bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Disentangling temporal associations in marine microbial networks

Ina Maria Deutschmann[1*], Anders K. Krabberød[2], L. Felipe Benites[3], Francisco Latorre[1], Erwan Delage[4,5], Celia Marrasè[1], Vanessa Balagué[1], Josep M. Gasol[1,6], Ramon Massana[1], Damien Eveillard[4,5], Samuel Chaffron[4,5] and Ramiro Logares[1*]

[1]Institute of Marine Sciences, CSIC, Passeig Marítim de la Barceloneta, 37, 08003, Barcelona, Spain.
[2]Department of Biosciences/Section for Genetics and Evolutionary Biology (EVOGENE), University of Oslo, p.b. 1066 Blindern, N-0316, Oslo, Norway.
[3]Integrative Biology of Marine Organisms (BIOM), CNRS, Oceanological Observatory of Banyuls, Sorbonne University, Avenue Pierre Fabre, 66650, Banyuls-sur-Mer, France.
[4]Université de Nantes, CNRS UMR 6004, LS2N, F-44000, 2 rue de la Houssinière, 44322, Nantes, France.
[5]Research Federation (FR2022) Tara Océan GO-SEE, 3 rue Michel-Ange, 75016, Paris, France.
[6]Center for Marine Ecosystems Research, School of Sciences, Edith Cowan University, Joondalup, WA Australia.

[*]Corresponding authors
Ina Maria Deutschmann (ina.m.deutschmann@gmail.com)
Ramiro Logares (ramiro.logares@icm.csic.es)

Running title: Temporal marine microbial network

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Microbial interactions are fundamental for Earth's ecosystem functioning and biogeochemical cycling. Nevertheless, they are challenging to identify and remain barely known. The omics-based censuses are helpful to predict microbial interactions through the inference of static association networks. However, since microbial interactions are highly dynamic, we have developed a post-network-construction approach to generate a temporal network from a single static network. We applied it to understand the monthly microbial associations' dynamics occurring over ten years in the Blanes Bay Microbial Observatory (Mediterranean Sea). For the decade, we identified persistent, seasonal, and temporary microbial associations. Moreover, we found that the temporal network appears to follow an annual cycle, collapsing and reassembling when transiting between colder and warmer waters. We observed higher repeatability in colder than warmer months. Altogether, our results indicate that marine microbial networks follow recurrent temporal dynamics, which need to be accounted to better understand the dynamics of the ocean microbiome.

Keywords: association network; temporal network; time series; microbial interactions; microorganisms; ocean; plankton

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

Microorganisms are the most abundant life forms on Earth and are fundamental for global ecosystem functioning (Falkowski *et al.*, 2008; DeLong, 2009; Krabberød *et al.*, 2017). The number of microorganisms on Earth is estimated to be $\approx 10^{12}$ species (Locey & Lennon, 2016), comprising $\approx 10^{30}$ cells (Whitman *et al.*, 1998; Kallmeyer *et al.*, 2012). The oceans harbor $\approx 10^{29}$ microbial cells (Whitman *et al.*, 1998) accounting for ~70% of the total marine biomass (Bar-On *et al.*, 2018; Bar-On & Milo, 2019). These cell numbers are known to be dynamic.

Microbial ecosystems are dynamic and their community composition is determined through a combination of ecological processes: selection, dispersal, drift, and speciation (Vellend, 2020). Selection is a prominent community structuring force that is exerted via multiple abiotic and biotic environmental factors (Lindström & Langenheder, 2012; Mori *et al.*, 2018). Several studies have addressed the role of *abiotic* factors in structuring microbial communities. For example, temperature, one of the primary environmental variables, exerts selection in the ocean microbiome over spatiotemporal scales (Bunse & Pinhassi, 2017; Giner *et al.*, 2019; Lambert *et al.*, 2019; Logares *et al.*, 2020). *Biotic* factors can also exert a strong selection on microbial communities (Barraclough, 2015). However, a mechanistic understanding of how they affect community structure is currently lacking, as the diversity of microbial interactions is barely known (Krabberød *et al.*, 2017; Bjorbækmo *et al.*, 2019).

The vast microbial diversity and the fact that most microorganisms are still uncultured (Baldauf, 2008; Lewis *et al.*, 2020) make it impossible to experimentally test all potential interactions. However, omics-technologies allow estimating microbial sequence abundances over spatiotemporal scales, which permit determining (statistical) associations between microorganisms. These associations can be summarized as a network with nodes representing microorganisms and edges representing potential interactions (Weiss *et al.*, 2016; Layeghifard *et al.*, 2017).

As microorganisms are highly interconnected (Layeghifard *et al.*, 2017), association networks provide a general overview of the entire microbial system and have been tremendously valuable for generating interaction hypotheses. In particular, several time-series have allowed the investigation of possible ecological interactions among marine microorganisms (Steele *et al.*, 2011; Chow *et al.*, 2013, 2014; Cram *et al.*, 2015; Needham *et al.*, 2017; Parada & Fuhrman, 2017; Krabberød *et al.*, 2021). For example, previous work characterized ecological links between marine archaea, bacteria, and eukaryotes (Steele *et al.*, 2011), including links with viruses (Chow *et al.*, 2014; Needham *et al.*, 2017), also investigating within- and between ocean-depth relationships (Cram *et al.*, 2015; Parada & Fuhrman,

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2017). Not only time-dependent associations among ecologically important taxa were identified, but also potential synergistic or antagonistic relationships, as well as possible 'keystone' species and potential niches (Steele *et al.*, 2011; Chow *et al.*, 2013). Moreover, studies found more associations between microorganisms than between the microorganisms and environmental factors, which would suggest the dominance of microbial relationships over associations between microorganisms and environmental factors (Steele *et al.*, 2011; Krabberød *et al.*, 2021).

Previous studies used temporal microbial-abundance data to build static networks. This static abstraction is based on several assumptions (Blonder *et al.*, 2012), principally that the network topology does not change (static) and edges represent persistent associations assumed as interactions, that is, edges are present throughout time-space. This assumption cannot represent the reality for most microbial interactions. Thus, a single static network usually captures persistent, temporary, and recurring (including seasonal) associations, which need to be disentangled.

Despite the contribution of static networks to our understanding of microbial interactions in the ocean, it is necessary to incorporate the temporal dimension. Using a temporal network instead of a single static network would allow investigating the dynamic nature of microbial associations, contributing to comprehend how they change over time, whether their change is deterministic or stochastic, and how environmental selection influences network architecture. Addressing these questions is fundamental for a better understanding of the dynamic interactions that underpin microbial ecosystem function. Here, we investigate marine microbial associations through time using an approach developed to determine a temporal network from a single static network.

RESULTS

Extracting a temporal network from a single static association network

Leveraging ten years of monthly samples from the Blanes Bay Microbial Observatory (BBMO) in the Mediterranean Sea (Gasol *et al.*, 2016), we computed sequence abundances for 488 bacteria and 1005 microbial eukaryotes from two organismal size-fractions: picoplankton (0.2 – 3 µm) and nanoplankton (3 – 20 µm). We removed Archaea since they are not very abundant in the BBMO surface and, additionally, primers were not optimal to quantify them. We inferred Amplicon Sequence Variants (ASVs) using the 16S and 18S rRNA-gene. After filtering the initial ASV table for sequence abundance and shared taxa among size fractions, we kept 285 and 417 bacterial, 526 and 481 eukaryotic ASVs in the pico- and nanoplankton size-fractions, respectively. We found 214 bacterial ASVs that appeared in both size

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fractions, but only two eukaryotic ASVs: a *Cryothecomonas* (Cercozoa) and a dinoflagellate (Alveolate).

We used a total of 1709 ASVs to infer a preliminary association network with the tool eLSA (Xia *et al.*, 2011, 2013). Next, we removed environmentally-driven edges with EnDED (Deutschmann *et al.*, 2020) and only considered edges which association partners co-occurred more than half of the times together than alone (see methods and Figure 1A-B). Our filtering strategy removed a higher fraction of negative than positive edges (see methods and Supplementary Table 1). The resulting network is our single static network connecting 709 nodes via 16626 edges (16481 edges, 99.1%, positive and 145, 0.9% negative).

Next, we developed a post-network-construction approach to determine a temporal network from a single static network. Building upon the single static network, we determined 120 sample-specific (monthly) subnetworks (see methods for details). These monthly subnetworks represent the 120 months of the time series and together comprise the temporal network. Each monthly subnetwork contains a subset of the nodes and a subset of the edges of the single static network. To determine which nodes and edges are present each month, we used the ASV abundances indicating the presence (ASV abundance > 0) or absence (ASV abundance = 0) as well as the estimated start and duration of associations inferred with the network construction tool eLSA (Xia *et al.*, 2011, 2013) (Figure 1, see Methods).

The single static network metrics differed from most monthly subnetworks

Since each monthly subnetwork was derived from the single static network, they were smaller, containing between 141 (August 2005) and 571 (January 2012) nodes, median ≈354 (Figure 2A), and between 560 (April 2006) to 15704 (January 2012) edges, median ≈6052 (Figure 2B). For further characterization, we computed six global network metrics (Figure 2C and Methods). The results indicated that the single static network differed from most monthly subnetworks and it also differed from the average. In general, the single static network was less connected (edge density) and more clustered (transitivity) with higher distances between nodes (average path length) and stronger associations (average positive association score) than most monthly subnetworks (Figure 2C). In addition, the single static network was usually more assortative according to the node degree but less assortative according to the domain (bacteria vs. eukaryote) than most monthly subnetworks (Figure 2C). High assortativity indicates that nodes tend to connect to nodes of similar degree and domain, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Monthly subnetworks display seasonal behavior with yearly periodicity

Over the analyzed decade, the network became more connected and clustered in colder months, with stronger associations and shorter distances between nodes (Figure 2C, Supplementary Figures 1 and 2). Most global network metrics indicated seasonal behavior with yearly periodicity (Figure 2C). For instance, edge density, average positive association score, and transitivity were highest at the beginning and end of each year, while average path length and assortativity (bacteria vs. eukaryotes) were highest in the middle of each year. Assortativity (degree), in contrast to other metrics, usually had two peaks per year corresponding to April or May, and November (Figure 2C).

We found that mainly temperature and day length, and to a lesser extent nutrient concentrations (mainly SiO_2, NO_3^- and NO_2^-, less PO_4^{3-}), and total chlorophyll-a concentration affected network topologies as indicated by correlation analysis (Supplementary Figure 2). For example, edge density was highest and temperature lowest in January-March. Then, the edge density dropped as the temperature increased. April-June displayed edge densities slightly above or similar to the warmest months July-September, while October-December had similar or slightly lower edge densities than the coldest months January-March. Edge density vs. hours of light (day length) indicated a yearly recurrent circular pattern for September-April (Supplementary Figure 1). May-August were not part of the circular pattern and had the highest day length and lowest edge density (Supplementary Figure 1).

Next, we quantified how many edges are preserved (kept), lost, and gained (new) in consecutive months. We found the highest loss of edges in April. The overall number of edges (preserved and gained) were lowest during April-September and increased towards the end of each year (Figure 2B). The number of associations changed over time in a yearly recurring pattern with few associations being preserved when transitioning from colder to warmer waters. We see a clear network change from colder to warmer months, similar to a crash. In turn, the network change from warmer to colder months is less abrupt, similar to a reassembling. Thus, network change was not symmetrical over the studied decade at BBMO. Moreover, we defined summer and winter as in (Krabberød *et al.*, 2021), and compared both seasons between consecutive years in terms of preserved, gained and lost associations and ASVs. We observed higher repeatability in terms of edges (Supplementary Figure 3) and ASVs (results not shown) in colder than in warmer months, indicating higher predictability during low temperature seasons.

Potential core associations

A single static network can comprise permanent, seasonal, and temporary associations. By comparing

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

monthly subnetworks, we identified edges that remain (preserved), appear (gained), or disappear (lost) over time (Figure 2B). Intuitively, we would classify permanent associations through 100% recurrence. However, no association fulfilled the 100% criteria. Most associations had a low recurrence with three-quarters of the associations present in no more than 38% (46 monthly subnetworks). The average association prevalence increased slightly for taxonomically more related microorganisms (Supplementary Figure 4). Considering the 100 most prevalent associations, which appeared in 71.7-98.3% (86-118) monthly subnetworks, 87 were bacterial associations (Supplementary Table 2).

Although temporal recurrence of associations over the ten years was low, we found high recurrence in corresponding months from different years. We quantified the fraction of subnetworks in which each association appeared (Supplementary Figure 5). We observed the highest prevalence from December to March, and the lowest prevalence from June to August (Supplementary Figure 5). For each month, we taxonomically characterized prevalent associations appearing in at least nine out of ten monthly subnetworks (Figure 3). We found more association partners in colder waters compared to warmer waters. *Alphaproteobacteria* associations dominated, especially in April and May. The *Alphaproteobacteria* ASVs having highly prevalent associations belonged to *Pelagibacter ubique* (SAR11 Clades Ia & II), *Rhodobacteraceae*, *Amylibacter*, *Puniceispirillales (SAR116)*, *Ascidiaceihabitans*, *Planktomarina*, *Parvibaculales (OCS116)*, and *Kiloniella*. Between April and May, we noticed a large increase in the fraction of associations including *Cyanobacteria* or *Bacteroidetes* as association partners. While *Cyanobacteria* associations were a small fraction during November-April, they had a dominant role from May-October along with *Bacteroidetes* and *Alphaproteobacteria* associations (Figure 3).

Dynamic associations within main taxonomic groups: the case of Cyanobacteria

Our results indicated that associations are dynamic within specific taxonomic groups. Therefore, we investigated their behavior in *Cyanobacteria* given the importance of this group as primary producers in the ocean. We found 661 associations for *Cyanobium*, *Prochlorococcus*, and *Synechococcus* ASVs (Figure 4 and Supplementary Figure 6). Most associations between cyanobacterial ASVs were positive (63 of 65), only a *Synechococcus* (referred to as bn_ASV_5) was negatively associated (association score measured -0.5) to other *Synechococcus* (bn_ASV_1 and bn_ASV_25), which were positively associated (association score of 0.8). While bn_ASV_5 appeared mainly in colder months, the other two appeared mainly in warmer months (Supplementary Figure 6). All *Cyanobacteria* had more associations to other bacteria (in total 433) than eukaryotes (in total 163), which were dinoflagellate (103), Chlorophyta (25),

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ochrophyta (12), Cryptophyta (11), Stramenopiles (5), Ciliophora (5), and Cercozoa (2).

Within the temporal network, the fraction of *Cyanobacteria* associations was highest in April-October (Figure 4A), which are the months with the fewest edges in the entire temporal network (Figure 2B), e.g., in the year 2011 (Figure 4B). We found that cyanobacterial ASVs, although being evolutionarily related, behaved differently in terms of number of associations over time, and association partners (Supplementary Figure 6). For example, *Synechococcus* bn_ASV_5 had less partner than bn_ASV_1 according to numbers of associations but more according to taxonomic variety; both belonged to the most abundant ASVs (Supplementary Figure 6). Only a tiny fraction of *Prochlorococcus* (e.g. bp_ASV_18) association partners were other *Cyanobacteria*, which contrasted to *Synechococcus* and *Cyanobium* (Supplementary Figure 6). Moreover, we observed that *Cyanobium* (bn_ASV_20) connected to one *Deltaproteobacteria* (SAR324) ASV during the first eight years, but the association disappeared in the last two years. In particular, the inferred association duration was 101 months, starting March 2004 and ending with July 2012. After summer 2012, the Deltaproteobacteria ASV was not detected except from a few reads in November and December of 2012 and 2013. This *Cyanobacteria* example is likely representative of the dynamics of associations within other main taxonomic groups.

DISCUSSION

Previous work found yearly recurrence in microbial community composition at the BBMO (Giner *et al.*, 2019; Auladell *et al.*, 2020; Krabberød *et al.*, 2021), and at the Bay of Banyuls (Lambert *et al.*, 2019), both in the NW Mediterranean Sea. Our approach focused in the connectivity of microorganisms and how they organize themselves from a network perspective. Similar to previous studies (Giner *et al.*, 2019; Lambert *et al.*, 2019; Auladell *et al.*, 2020; Krabberød *et al.*, 2021), our temporal network displayed seasonality with annual periodicity for most global network metrics. In general, our measured global network metrics are within previous work range (Steele *et al.*, 2011; Chow *et al.*, 2013, 2014; Cram *et al.*, 2015; Lima-Mendez *et al.*, 2015; Zhao *et al.*, 2016; Chaffron *et al.*, 2020) (Table 2 for edge density, transitivity, and average path length). Contrary to early works reporting biological networks generally being disassortative (negative assortativity based on degree) (Newman, 2002), our single static network and monthly subnetworks were assortative. Microorganisms had more and stronger connections and a tighter clustering in colder than in warmer waters. Seasonal bacterial freshwater networks (Zhao *et al.*, 2016) also showed higher clustering in fall and winter than spring and summer, but in contrast to our work, networks were biggest in summer and smallest in winter. In agreement with our results, Chaffron et al.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reported higher association strength, edge density, and transitivity in polar regions (colder) compared to other regions (warmer) of the global ocean (Chaffron *et al.*, 2020). Colder waters in the Mediterranean Sea are milder than polar waters, but together, these results suggest that either microorganisms interact more in colder environments, or that their recurrence is higher due to higher environmental selection exerted by low temperatures and therefore, they tend to co-occur. Alternatively, lack of resources (mostly nutrients) in summer or in the tropical and subtropical ocean may prevent the establishment of several microbial interactions. In any case, temperature may not be the only driver of network architecture.

The effects of environmental variables on network metrics are unclear (Röttjers & Faust, 2018), yet, our approach allowed identifying potential environmental drivers of network architecture. Correlation analyses pointed to the usual suspects that have been already found to influence microbial abundances. For instance, our results indicated that temperature and day length, key variables driving microbial assemblages in seasonal time-series (Bunse & Pinhassi, 2017; Giner *et al.*, 2019; Lambert *et al.*, 2019), and to a lesser extent inorganic nutrients, were the main factors influencing global network metrics. This is also in agreement with earlier works indicating that phosphorus and nitrogen are the primary limiting nutrients in the Western Mediterranean Sea (Estrada, 1996; Sala *et al.*, 2002). Altogether, our correlation analysis is a step forward to elucidate the effects of environmental variables on network metrics, although we did not consider several other variables that could affect networks (e.g. organic matter).

Our preliminary network (significant associations derived with eLSA) contained 18% negative edges compared to 0.9% in the single static network (after applying EnDED and Jaccard index). Thus, our filtering strategy removed proportionally more negative edges. Associations may represent positive or negative interactions, but they can also indicate high niche overlap (positive association) or divergent niches (negative association) between microorganisms (Hernandez *et al.*, 2021). We hypothesize that most of the removed negative edges represented associations between microorganisms from divergent niches, most likely corresponding to colder or warmer months.

We found more highly prevalent associations within specific months, than when considering all ten-years of data. Furthermore, our results indicate a potentially low number of core interactions and a vast number of non-core ones. Usually, core microorganisms are defined based on sequence abundances, as microorganisms (or taxonomical groups) appearing in all samples or habitats being under investigation (Shade & Handelsman, 2012). Shade & Handelsman (Shade & Handelsman, 2012) suggested other parameters, including connectivity, will create a more complex portrait of the core microbiome and advance our understanding of the role of key microorganisms and functions within and across ecosystems

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Shade & Handelsman, 2012). Using a temporal network, we identified core associations based on recurrence, which contributes to our understanding of key interactions underpinning microbial ecosystem function. Considering associations within each month, we found more highly-prevalent associations in colder than in warmer months. Our results indicated microbial connectivity is more repeatable (indicating higher predictability) in colder than in warmer waters. On one hand, the microbial community in colder waters being more recurrent (Giner *et al.*, 2019) may explain our observations indicating a more robust connectivity. On the other hand, it may be the stronger connectivity that leads to more similar communities in colder waters in BBMO. Last but not least, the interplay of both species dynamics and interactions may determine community turnover in the studied ecosystem. From a technical viewpoint, the overall single static network may have missed to capture summer associations resulting in smaller monthly subnetworks. For instance, a previous work in freshwater lakes constructed season specific networks and found more associations in summer than winter with *Cyanobacteria* dominating in summer, which may be due to strong co-occurrence patterns and suitable living conditions (Zhao *et al.*, 2016).

Several network-based analyses have been used to study *Cyanobacteria* associations. For example, Chow et al. (Chow *et al.*, 2014) determined for 12 *Cyanobacteria* (*Prochlorococcus* and *Synechoccus*) 44 potential relationships with two potential eukaryote grazers (a ciliate and a dinoflagellate), 39 to other bacteria and three between *Cyanobacteria*, which were all positive. Similarly, all cyanobacterial ASVs in our study connected primarily to other bacterial ASVs, and exerted mainly positive associations. In agreement, Cyanobacteria also displayed primarily positive associations in a network determined for the global ocean (Lima-Mendez *et al.*, 2015).

Identifying different potential association partners of closely related *Cyanobacteria*, may indicate adaptations to different niches. A recent study found distinct seasonal patterns of closely related taxa indicating niche partitioning at the BBMO, including *Synechococcus* ASVs (Auladell *et al.*, 2020). Our approach can complement and further characterize “sub”-niches by providing association partners for different ASVs. Moreover, in contrast to a single static network, temporal networks allow identifying associated partners in time (Supplementary Figure 6). An increase in abundance of a microorganism may promote the growth of associated partners and a decrease may hinder the growth of partners or cause predators to prey on other microorganisms. Moreover, given the majority of association partners being other bacteria, the growth of *Cyanobacteria* may affect other bacteria and their growth, which is why it is necessary to explore potential interaction partners (Zhao *et al.*, 2016).

From a technical perspective, our approach allowed us to see what the single static network

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

captured since all our temporal network observations are linked to it. Thus, future studies with higher sampling frequency may be able to construct networks within a month. However, our approach is a good starting point that allows us to move forward, but still, it has limitations, suggesting caution when making biological interpretations from the temporal network. Another limitation is that we disregarded local network patterns by using global network metrics. Future work could use the local-topological metric based on graphlets (Pržulj *et al.*, 2004). Counting the number of graphlets a node is part of quantifies their local connection patterns, which allows to infer seasonal microorganisms through recurring connection patterns in a temporal network. Such a network-based approach would complement the detection of the seasonal microorganisms based on sequence abundances (Giner *et al.*, 2019).

CONCLUSION

Incorporating the temporal dimension in the microbial association analysis unveiled multiple patterns that often remain hidden when using static networks. We developed a post-network-construction approach to generate a temporal network from a single static network that represents a step forward for disentangling the temporal nature of microbial associations. Yet, this approach has limitations, such as the monthly sampling frequency in our study. Using a higher sampling frequency would be the main solution. Investigating a coastal marine microbial ecosystem over ten years revealed a one-year-periodicity in the network topology. The temporal architecture was not stochastic, but displayed a modest amount of recurrence over time, especially in winter. Altogether, our approach allows comparing (sub)networks across spatiotemporal scales. Future efforts to understand the ocean microbiome should consider the dynamics of microbial interactions as these can be basis of ecosystem function.

MATERIALS AND METHODS

The Blanes Bay Microbial Observatory (BBMO)

BBMO is a coastal oligotrophic site in the North-Western Mediterranean Sea (41°40′N 2°48′E) with not many identified natural disturbances and little anthropogenic pressures, with the exception of the construction of a nearby harbor from 2010 to 2012 (Gasol *et al.*, 2016; Ferrera *et al.*, 2020). The seasonal cycle is typical for a temperate coastal system (Gasol *et al.*, 2016), and the main environmental factors influencing microbial seasonal succession in temperate waters have been well studied and are known (Bunse & Pinhassi, 2017). Shortly, the water column is slightly stratified in summer before it destabilizes

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and mixes in late fall, increasing the availability of inorganic nutrients with maximum concentrations in winter, between November and March. The high amounts of nutrients and increasing light induce phytoplankton blooms, mostly in late winter-early spring. During summer, inorganic nutrients become limiting, primary production is minimal, and dissolved organic carbon accumulates (Gasol *et al.*, 2016).

From sampling to sequence abundances

We sampled surface water (≈ 1m depth) monthly from January 2004 to December 2013 to determine microbial community composition and also measured ten environmental variables, which were previously described (Gasol *et al.*, 2016; Giner *et al.*, 2019): water temperature (°C) and salinity (obtained in situ with a SAIV-AS-SD204 Conductivity-Temperature-Depth probe), day-length (hours of light), turbidity (Secchi depth in meters), total chlorophyll-a concentration (μg/l, fluorometry of acetone extracts after 150 ml filtration on GF/F filters), and five inorganic nutrients: PO_4^{3-}, NH_4^+, NO_2^-, NO_3^- and SiO_2 (μM, determined with an Alliance Evolution II autoanalyzer (Grasshoff *et al.*, 2009)).

Sampling of microbial communities, DNA extraction, rRNA-gene amplification, sequencing and bioinformatic analyses are explained in detail in (Krabberød *et al.*, 2021). In short, 6 L of water were prefiltered through a 200 µm nylon mesh and subsequently filtered through another 20 µm nylon mesh and separated into nanoplankton (3 – 20 μm) and picoplankton (0.2 – 3 μm) using a 3 µm and 0.2 µm pore-size polycarbonate and Sterivex filters, respectively. Then, the DNA was extracted from the filters using a phenol-chloroform protocol (Schauer *et al.*, 2003), which has been modified for using Amicon units (Millipore) for purification. We amplified the 18S rRNA genes (V4 region) with the primers TAReukFWD1 and TAReukREV3 (Stoeck *et al.*, 2010), and the 16S rRNA genes (V4 region) with Bakt 341F (Herlemann *et al.*, 2011) and 806RB (Apprill *et al.*, 2015). Amplicons were sequenced in a MiSeq platform (2x250bp) at RTL Genomics (Lubbock, Texas). Read quality control, trimming, and inference of Operational Taxonomic Units (OTUs) delineated as Amplicon Sequence Variants (ASVs) was made with DADA2 (Callahan *et al.*, 2016), v1.10.1, with the maximum number of expected errors set to 2 and 4 for the forward and reverse reads, respectively.

Microbial sequence abundance tables were obtained for each size fraction for both microbial eukaryotes and prokaryotes. Before merging the tables, we subsampled each table to the lowest sequencing depth of 4907 reads with the *rrarefy* function from the Vegan R-package (Oksanen *et al.*, 2019), v2.4-2, (see details in (Krabberød *et al.*, 2021)). We excluded 29 nanoplankton samples (March 2004, February 2005, May 2010 - July 2012) due to suboptimal amplicon sequencing. In these, abundances were estimated

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

using seasonally aware missing value imputation by the weighted moving average for time series as implemented in the imputeTS R-package (Moritz & Gatscha, 2017), v2.8.

Sequence taxonomy was inferred using the naïve Bayesian classifier method (Wang *et al.*, 2007) together with the SILVA database (Quast *et al.*, 2012), v.132, as implemented in DADA2 (Callahan *et al.*, 2016). Additionally, eukaryotic microorganisms were BLASTed (Altschul *et al.*, 1990) against the Protist Ribosomal Reference (PR2) database (Guillou *et al.*, 2012), v4.10.0. The PR2 classification was used when the taxonomic assignment from SILVA and PR2 disagreed. We removed ASVs that identified as Metazoa, Streptophyta, plastids, mitochondria, and Archaea since the 341F-primer was not optimal for recovering this domain (McNichol *et al.*, 2020).

The resulting table contained 2924 ASVs, Table 1A. Next, we removed rare ASVs, keeping ASVs with sequence abundance sums above 100 reads and prevalence above 15% of the samples, i.e., we considered taxa present in at least 19 months. The resulting table contained 1782 ASVs, Table 1B. An ASV can appear twice, in the nano and pico size fractions due to dislodging cells or particles and filter clogging. This can introduce biases in our analysis. To reduce these biases, as done previously (Krabberød *et al.*, 2021), we divided the abundance sum of the bigger by the smaller size-fraction for each ASV appearing in both size fractions and set the picoplankton abundances to zero if the ratio exceeded 2. Likewise, we set the nanoplankton abundances to zero if the ratio was below 0.5. This operation removed two eukaryotic ASVs and 41 bacterial ASVs from the nanoplankton, and 30 bacterial ASVs from the picoplankton (Table 1C). The resulting table was used for network inference.

From sequence abundances to the single static network

First, we constructed a preliminary network using the tool eLSA (Xia *et al.*, 2011, 2013), as done in (Deutschmann *et al.*, 2020; Krabberød *et al.*, 2021), including default normalization and z-score transformation, using median and median absolute deviation. Although we are aware of time-delayed interactions, we considered our 1-month sampling interval as too large for inferring time-delayed associations with a solid ecological basis, and focused on contemporary interactions between co-occurring microbes. Using 2000 iterations, we estimated p-values with a mixed approach that performs a random permutation test of a co-occurrence if the comparison's theoretical *p*-values are below 0.05. The Bonferroni false discovery rate (*q*) was calculated based on the *p*-values using the *p.adjust* function from the stats R-package (R Core Team, 2019). We used the 0.001 significance threshold for the *p* and *q* values, as suggested in other studies (Weiss *et al.*, 2016). We refrained from using an association strength

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

threshold since it may not be appropriate to differentiate between true interactions and environmentally-driven associations (Deutschmann *et al.*, 2020), and changing thresholds have been shown to lead to different network properties (Connor *et al.*, 2017). The preliminary network contained 754 nodes and 29820 edges (24458, 82% positive, and 5362, 18% negative).

Second, for environmentally-driven edge detection, we applied EnDED (Deutschmann *et al.*, 2020), combining the methods Interaction Information (with a 0.05 significance threshold and 10000 iterations) and Data Processing Inequality. We inserted artificial edges connecting each node to each environmental parameter. We identified and removed 3315 (11.12%) edges that were environmentally-driven, i.e., 26505 edges (23405, 88.3% positive, and 3100, 11.7% negative) remained, Supplementary Table 3 and 4.

Third, we determined the Jaccard index, J, for each microorganisms pair associated through an edge. Let S_i be the set of samples in which both microorganisms are present (sequence abundance above zero), and S_u be the set of samples in which one or both microorganisms are present. Then, we can calculate the Jaccard index as the fraction of samples in which both appear (intersection) from the number of samples in which at least one appears (union): $J = S_i/S_u$. We chose $J > 0.5$, which removed 9879 edges and kept 16626 edges (16481, 99.1% positive and 145, 0.9% negative). We removed isolated nodes, i.e., nodes without an associated partner in the network. The number and fraction of retained reads are listed in Table 1. The resulting network is our single static network.

From the single static network to the temporal network

We determined the temporal network comprising 120 sample-specific (monthly) subnetworks through the three conditions indicated below and visualized in Figure 1. The subnetworks are derived from the single static network and contain a node subset and an edge subset of the static network. Let *e* be an association between microorganisms *A* and *B*, with association duration $d = (t_1, t_2)$, i.e., the association starts at time point t_1 and ends at t_2. Then, considering month *m*, the association *e* is present in the monthly subnetwork N_m, if

1) *e* is an association in the single static network,
2) the microorganisms *A* and *B* are present within month *m*, and
3) *m* is within the duration of association, i.e., $t_1 \leq m \leq t_2$.

With the 2nd condition, we assumed that an association was present in a month if both microorganisms were present, i.e., the microbial abundances were non-zero for that month. However, we cannot assume

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

411 that microbial co-occurrence is a sufficient condition for a microbial interaction because different
412 mechanisms influence species and interactions, and the environmental filtering of species and interactions
413 can be different (Poisot *et al.*, 2012). Using only the species occurrence assumption would increase
414 association prevalence. To lower this bias, we also required that the association was present in the static
415 network, 1st condition, and within the association duration, 3rd condition, both inferred by eLSA (Xia *et*
416 *al.*, 2011, 2013). Lastly, we removed isolated nodes from each monthly subnetwork.
417
418 *Network analysis*
419 We computed global network metrics to characterize the single static network and each monthly
420 subnetwork, using the igraph R-package (Csardi & Nepusz, 2006). Some metrics tend to be more
421 correlated than others implying redundancy between them and clustering them into four groups
422 (Jamakovic & Uhlig, 2008). Thus, we selected one metric from each group: *edge density*, *average path*
423 *length*, *transitivity*, and *assortativity* based on node degree. In addition, we also computed the *average*
424 *strength of positive associations* between microorganisms using the mean, and *assortativity* based on the
425 nominal classification of nodes into bacteria and eukaryotes. Assortativity (bacteria vs. eukaryotes) is
426 positive if bacteria tend to connect with bacteria and eukaryotes tend to connect with eukaryotes. It is
427 negative if bacteria tend to connect to eukaryotes and vice versa. We also quantified associations by
428 calculating their prevalence as the fraction of monthly subnetworks in which the association was present
429 for all ten years (recurrence), and monthly. We visualized highly prevalent associations with the circlize
430 R-package (Gu *et al.*, 2014). We tested our hypotheses of environmental factors influencing network
431 topology by calculating the Spearman correlations between global network metrics and environmental
432 data, using Holm's multiple test correction to adjust p-values (Holm, 1979), with the function *corr.test* in
433 psych R-package (Revelle, 2020). We used Gephi (Bastian *et al.*, 2009), v.0.9.2, and the Fruchterman
434 Reingold Layout (Fruchterman & Reingold, 1991) for network visualizations.
435
436 *Cyanobacteria*
437 Our dataset contained 19 cyanobacterial ASVs, which all appeared in the nano-, and nine in the
438 picoplankton. This is against expectations, as *Cyanobacteria* are part of the pico-plankton. Yet, they have
439 been observed in fractions above 3 µm at BBMO (Mestre *et al.*, 2020). Recovering ASVs in the
440 nanoplankton may be due to cell aggregation, particle attachment, clogging of filters or being prey to
441 bigger microorganisms. We blasted the sequences against the Cyanorak database (Garczarek *et al.*, 2021),

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

v.2. We performed BLASTN matches against the nucleotide database containing all *Synechococcus* and *Prochlorococcus* RNAs with the option -evalue 1.0e-5. We found 2812 sequences comprising 95 different ecotypes (considering name, clade and subclade), with 93.84-100% identity. There were 63 sequences (34 different microorganisms) with a similarity of 100% for 11 ASVs. Most matching sequences were found for *Synechococcus* ASV_1. While *Synechococcus* ASV_5 had only two 100% hits, they did not 100% match to ASV_1 (Supplementary Table 5).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The BBMO microbial sequence abundances (ASV tables), taxonomic classifications, environmental data including nutrients will be publicly available after acceptance. The data are of course available to editors and reviewers upon request. Networks are already available. R-Markdowns for data analysis including commands to run eLSA and EnDED (environmentally-driven-edge-detection and computing Jaccard index) are publicly available: https://github.com/InaMariaDeutschmann/TemporalNetworkBBMO.

Competing interests

The authors declare that they have no competing interests.

Funding

This project and IMD received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 675752 (ESR2, http://www.singek.eu) to RL. RL was supported by a Ramón y Cajal fellowship (RYC-2013-12554, MINECO, Spain). This work was also supported by the projects INTERACTOMICS (CTM2015-69936-P, MINECO, Spain), MicroEcoSystems (240904, RCN, Norway) and MINIME (PID2019-105775RB-I00, AEI, Spain) to RL. FL was supported by the Spanish National Program FPI 2016 (BES-2016-076317, MICINN, Spain). SC was supported by the CNRS MITI through the interdisciplinary program Modélisation du Vivant (GOBITMAP grant), and the H2020 project AtlantECO (award number 862923). A range of projects from the EU and the Spanish Ministry of Science funded data collection and ancillary analyses at the BBMO.

Author's contributions

The overall project was conceived and designed by RL and AKK. VB, JMG, and RM were responsible for the sampling and contextual data at the BBMO. RL processed the amplicon data from BBMO generating ASV tables. AKK constructed the initial preliminary network. It was the starting point of the present study, which is part of the overall project. IMD developed the conceptual approach and DE, SC, and RL contributed to its finalization. IMD performed the data analysis. ED, DE, SC, LFB, FL, AKK, CM, and RL contributed with biological interpretation of the results. IMD wrote the original draft. All authors contributed to manuscript revisions and approved the final version of the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

We thank all members of the Blanes Bay Microbial Observatory team with the multiple projects funding this collaborative effort. Part of the analyses have been performed at the Marbits bioinformatics core at ICM-CSIC(https://marbits.icm.csic.es).

References

ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W., & LIPMAN, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

APPRILL, A., MCNALLY, S., PARSONS, R., & WEBER, L. (2015) Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, **75**, 129–137.

AULADELL, A., BARBERÁN, A., LOGARES, R., GARCÉS, E., GASOL, J.M., & FERRERA, I. (2020) Seasonal niche differentiation between evolutionary closely related marine bacteria. *bioRxiv*, 2020.12.17.423265.

BALDAUF, S.L. (2008) An overview of the phylogeny and diversity of eukaryotes. *Journal of Systematics and Evolution*, **46**, 263.

BAR-ON, Y.M. & MILO, R. (2019) The Biomass Composition of the Oceans: A Blueprint of Our Blue Planet. *Cell*, **179**, 1451–1454.

BAR-ON, Y.M., PHILLIPS, R., & MILO, R. (2018) The biomass distribution on Earth. *Proceedings of the National Academy of Sciences*, **115**, 6506–6511.

BARRACLOUGH, T.G. (2015) How Do Species Interactions Affect Evolutionary Dynamics Across Whole Communities? *Annu. Rev. Ecol. Evol. Syst.*, **46**, 25–48.

BASTIAN, M., HEYMANN, S., & JACOMY, M. (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks. *ICWSM*, **3**.

BJORBÆKMO, M.F.M., EVENSTAD, A., RØSÆG, L.L., KRABBERØD, A.K., & LOGARES, R. (2019) The planktonic protist interactome: where do we stand after a century of research? *The ISME Journal*, DOI: 10.1038/s41396-019-0542-5.

BLONDER, B., WEY, T.W., DORNHAUS, A., JAMES, R., & SIH, A. (2012) Temporal dynamics and network analysis. *Methods in Ecology and Evolution*, **3**, 958–972.

BUNSE, C. & PINHASSI, J. (2017) Marine Bacterioplankton Seasonal Succession Dynamics. *Trends in Microbiology*, **25**, 494–505.

CALLAHAN, B.J., MCMURDIE, P.J., ROSEN, M.J., HAN, A.W., JOHNSON, A.J.A., & HOLMES, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, **13**, 581–583.

CHAFFRON, S., DELAGE, E., BUDINICH, M., VINTACHE, D., HENRY, N., NEF, C., ARDYNA, M., ZAYED, A.A., JUNGER, P.C., GALAND, P.E., LOVEJOY, C., MURRAY, A., SARMENTO, H., ACINAS, S., BABIN, M., IUDICONE, D., JAILLON, O., KARSENTI, E., WINCKER, P., KARP-BOSS, L., SULLIVAN, M.B., BOWLER, C., DE VARGAS, C., & EVEILLARD, D. (2020) Environmental vulnerability of the global ocean plankton community interactome. *bioRxiv*, 2020.11.09.375295.

CHOW, C.-E.T., KIM, D.Y., SACHDEVA, R., CARON, D.A., & FUHRMAN, J.A. (2014) Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. *The ISME Journal*, **8**, 816–829.

CHOW, C.-E.T., SACHDEVA, R., CRAM, J.A., STEELE, J.A., NEEDHAM, D.M., PATEL, A., PARADA, A.E., & FUHRMAN, J.A. (2013) Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. *The ISME Journal*, **7**, 2259–2273.

CONNOR, N., BARBERÁN, A., & CLAUSET, A. (2017) Using null models to infer microbial co-occurrence

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

networks. *PLOS ONE*, **12**, 1–23.

CRAM, J.A., XIA, L.C., NEEDHAM, D.M., SACHDEVA, R., SUN, F., & FUHRMAN, J.A. (2015) Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes. *The ISME Journal*, **9**, 2573–2586.

CSARDI, G. & NEPUSZ, T. (2006) The igraph software package for complex network research. *InterJournal*, **Complex Systems**, 1695.

DELONG, E.F. (2009) The microbial ocean from genomes to biomes. *Nature*.

DEUTSCHMANN, I.M., LIMA-MENDEZ, G., KRABBERØD, A.K., RAES, J., VALLINA, S.M., FAUST, K., & LOGARES, R. (2020) Disentangling environmental effects in microbial association networks. *PREPRINT (Version1) available at Research Square*, DOI: 10.21203/rs.3.rs-57387/v1.

ESTRADA, M. (1996) Primary production in the northwestern Mediterranean.

FALKOWSKI, P.G., FENCHEL, T., & DELONG, E.F. (2008) The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science*.

FERRERA, I., REÑÉ, A., FUNOSAS, D., CAMP, J., MASSANA, R., GASOL, J.M., & GARCÉS, E. (2020) Assessment of microbial plankton diversity as an ecological indicator in the NW Mediterranean coast. *Marine Pollution Bulletin*, **160**, 111691.

FRUCHTERMAN, T.M.J. & REINGOLD, E.M. (1991) Graph drawing by force-directed placement. *Software: Practice and Experience*, **21**, 1129–1164.

GARCZAREK, L., GUYET, U., DORÉ, H., FARRANT, G.K., HOEBEKE, M., BRILLET-GUÉGUEN, L., BISCH, A., FERRIEUX, M., SILTANEN, J., CORRE, E., LE CORGUILLÉ, G., RATIN, M., PITT, F.D., OSTROWSKI, M., CONAN, M., SIEGEL, A., LABADIE, K., AURY, J.-M., WINCKER, P., SCANLAN, D.J., & PARTENSKY, F. (2021) Cyanorak v2.1: a scalable information system dedicated to the visualization and expert curation of marine and brackish picocyanobacteria genomes. *Nucleic Acids Research*, **49**, D667–D676.

GASOL, J.M., CARDELÚS, C., G MORÁN, X.A., BALAGUÉ, V., FORN, I., MARRASÉ, C., MASSANA, R., PEDRÓS-ALIÓ, C., MONTSERRAT SALA, M., SIMÓ, R., VAQUÉ, D., & ESTRADA, M. (2016) Seasonal patterns in phytoplankton photosynthetic parameters and primary production at a coastal NW Mediterranean site. *Scientia Marina*.

GINER, C.R., BALAGUÉ, V., KRABBERØD, A.K., FERRERA, I., REÑÉ, A., GARCÉS, E., GASOL, J.M., LOGARES, R., & MASSANA, R. (2019) Quantifying long-term recurrence in planktonic microbial eukaryotes. *Molecular Ecology*, **28**, 923–935.

GRASSHOFF, K., KREMLING, K., & EHRHARDT, M. (2009) *Methods of seawater analysis*. John Wiley & Sons.

GU, Z., GU, L., EILS, R., SCHLESNER, M., & BRORS, B. (2014) circlize implements and enhances circular visualization in R. *Bioinformatics*, **30**, 2811–2812.

GUILLOU, L., BACHAR, D., AUDIC, S., BASS, D., BERNEY, C., BITTNER, L., BOUTTE, C., BURGAUD, G., DE VARGAS, C., DECELLE, J., DEL CAMPO, J., DOLAN, J.R., DUNTHORN, M., EDVARDSEN, B., HOLZMANN, M., KOOISTRA, W.H.C.F., LARA, E., LE BESCOT, N., LOGARES, R., MAHÉ, F., MASSANA, R., MONTRESOR, M., MORARD, R., NOT, F., PAWLOWSKI, J., PROBERT, I., SAUVADET, A.-L., SIANO, R., STOECK, T., VAULOT, D., ZIMMERMANN, P., & CHRISTEN, R. (2012) The Protist Ribosomal Reference database (PR^2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, **41**, D597–D604.

HERLEMANN, D.P., LABRENZ, M., JÜRGENS, K., BERTILSSON, S., WANIEK, J.J., & ANDERSSON, A.F. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *The ISME Journal*, **5**, 1571–1579.

HERNANDEZ, D.J., DAVID, A.S., MENGES, E.S., SEARCY, C.A., & AFKHAMI, M.E. (2021) Environmental

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

stress destabilizes microbial networks. *The ISME Journal*, DOI: 10.1038/s41396-020-00882-x.

HOLM, S. (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

JAMAKOVIC, A. & UHLIG, S. (2008) On the relationships between topological measures in real-world networks. *Networks & Heterogeneous Media*, **3**, 345–359.

KALLMEYER, J., POCKALNY, R., ADHIKARI, R.R., SMITH, D.C., & D'HONDT, S. (2012) Global distribution of microbial abundance and biomass in subseafloor sediment. *Proceedings of the National Academy of Sciences*, **109**, 16213–16216.

KRABBERØD, A.K., BJORBÆKMO, M.F.M., SHALCHIAN-TABRIZI, K., & LOGARES, R. (2017) Exploring the oceanic microeukaryotic interactome with metaomics approaches. *Aquatic Microbial Ecology*, **79**, 1–12.

KRABBERØD, A.K., DEUTSCHMANN, I.M., BJORBÆKMO, M.F.M., BALAGUÉ, V., GINER, C.R., FERRERA, I., GARCÉS, E., MASSANA, R., GASOL, J.M., & LOGARES, R. (2021) Long-term patterns of an interconnected core marine microbiota. *bioRxiv*, 2021.03.18.435965.

LAMBERT, S., TRAGIN, M., LOZANO, J.-C., GHIGLIONE, J.-F., VAULOT, D., BOUGET, F.-Y., & GALAND, P.E. (2019) Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular environmental perturbations. *The ISME Journal*, **13**, 388.

LAYEGHIFARD, M., HWANG, D.M., & GUTTMAN, D.S. (2017) Disentangling Interactions in the Microbiome: A Network Perspective. *Trends in Microbiology*.

LEWIS, W.H., TAHON, G., GEESINK, P., SOUSA, D.Z., & ETTEMA, T.J.G. (2020) Innovations to culturing the uncultured microbial majority. *Nature Reviews Microbiology*, DOI: 10.1038/s41579-020-00458-8.

LIMA-MENDEZ, G., FAUST, K., HENRY, N., DECELLE, J., COLIN, S., CARCILLO, F., CHAFFRON, S., IGNACIO-ESPINOSA, J.C., ROUX, S., VINCENT, F., BITTNER, L., DARZI, Y., WANG, J., AUDIC, S., BERLINE, L., BONTEMPI, G., CABELLO, A.M., COPPOLA, L., CORNEJO-CASTILLO, F.M., D'OVIDIO, F., DE MEESTER, L., FERRERA, I., GARET-DELMAS, M.-J., GUIDI, L., LARA, E., PESANT, S., ROYO-LLONCH, M., SALAZAR, G., SÁNCHEZ, P., SEBASTIAN, M., SOUFFREAU, C., DIMIER, C., PICHERAL, M., SEARSON, S., KANDELS-LEWIS, S., GORSKY, G., NOT, F., OGATA, H., SPEICH, S., STEMMANN, L., WEISSENBACH, J., WINCKER, P., ACINAS, S.G., SUNAGAWA, S., BORK, P., SULLIVAN, M.B., KARSENTI, E., BOWLER, C., DE VARGAS, C., & RAES, J. (2015) Determinants of community structure in the global plankton interactome. *Science*, **348**, 1262073.

LINDSTRÖM, E.S. & LANGENHEDER, S. (2012) Local and regional factors influencing bacterial community assembly. *Environmental Microbiology Reports*, **4**, 1–9.

LOCEY, K.J. & LENNON, J.T. (2016) Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences*, **113**, 5970–5975.

LOGARES, R., DEUTSCHMANN, I.M., JUNGER, P.C., GINER, C.R., KRABBERØD, A.K., SCHMIDT, T.S.B., RUBINAT-RIPOLL, L., MESTRE, M., SALAZAR, G., RUIZ-GONZÁLEZ, C., SEBASTIÁN, M., DE VARGAS, C., ACINAS, S.G., DUARTE, C.M., GASOL, J.M., & MASSANA, R. (2020) Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome*, **8**, 55.

MCNICHOL, J., BERUBE, P.M., BILLER, S.J., & FUHRMAN, J.A. (2020) Evaluating and Improving SSU rRNA PCR Primer Coverage via Metagenomes from Global Ocean Surveys. *bioRxiv*, 2020.11.09.375543.

MESTRE, M., HÖFER, J., SALA, M.M., & GASOL, J.M. (2020) Seasonal Variation of Bacterial Diversity Along the Marine Particulate Matter Continuum. *Frontiers in Microbiology*, **11**, 1590.

MORI, A.S., ISBELL, F., & SEIDL, R. (2018) β-Diversity, Community Assembly, and Ecosystem Functioning. *Trends in Ecology & Evolution*, **33**, 549–564.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MORITZ, S. & GATSCHA, S. (2017) *imputeTS: Time Series Missing Value Imputation*.

NEEDHAM, D.M., SACHDEVA, R., & FUHRMAN, J.A. (2017) Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *The ISME Journal*, **11**, 1614–1629.

NEWMAN, M.E.J. (2002) Assortative Mixing in Networks. *Phys. Rev. Lett.*, **89**, 208701.

OKSANEN, J., BLANCHET, F.G., FRIENDLY, M., KINDT, R., LEGENDRE, P., MCGLINN, D., MINCHIN, P.R., O'HARA, R.B., SIMPSON, G.L., SOLYMOS, P., STEVENS, M.H.H., SZOECS, E., & WAGNER, H. (2019) *vegan: Community Ecology Package*.

PARADA, A.E. & FUHRMAN, J.A. (2017) Marine archaeal dynamics and interactions with the microbial community over 5 years from surface to seafloor. *The ISME Journal*, **11**, 2510–2525.

POISOT, T., CANARD, E., MOUILLOT, D., MOUQUET, N., & GRAVEL, D. (2012) The dissimilarity of species interaction networks. *Ecology Letters*, **15**, 1353–1361.

PRŽULJ, N., CORNEIL, D.G., & JURISICA, I. (2004) Modeling interactome: scale-free or geometric? *Bioinformatics*, **20**, 3508–3515.

QUAST, C., PRUESSE, E., YILMAZ, P., GERKEN, J., SCHWEER, T., YARZA, P., PEPLIES, J., & GLÖCKNER, F.O. (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, **41**, D590–D596.

R CORE TEAM (2019) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

REVELLE, W. (2020) *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University.

RÖTTJERS, L. & FAUST, K. (2018) From hairballs to hypotheses–biological insights from microbial networks. *FEMS Microbiology Reviews*, **42**, 761–780.

SALA, M.M., PETERS, Francesc, GASOL, J.M., PEDRÓS-ALIÓ, C., MARRASÉ, C., & VAQUÉ, D. (2002) Seasonal and spatial variations in the nutrient limitation of bacterioplankton growth in the northwestern Mediterranean. *Aquatic Microbial Ecology*, **27**, 47–56.

SCHAUER, M., BALAGUÉ, V., PEDRÓS-ALIÓ, C., & MASSANA, R. (2003) Seasonal changes in the taxonomic composition of bacterioplankton in a coastal oligotrophic system. *Aquatic Microbial Ecology*, **31**, 163–174.

SHADE, A. & HANDELSMAN, J. (2012) Beyond the Venn diagram: the hunt for a core microbiome. *Environmental Microbiology*, **14**, 4–12.

STEELE, J.A., COUNTWAY, P.D., XIA, L., VIGIL, P.D., BEMAN, J.M., KIM, D.Y., CHOW, C.-E.T., SACHDEVA, R., JONES, A.C., SCHWALBACH, M.S., ROSE, J.M., HEWSON, I., PATEL, A., SUN, F., CARON, D.A., & FUHRMAN, J.A. (2011) Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *The ISME Journal*, **5**, 1414–1425.

STOECK, T., BASS, D., NEBEL, M., CHRISTEN, R., JONES, M.D.M., BREINER, H.-W., & RICHARDS, T.A. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, **19**, 21–31.

VELLEND, M. (2020) *The theory of ecological communities (MPB-57)*. Princeton University Press.

WANG, Q., GARRITY, G.M., TIEDJE, J.M., & COLE, J.R. (2007) Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology*, **73**, 5261–5267.

WEISS, S., VAN TREUREN, W., LOZUPONE, C., FAUST, K., FRIEDMAN, J., DENG, Y., XIA, L.C., XU, Z.Z., URSELL, L., ALM, E.J., BIRMINGHAM, A., CRAM, J.A., FUHRMAN, J.A., RAES, J., SUN, F., ZHOU, J., & KNIGHT, R. (2016) Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *The ISME Journal*, **10**, 1669–1681.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

669 WHITMAN, W.B., COLEMAN, D.C., & WIEBE, W.J. (1998) Prokaryotes: The unseen majority. *Proceedings*
670 *of the National Academy of Sciences*, **95**, 6578–6583.
671 XIA, L.C., AI, D., CRAM, J., FUHRMAN, J.A., & SUN, F. (2013) Efficient statistical significance
672 approximation for local similarity analysis of high-throughput time series data. *Bioinformatics*, **29**,
673 230–237.
674 XIA, L.C., STEELE, J.A., CRAM, J.A., CARDON, Z.G., SIMMONS, S.L., VALLINO, J.J., FUHRMAN, J.A., &
675 SUN, F. (2011) Extended local similarity analysis (eLSA) of microbial community and other time
676 series data with replicates. *BMC Systems Biology*, **5**, S15.
677 ZHAO, D., SHEN, F., ZENG, J., HUANG, R., YU, Z., & WU, Q.L. (2016) Network analysis reveals seasonal
678 variation of co-occurrence correlations between Cyanobacteria and other bacterioplankton.
679 *Science of The Total Environment*, **573**, 817–825.
680

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE LEGENDS

Figure 1: Conceptual idea on how we determine a temporal network from a single static network via subnetworks. A) A complete network would contain all possible associations (edges) between microorganisms (nodes). B) The single static network is inferred with the network construction tool eLSA and a filtering strategy considering association significance, the removal of environmentally-driven associations, and associations whose partners appeared in more samples together than alone, i.e., Jaccard index being above 0.5. An association having to be present in the single static network is the first out of three conditions for an association to be present in a monthly subnetwork. C) In order to determine monthly subnetworks, we established two further conditions for each edge. First, both microorganisms need to be present in the sample taken in the specific month. Second, the month lays within the time window of the association inferred through the network construction tool. Here, three months are indicated as an example. D) Example of monthly subnetworks for the three months. The colored nodes correspond to the abundances depicted in C).

Figure 2: Global (sub)network metrics. A) Number of ASVs (counting an ASV twice if it appears in both size fractions) for each of the 120 months of the Blanes Bay Microbial Observatory time series. There are 1709 ASVs, of which 709 ASVs are connected in the static network. In black, we show the number of nodes connected in the temporal network, and in red the number of nodes that are isolated in the temporal network, i.e., they are connected in the static network and have a sequence abundance above zero for that month ("non-zero"). In dark gray, we show the number of ASVs that are non-zero in a given month but were not connected in the static and subsequently temporal network. In light gray, we show the number of ASVs with zero-abundance in a given month. The sum of connected and isolated nodes and non-zero ASVs represents each month's richness (i.e., number of ASVs). B) By comparing the edges of two consecutive months, i.e., two consecutive monthly subnetworks, we indicate the number of edges that have been lost (red), preserved (black), and those that are gained (blue), compared to the previous month. C) Six selected global network metrics for each sample-specific subnetwork of the temporal network. The colored line indicates the corresponding metric for the single static network.

Figure 3: Associations with a monthly prevalence of at least 90%. Bacteria and eukaryotes are separated and ordered alphabetically. We provide in parentheses the number of associations that appeared in at least nine out of ten monthly subnetworks.

Figure 4: *Cyanobacteria* associations. A) Fraction of edges in the temporal network containing at least one *Cyanobacteria*. B) Location of *Cyanobacteria* associations in the temporal network and the single static network. Here we show, as an example, selected months of year 2011. The number and fraction of cyanobacterial edges and total number of edges is listed below each monthly subnetwork and the single static network.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TABLES

Table 1: Number and fraction of ASVs and reads (total, bacterial and eukaryotic) for the sequence abundance tables (A, B, and C), the preliminary network with significant edges (D), and the single static network (E) obtained after removing environmentally-driven edges and edges with association partners appearing more often alone than with the partner. If an ASV appeared in the nano- and pico-plankton size fractions, it was counted twice.

Count tables	ASVs	Reads	Eukaryote	Eukaryotic reads	Bacteria	Bacterial reads
A	2 924	2 273 548	1 365	1 121 855	1 559	1 151 693
B	1 782	2 155 318	1 009	1 057 599	773	1 097 719
C	1 709	2 062 866	1 007	1 057 263	702	1 005 603
D	754	1 657 885	306	730 025	448	927 860
E	709	1 621 959	294	719 558	415	902 401

Fractions	ASV	Reads	Eukaryote	Eukaryotic reads	Bacteria	Bacterial reads
B/A*100	60.94	94.80	73.92	94.27	49.58	95.31
C/A*100	58.45	90.73	73.77	94.24	45.03	87.32
D/C*100	44.12	80.37	30.39	69.05	69.05	92.27
E/C*100	41.49	78.63	29.20	68.06	59.12	89.74

A – raw sequence abundance table; B – sequence abundance table without rare ASVs; C –sequence abundance table after size-fraction filtering; D – preliminary network with significant edges; E – single static network

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

726 **Table 2:** Global network metrics of previously described microbial association networks

Edge density	Transitivity	Average path length	Sampling	Location	Domains	Notes	Reference
0.04	0.26	3.05	Monthly samples August 2000 - March 2004	Subsurface deep chlorophyll maximum depth off the southern California coast (SPOT)	Archaea, bacteria, and eukaryotes	Edge density for microbial network including environmental factors. Transitivity and average path length for microbial network.	(Steele *et al.*, 2011)
0.14	0.33	1.94	Monthly samples August 2000 - January 2011	Two depths at SPOT	Free-living bacteria and some picoeukaryotes	Metrics from surface layer network.	(Chow *et al.*, 2013)
0.02	0.24		Monthly samples March 2008 - January 2011	surface ocean (0-5m) at SPOT	Free-living eukaryotes (0.7–20 µm), bacteria (0.22–1 µm) and viruses (30 kDa–0.22 µm)		(Chow *et al.*, 2014)
0.04	0.28	2.07	Monthly samples August 2003 - January 2011	Five depths at SPOT	Free-living bacteria	Metrics from 5 m layer network.	(Cram *et al.*, 2015)
(0.023) W:0.033 Sp:0.032 S:0.036 F:0.029	(0.472) W:0.518 Sp:0.480 S:0.475 F:0.573	(4.84) W:2.16 Sp:5.03 S:7.26 F:3.04	Spatial samples	52 samples from freshwater lakes (surface) in China	Bacteria	Metrics for (whole network) and seasonal networks: W: winter, Sp: spring, S: summer, and F: fall	(Zhao *et al.*, 2016)
E:0.005 EP:0.003 P:0.008	E:0.2 EP:0.0 P:0.43	E:3.05, EP:3.02 P:2.56	Spatial sampling	68 stations from the Tara Oceans expeditions (TARA) at two depths across eight oceanic provinces	Organisms from seven size fractions spanning from viruses to small metazoans	Metrics from surface networks including E; eukaryotes only, EP: eukaryotes and prokaryotes (0.5-5 µm), and P: prokaryotes only (0.2-1.6 µm)	(Lima-Mendez *et al.*, 2015)
0.002	0.036		Spatial sampling	Samples from 115 stations from the TARA at two depths covering all major oceanic provinces from pole to pole	Bacteria, archaea, and eukaryotes from six size fractions.	Metrics represent the means of sample-specific subnetworks.	(Chaffron *et al.*, 2020)

727

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SUPPLEMENTARY MATERIAL

FIGURES LEGENDS

Supplementary Figure 1: Correlation Analysis. Using the temporal network, we correlated six global network metrics with environmental factors including the nutrients PO_4^{3-}, NH_4^+, NO_2^-, NO_3^- and SiO_2. The global network metrics were: Edge density, Average positive association (Avg. pos. ass.) score, Transitivity, Average path length (Avg. path length), Assortativity (degree), and Assortativity (bacteria vs. eukaryote). Each dot is a sample-specific subnetwork and its color indicates the month it represents. Also, the linear regression line with a 0.95 confidence interval is shown in gray.

Supplementary Figure 2: Correlation Analysis through linear regression. Using the temporal network, we correlated six global network metrics with environmental factors including the nutrients PO_4^{3-}, NH_4^+, NO_2^-, NO_3^- and SiO_2. The global network metrics were: Edge density, Average positive association (Avg. pos. ass.) score, Transitivity, Average path length (Avg. path length), Assortativity (degree), and Assortativity (bacteria vs. eukaryote). The number, circle's size and color in the square correspond to the Spearman correlation scores, no circle indicates non-significance.

Supplementary Figure 3: Number of preserved, gained, and lost edges in summer and winter. A) Indicates how we determined summer indicated with red dots (temperature above 17 ºC and day length above 14 hours) and winter indicated with blue dots (temperature below 17 ºC and day length below 11 hours); gray dots indicate months that are neither summer nor winter. B) accumulation curve of ASVs per year for winter (blue) and summer (red). C) and D) number of preserved, gained, and lost edges for winter and summer, respectively. The colors of flows indicate the prevalence of an edge with 10 (light blue) being present in each year, and 1 (dark blue) appearing in only one year. An edge appears in a year if it appears in at least one monthly subnetwork in the corresponding season. In winter, most edges appear in all years (light blue indicating 100% prevalence with edges present in all ten years), i.e. most edges are preserved in the consecutive months (we see a flow from the white preserved box to the next white preserved box). In summer, compared to winter, less edges are present in a month (combination of boxes indicating preserved, first time gained, and gained), and more edges are (re)gained and lost throughout the years (subsequently prevalence is lower indicated through darker blue).

Supplementary Figure 4: Association prevalence increases slightly when microorganisms are taxonomically more related. We grouped the associations according to the taxonomic classification of association partners (columns) and size fractions (rows). For example “Class” groups associations between bacteria and eukaryotes, respectively, which were assigned to the same class. The gray column groups associations between bacteria and eukaryotes. The boxplot shows the association prevalence over a decade, i.e. in how many monthly subnetworks an association appears (given as fraction from 0 to 100% = 120 networks).

Supplementary Figure 5: Association prevalence per month. Big bar plots: distribution of associations' prevalence for each month. For example, the bar at 100 for January indicates the number of edges that have been present in all Januarys of the ten -year time series. Small bar plots: number of nodes forming the associations with a 100% prevalence. For example, only bacteria were responsible for the edges during May, with an association prevalence of 100%. Bacteria are indicated with B or b, eukaryote with E or e. ASVs from the nano size-fraction have a capital letter (B, E), and ASVs from the pico size-fraction have a small letter (b, e).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figure 6: Association Partners of Cyanobacteria. Number of Cyanobacteria associations in the temporal network (stacked bars) and the cyanobacterial sequence abundance in each month (black dashed line). Within the box, figures are split by ASVs (rows) and size fraction: picoplankton (left column) and nanoplankton (right column). The unboxed plots on the right are ASVs detected only in the nanoplankton. The height of the bar indicates the number of edges in each month for each cyanobacterial ASV. The color indicates the taxonomy of the association partner. From bottom to top, first appear bacteria and then eukaryotes, both sorted alphabetically. The subtitle shows the number of association partners followed by their identifiers (first 3 letters) for bacteria and eukaryotes.

SUPPLEMENTARY TABLES

Supplementary Table 1: Number of nodes, removed isolated nodes, and number and fraction of edges in the preliminary network (A), and network obtained after removing environmentally-driven edges (B) and edges with association partners appearing more often alone than with the partner (C), which is the single static network. For comparison, we also give the minimum and maximum number of nodes and edges for the temporal network (D). We did not determine the union and intersection for the temporal network. If an ASV appeared in the nano and pico size fraction, it is counted twice. Therefore, for A-C) we also determined the number of microorganisms not considering size fraction (union) and being present in both size fractions (both, i.e. intersection).

	A) eLSA	B) EnDED	C) Static network	D) Range in Temporal network
Connected nodes	754	754	709	130-542
Bacteria (pico)	169	169	164	13-148
Bacteria (nano)	279	279	251	31-204
Bacteria (union)	309	309	281	
Bacteria (both)	139	139	134	
Eukaryote (pico)	150	150	141	7-124
Eukaryote (nano)	156	156	153	2-138
Eukaryote (union)	306	306	294	
Eukaryote (both)	0	0	0	
Isolated nodes	1000	0	45	6-38
Edges	29820	26505	16626	538-15083
Positive edges	24458	23405	16481	523-14940
(%)	82.0	88.3	99.1	92.2-99.7
Negative edges	5362	3100	145	12-143
(%)	18.0	11.7	0.9	0.3-7.8

pico and nano – microorganism detected in the picoplankton and nanoplankton, respectively, union – how many microorganisms when not considering size-fraction, both – how many microorganims appear in both size fractions

Supplementary Table 2: Top 100 most prevalent/recurring associations. Associations were classified based on the domain of association partners.

Association partners	Number of associations
Bacterial association in picoplankton	42
Bacterial association in nanoplankton	35
Bacterial associations between size fractions	10
Bacteria associated to Eukaryote in nanoplankton	4
Eukaryotic association in nanoplankton	3
Bacteria associated to Eukaryote in picoplankton	3
Bacteria in nanoplankton associated to Eukaryotic picoplankton	2
Eukaryotic association in picoplankton	1

Supplementary Table 3: Number of environmental factors leading to the removal of edges.

Number of environmental factors	Edges	Positive edges	Negative edges
0, i.e. not environmentally-driven edges	26505	23405 (88.3%)	3100 (11.7%)
1	2747	1019 (37.1%)	1728 (62.9%)
2	506	33 (6.5%)	473 (93.5%)
3	61	1 (1.6%)	60 (98.4%)
4	1	0 (0%)	1 (100%)

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table 4: Number of environmentally-driven edges for each environmental factor and fraction considering the total number of edges (29820) in the network. In addition, we present the number of positive and negative edges and the fraction considering number of edges removed through an environmental factor.

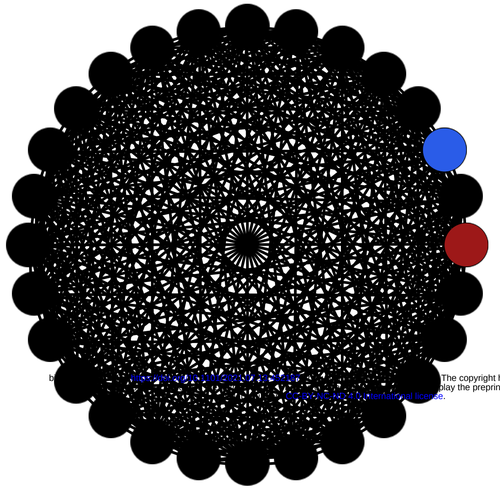
Environmental factor	Edges	Positive edges	Negative edges
Temperature	1920 (6.44%)	725 (37.8%)	1195 (62.2%)
Total chlorophyll-a concentration	838 (2.81%)	82 (9.8%)	756 (90.2%)
Day length	730 (2.45%)	237 (32.5%)	493 (67.5%)
NO_2^-	192 (0.64%)	26 (13.5%)	166 (86.5%)
SiO_2	162 (0.54%)	6 (3.7%)	156 (96.3%)
NO_3^-	57 (0.19%)	12 (21.1%)	45 (78.9%)
Turbidity	47 (0.16%)	0	47 (100%)
Salinity, NH_4^+, and PO_4^{3-}	0	0	0

Supplementary Table 5: 100% Matching sequences from Cyanorak database for selected cyanobacterial ASVs

ASV	Number	Matching sequence name with clade and subclade
Synechococcus #1	38	2x A15-24 III IIIa, 2x A15-28 III IIIb, 3x A15-44 II IIa, 2x A15-62 II IIc, 2x A18-40 III IIIa, 2x A18-46.1 III IIIa, 2x BOUM118 III IIIa, 2x CC9605 II IIc, 2x M16.1 II IIa, 2x PROS-U-1 II IIh, 2x ROS8604 I Ib, 3x RS9902 II IIa, 3x RS9907 II IIa, 2x RS9915 III IIIa, 2x TAK9802 II IIa, 1x WH8016 I Ib, 2x WH8103 III IIIa, 2x WH8109 II IIa
Synechococcus #5	2	2x PROS-9-1 I Ib
Prochlorococcus #18	2	1x EQPAC1 HLI HLI, 1x MED4 HLI HLI
Cyanobium #20	2	1x MINOS11 5.3 5.3, 1x RCC307 5.3 5.3

The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.
It is made available under a CC-BY-NC-ND 4.0 International license.

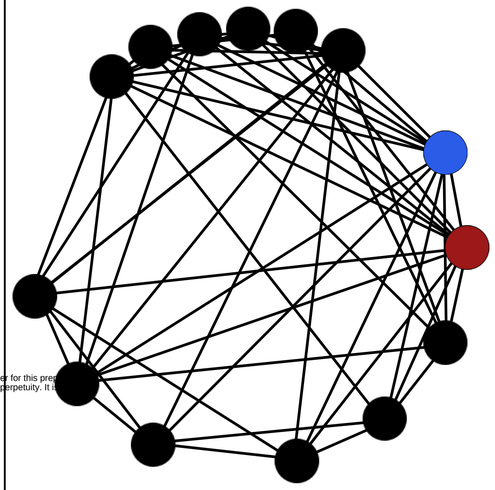
A) all potential associations



Microorganism A

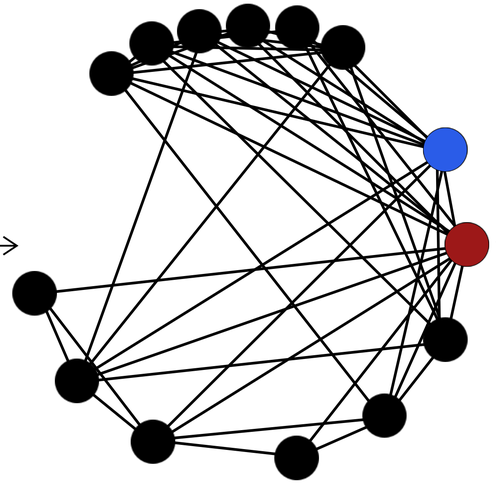
Microorganism B

B) determing single static network (1st condition)

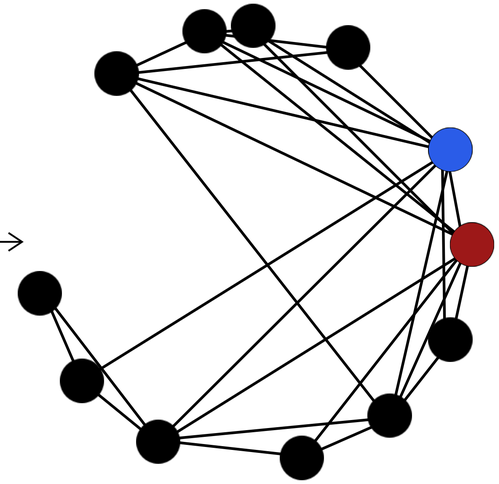


Preliminary network inferred through eLSA with significance threshold (p and q < 0.001)

removing environmentally-driven associations

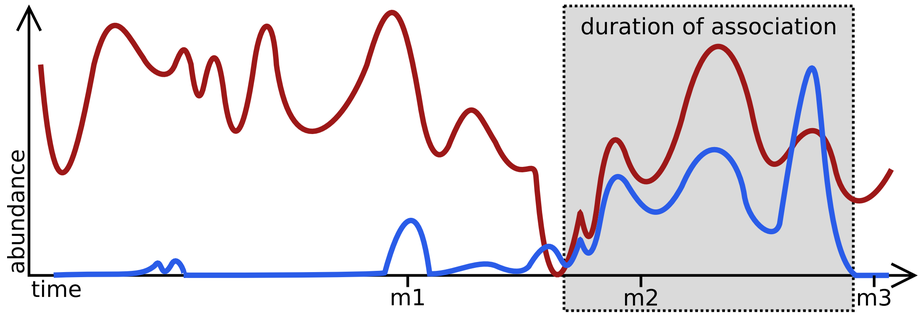


only keeping associations with Jaccard index >0.5



Single static network

C) determining subnetworks



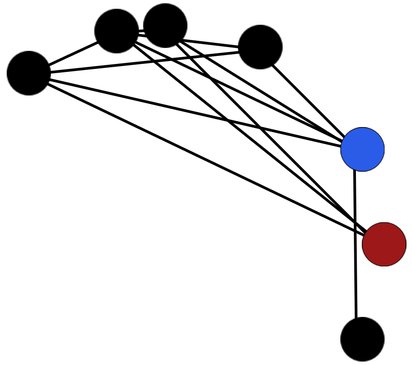
2nd condition: both microorganisms are present

✓m1 ✓m2 ✗m3

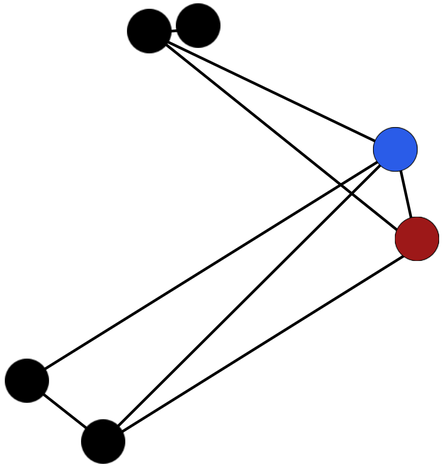
3rd condition: month is within duration of association

✗m1 ✓m2 ✗m3

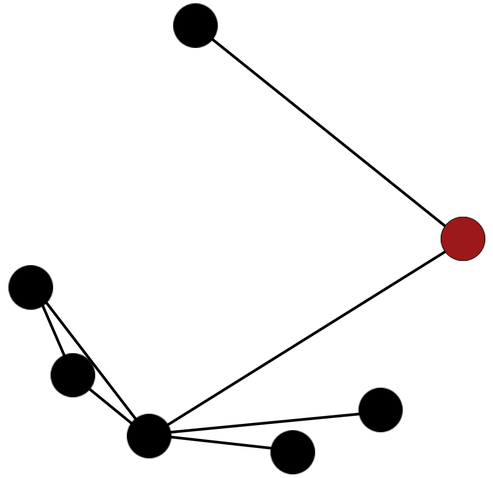
D) temporal network constituted from monthly subnetworks



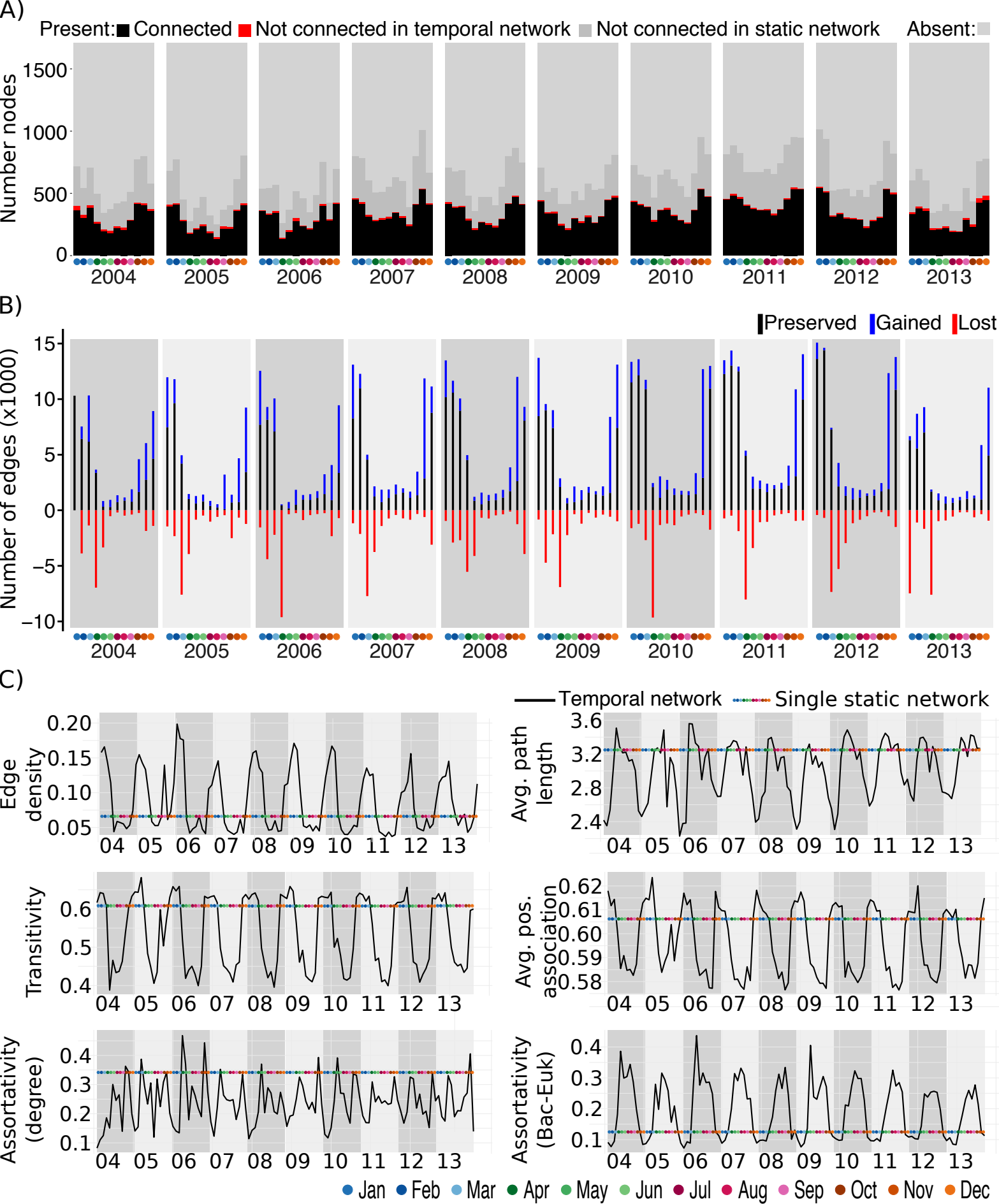
Month m1



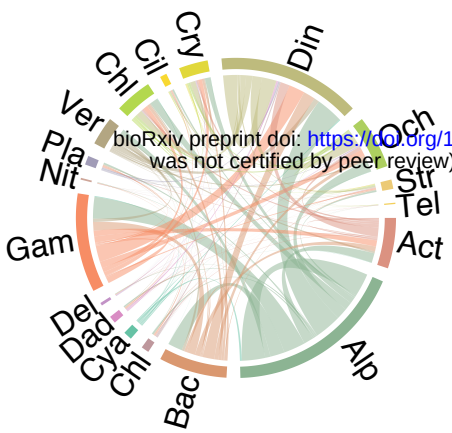
Month m2



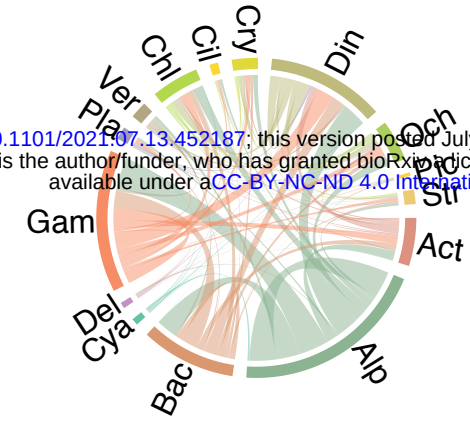
Month m3


A)
Present: Connected
Not connected in temporal network
Not connected in static network
Absent:
Number nodes
1500
1000
500
0
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
B)
Preserved
Gained
Lost
Number of edges (x1000)
15
10
5
0
−5
−10
C)
Temporal network
Single static network
Edge density
0.20
0.15
0.10
0.05
Avg. path length
3.6
3.2
2.8
2.4
Transitivity
0.6
0.5
0.4
Avg. pos. association
0.62
0.61
0.60
0.59
0.58
Assortativity (degree)
0.4
0.3
0.2
0.1
Assortativity (Bac-Euk)
0.4
0.3
0.2
0.1
04 05 06 07 08 09 10 11 12 13
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

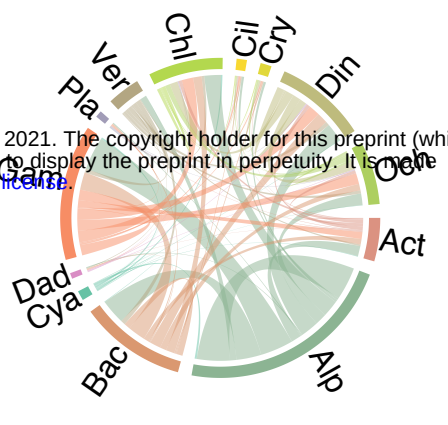
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



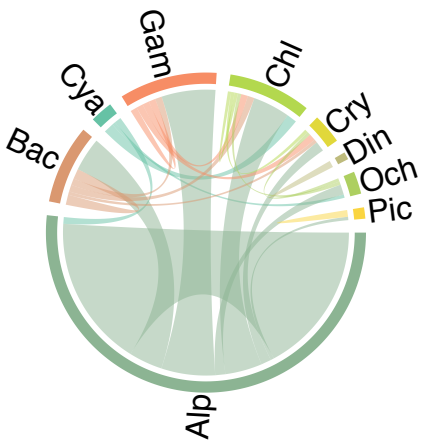
January (7441)



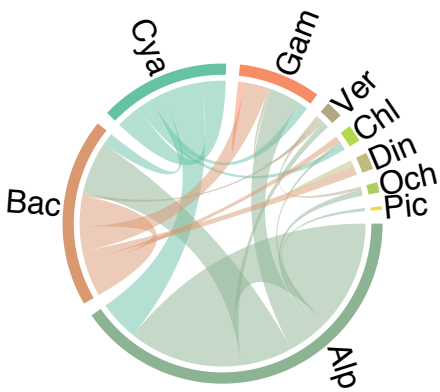
February (5423)



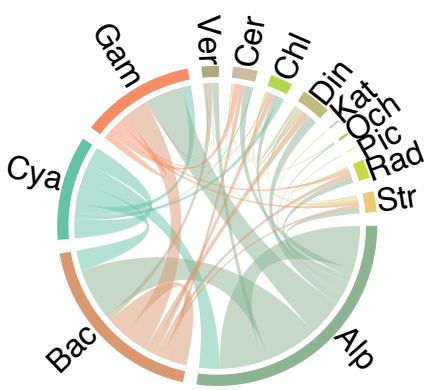
March (1965)



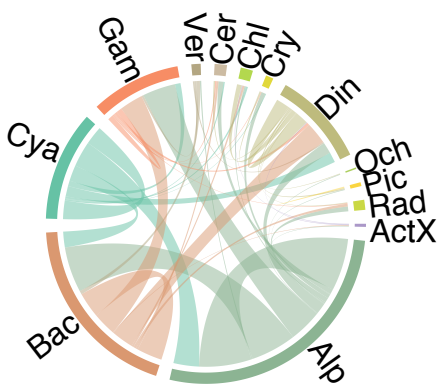
April (115)



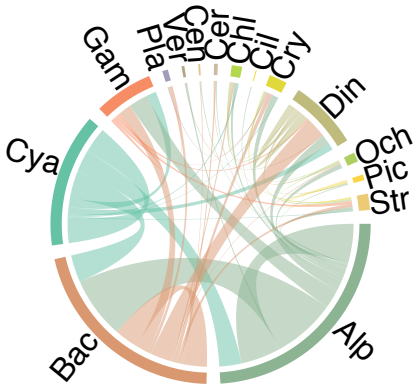
May (126)



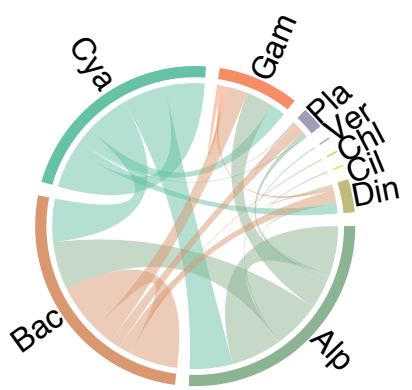
June (489)



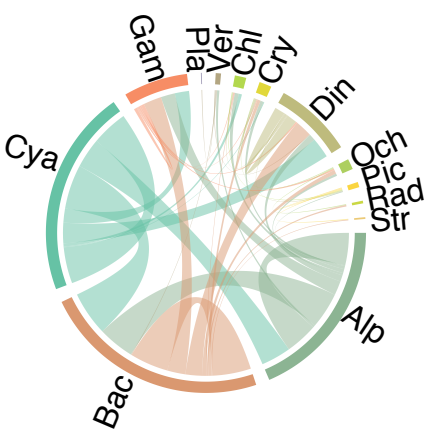
July (526)



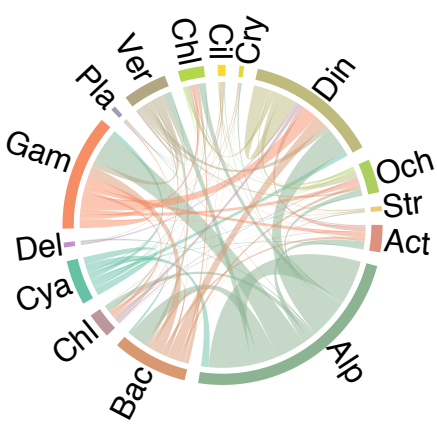
August (412)



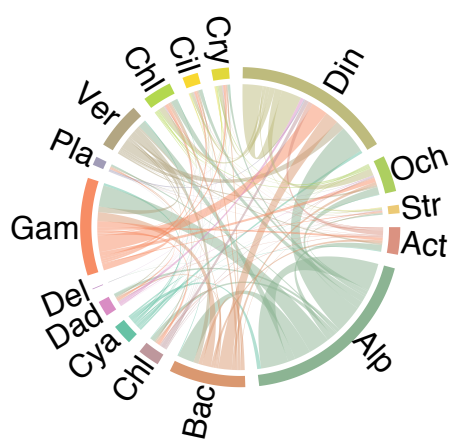
September (314)



October (460)

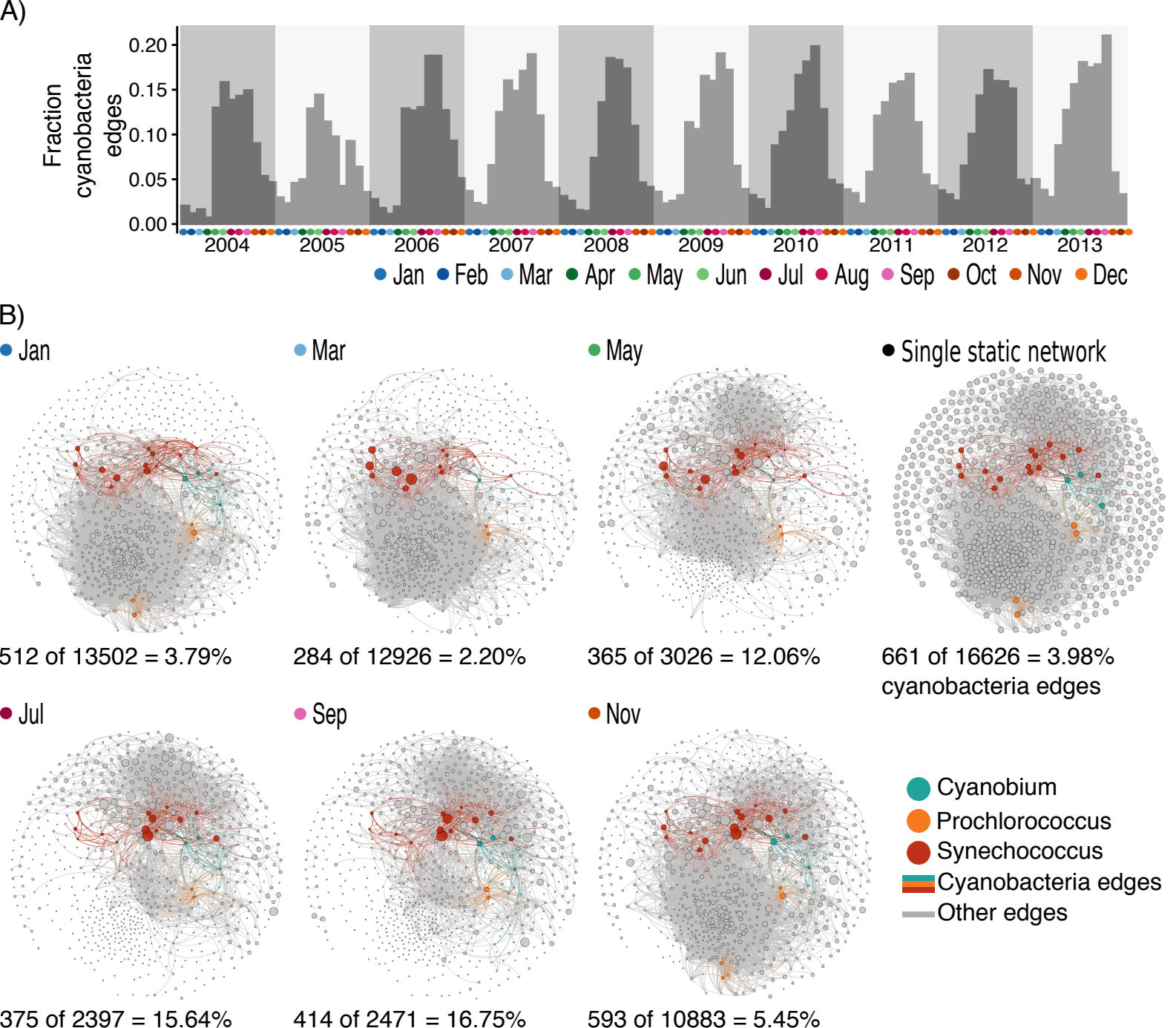


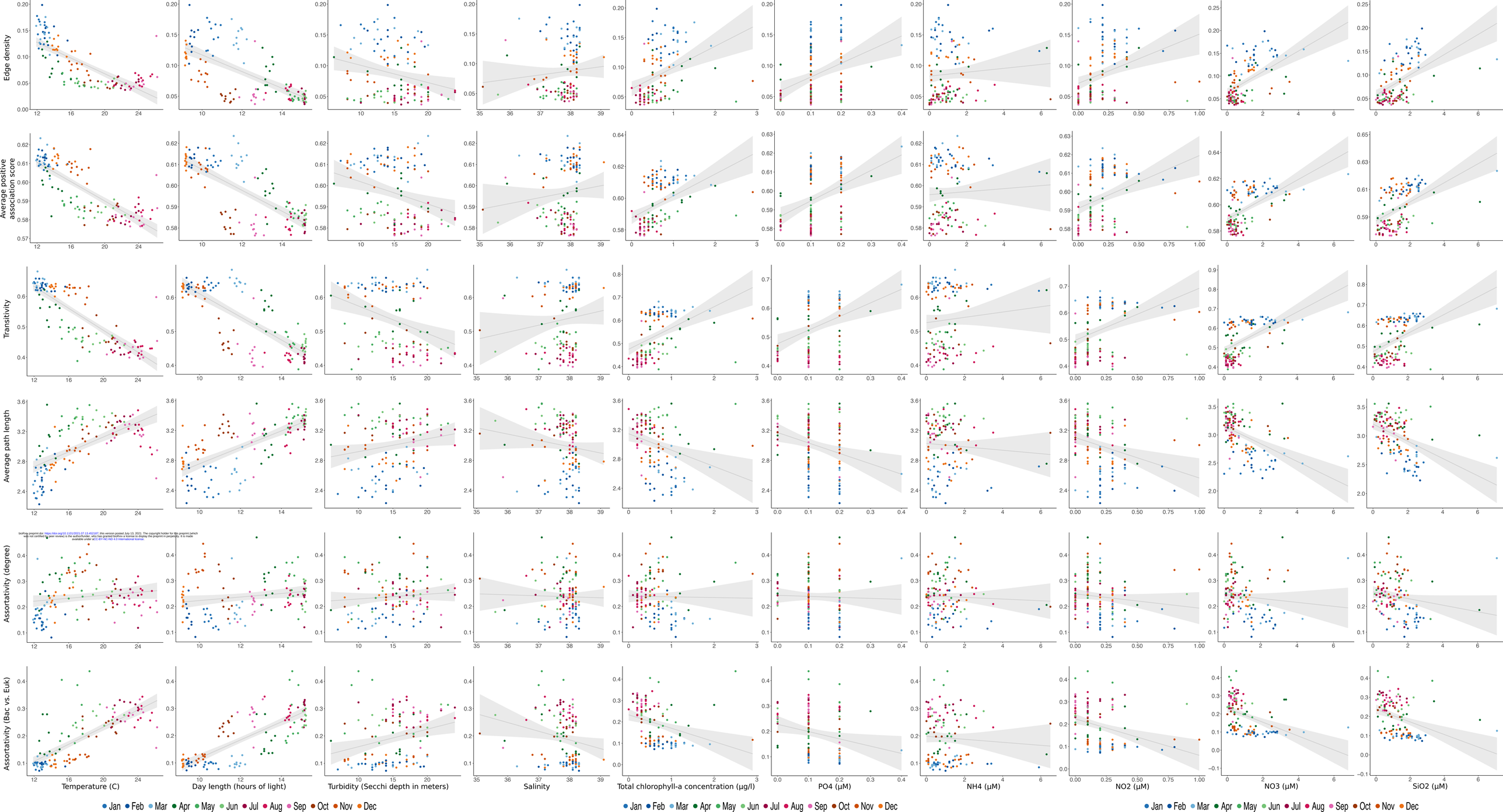
November (1737)



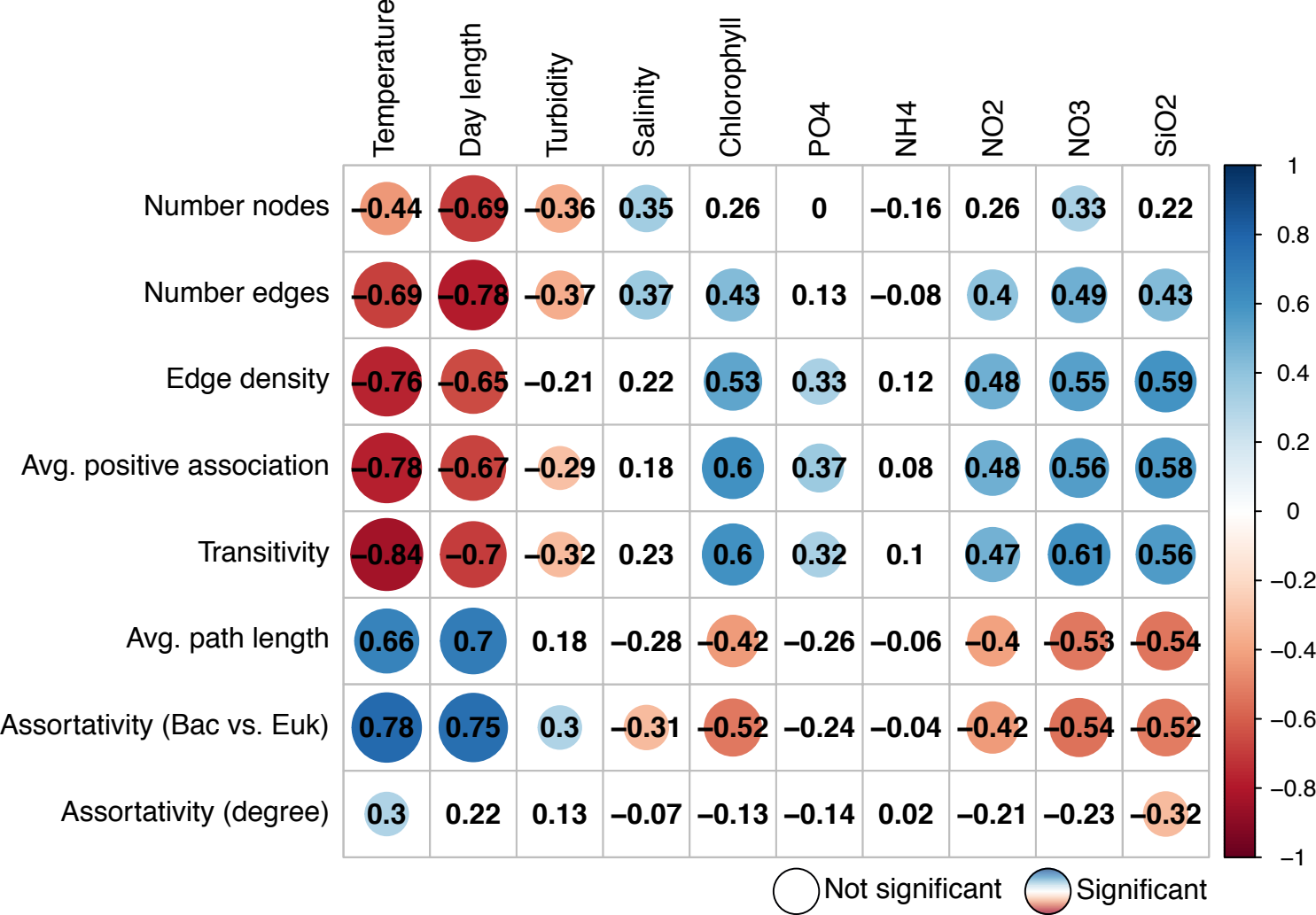
December (4633)

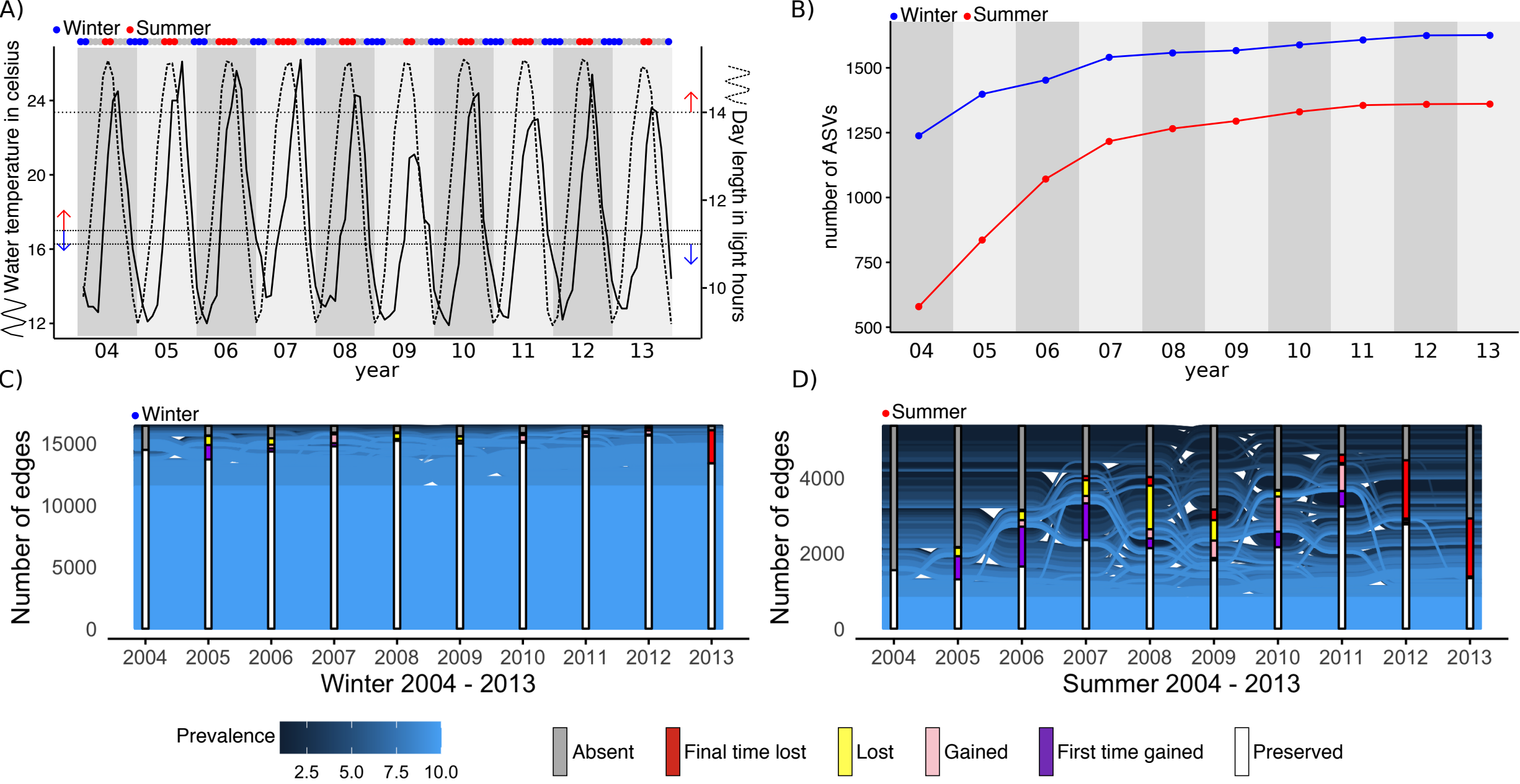
- Actinobacteria
- Actinobacteria_X
- Alphaproteobacteria
- Bacteroidetes
- Chloroflexi
- Cyanobacteria
- Dadabacteria
- Deltaproteobacteria
- Gammaproteobacteria
- Planctomycetes
- Verrucomicrobia
- Centroheliozoa
- Cercozoa
- Chlorophyta
- Ciliophora
- Cryptophyta
- Dinoflagellata
- Katablepharidophyta
- Ochrophyta
- Picozoa
- Radiolaria
- Stramenopiles_X
- Telonemia

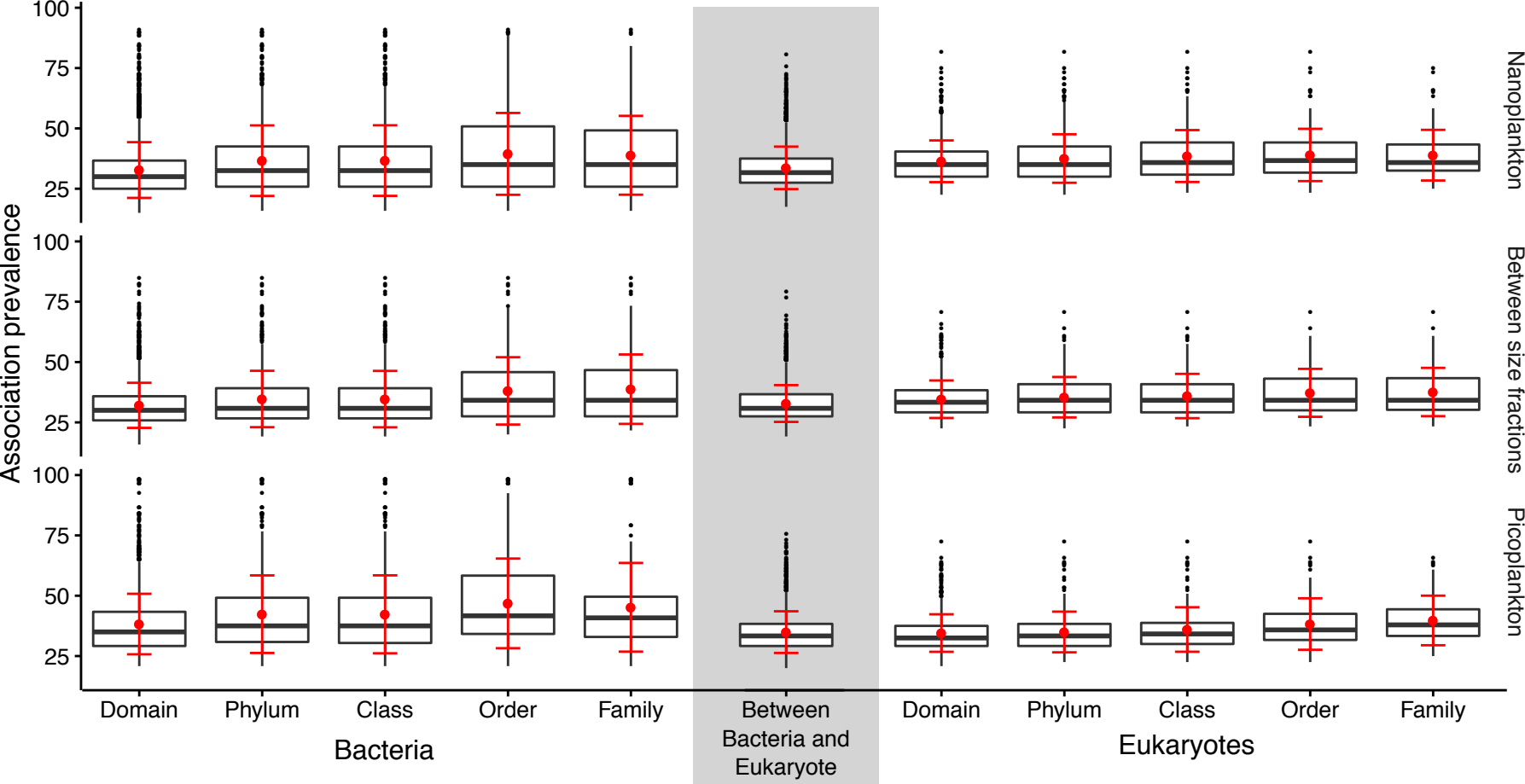

A)
Fraction cyanobacteria edges
0.20
0.15
0.10
0.05
0.00
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
B)
Jan
Mar
May
Single static network
512 of 13502 = 3.79%
284 of 12926 = 2.20%
365 of 3026 = 12.06%
661 of 16626 = 3.98%
cyanobacteria edges
Jul
Sep
Nov
375 of 2397 = 15.64%
414 of 2471 = 16.75%
593 of 10883 = 5.45%
Cyanobium
Prochlorococcus
Synechococcus
Cyanobacteria edges
Other edges

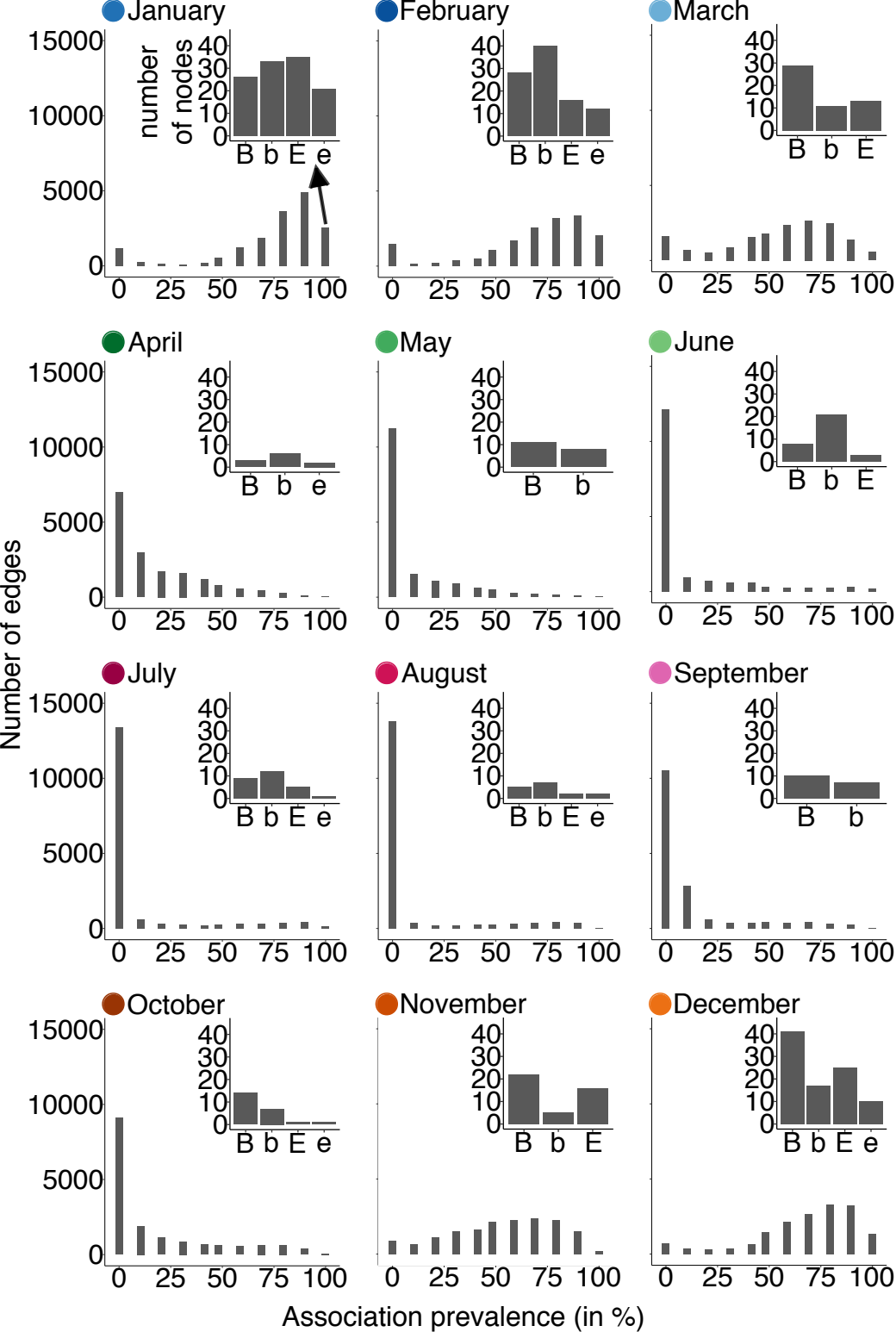


bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452287; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


Temperature
Day length
Turbidity
Salinity
Chlorophyll
PO4
NH4
NO2
NO3
SiO2
Number nodes
Number edges
Edge density
Avg. positive association
Transitivity
Avg. path length
Assortativity (Bac vs. Euk)
Assortativity (degree)
−0.44 −0.69 −0.36 0.35 0.26 0 −0.16 0.26 0.33 0.22
−0.69 −0.78 −0.37 0.37 0.43 0.13 −0.08 0.4 0.49 0.43
−0.76 −0.65 −0.21 0.22 0.53 0.33 0.12 0.48 0.55 0.59
−0.78 −0.67 −0.29 0.18 0.6 0.37 0.08 0.48 0.56 0.58
−0.84 −0.7 −0.32 0.23 0.6 0.32 0.1 0.47 0.61 0.56
0.66 0.7 0.18 −0.28 −0.42 −0.26 −0.06 −0.4 −0.53 −0.54
0.78 0.75 0.3 −0.31 −0.52 −0.24 −0.04 −0.42 −0.54 −0.52
0.3 0.22 0.13 −0.07 −0.13 −0.14 0.02 −0.21 −0.23 −0.32
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
Not significant
Significant


A)
Winter
Summer
Water temperature in celsius
Day length in light hours
year
B)
Winter
Summer
number of ASVs
year
C)
Winter
Number of edges
Winter 2004 - 2013
D)
Summer
Number of edges
Summer 2004 - 2013
Prevalence
2.5
5.0
7.5
10.0
Absent
Final time lost
Lost
Gained
First time gained
Preserved


Association prevalence
100
75
50
25
Nanoplankton
Between size fractions
Picoplankton
Domain
Phylum
Class
Order
Family
Bacteria
Between Bacteria and Eukaryote
Eukaryotes



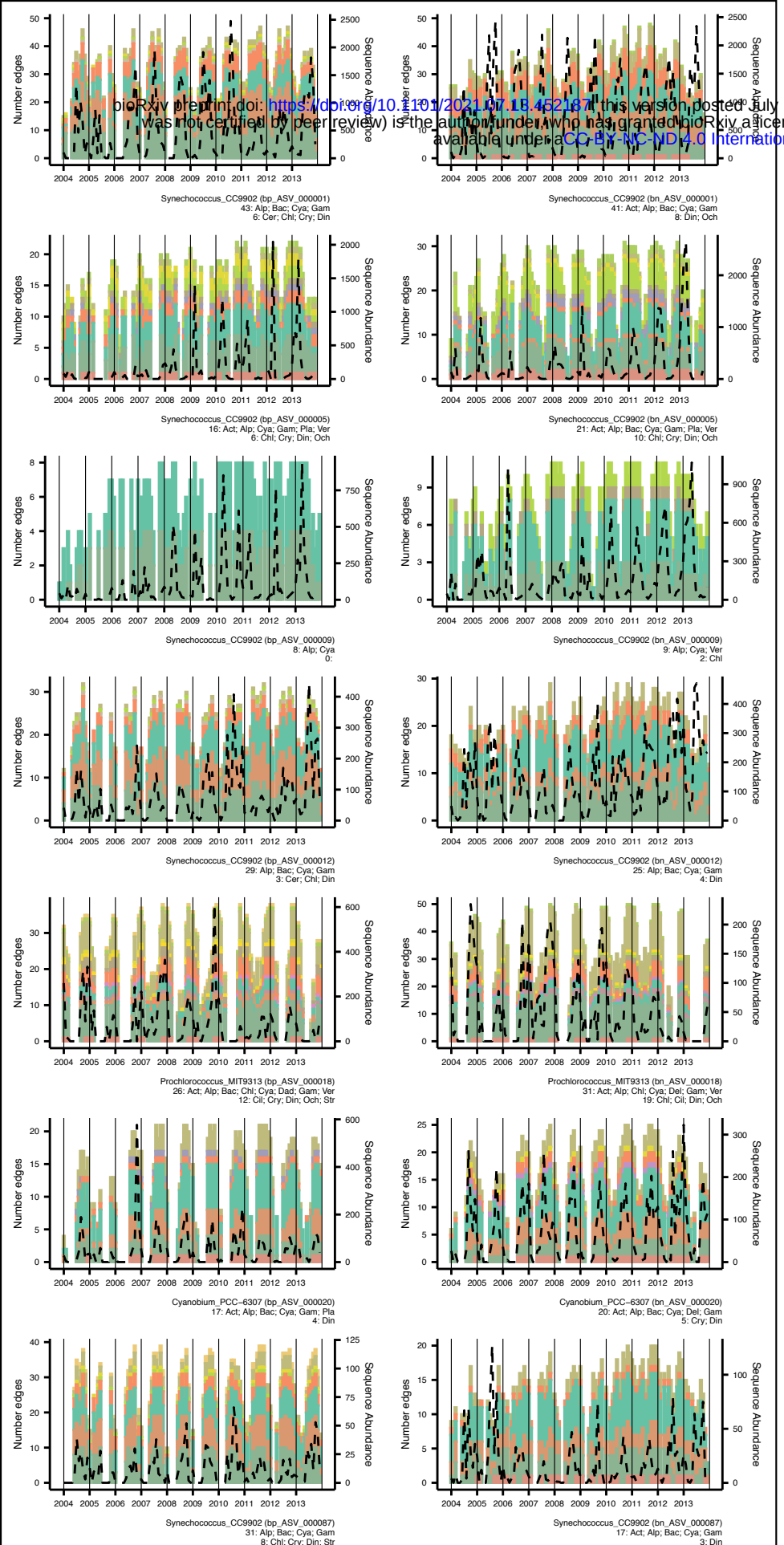
Big barplots:
associations' prevalence within a month.

Little barplots:
number of nodes of associations with a 100% prevalence.

B, b - Bacteria
E, e - Eukaryotes

B, E - Nano size fraction
b,e - Pico size fraction

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452187 this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



Bacteria
- Actinobacteria
- Alphaproteobacteria
- Bacteroidetes
- Chloroflexi
- Cyanobacteria
- Dadabacteria
- Deltaproteobacteria
- Gammaproteobacteria
- Planctomycetes
- Verrucomicrobia

Eukaryote
- Cercozoa
- Chlorophyta
- Ciliophora
- Cryptophyta
- Dinoflagellata
- Ochrophyta
- Stramenopiles_X