limit that will come in and out of detection randomly. We attempted to account for this randomness by restricting our analysis to OTUs that were present on at least ~20% of the sampling dates, as well as through permutation testing in the local similarity analysis program to determine the validity of the local similarity correlations. We feel that we have taken a conservative approach and are confident that we are able to compare changes in relative abundance over time and that we can statistically relate these changes to environmental data to describe the ecology of the microbes in this system.

Table 1 Abbreviations and identifications for organisms and physico-chemical variables included in the microbial association network

Abbreviation	*Translation*	*Abbreviation*	*Translation*
Aca	Acantharea	Lin	Linglulodinium
Act	Actinobacterium	NO_2	Dissolved nitrite
Alp	α-Proteobacterium	NO_3	Dissolved nitrate
Alt	Alteromonas	O_2	Dissolved oxygen
Alv	Alveolate	Ost	Ostreococcus
AmoA	amoA gene abundance	OTU	Unidentified bacterial taxon
Arch Tot	Archaea abundance	Phaeo	Phaeophytin concentration
Art	Arthropod	Plst	Plastid
Bac	Bacteroidetes	PO_4	Dissolved phosphate
Bact Tot	Bacteria abundance	Pro	Prochlorococcus
Bet	β-Proteobacterium	B Prod (leu)	Heterotrophic bacterial production (leucine incorporation)
Cfl	Choanoflagellate	B Prod (thy)	Heterotrophic bacterial production (thymidine incorporation)
CHB	CHAB1-7	Rho	Rhodophyte
Chl	Chlorophyte	Ros	Roseobacter
Chlor a	Chlorophyll *a* concentration	S11	SAR11
Cil	Ciliate	S116	SAR116
Cnd	Cnidarian	S406	SAR406
Cnt	Centroheliozoa	S86	SAR86
Crc	Cercozoid	S92	SAR92
Cren	Crenarchaea abundance	Sal	Water salinity
Crp	Cryptophyte	SiO3	Dissolved silicate
Del	δ-Proteobacteria	Sph	Sphingobacter
Den	Water density	Str	Stramenopile
Dep	Depth of chlorophyll maximum layer	Syn	Synechococcus
Dia	Diatom	Tel	Telonema
Din	Dinoflagellate	Temp	Water temperature
EOTU	Unidentified eukaryote taxon	Ver	Verrucomicrobium
Eury	Euryarchaea abundance	Vir Tot	Virus abundance
Fla	Flavobacterium		
Gam	γ-Proteobacterium		
Hap	Haptophyte		

Comparing relative abundance by fragment analysis has important consequences to bear in mind. We restricted our comparisons to the OTUs that were amenable to detection by these methods, which is naturally a subset of the community. Also, we binned and standardized the fluorescent peaks in order to create fair comparisons between all samples. This meant we estimated changes in relative abundance of each OTU reflecting a shift in the proportion of that particular OTU within the community, and not necessarily a change in its absolute abundance. We interpret the local similarity rank correlations in these proportions as indicative of similarities in overall environmental responses over time, for example, OTUs that go up or down together over time in relative proportions (or are both unchanging) are mathematically responding (or 'behaving') like each other; in contrast, OTUs whose proportions change unrelated to each other do not appear to 'behave' like each other.

Organisms associated with the phylogenetically related 'hubs' in the subnetworks reveal complex community interactions. The alveolate (for example, dinoflagellates) subnetwork (Figure 2a) includes a highly interconnected group of 15 unknown eukaryotic taxa that correlate with both Alveolate-562 and Dinoflagellate-198. Local similarity correlations among all 15 taxa range from 0.51 to 0.90 (Figure 2b). We speculate that the highest correlations (such as those between Alveolate-562 and EOTU-642 ($ls = 0.90$, $P < 0.001$), EOTU-622 ($ls = 0.92$, $P < 0.001$) and EOTU-522 ($ls = 0.88$, $P < 0.001$)) may imply either multiple rRNA operons from a single taxon or direct symbiotic dependence such as mutualism or parasitism (Chambouvet *et al.*, 2008). Thus, these 15 nodes may not all be unique or independent organisms, however, it is equally likely that these are distinct organisms that are closely bound ecologically.

Correlations with environmental factors provide awareness of the conditions that favor or disfavor particular collections of organisms. In general, correlations between microbes dominated the network, rather than those between microbes and abiotic or biotic environmental parameters (Supplementary Figure S1). This may relate to the relatively stable environment in the deep chlorophyll maximum layer, hence changes in community composition were driven more by biological interactions than significant changes in the physico-chemical environment. It is also possible that there were influential environmental parameters which were not measured but could have explained instances of a microbe's occurrence or shifts in the community. Correlations with biotic parameters generate hypotheses to help explain microbial niches. For example, in the SAR11 subnetwork, we observed a negative correlation ($ls = -0.576$, $P < 0.001$) between the relative abundance of SAR11 Surface group 3-719 and total bacterial counts (Figure 1a). This may

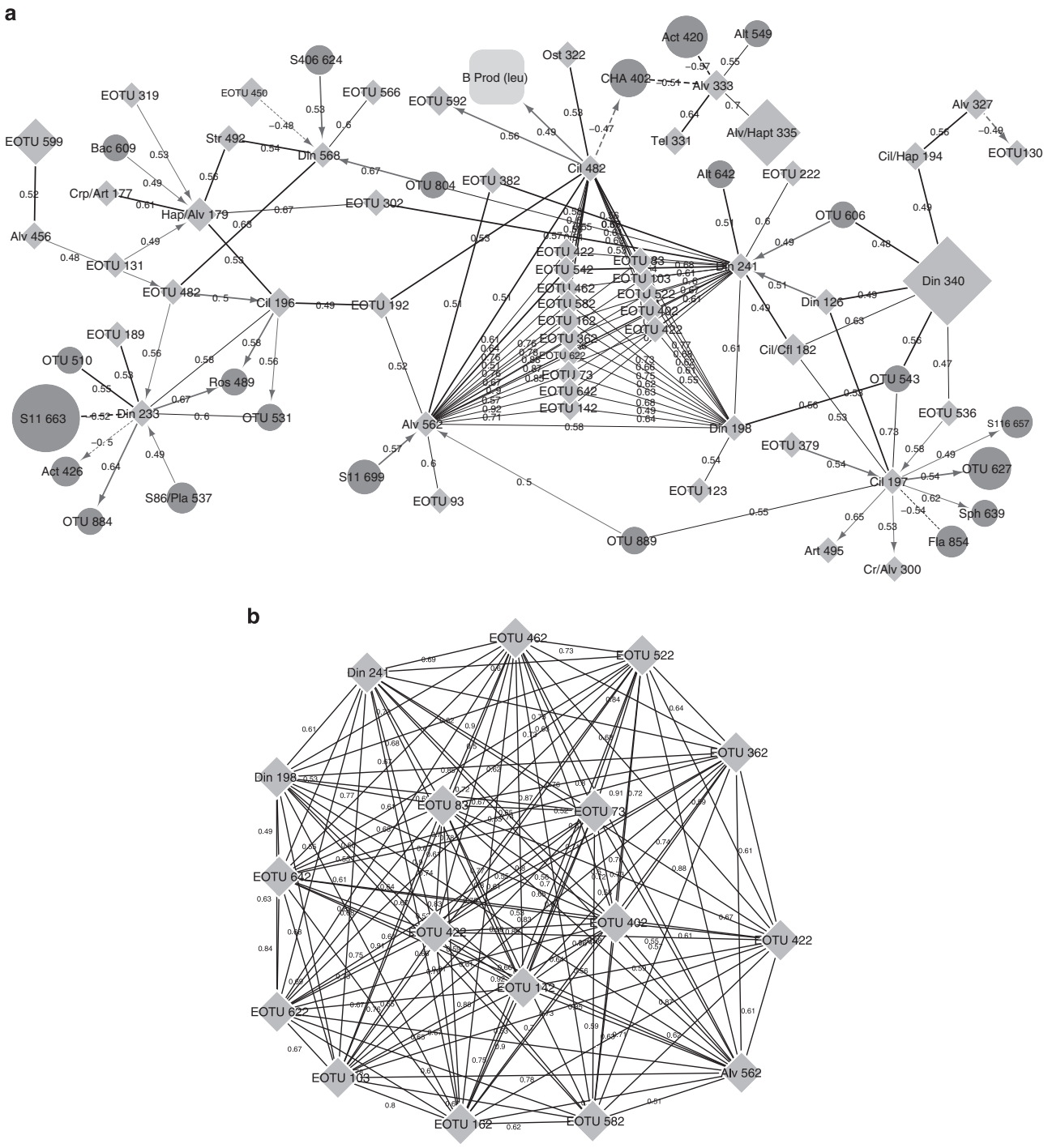

Figure 2 Alveolate subnetwork and highly connected eukaryotic cluster. The subnetwork is organized around 17 alveolates (abbreviated Alv: alveolate, Cil: ciliate, Din: dinoflagellate) as central nodes (**a**). Fifteen eukaryotic OTUs are highly correlated with Alv 562, Din 198 and Din 241. When visualized separately, these OTUs form a highly inter-connected cluster (**b**). This cluster accounts for the eukaryotes with the highest connectedness in the overall network.

reflect superior competition for growth under low abundance conditions, or may be the result of better resistance to losses by grazing or viral lysis compared with other taxa. In contrast, γ-proteobacterium SAR92-749 (Figure 3) more likely is a weedy or opportunistic species, as the relative abundance of SAR92-749 positively correlated with bacterial production measured by leucine and thymidine incorporation ($ls = 0.54$, $P = 0.003$ and $ls = 0.495$, $P = 0.005$, respectively).

Correlations with abiotic parameters revealed putative associations between the connected taxa and particular environmental measurements. The positive correlation between Stramenopile-490 and

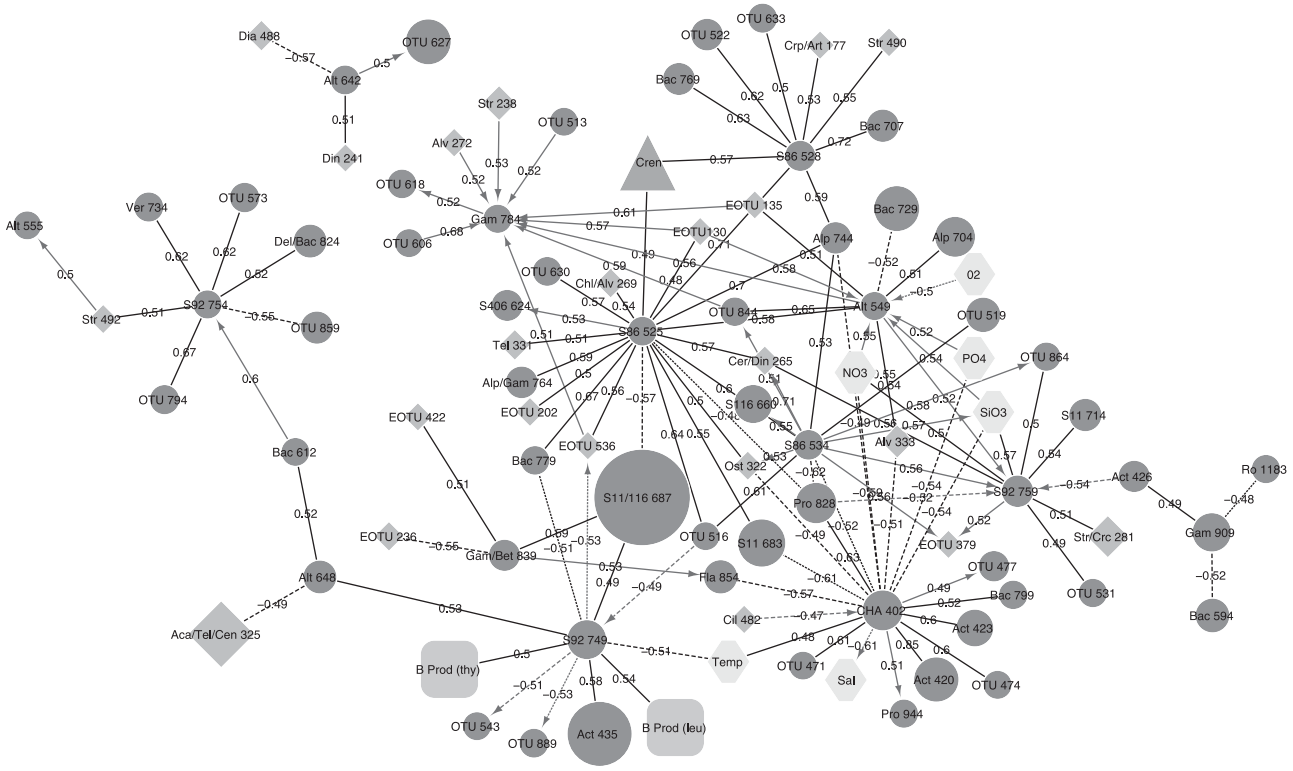

Figure 3 Subnetwork built around γ-proteobacteria OTUs as central nodes (abbreviated Alt: alteromonas, CHB: CHABI-7, Gam: γ-proteobacterium, S86: SAR86, S92: SAR92). This subnetwork identifies 12 γ-proteobacterial OTUs. γ-proteobacteria OTUs correlate with eukaryotes and Crenarchaea (Cren), as well as environmental parameters and bacterial production.

density (ls = 0.54, $P = 0.001$) associates this stramenopile with colder, saltier water (Figure 1b), perhaps indicating an association with a deeper chlorophyll maximum layer, physical forces affecting temperature or salinity (for example, upwelling), or wintertime southward flow of the Southern California current system (Fuhrman *et al.*, 2006). SAR11 Surface group 3-719 was negatively correlated with density at a 1-month time lag (ls = −0.5, $P = 0.004$), that is, a month after the water becomes warmer and less saline, the relative abundance of SAR11 Surface group 3-719 increases (Figure 1a). Although a direct connection is not identified by the analysis, this may suggest that SAR11 Surface group 3-719 and Stramenopile-492 were part of an ecological progression tied to seasonal changes in the water or to physical disruption.

Correlations between domains were common in the network. Although the subnetworks with bacteria at the center (Figures 1a, 3 and 4a) tended to have more connections to bacteria, and those with eukaryotes at the center (Figures 1b, 2a and 4b) tended to have more connections to eukaryotes, each subnetwork had multiple instances where bacteria and eukaryotes were connected. Specific bacteria and eukaryotes were also connected to archaea (for example, Euryarchaea, Figures 1b and 4a; Crenarchaea, Figure 3). Some of these co-occurrences may represent guilds of organisms performing similar or complementary functions to each other, while others may co-occur because of shared, preferred environmental conditions. When attempting to describe these collections, it is tempting to look for positive correlations among the phylogenetically related 'hubs' in these subnetworks that may indicate OTUs that are performing similar ecological roles, for example, the correlation between SAR11 Surface group 3-719 and SAR11 Surface group 1-681 (ls = 0.54, $P = 0.003$; Figure 1a) and between Stramenopile-492 and Stramenopile-490 (ls = 0.64, $P < 0.001$; Figure 1b). It is possible that some co-occurring organisms are in different niches within our samples but that those niches tend to co-occur. Note that if these organisms were truly redundant (that is, ecologically identical), other taxa should have responded to both similarly, creating many shared neighbors. While the Stramenopile-490 and Stramenopile-492 shared four neighbors, SAR11 Surface group 3-719 and SAR11 Surface group 1-681 did not share neighbors. Furthermore, there were distinct, unshared neighbors in each of these cases: six neighbors in SAR11 example (Figure 1a) and eight neighbors in Stramenopile example (Figure 1b). While functional redundancy presumably occurs among microbes, it is unclear to what extent the collective suite of functions of particular microorganisms overlap—that is, the extent of distinctiveness of individual

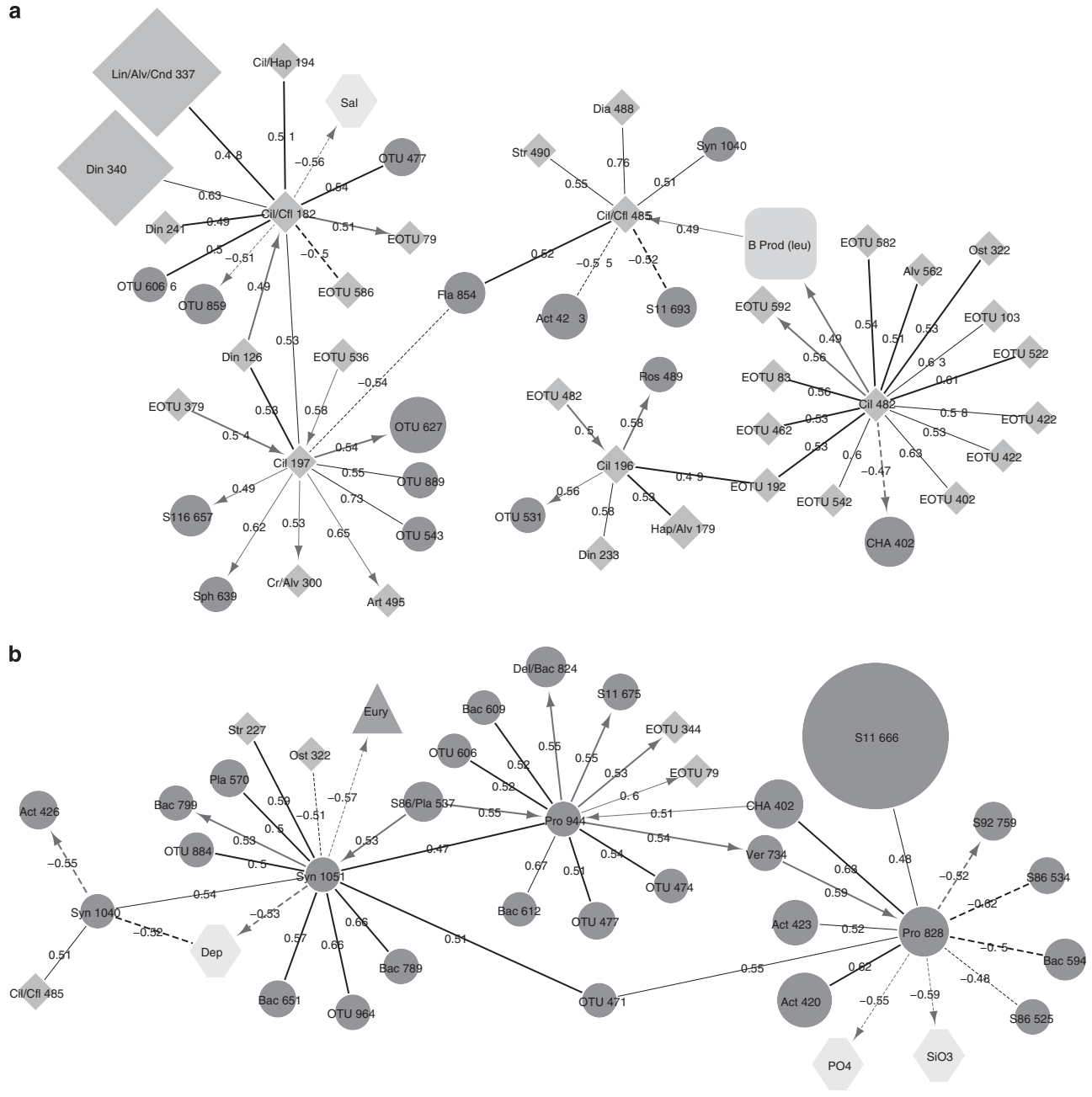

Figure 4 Subnetworks built around four cyanobacteria OTUs (abbreviated Pro: Prochlorococcus and Syn: Synechococcus, **a**) and built around three ciliate OTUs and two OTUs identified as ciliates or choanoflagellates OTU (abbreviated Cil: ciliate, Cfl: choanoflagellate, **b**). Correlations among physico-chemical variables and archaea, bacteria and eukarya show the potential for this network visualization to provide information about ecological relationships of the targeted OTUs.

microbial niches (Achtman and Wagner, 2008). In an earlier study, Fuhrman *et al*. (2006) concluded that the predictability of most bacterial OTUs at this location suggested little functional redundancy among them. It is interesting to note that while SAR11 Surface 3-719 correlates with many physico-chemical parameters and fewer OTUs, SAR11 Surface 1-681 correlates with no physico-chemical parameters. In this case, it may be that SAR11 Surface 3-719 is serving as a signal or proxy for the combination of environmental parameters which would correlate with SAR11 Surface 1-681. Similarly, the correlation between Stramenopile-490 and Stramenopile-492 may also be due to these OTUs standing in as a representative correlation for the collection of OTUs to which they are connected (that is, a proxy correlation). Previous studies have shown that combinations of environmental variables are more predictive of the microbial community changes over time compared with single environmental variables (Fuhrman *et al*., 2006; Vigil *et al*., 2009).

Negative correlations may indicate competition or predation among the taxa. The negative correlation between Ciliate/Choanoflagellate-485 with actinobacter-423 ($ls = -0.55$, $P < 0.001$) and SAR11 693 ($ls = -0.52$, $P = 0.003$) or between Ciliate-197 with flavobacter-854 ($ls = -0.54$, $P = 0.003$; Figure 4b) with no time delay could point to a predator–prey relationship, especially considering that these protists are phagotrophic. A negative correlation ($ls = -0.51$, $P = 0.003$) with no time delay between *Synechococcus*-1051 and *Ostreococcus*-322 (a phototrophic prokaryote and a phototrophic eukaryote of similar size; Figure 4a) may reflect competition. A positive correlation with a time delay may reflect a change in the bacterial community as the environment changes. *Prochlorococcus* grpII-944 (a low light-adapted group) and CHABI-7-402 exhibited a positive correlation with a 1-month time lag ($ls = 0.51$, $P = 0.002$), and *Prochlorococcus* grpI-828 (a high light-adapted group) correlated positively with CHABI-7-402 ($ls = 0.63$, $P < 0.001$; Figure 4a) with no time delay, suggesting a progression where CHABI-7-402 and *Prochlorococcus* grpI-828 give way to the low light-adapted *Prochlorococcus* grpII-944. We suggest that positive correlations without delay indicate organisms that thrive under similar conditions, while delayed positive correlations reflect a shift of dominance during a monthly progression. These possibilities raise many hypotheses for further examination using approaches targeting particular taxa.

Observing the network as a whole, it is interesting to note that many of the physico-chemical and biotic parameters were not as highly connected as the OTUs, shown clearly when they are sorted by their number of connections (Supplementary Figure S1). In fact, relatively invariant biotic parameters, such as total virus counts, and pigments did not have as many significant correlations as seasonally variable parameters such as depth of the chlorophyll maximum layer, NO_3 and salinity. Interestingly, heterotrophic bacterial production by thymidine incorporation had fewer connections than production measured by leucine incorporation (Supplementary Figure S1). This may be due to leucine, but not thymidine, incorporation from nanomolar concentrations by cyanobacteria and pico-eukaryotes in addition to heterotrophic bacteria in surface ocean waters (Zubkov *et al.*, 2003; Michelou *et al.*, 2007), with nanomolar thymidine believed to be taken up effectively only by heterotrophic bacteria (Fuhrman and Azam, 1982). Some connections to production are intriguing and call for further study, for example, the positive correlation lagged 1 month between two different ciliates (482 and 485) and bacterial production measured by leucine incorporation (Figure 4a). We also note that our pairwise analysis could miss interactions where individual parameters do not relate consistently to changes in microbial community composition even though particular combinations of parameters might.

Properties of the whole microbial association network can reveal patterns that can be used to place the microbial network in context with other ecological, biological and non-biological networks. We calculated the clustering coefficient, the characteristic path length and the node degree distribution (Watts and Strogatz, 1998; Albert and Barabási, 2002; Dunne *et al.*, 2002; Oleson *et al.*, 2006) of the entire network (microbial and environmental nodes) and just the microbial nodes and compared the structure of the association network with a random network of the same size as well as networks from previous studies (Table 2). As restricting these analyses to just the microbial parameters did not fundamentally change the network metrics we will discuss the more inclusive values. For the whole network, the observed characteristic path length of 2.99 and clustering coefficient of 0.27 were both greater than the random characteristic path length of 2.66 and random clustering coefficient of 0.044 (Table 2). The observed:random network clustering coefficient ratio of 6.1 (log response ratio of 1.81) shows that the association network has 'small world' properties (that is, nodes are more connected than a random network of similar size; Watts and Strogatz, 1998), although the effect is much smaller than the social networks (log response ratio 5.19–5.7 for the internet and 7.98 for actors; Watts and Strogatz, 1998; Albert and Barabási, 2002; Dunne *et al.*, 2002; Oleson *et al.*, 2006). The observed clustering coefficient of 0.27 is not as high as clustering coefficients for pollination (Oleson *et al.*, 2006), social networks (Watts and Strogatz, 1998) or the global microbial 16S/genome database network (Chaffron *et al.*, 2010). However, it is within the range of biological networks (for example, *Escherichia coli* substrate and *Caenorhabditis elegans* neural networks; Watts and Strogatz, 1998; Albert and Barabási, 2002), ecological networks (for example, food webs, Albert *et al.*, 2000; Montoya *et al.*, 2006) and the functional gene network from soil bacteria (Table 2; Zhou *et al.*, 2010). It is interesting to note that the larger metadata focused microbial study which spanned multiple environments (Chaffron *et al.*, 2010) found higher clustering compared with the Zhou *et al.* (2010) study of the bacterial community in soil plots and this study at an ocean time series. While our slightly lower clustering coefficient indicates our three-domain study included less highly correlated microbes compared with the clusters in the global 16S bacterial network, the overall comparisons with other biological/ecological networks suggest that the higher than random clustering we report here is not unreasonable and may be more characteristic of bacterial/archaeal/protistan microbial communities in general. In any case, it suggests that further investigation into the microbial community in the ocean and elsewhere will reveal more potential interactions.

Node connectivity distribution can be used to compare hierarchical structure between systems.

Table 2 Clustering coefficient and average shortest path length from the microbial association network and comparisons to other real and random networks

	Cl[a]	Cl_{random}	Cl/Cl_{random}	L	L_{random}	lr_{Cl}	lr_L
Ecological networks							
Microbial association (this study)							
All nodes ($n=212$)	0.27 (±0.02)	0.044 (±0.003)	6.14	2.99 (±0.03)	2.62 (±0.01)	1.81	0.13
Microbial nodes ($n=197$)	0.26 (±0.02)	0.047 (±0.003)	5.53	3.05 (±0.03)	2.64 (±0.012)	1.71	0.14
Food webs[b,c,d]	0.02 to 0.43	0.03 to 0.33	0.30 to 3.80	1.33 to 3.74	1.41 to 3.73	−1.20 to 1.34	0.34 to 1.32
Pollinator-plant networks[c,d]	0.72 to 1.00	0.08 to 1.0	1.0 to 10.9	1.0 to 2.31	ND	0.0 to 2.39	ND
Microbial database network[e]	0.501	ND	ND	6.30	ND	ND	ND
Functional microbial networks[f]	0.10 to 0.22	0.028 to 0.099	2.22 to 3.57	3.09 to 4.21	3.00 to 3.84	0.79 to 1.28	0.030 to 0.091
Biological networks							
Caenorhabditis elegans, neural network[g,h]	0.28	0.05	5.6	2.65	2.25	1.71	0.16
Escherichia coli, metabolic network[h,i]	0.32 to 0.59	0.026 to 0.09	6.56 to 12.3	2.62 to 2.90	1.98 to 3.04	1.88 to 2.51	−0.05 to 0.27
Social networks							
Power grid[g,h]	0.08	0.005	16	18.7	12.4	2.77	0.41
Actors[g,h]	0.79	0.00027	2925	3.65	2.99	7.98	0.20
Internet, domain level[h,j,k,l]	0.18 to 0.3	0.001	180 to 300	3.70 to 3.76	6.36 to 6.18	5.19 to 5.70	−0.50 to −0.54

[a]Cl is the average clustering coefficient, Cl_{random} is the clustering coefficient from an identically sized random network, Cl/Cl_{random} is the ratio of the clustering coefficient from the real network to the clustering coefficient of the random network, L is the average shortest path length, L_{random} is the average shortest path length from the random network, lr_{CL} and lr_L are the log response ratio for the average clustering coefficient and average shortest path length between the observed and random networks; values in parentheses are standard error; ND indicates missing data.
[b]Dunne *et al.* (2002).
[c]Oleson *et al.* (2006).
[d]Montoya *et al.* (2006).
[e]Chaffron *et al.* (2010).
[f]Zhou *et al.* (2010).
[g]Watts and Strogatz (1998).
[h]Albert and Barabási (2002).
[i]Wagner and Fell (2000).
[j]Albert *et al.* (2000).
[k]Yook *et al.* (2002).
[l]Pastor-Satorras *et al.* (2001).

For the entire microbial network it did not resemble the Poisson shape of the random network (best fit by a Gaussian curve, $r^2=0.91$; Supplementary Figure S2), but was best described by a truncated power law function ($P(k)=k^{-a}e^{-bk}+c$; Supplementary Figure S2); fitting this function yielded $a=0.014$ and $b=0.004$ ($r^2=0.8$), which is characteristic of some ecological networks (for example, plants in plant–frugivore interactions and some food webs) and like other ecological networks these coefficient values are much smaller than the power law exponents of 2–3 found in large social or protein interaction networks (Watts and Strogatz, 1998; Albert and Barabási, 2002; Dunne *et al.*, 2002; Montoya *et al.*, 2006). These structural similarities to many other ecological, microbial and self-organizing networks (Watts and Strogatz, 1998; Albert and Barabási, 2002; Dunne *et al.*, 2002; Montoya *et al.*, 2006; Oleson *et al.*, 2006), allow us to infer that we are observing meaningful, non-random relationships over time. The small world pattern of few highly connected nodes, as opposed to an even distribution of connectivity, makes the network more robust to change (Albert *et al.*, 2000; Montoya *et al.*, 2006) with an important caveat: if highly connected nodes are lost, the network would change dramatically. These highly connected nodes might be analogous to microbial 'keystone species.'

Although we cannot claim that we have a comprehensive view of interactions within marine microbial communities, our correlation network provides information on the natural history of bacteria, eukaryotes and archaea, and their interactions with the environment. Future studies comparing networks for different communities will allow for exploration of the resilience and resistance of those communities (Albert *et al.*, 2000; Montoya *et al.*, 2006) and yield insight into the effects of future environmental changes. We feel that these analyses are useful not only for this type of community analysis but are able to extend to other datasets (for example, new studies based on deep DNA sequencing across all three domains of life) for further exploration and generation of testable hypotheses and, someday, predictive models of microbial diversity that will deepen our understanding of microbial interactions in the ocean.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank the captain and crew of the R/V Seawatch, M Neumann and R Smith for oxygen, salinity and chlorophyll data; Z Zheng for nutrient data. We thank D Capone, W Ziebis, J Devinny D Keifer and E Berlow for helpful comments on earlier versions of this manuscript, and anonymous reviewers for their careful reading and excellent comments that improved this paper. The US National Science Foundation Microbial Observatories and Biological Oceanography programs funded this research, grants 0703159, 0623575 and 0648581.

References

Achtman M, Wagner M. (2008). Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol* **6**: 431–440.

Albert R, Barabási AL. (2002). Statistical mechanics of complex networks. *Rev Mod Phys* **74**: 47–97.

Albert R, Jeong H, Barabasi AL. (2000). Error and attack tolerance of complex networks. *Nature* **406**: 378–382.

Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M. (2008). Computing topological parameters of biological networks. *Bioinformatics* **24**: 282–284.

Beman JM, Sachdeva R, Fuhrman JA. (2010). Population ecology of nitrifying archaea and bacteria in the Southern California Bight. *Environ Microbiol* **12**: 1282–1292.

Brown MV, Fuhrman JA. (2005). Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat Microb Ecol* **41**: 15–23.

Brown MV, Schwalbach MS, Hewson I, Fuhrman JA. (2005). Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Environ Microbiol* **7**: 1466–1479.

Carlson CA, Morris R, Parsons R, Treusch AH, Giovannoni SJ *et al.* (2009). Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J* **3**: 283–295.

Caron DA. (2009). New accomplishments and approaches for assessing protistan diversity and ecology in natural ecosystems. *Bioscience* **59**: 287–299.

Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD *et al.* (2009). Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Appl Environ Microbiol* **75**: 5797–5808.

Chaffron S, Rehrauer H, Pernthaler J, Von Mering C. (2010). A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res* **20**: 947–959.

Chambouvet A, Morin P, Marie D, Guillou L. (2008). Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science* **322**: 1254–1257.

Countway PD, Caron DA. (2006). Abundance and distribution of *Ostreococcus* sp. in the San Pedro Channel, California, as revealed by quantitative PCR. *Appl Environ Microbiol* **72**: 2496–2506.

Countway PD, Gast RJ, Savai P, Caron DA. (2005). Protistan diversity estimates based on 18S rDNA from seawater incubations in the western North Atlantic. *J Eukaryot Microbiol* **52**: 95–106.

Countway PD, Vigil PD, Schnetzer A, Moorthi S, Caron DA. (2010). Seasonal analysis of protistan community structure and diversity at the USC Microbial Observatory (San Pedro Channel, North Pacific Ocean). *Limnol Oceanogr* **55**: 2381–2396.

Crosby LD, Criddle CS. (2003). Understanding bias in microbial community analysis techniques due to *rrn* operon copy number heterogeneity. *Biotechniques* **34**: 790–802.

DeLong EF. (2009). The microbial ocean from genomes to biomes. *Nature* **459**: 200–206.

DeSantis TZ, Brodie EL, Moberg JP, Zubieta IX, Piceno YM *et al.* (2007). High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol* **53**: 371.

Diez B, Pedros-Alió C, Massana R. (2001). Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl Environ Microbiol* **67**: 2932–2941.

Ducklow HW. (2000). *Bacterial production and biomass in the oceans in microbial ecology of the oceans*. Kirchman DL (ed.). Wiley-Liss: New York. pp, 85–120.

Dunne JA, Williams RJ, Martinez ND. (2002). Food-web structure and network theory: the role of connectance and size. *Proc Natl Acad Sci USA* **99**: 12917–12922.

Erdös P, Rényi A. (1960). On the evolution of random graphs. *Publ Math Inst Hung Acad Sci* **5**: 17–61.

Fisher MM, Triplett EW. (1999). Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl Environ Microbiol* **65**: 4630–4636.

Fogel GB, Collins CR, Li J, Brunk CF. (1999). Prokaryotic genome size and SSU rDNA copy number: estimation of microbial relative abundance from a mixed population. *Microb Ecol* **38**: 93–113.

Fuhrman JA. (2009). Microbial community structure and its functional implications. *Nature* **459**: 193–199.

Fuhrman JA, Azam F. (1982). Thymidine incorporation as a measure of heterotrophic bacterioplankton production in marine surface waters—evaluation and field results. *Mar Biol* **66**: 109–120.

Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV *et al.* (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* **103**: 13104–13109.

Fuhrman JA, Steele JA. (2008). Community structure of marine bacterioplankton: patterns, networks, and relationships to function. *Aquat Microb Ecol* **53**: 69–81.

Giovannoni S, Stingl U. (2007). The importance of culturing bacterioplankton in the 'omics' age. *Nat Rev Microbiol* **5**: 820–826.

Gurevitch J, Hedges LV. (1999). Statistical issues in ecological meta-analysis. *Ecology* **80**: 1142–1149.

Gurevitch J, Hedges LV. (2001). *Meta-analysis: combining the results of independent experiments in design and analysis of ecological experiments*. Scheiner SM and Gurevitch J (eds). Oxford University Press: New York.

Hedges LV, Gurevitch J, Curtis P. (1999). The meta-analysis of response ratios in experimental ecology. *Ecology* **80**: 1150–1156.

Hewson I, Fuhrman JA. (2006). Improved strategy for comparing microbial assemblage fingerprints. *Microb Ecol* **51**: 147–153.

Huber JA, Mark Welch D, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML. (2007). Microbial population structures in the deep marine biosphere. *Science* **318**: 97–100.

Koeppel A, Perry EB, Sikorski J, Krizanc D, Warner A *et al.* (2008). Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc Natl Acad Sci USA* **105**: 2504–2509.

Michelou VK, Cottrell MT, Kirchman DL. (2007). Light-stimulated bacterial production and amino acidassimilation by cyanobacteria and other microbes in the North Atlantic Ocean. *Appl Environ Microbiol* **73**: 5539–5546.

Montoya JM, Pimm SL, Sole RV. (2006). Ecological networks and their fragility. *Nature* **442**: 259–264.

Oleson JM, Bascompte J, Dupont YL, Jordano P. (2006). The smallest of all worlds: pollination networks. *J Theor Biol* **240**: 270–276.

Orphan VJ, House CH, Hinrichs KU, McKeegan KD, DeLong EF. (2001). Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* **293**: 484–487.

Pace NR. (1997). A molecular view of microbial diversity and the biosphere. *Science* **276**: 734–740.

Pastor-Satorras R, Vazquez A, Vespignani A. (2001). Dynamical and correlation properties of the Internet, Phys. *Rev Lett* **87**: 258701.

Polz MF, Cavanaugh CM. (1998). Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64**: 3724–3730.

Rocap G, Distel DL, Waterbury JB, Chisolm SW. (2002). Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* **68**: 1180–1191.

Rodriguez F, Derelle E, Guillou L, Le Gall F, Vaulot D *et al.* (2005). Ecotype diversity in the marine picoeukaryote *Ostreococcus* (*Chlorophyta, Prasinophyceae*). *Environ Microbiol* **7**: 853–859.

Ruan QS, Dutta D, Schwalbach MS, Steele JA, Fuhrman JA *et al.* (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* **22**: 2532–2538.

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S *et al.* (2007). The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5**: 398–431.

Schütte UM, Abdo EZ, Bent SJ, Shyu C, Williams CJ, Pierson JD, Forney LJ. (2008). Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Appl Microbiol Biotechnol* **80**: 365–380.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT *et al.* (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498–2503.

Sherr EB, Sherr BF. (2008). *Understanding roles of microbes in marine pelagic food webs: a brief history in microbial ecology of the oceans*, Kirchman DL (ed.). Wiley-Liss: New York. pp, 27–44.

Storey JD. (2002). A direct approach to false discovery rates. *J Roy Stat Soc B* **64**: 479–498.

Suzuki MT, Giovannoni SJ. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62**: 625–630.

Suzuki MT, Rappé M, Giovannoni SJ. (1998). Kinetic bias in estimates of picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. *Appl Environ Microbiol* **64**: 4522–4529.

Vigil P, Countway PD, Rose J, Lonsdale DJ, Gobler CJ *et al.* (2009). Rapid shifts in dominant taxa among microbial eukaryotes in estuarine ecosystems. *Aquat Microb Ecol* **54**: 83–100.

Wagner A, Fell D. (2000) Technical Report No. 00-07-041, Santa Fe Institute.

Watts DJ, Strogatz SH. (1998). Collective dynamics of 'small-world' networks. *Nature* **393**: 440–442.

West NJ, Scanlan DJ. (1999). Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean. *Appl Environ Microbiol* **65**: 2585–2591.

Yannarell AC, Triplett EW. (2005). Geographic and environmental sources of variation in lake bacterial community composition. *Appl Environ Microbiol* **71**: 227–239.

Yook SH, Jeong H, Barabási AL. (2002). Modeling the Internet's large-scale topology. *Proc Natl Acad Sci USA* **99**: 13382–13386.

Zhou J, Deng Y, Luo F, He Z, Tu Q *et al.* (2010). Functional molecular ecological networks. *MBio* **1**: e00169–e00110.

Zubkov MV, Fuchs BM, Tarran GA, Burkill PH, Amann R. (2003). High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Appl Environ Microbiol* **69**: 1299–1304.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)