Disentangling marine microbial networks across space

Ina Maria Deutschmann[1*], Erwan Delage[2,3], Caterina R. Giner[1], Marta Sebastián[1], Julie Poulain[4], Javier Arístegui[5], Carlos M. Duarte[6], Silvia G. Acinas[1], Ramon Massana[1], Josep M. Gasol[1,7], Damien Eveillard[2,3], Samuel Chaffron[2,3] and Ramiro Logares[1*]

[1]Institute of Marine Sciences, CSIC, Passeig Marítim de la Barceloneta, 37-49, 08003, Barcelona, Spain

[2]Université de Nantes, CNRS UMR 6004, LS2N, F-44000, 2 rue de la Houssinière, 44322, Nantes, France

[3]Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022 / Tara Oceans GOSEE, 3 rue Michel-Ange, 75016 Paris, France

[4]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France

[5]Instituto de Oceanografía y Cambio Global, IOCAG, Universidad de Las Palmas de Gran Canaria, ULPGC, Gran Canaria, Spain

[6] King Abdullah University of Science and Technology (KAUST), Red Sea Research Center (RSRC), Thuwal, Saudi Arabia

[7]Center for Marine Ecosystems Research, School of Sciences, Edith Cowan University, Joondalup, WA Australia

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

Short title: Marine microbial networks across space

ABSTRACT

Although microbial interactions underpin ocean ecosystem functions, they remain barely known. Different studies have analyzed microbial interactions using static association networks based on omics-data. However, microbial associations are dynamic and can change across physicochemical gradients and spatial scales, which needs to be considered to understand the ocean ecosystem better. We explored associations between archaea, bacteria, and picoeukaryotes along the water column from the surface to the deep ocean across the northern subtropical to the southern temperate ocean and the Mediterranean Sea by defining sample-specific subnetworks. Quantifying spatial association recurrence, we found the lowest fraction of global associations in the bathypelagic zone, while associations endemic of certain regions increased with depth. Overall, our results highlight the need to study the dynamic nature of plankton networks and our approach represents a step forward towards a better comprehension of the biogeography of microbial interactions across ocean regions and depth layers.

INTRODUCTION

Microorganisms play fundamental roles in ecosystem functioning (DeLong, 2009; Krabberød *et al.*, 2017) and ocean biogeochemical cycling (Falkowski *et al.*, 2008). The main processes shaping microbial community composition are selection, dispersal, and drift (Vellend, 2020). Selection exerted via environmental conditions and biotic interactions are essential in structuring the ocean microbiome (Logares *et al.*, 2020), leading to heterogeneities reflecting those in the ocean environment, mainly in terms of temperature, light, pressure, nutrients and salinity. In particular, global-scale studies of the surface ocean reported strong associations between microbial community composition and diversity with temperature (Sunagawa *et al.*, 2015; Ibarbalz *et al.*, 2019; Salazar *et al.*, 2019; Logares *et al.*, 2020). Marked changes in microbial communities with ocean depth have also been reported (Cram *et al.*, 2015; Parada & Fuhrman, 2017; Mestre *et al.*, 2018; Peoples *et al.*, 2018; Xu *et al.*, 2018; Giner *et al.*, 2020), reflecting the steep vertical gradients in light, temperature, nutrients and pressure.

Prokaryotes (bacteria and archaea) and unicellular eukaryotes are fundamentally different in terms of ecological roles, functional versatility, and evolutionary history (Massana & Logares, 2013) and are connected through biogeochemical and food web interaction networks (Layeghifard *et al.*, 2017; Seymour *et al.*, 2017). Still, knowledge about these interactions remains limited despite their importance to understand better microbial life in the oceans (Krabberød *et al.*, 2017; Bjorbækmo *et al.*, 2019). Such interactions are very difficult to resolve experimentally, mainly because most microorganisms are hard to cultivate (Baldauf, 2008; Lewis *et al.*, 2020) and synthetic laboratory communities are unlikely to mirror the complexity of wild communities. However, metabarcoding approaches to identify and quantify marine microbial taxa allow to infer association networks, where nodes represent microorganisms and edges potential interactions.

Association networks provide a general overview of the microbial ecosystem aggregated over a given period of time (Steele *et al.*, 2011; Chow *et al.*, 2013, 2014; Cram *et al.*, 2015; Needham *et al.*, 2017; Parada & Fuhrman, 2017) or through space (Lima-Mendez *et al.*, 2015; Milici *et al.*, 2016;

Chaffron *et al.*, 2020). Previous work characterized potential marine microbial interactions, including associations within and across depths. For example, monthly sampling allowed investigating prokaryotic associations in the San Pedro Channel, off the coast of Los Angeles, California, covering the water column from the surface (5 m) to the seafloor (890 m) (Cram *et al.*, 2015; Parada & Fuhrman, 2017). Furthermore, a global spatial survey occurring within the TARA Oceans expedition, allowed to investigate planktonic associations between a range of organismal size fractions in the epipelagic zone, from pole to pole (Lima-Mendez *et al.*, 2015; Chaffron *et al.*, 2020). However, these studies did not include the bathypelagic realm, below 1000 m depth, which represents the largest microbial habitat in the biosphere (Arístegui *et al.*, 2009).

A single static network determined from spatially distributed samples over the global ocean captures global, regional and local associations. Also, given that global-ocean expeditions collect samples over several months, networks could include temporal associations, yet disentangling them from spatial associations is normally complicated and not considered. Global associations may constitute the core interactome, that is, the set of microbial interactions essential for the functioning of the ocean ecosystem (Shade & Handelsman, 2012). Core associations may be detected by constructing a single network from numerous locations and identifying the most significant associations and strongest associations (Coutinho *et al.*, 2015). On the other hand, regional and local associations may point to interactions occurring in specific spatial areas of different sizes due to particular taxa distributions resulting from environmental selection, dispersal limitation, specific ecological niches or biotic/abiotic filtering. The fraction of regional associations may be determined by excluding all samples belonging to one region and recomputing network inference with the reduced dataset (Lima-Mendez *et al.*, 2015). Alternatively, regional networks can be built allowing to determine both, global and regional associations (Mandakovic *et al.*, 2018) by investigating which edges networks have in common and which are unique. Such regional networks could contribute to understanding how the architecture of potential microbial interactions changes with environmental

heterogeneity, also helping to comprehend associations that are stable (i.e., two partners always together) or variable (one partner able to interact with multiple partners across locations).

Regional networks, however, require a high number of samples per delineated zone, but these may not be available due to logistic or budgetary limitations. Recent approaches circumvent this limitation by deriving sample-specific subnetworks from a single static, i.e., all-sample network, which allows quantifying association recurrence over spatiotemporal scales (Chaffron *et al.*, 2020; Deutschmann *et al.*, 2021). Here, we adjusted this approach and used it to determine global and regional associations along vertical and horizontal ocean scales, which allowed us determining the biogeography of marine microbial associations. We analyzed associations between archaea, bacteria, and picoeukaryotes covering the water column, from surface to deep waters, in the Mediterranean Sea (hereafter MS) and five ocean basins: North and South Atlantic Ocean, North and South Pacific Ocean, and Indian Ocean (hereafter NAO, SAO, NPO, SPO, and IO). We estimated microbial taxa abundances using 397 globally distributed samples from the epipelagic to the bathypelagic zone in six ocean regions (Figure 1). We separated most epipelagic samples into surface and deep-chlorophyll maximum (DCM) samples. Next, we constructed a first global network comprising 5457 nodes and 31966 edges, 30657 (95.9%) positive and 1309 (4.1%) negative. Then, we applied a filter strategy including the removal of environmentally-driven edges due to nutrients (4.9% NO_3^-, 4.2% PO_4^{3-}, 2.0% SiO_2), temperature (1.9%), salinity (0.2%), and Fluorescence (0.01%) (Supplementary Table 1). Altogether, our sample-specific network-based exploration allowed us to determine core associations in the global ocean and specific regions, analyze changes in associations and network topology with depth and regions, and to investigate the vertical connectivity of planktonic associations.

RESULTS

From a global static network to sample-specific subnetworks

The resulting global static network contained 5448 nodes and 29118 edges, 28178 (96.8%) positive and 940 (3.2%) negative. It served as the underlying structure from which we generated 397 sample-specific subnetworks following three criteria. First, we required that an edge must be present in the global static network. Second, an edge can only be present within a subnetwork if both microorganisms associated with the edge have a sequence abundance above zero in the corresponding sample. Third, microorganisms associated need to appear together (intersection) in more than 20% of the samples, in which one or both appear (union) for that specific region and depth. This third condition was robust since random subsets retained most associations compared with the associations obtained when using all samples (Supplementary Figure 1). In addition to these three conditions, a node is present in a subnetwork if it has at least one association partner. Consequently, each subnetwork is included in the global static network.

Spatial recurrence

We determined the spatial recurrence of each association using its prevalence computed as the fraction of subnetworks in which a given association was present across the 397 samples (Figure 2A) and within each region-depth-layer combination (Figure 2B). The global ocean surface layer (contributing with 40% of samples) had more associations compared to the other depths (Figure 2B). Remarkably, 14971 of 18234 (82.1%) global ocean surface associations were absent from the MS. In turn, the number of surface associations was similar across ocean basins (Figure 2B).

Considering the most prevalent associations (those found in over 70% of subnetworks), we found that major vertical taxonomic patterns were conserved across regions: the epipelagic layers (surface and DCM) and the two lower layers (meso- and bathypelagic zones) were more similar to each other, respectively (Figure 3). The fraction of associations including *Alphaproteobacteria* was moderate to high in all zones in contrast to *Cyanobacteria* appearing mainly, as expected, in the epipelagic zone (Figure 3). The fraction of *Dinoflagellata* associations was moderate to high in the

139 epipelagic zone and lower in the meso- and bathypelagic zones. While *Dinoflagellata* associations
140 dominated most epipelagic layers, fewer were found in the MS and SAO surface and NAO DCM
141 (Figure 3). *Thaumarchaeota* associations were moderate to high especially in the mesopelagic
142 (dominant in the MS), moderate in the bathypelagic, and lower in the epipelagic zone (Figure 3).
143 Another interesting pattern is the increase in associations including *Gammaproteobacteria* with depth
144 being higher in the meso- and bathypelagic than in the epipelagic, especially in the SAO, SPO, NPO
145 and IO.

146 Highly prevalent associations present across all regions are candidates to represent putative
147 core interactions in the global ocean, which are likely to perform processes crucial for ecosystem
148 functioning. We defined global associations as those appearing in more than 70% of subnetworks in
149 each region. While we found several (21-26) global associations in the epi- and mesopelagic zones, no
150 global associations were identified in the bathypelagic zone (Table 1, Supplementary Figure 2). In
151 addition, we resolved prevalent (>50%) and low-frequency (>20%) associations. These three types of
152 associations are distinct by definition, i.e., a global association cannot be assigned to another type. The
153 fraction of global, prevalent, and low-frequency associations was highest in the DCM layer and lowest
154 in the bathypelagic zone (third and fifth column in Table 1, Supplementary Figure 2B, 2D). Given that
155 the MS bathypelagic is warmer (median temperature of 13.78°C) than the global ocean bathypelagic
156 (median temperature between 1.4°C in SPO and 4.41°C in NAO), we calculated these associations for
157 the global ocean only. We found slightly to moderately more global, prevalent, and low-frequency
158 associations in the global ocean when not considering the MS (fifth to seventh row in Table 1,
159 Supplementary Figure 2E-H).

160 Next, we determined regional associations within each depth layer. A regional association was
161 defined as detected in at least one sample-specific subnetwork of one region and absent from all
162 subnetworks of the other five regions. Results indicated an increasing proportion of regional
163 associations with depth (Table 1, Figure 4A-B, Supplementary Figure 3). We found substantially more

associations in the DCM and mesopelagic layers of the MS than corresponding layers of the global ocean. This may reflect the different characteristics of these layers in the MS vs. the global ocean or the massive differences in spatial dimensions between the global ocean and the MS. More surface and bathypelagic regional associations corresponded to the MS and NAO than in other regions (Table 1). Most regional associations had low prevalence, i.e., they were present in a few sample-specific subnetworks within the region (Figure 4C). We found 235 prokaryotic highly prevalent (>70%) regional associations in contrast to 89 eukaryotic and 24 associations between domains (Supplementary Material 1).

Previous studies have found a substantial vertical connectivity in the ocean microbiota, with surface microorganisms having an impact in deep sea counterparts (Mestre *et al.*, 2018; Ruiz-González *et al.*, 2020). Thus, here, we analyzed the vertical connectivity of microbial associations. Few associations appeared throughout the water column within a region: 327 prokaryotic, 119 eukaryotic, and 13 associations between domains (Supplementary Material 2). In general, most associations appearing in the meso- and bathypelagic did not appear in upper layers except for the MS and NAO where most and about half, respectively, of the bathypelagic associations already appeared in the mesopelagic (Figure 5). Specifically, 81.77 – 90.90% mesopelagic and 43.54-72.71% bathypelagic associations appeared for the first time in the five ocean basins (Supplementary Table 2). In the MS, 71.24% mesopelagic and 22.44% bathypelagic associations appeared for the first time and 69.71% of bathypelagic associations already appeared in the mesopelagic (Supplementary Table 2). This points to specific microbial interactions occurring in the deep ocean that do not occur in upper layers. In addition, while most surface associations also appeared in the DCM in the MS, most surface associations disappeared with depth in the five ocean basins (Figure 5) suggesting that most surface ocean associations are not transferred to the deep sea, despite microbial sinking (Mestre *et al.*, 2018). In fact, we observed that most deep ocean ASVs already appeared in the upper layers (Supplementary Figure 4), in agreement with previous work that has shown that a large proportion of deep sea microbial

taxa are also found in surface waters, and that their presence in the deep sea is related to sinking processes (Mestre *et al.*, 2018).

Comparing subnetworks

Vertical and horizontal spatial variability is expected to affect network topology via biotic and abiotic variables as well as through dispersal processes (e.g., dispersal limitation). Yet, we have a limited understanding on how much marine microbial networks change due to these processes, thus analyzing the topology of subnetworks from specific ocean regions and depths is a first step to address this question. We compared the subnetworks of the six regions and depth layers using eight global network metrics (see Methods). We found that global network metrics change along the water column (Supplementary Figure 5). As a general trend, subnetworks from deeper zones were more clustered (transitivity) with higher average path length, stronger associations (average positive association scores) and lower assortativity (based on degree) compared to those in surface waters. Most DCM and bathypelagic subnetworks had the highest connectivity (edge density). Contrarily, in the MS, the surface subnetworks had the highest connectivity (Supplementary Figure 5).

To avoid predefined groupings into regions and depth layers, we grouped similar subnetworks via a local network metric (see Methods) and identified 36 clusters of 5 to 28 subnetworks (Supplementary Table 3). We found 13 (36.1%) clusters that were dominated by surface subnetworks: six clusters (100% surface subnetworks) from three to five oceans but not MS and seven clusters with 55-86% surface networks from two to five of the six ocean regions. In turn, 11 clusters were dominated by a deeper layer: two DCM (64-90%), five mesopelagic (62-83%) and four bathypelagic dominated clusters (60-69%). Nine of these 11 clusters combined different regions except for one mesopelagic and one bathypelagic dominated cluster representing exclusively the MS (Supplementary Table 3). Furthermore, we found 11 clusters containing exclusively or mainly MS subnetworks in contrast to only one cluster dominated by an ocean basin (NAO).

Next, we built a more comprehensive representation of network similarities between subnetworks via a minimal spanning tree (MST, see Methods) to underline the pervasive connectivity of associations across depth and environmental gradients. The depth layers, ocean regions, location of clusters, and environmental factors were projected onto the MST (Figure 6). Most surface subnetworks were centrally located, while subnetworks from other depths appeared in different MST areas. Most MS subnetworks were located in a specific branch of the MST, while the five oceans were mixed, indicating homogeneity within oceans but network-based differences between the oceans and the MS. However, subnetworks in the MST tended to connect to subnetworks from the same depth layer, cluster or similar environmental conditions. All in all, the above results suggest a strong influence of environmental gradients in shaping network topology and plankton associations, as previously observed in epipelagic communities at global scale (Chaffron *et al.*, 2020).

DISCUSSION

In this work, we disentangled and analyzed global and regional microbial associations across the oceans' vertical and horizontal dimensions. We found a low number of global associations indicating a potentially small global core interactome within each depth layer across six oceanic regions. Core microorganisms are often defined as those appearing in most or all samples from similar habitats (Shade & Handelsman, 2012). We previously identified a core microbiota in a coastal MS observatory based on both association patterns (Krabberød *et al.*, 2021) and temporal recurrence of associations (Deutschmann *et al.*, 2021). Both studies indicate more robust microbial connectivity, suggesting a broader core, in colder than in warmer seasons. In contrast, within each region, we found less highly prevalent associations in the bathypelagic zone of the global ocean (pointing to a smaller regional core) than in the upper layers, except from the NPO, having less highly prevalent associations in the meso- than in the bathypelagic. In agreement, we found more regional bathypelagic associations than in upper layers. Thus, associations may reflect the heterogeneity and isolation of the deep ocean regions due to

239 deep currents, water masses, or the topography of the seafloor that may prevent microbial dispersal.
240 Moreover, the higher complexity of the deep ocean ecosystem may provide a higher number of
241 ecological niches potentially resulting in more regional associations and agreeing with our
242 observations. A high diversification of niches may be associated to the different quality and types
243 (labile, recalcitrant, etc.) of organic matter reaching the deep ocean from the epipelagic zone (Arístegui
244 *et al.*, 2009), which is significantly different across oceanic regions (Hansell & Carlson, 1998). In an
245 exploration of generalists versus specialist prokaryotic metagenome-assembled genomes (MAGs) in
246 the arctic Ocean, most of the specialists were linked to mesopelagic samples indicating that their
247 distribution was uneven across depth layers (Royo-Llonch *et al.*, 2020). This is in agreement with
248 putatively more niches in the deep ocean than in upper ocean layers leading to more specialist taxa and
249 subsequently more regional associations.

250 Vertical connectivity in the ocean microbiome is partially modulated by surface productivity
251 through sinking particles (Mestre *et al.*, 2018; Boeuf *et al.*, 2019; Ruiz-González *et al.*, 2020). An
252 analysis of eight stations, distributed across the Atlantic, Pacific and Indian oceans (including 4 depths:
253 Surface, DCM, meso- and bathypelagic), indicated that bathypelagic communities comprise both
254 endemic taxa as well as surface-related taxa arriving via sinking particles (Mestre *et al.*, 2018). Ruiz-
255 González *et al.* (Ruiz-González *et al.*, 2020) identified for both components (i.e. surface-related and
256 deep-endemic) the dominating phylogenetic groups: while *Thaumarchaeota*, *Deltaproteobacteria*,
257 *OM190* (*Planctomycetes*) and *Planctomycetacia* (*Planctomycetes*) dominated the endemic
258 bathypelagic communities, *Actinobacteria*, *Alphaproteobacteria*, *Gammaproteobacteria* and
259 *Flavobacteriia* (*Bacteroidetes*) dominated the surface-related taxa in the bathypelagic zone. We found
260 association partners for each dominating phylogenetic group within each investigated type of
261 association, i.e., highly prevalent, regional, global, prevalent, and low-frequency associations. While
262 ASVs belonging to these taxonomic groups were present throughout the water column, specific
263 associations were observed especially in the mesopelagic and bathypelagic zones, which suggests

specific associations between deep-sea endemic taxa. This is in agreement with a recent study that found a remarkable taxonomic novelty in the deep ocean after analyzing 58 microbial metagenomes from global samples, unveiling ~68% archaeal and ~58% bacterial novel species (Acinas *et al.*, 2021).

Less is known about associations found along the entire or a substantial fraction of the water column, suggesting consortia of associated microorganisms that sink together or that populate large vertical ranges of the water column. Associations present across all layers were few but may represent interacting taxa that populate the entire water column or that sink together. However, given that we targeted mainly picoplankton, we would not expect a considerable influence of sinking particles in the vertical distribution of associations in this study. Some associations observed in the deep ocean may correspond to consortia of taxa degrading sinking particles, or taxa that might have detached from sinking particles, i.e., dual life-style taxa as observed in (Sebastián, Sánchez, *et al.*, 2021). Alternatively, microorganisms may have reached bathypelagic waters via fast-sinking processes, embedded in (larger) particles (Agusti *et al.*, 2015). By following this observation, a previous study found that the abundances of microorganisms in deeper layers mirrored the changes in abundance of microorganisms in shallower layers, at a single sampling station, indicating that communities populating different ocean depths are not isolated from each other but linked, possibly through sinking particles or migrating organisms transporting nutrients through the water column (Cram *et al.*, 2015). However, microbial co-occurrence alone does not suffice to infer microbial interactions, because different mechanisms, such as selection or dispersal, influence species as well as their interactions (Poisot *et al.*, 2012). Our results suggest that microorganisms can potentially change their interaction partners along vertical (and horizontal) scales and, to a lesser extent, maintain interactions along the water column.

A study of global-ocean picoplanktonic eukaryotes through the water column (from the Epi- to the Bathypelagic zone) found the highest and lowest relative metabolic activity for most eukaryotes in the meso- and bathypelagic zones, respectively (Giner *et al.*, 2020). Thus, we could hypothesize more

competition in the mesopelagic zone and more beneficial interactions in the bathypelagic zone. In our study, mesopelagic subnetworks displayed the lowest connectivity in most regions on average, and we found the strongest associations among both meso- and bathypelagic subnetworks. Moreover, we found the highest clustering (transitivity) in the meso- and bathypelagic zones (relatively colder waters) compared to the epipelagic zone (warmer waters). Similarly, a previous global-scale study (Chaffron *et al.*, 2020) concentrating on the epipelagic zone and including polar waters, found higher edge density, association strength and clustering in polar (colder waters) compared to warmer waters. These results suggest that either microorganisms interact more in colder and darker environments or that their recurrence is higher due to a higher environmental selection exerted by low temperatures and no light. Alternatively, limited resources (primarily nutrients) in the surface versus deep ocean may prevent the establishment of specific microbial interactions. Furthermore, another explanation could be the higher diversity of ecological niches and, thus, a higher diversity of associations in the meso- and bathypelagic.

Through quantifying regional associations, our results indicated distinct associations in the MS, where most regional associations were observed compared to the global ocean, as previously shown in an epipelagic network (Lima-Mendez *et al.*, 2015). Furthermore, we found a substantial number of regional associations in the NAO compared to other ocean basins, contrasting with the NAO having the lowest number of regional associations in a previous epipelagic network (Lima-Mendez *et al.*, 2015).

To conclude, our network-based exploration disentangles the spatial distribution of associations of the global ocean microbiome, from top to bottom layers, suggesting both global and regional interactions. Our analysis demonstrated the change of network topology across vertical (water column) and horizontal (different regions) dimensions of the ocean. Furthermore, our results indicate that associations have specific spatial distributions that are not just mirroring ASV distributions.

METHODS

Dataset

Samples originated from two expeditions, Malaspina-2010 (Duarte, 2015) and Hotmix (Martínez-Pérez *et al.*, 2017). The former was onboard the R/V Hespérides and most ocean basins were sampled between December 2010 and July 2011. Malaspina samples included i) *MalaSurf*, surface samples (Ruiz-González *et al.*, 2019; Logares *et al.*, 2020), ii) *MalaVP*, vertical profiles (Giner *et al.*, 2020), and iii) *MalaDeep*, deep-sea samples, (Pernice *et al.*, 2016; Salazar *et al.*, 2016; Sanz-Sáez, 2021). For the Hotmix expedition, sampling took place onboard the R/V Sarmiento de Gamboa between 27th April and 29th May 2014 and represented a quasi-synoptic transect across the MS and the adjacent North-East of the NAO. See details in Table 2.

DNA extractions are indicated in the papers associated with each dataset (Table 2). From the DNA extractions, the 16S and 18S rRNA genes were amplified and sequenced. PCR amplification and sequencing of *MalaSurf, MalaVP* (18S), and *Hotmix* (16S) are indicated in the papers associated with each dataset in Table 2. *MalaVP* (16S) and *Hotmix* (18S) were PCR-amplified and sequenced following the same approach as in (Logares *et al.*, 2020). *MalaDeep* samples were obtained from (Pernice *et al.*, 2016; Salazar *et al.*, 2016) but re-sequenced in Genoscope (France) with different primers, as described below. *MalaSurf, MalaVP* and *Hotmix* datasets were sequenced at RTL Genomics (Texas, USA).

We used the same amplification primers for all samples. For the 16S, we amplified the V4-V5 hypervariable region using the primers 515F-Y and 926R (Parada *et al.*, 2016). For the 18S, we amplified the V4 hypervariable region with the primers TAReukFWD1 and TAReukREV3 (Stoeck *et al.*, 2010). See more details in (Logares *et al.*, 2020). Amplicons were sequenced in *Illumina* MiSeq or HiSeq2500 platforms (2x250 or 2x300 bp reads). Operational Taxonomic Units were delineated as Amplicon Sequence Variants (ASVs) using DADA2 (Callahan *et al.*, 2016), running each dataset separately before merging the results. ASVs were assigned taxonomy using SILVA (Quast *et al.*,

2012), v132, for prokaryotes, and PR2 (Guillou *et al.*, 2012), v4.11.1, for eukaryotes. ASVs corresponding to Plastids, Mitochondria, Metazoa, and Plantae, were removed. Only samples with at least 2000 reads were kept. The dataset contained several *MalaDeep* replicates, which we merged, and two filter sizes: given the cell sizes of prokaryotes versus microeukaryotes, we selected the smallest available filter size (0.2-0.8 µm) for prokaryotes and the larger one (0.8-20 µm) for microeukaryotes. The other three datasets used filter sizes of 0.2-3 µm. Additionally, we required that samples had eukaryotic and prokaryotic data, resulting in 397 samples for downstream analysis: 122 *MalaSurf*, 83 *MalaVP*, 13 *MalaDeep*, and 179 *Hotmix*. We separated the samples into epipelagic, mesopelagic and bathypelagic zone (Figure 1). Furthermore, we separated most epipelagic samples into surface and deep-chlorophyll maximum (DCM) samples, but 18 MS and 4 NAO samples belonged to neither. We also considered environmental variables: Temperature (2 missing values = mv), salinity (2 mv), fluorescence (3 mv), and inorganic nutrients NO_3^- (36 mv), PO_4^{3-} (38 mv), and SiO_2 (37 mv), which were measured as indicated elsewhere (Giner *et al.*, 2020; Logares *et al.*, 2020; Sebastián, Ortega-Retuerta, *et al.*, 2021). In specific samples, missing data on nutrient concentrations were estimated from the World Ocean Database (Boyer *et al.*, 2013).

Single static network

We constructed the single static network in four steps. First, we prepared the data for network construction. We excluded rare microorganisms by keeping ASVs with a sequence abundance sum above 100 reads and appearing in at least 20 samples (>5%). The latter condition removes bigger eukaryotes only appearing in the 13 *MalaDeep* eukaryotic samples of a bigger size fraction. To control for data compositionality (Gloor *et al.*, 2017), we applied a centered-log-ratio transformation separately to the prokaryotic and eukaryotic tables before merging them.

Second, we inferred a (preliminary) network using FlashWeave (Tackmann *et al.*, 2019), selecting the options “heterogeneous” and “sensitive”. FlashWeave was chosen as it can handle sparse

datasets like ours, taking zeros into account and avoiding spurious correlations between ASVs that share many zeros.

Third, we aimed to remove environmentally-driven edges. FlashWeave could detect indirect edges and allows to supply additional metadata such as environmental variables, but currently does not support missing data. Thus, we applied EnDED (Deutschmann et al. 2020), combining the methods Interaction Information (with 0.05 significance threshold and 10000 iterations) and Data Processing Inequality as done previously via artificially-inserted edges to connect all microbial nodes to the six environmental parameters (Deutschmann *et al.*, 2021). Although EnDED can handle missing environmental data when calculating intermediate values relating ASV and environmental factors, it would compute intermediate values for microbial edges using all samples. Thus, to avoid a possible bias and speed up the calculation process, we applied EnDED individually for each environmental factor, using only the samples containing values for the specific environmental factor.

Fourth, we removed isolated nodes, i.e., nodes without any edge. The resulting network represented the single static network in our study.

Sample-specific subnetwork

We constructed 397 sample-specific subnetworks. Each subnetwork represented one sample and was derived from the single static network, i.e., a subnetwork contained nodes and edges present in the single static network but not vice versa. Consider sample s_{RL} with R being the marine region, and L the sample's depth layer. Let e be an association between microorganisms A and B. Then, association e is present in the sample-specific subnetwork N_s, if

i. e is an association in the single static network,
ii. the microorganisms A and B are present within sample s, i.e., the abundances are above zero within that particular sample, and
iii. the association has a region and depth specific Jaccard index, J_{RL}, above 20% (see below).

In addition to these three conditions, a node is present in a sample-specific subnetwork when connected to at least one edge, i.e., we removed isolated nodes.

Regarding the third condition, we determined J_{RL} for each association pair by computing within each region and depth layer, the fraction of samples two microorganisms appeared together (intersection) from the total samples at least one microorganism appears (union). Supplementary Table 4 shows the number of edges using different thresholds. Given the heterogeneity of the dataset within regions and depth layers, we decided to use a low threshold, keeping edges with a Jaccard index above 20% and removed edges below or equal to 20%. We tested robustness by randomly drawing a subset of samples from each region and depth combination. The subset contained between 10% and 90% of the original samples. We rounded up decimal numbers to avoid zero sample subsets, e.g., 10% of 7 samples results in a subset of 1 sample. We excluded the DCM of the SPO because it contained only one sample. Next, we recomputed the Jaccard index for the random subset. Lastly, requiring J>20%, we evaluated robustness determining i) how many edges were kept in the random subsamples compared to all samples, and ii) how many edges were kept in the random subset that were also kept when all samples were used. We repeated the procedure for each region-depth combination 1000 times.

Spatial recurrence

To determine an association's spatial recurrence, we calculated its prevalence as the fraction of subnetworks in which the association was present. We determined association prevalence across the 397 samples and each region-layer combination. We mapped the scores onto the single static network, visualized in Gephi (Bastian *et al.*, 2009), v.0.9.2, using the Fruchterman Reingold Layout (Fruchterman & Reingold, 1991) with a low gravity score of 0.5. We used the region-layer prevalence to determine global and regional associations. We considered an association to be global within a specific depth layer if its prevalence was above 70% in all regions. In turn, a regional association had an association prevalence above 0% within a particular region-layer (present, appearing in at least one

subnetwork) and 0% within other regions of the same layer (absent, appearing in no subnetwork). In addition, associations that are not global but appear in all regions over 50% are considered prevalent. Similarly, associations that are not global nor prevalent but appear in all regions over 20% are considered low-frequency. Thus, an association can be classified as i) global, ii) regional, iii) prevalent, iv) low-frequency, and v) "other", i.e., associations that have not been classified into the previous categories.

Global network metrics

We considered the *number of nodes* and *edges* and six other global network metrics of which most were computed with functions of the igraph R-package (Csardi & Nepusz, 2006). *Edge density* indicating connectivity is computed through the number of actual edges divided by the number of possible edges. The *average path length* is the average length of all shortest paths between nodes in a network. *Transitivity* indicating how well a network is clustered is the probability that the nodes' neighbors are connected. *Assortativity* measures if similar nodes tend to be connected, i.e., *assortativity (degree)* is positive if high degree nodes tend to connect to other high degree nodes and negative otherwise. Similarly, *assortativity (Euk-Prok)* is positive if eukaryotes tend to connect to other eukaryotes and prokaryotes tend to connect to other prokaryotes. Lastly, we computed the *average positive association strength* as the mean of all positive association scores provided by FlashWeave.

Local network metric

The previous global metrics disregard local structures' complexity, and topological analyses should include local metrics (Espejo *et al.*, 2020), e.g., graphlets (Pržulj *et al.*, 2004). Here, we determined network-dissimilarity between each pair of sample-specific subnetworks as proposed in (Yaveroğlu *et al.*, 2014), comparing network topology without considering specific ASVs. The network-dissimilarity

is a distance measurement that is always positive: 0 if networks are identical and greater numbers indicate greater dissimilarity.

Next, we constructed a Network Similarity Network (NSN), where each node is a subnetwork and each node connects with all other nodes, i.e., the NSN was a complete graph. We assigned the network-dissimilarity score as edge weight within the NSN. To simplify the NSN while preserving its main patterns, we determined the minimal spanning tree (MST) of the NSN. The MST had 397 nodes and 396 edges. The MST is a backbone, with no circular path, in which the edges are chosen so that the edge weights sum is minimal and all nodes are connected, i.e., a path exists between any two nodes. We determined the MST using the function *mst* in the igraph package in R (Prim, 1957; Csardi & Nepusz, 2006).

Using the network-dissimilarity (distance) matrix, we determined clusters of similar subnetworks in python. First, we reduced the matrix to ten dimension using *umap* (McInnes *et al.*, 2018) with the following parameter settings: n_neighbors=3, min_dist=0, n_components=10, random_state=123, and metric='precomputed'. Second, we clustered the subnetworks (represented via ten dimensions) with *hdbscan* (McInnes *et al.*, 2017) setting the parameters to min_samples=3 and min_clusters=5.

Reproducibility

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

Acknowledgements

We thank all members of the Malaspina and Hotmix expeditions with the multiple projects funding these collaborative efforts. Sampling was carried out thanks to the Consolider-Ingenio programme (project Malaspina 2010 Expedition, ref. CSD2008–00077) and HOTMIX project (CTM2011-30010/MAR), funded by the Spanish Ministry of Economy and Competitiveness Science and Innovation. Part of the analyses have been performed at the Marbits bioinformatics core at ICM-CSIC (https://marbits.icm.csic.es). 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. SC was supported by the CNRS MITI through the interdisciplinary program Modélisation du Vivant (GOBITMAP grant). SC, DE and SGA were funded by the H2020 project AtlantECO (award number 862923).

Author's contributions

The overall project was conceived and designed by RL. JMG, CMD, SGA, RM, JA were responsible for the sampling and acquisition of contextual data. CRG, JP and MS processed specific samples in the laboratory. RL processed the amplicon data generating the two ASV tables. They were 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, MS, CMD, SGA, RM, JMG, DE, SC, and RL contributed with interpretation of the results. IMD wrote the original draft. All authors contributed to manuscript revisions and approved the final version of the manuscript.

Competing interests: The authors declare that they have no competing interests.

REFERENCES

ACINAS, S.G., SÁNCHEZ, P., SALAZAR, G., CORNEJO-CASTILLO, F.M., SEBASTIÁN, M., LOGARES, R., ROYO-LLONCH, M., PAOLI, L., SUNAGAWA, S., HINGAMP, P., OGATA, H., LIMA-MENDEZ, G., ROUX, S., GONZÁLEZ, J.M., ARRIETA, J.M., ALAM, I.S., KAMAU, A., BOWLER, C., RAES, J., PESANT, S., BORK, P., AGUSTÍ, S., GOJOBORI, T., VAQUÉ, D., SULLIVAN, M.B., PEDRÓS-ALIÓ, C., MASSANA, R., DUARTE, C.M., & GASOL, J.M. (2021) Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Communications Biology*, **4**, 604.

AGUSTI, S., GONZÁLEZ-GORDILLO, J.I., VAQUÉ, D., ESTRADA, M., CEREZO, M.I., SALAZAR, G., GASOL, J.M., & DUARTE, C.M. (2015) Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. *Nature Communications*, **6**, 7608.

ARÍSTEGUI, J., GASOL, J.M., DUARTE, C.M., & HERNDLD, G.J. (2009) Microbial oceanography of the dark ocean's pelagic realm. *Limnology and Oceanography*, **54**, 1501–1529.

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

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.

BOEUF, D., EDWARDS, B.R., EPPLEY, J.M., HU, S.K., POFF, K.E., ROMANO, A.E., CARON, D.A., KARL, D.M., & DELONG, E.F. (2019) Biological composition and microbial dynamics of sinking particulate organic matter at abyssal depths in the oligotrophic open ocean. *Proc Natl Acad Sci USA*, **116**, 11824.

BOYER, T.P., ANTONOV, J.I., BARANOVA, O.K., GARCIA, H.E., JOHNSON, D.R., MISHONOV, A.V., O'BRIEN, T.D., SEIDOV, D., 1948-, SMOLYAR, I. (Igor), ZWENG, M.M., PAVER, C.R., LOCARNINI, R.A., REAGAN, J.R., FORGY, C. (Carla), GRODSKY, A., & LEVITUS, S. (2013) World ocean database 2013. NOAA atlas NESDIS ; 72, DOI: 10.7289/V5NZ85MT.

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.

COUTINHO, F.H., MEIRELLES, P.M., MOREIRA, A.P.B., PARANHOS, R.P., DUTILH, B.E., & THOMPSON, F.L. (2015) Niche distribution and influence of environmental parameters in marine microbial communities: a systematic review. *PeerJ*, **3**, e1008.

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.

534 Csardi, G. & Nepusz, T. (2006) The igraph software package for complex network research.
535 *InterJournal*, **Complex Systems**, 1695.
536 DeLong, E.F. (2009) The microbial ocean from genomes to biomes. *Nature*.
537 Deutschmann, I., Krabberød, A.K., Benites, L.F., Latorre, F., Delage, E., Marrasé, C., Balagué, V., Gasol,
538 J.M., Massana, R., Eveillard, D., Chaffron, S., & Logares, R. (2021) Disentangling temporal
539 associations in marine microbial networks. *Research Square*, DOI: 10.21203/rs.3.rs-
540 404332/v1.
541 Duarte, C.M. (2015) Seafaring in the 21St Century: The Malaspina 2010 Circumnavigation Expedition.
542 *Limnology and Oceanography Bulletin*, **24**, 11–14.
543 Espejo, R., Mestre, G., Postigo, F., Lumbreras, S., Ramos, A., Huang, T., & Bompard, E. (2020) Exploiting
544 graphlet decomposition to explain the structure of complex networks: the GHuST framework.
545 *Scientific Reports*, **10**, 12884.
546 Falkowski, P.G., Fenchel, T., & Delong, E.F. (2008) The Microbial Engines That Drive Earth's
547 Biogeochemical Cycles. *Science*.
548 Fruchterman, T.M.J. & Reingold, E.M. (1991) Graph drawing by force-directed placement. *Software:*
549 *Practice and Experience*, **21**, 1129–1164.
550 Giner, C.R., Pernice, M.C., Balagué, V., Duarte, C.M., Gasol, J.M., Logares, R., & Massana, R. (2020)
551 Marked changes in diversity and relative activity of picoeukaryotes with depth in the world
552 ocean. *The ISME Journal*, **14**, 437–449.
553 Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., & Egozcue, J.J. (2017) Microbiome Datasets Are
554 Compositional: And This Is Not Optional. *Frontiers in Microbiology*, **8**, 2224.
555 Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C.,
556 Decelle, J., del Campo, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra,
557 W.H.C.F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R.,
558 Not, F., Pawlowski, J., Probert, I., Sauvadet, A.-L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann,
559 P., & Christen, R. (2012) The Protist Ribosomal Reference database (PR^2): a catalog of
560 unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids*
561 *Research*, **41**, D597–D604.
562 Hansell, D.A. & Carlson, C.A. (1998) Deep-ocean gradients in the concentration of dissolved organic
563 carbon. *Nature*, **395**, 263–266.
564 Ibarbalz, F.M., Henry, N., Brandão, M.C., Martini, S., Busseni, G., Byrne, H., Coelho, L.P., Endo, H., Gasol,
565 J.M., Gregory, A.C., Mahé, F., Rigonato, J., Royo-Llonch, M., Salazar, G., Sanz-Sáez, I., Scalco,
566 E., Soviadan, D., Zayed, A.A., Zingone, A., Labadie, K., Ferland, J., Marec, C., Kandels, S., Picheral,
567 M., Dimier, C., Poulain, J., Pisarev, S., Carmichael, M., Pesant, S., Acinas, S.G., Babin, M., Bork,
568 P., Boss, E., Bowler, C., Cochrane, G., Vargas, C. de, Follows, M., Gorsky, G., Grimsley, N., Guidi,
569 L., Hingamp, P., Iudicone, D., Jaillon, O., Kandels, S., Karp-Boss, L., Karsenti, E., Not, F., Ogata,
570 H., Pesant, S., Poulton, N., Raes, J., Sardet, C., Speich, S., Stemmann, L., Sullivan, M.B.,
571 Sunagawa, S., Wincker, P., Babin, M., Boss, E., Iudicone, D., Jaillon, O., Acinas, S.G., Ogata, H.,
572 Pelletier, E., Stemmann, L., Sullivan, M.B., Sunagawa, S., Bopp, L., Vargas, C. de, Karp-Boss, L.,
573 Wincker, P., Lombard, F., Bowler, C., & Zinger, L. (2019) Global Trends in Marine Plankton
574 Diversity across Kingdoms of Life. *Cell*, **179**, 1084-1097.e21.
575 Krabberød, A.K., Bjorbækmo, M.F.M., Shalchian-Tabrizi, K., & Logares, R. (2017) Exploring the oceanic
576 microeukaryotic interactome with metaomics approaches. *Aquatic Microbial Ecology*, **79**, 1–
577 12.
578 Krabberød, A.K., Deutschmann, I.M., Bjorbækmo, M.F.M., Balagué, V., Giner, C.R., Ferrera, I., Garcés,
579 E., Massana, R., Gasol, J.M., & Logares, R. (2021) Long-term patterns of an interconnected
580 core marine microbiota. *bioRxiv*, 2021.03.18.435965.

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.

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.

MANDAKOVIC, D., ROJAS, C., MALDONADO, J., LATORRE, M., TRAVISANY, D., DELAGE, E., BIHOUÉE, A., JEAN, G., DÍAZ, F.P., FERNÁNDEZ-GÓMEZ, B., CABRERA, P., GAETE, A., LATORRE, C., GUTIÉRREZ, R.A., MAASS, A., CAMBIAZO, V., NAVARRETE, S.A., EVEILLARD, D., & GONZÁLEZ, M. (2018) Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. *Scientific Reports*, **8**, 5875.

MARTÍNEZ-PÉREZ, A.M., OSTERHOLZ, H., NIETO-CID, M., ÁLVAREZ, M., DITTMAR, T., & ÁLVAREZ-SALGADO, X.A. (2017) Molecular composition of dissolved organic matter in the Mediterranean Sea. *Limnology and Oceanography*, **62**, 2699–2712.

MASSANA, R. & LOGARES, R. (2013) Eukaryotic versus prokaryotic marine picoplankton ecology. *Environmental Microbiology*, **15**, 1254–1261.

MCINNES, L., HEALY, J., & ASTELS, S. (2017) hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, **2**, 205.

MCINNES, L., HEALY, J., SAUL, N., & GROSSBERGER, L. (2018) UMAP: Uniform Manifold Approximation and Projection. *The Journal of Open Source Software*, **3**, 861.

MESTRE, M., RUIZ-GONZÁLEZ, C., LOGARES, R., DUARTE, C.M., GASOL, J.M., & SALA, M.M. (2018) Sinking particles promote vertical connectivity in the ocean microbiome. *Proc Natl Acad Sci USA*, **115**, E6799.

MILICI, M., DENG, Z.-L., TOMASCH, J., DECELLE, J., WOS-OXLEY, M.L., WANG, H., JÁUREGUI, R., PLUMEIER, I., GIEBEL, H.-A., BADEWIEN, T.H., WURST, M., PIEPER, D.H., SIMON, M., & WAGNER-DÖBLER, I. (2016) Co-occurrence Analysis of Microbial Taxa in the Atlantic Ocean Reveals High Connectivity in the Free-Living Bacterioplankton. *Frontiers in Microbiology*, **7**, 649.

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.

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.

PARADA, A.E., NEEDHAM, D.M., & FUHRMAN, J.A. (2016) Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, **18**, 1403–1414.

627 Peoples, L.M., Donaldson, S., Osuntokun, O., Xia, Q., Nelson, A., Blanton, J., Allen, E.E., Church, M.J.,
628 & Bartlett, D.H. (2018) Vertically distinct microbial communities in the Mariana and
629 Kermadec trenches. *PLOS ONE*, **13**, 1–21.
630 Pernice, M.C., Giner, C.R., Logares, R., Perera-Bel, J., Acinas, S.G., Duarte, C.M., Gasol, J.M., & Massana,
631 R. (2016) Large variability of bathypelagic microbial eukaryotic communities across the
632 world's oceans. *The ISME Journal*, **10**, 945–958.
633 Poisot, T., Canard, E., Mouillot, D., Mouquet, N., & Gravel, D. (2012) The dissimilarity of species
634 interaction networks. *Ecology Letters*, **15**, 1353–1361.
635 Prim, R.C. (1957) Shortest connection networks and some generalizations. *The Bell System Technical*
636 *Journal*, **36**, 1389–1401.
637 Pržulj, N., Corneil, D.G., & Jurisica, I. (2004) Modeling interactome: scale-free or geometric?
638 *Bioinformatics*, **20**, 3508–3515.
639 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F.O. (2012)
640 The SILVA ribosomal RNA gene database project: improved data processing and web-based
641 tools. *Nucleic Acids Research*, **41**, D590–D596.
642 Royo-Llonch, M., Sánchez, P., Ruiz-González, C., Salazar, G., Pedrós-Alió, C., Labadie, K., Paoli, L.,
643 Chaffron, S., Eveillard, D., Karsenti, E., Sunagawa, S., Wincker, P., Karp-Boss, L., Bowler, C., &
644 Acinas, S.G. (2020) Ecogenomics of key prokaryotes in the arctic ocean. *bioRxiv*,
645 2020.06.19.156794.
646 Ruiz-González, C., Logares, R., Sebastián, M., Mestre, M., Rodríguez-Martínez, R., Galí, M., Sala, M.M.,
647 Acinas, S.G., Duarte, C.M., & Gasol, J.M. (2019) Higher contribution of globally rare bacterial
648 taxa reflects environmental transitions across the surface ocean. *Molecular Ecology*, **28**,
649 1930–1945.
650 Ruiz-González, C., Mestre, M., Estrada, M., Sebastián, M., Salazar, G., Agustí, S., Moreno-Ostos, E.,
651 Reche, I., Álvarez-Salgado, X.A., Morán, X.A.G., Duarte, C.M., Sala, M.M., & Gasol, J.M. (2020)
652 Major imprint of surface plankton on deep ocean prokaryotic structure and activity.
653 *Molecular Ecology*, **29**, 1820–1838.
654 Salazar, G., Cornejo-Castillo, F.M., Benítez-Barrios, V., Fraile-Nuez, E., Álvarez-Salgado, X.A., Duarte,
655 C.M., Gasol, J.M., & Acinas, S.G. (2016) Global diversity and biogeography of deep-sea pelagic
656 prokaryotes. *The ISME Journal*, **10**, 596–608.
657 Salazar, G., Paoli, L., Alberti, A., Huerta-Cepas, J., Ruscheweyh, H.-J., Cuenca, M., Field, C.M., Coelho,
658 L.P., Cruaud, C., Engelen, S., Gregory, A.C., Labadie, K., Marec, C., Pelletier, E., Royo-Llonch, M.,
659 Roux, S., Sánchez, P., Uehara, H., Zayed, A.A., Zeller, G., Carmichael, M., Dimier, C., Ferland, J.,
660 Kandels, S., Picheral, M., Pisarev, S., Poulain, J., Acinas, S.G., Babin, M., Bork, P., Boss, E.,
661 Bowler, C., Cochrane, G., Vargas, C. de, Follows, M., Gorsky, G., Grimsley, N., Guidi, L., Hingamp,
662 P., Iudicone, D., Jaillon, O., Kandels-Lewis, S., Karp-Boss, L., Karsenti, E., Not, F., Ogata, H.,
663 Pesant, S., Poulton, N., Raes, J., Sardet, C., Speich, S., Stemmann, L., Sullivan, M.B., Sunagawa,
664 S., Wincker, P., Acinas, S.G., Babin, M., Bork, P., Bowler, C., Vargas, C. de, Guidi, L., Hingamp, P.,
665 Iudicone, D., Karp-Boss, L., Karsenti, E., Ogata, H., Pesant, S., Speich, S., Sullivan, M.B., Wincker,
666 P., & Sunagawa, S. (2019) Gene Expression Changes and Community Turnover Differentially
667 Shape the Global Ocean Metatranscriptome. *Cell*, **179**, 1068-1083.e21.
668 Sanz-Sáez, I. (2021) Contribution of marine heterotrophic cultured bacteria to microbial diversity and
669 mercury detoxification.
670 Sebastián, M., Ortega-Retuerta, E., Gómez-Consarnau, L., Zamanillo, M., Álvarez, M., Arístegui, J., &
671 Gasol, J.M. (2021) Environmental and physical barriers drive the basin-wide spatial
672 structuring of Mediterranean Sea and adjacent Eastern Atlantic Ocean prokaryotic
673 communities. *Submitted*.

674 SEBASTIÁN, M., SÁNCHEZ, P., SALAZAR, G., ÁLVAREZ-SALGADO, X.A., RECHE, I., MORÁN, X.A.G., SALA, M.M.,
675 DUARTE, C.M., ACINAS, S.G., & GASOL, J.M. (2021) The quality of dissolved organic matter shapes
676 the biogeography of the active bathypelagic microbiome. *bioRxiv*, 2021.05.14.444136.
677 SEYMOUR, J.R., AMIN, S.A., RAINA, J.-B., & STOCKER, R. (2017) Zooming in on the phycosphere: the
678 ecological interface for phytoplankton–bacteria relationships. *Nature Microbiology*, **2**, 17065.
679 SHADE, A. & HANDELSMAN, J. (2012) Beyond the Venn diagram: the hunt for a core microbiome.
680 *Environmental Microbiology*, **14**, 4–12.
681 STEELE, J.A., COUNTWAY, P.D., XIA, L., VIGIL, P.D., BEMAN, J.M., KIM, D.Y., CHOW, C.-E.T., SACHDEVA, R., JONES,
682 A.C., SCHWALBACH, M.S., ROSE, J.M., HEWSON, I., PATEL, A., SUN, F., CARON, D.A., & FUHRMAN, J.A.
683 (2011) Marine bacterial, archaeal and protistan association networks reveal ecological
684 linkages. *The ISME Journal*, **5**, 1414–1425.
685 STOECK, T., BASS, D., NEBEL, M., CHRISTEN, R., JONES, M.D.M., BREINER, H.-W., & RICHARDS, T.A. (2010)
686 Multiple marker parallel tag environmental DNA sequencing reveals a highly complex
687 eukaryotic community in marine anoxic water. *Molecular Ecology*, **19**, 21–31.
688 SUNAGAWA, S., COELHO, L.P., CHAFFRON, S., KULTIMA, J.R., LABADIE, K., SALAZAR, G., DJAHANSCHIRI, B., ZELLER,
689 G., MENDE, D.R., ALBERTI, A., CORNEJO-CASTILLO, F.M., COSTEA, P.I., CRUAUD, C., D'OVIDIO, F., ENGELEN,
690 S., FERRERA, I., GASOL, J.M., GUIDI, L., HILDEBRAND, F., KOKOSZKA, F., LEPOIVRE, C., LIMA-MENDEZ, G.,
691 POULAIN, J., POULOS, B.T., ROYO-LLONCH, M., SARMENTO, H., VIEIRA-SILVA, S., DIMIER, C., PICHERAL, M.,
692 SEARSON, S., KANDELS-LEWIS, S., BOWLER, C., DE VARGAS, C., GORSKY, G., GRIMSLEY, N., HINGAMP, P.,
693 IUDICONE, D., JAILLON, O., NOT, F., OGATA, H., PESANT, S., SPEICH, S., STEMMANN, L., SULLIVAN, M.B.,
694 WEISSENBACH, J., WINCKER, P., KARSENTI, E., RAES, J., ACINAS, S.G., & BORK, P. (2015) Structure and
695 function of the global ocean microbiome. *Science*, **348**, 1261359.
696 TACKMANN, J., RODRIGUES, J.F.M., & VON MERING, C. (2019) Rapid Inference of Direct Interactions in
697 Large-Scale Ecological Networks from Heterogeneous Microbial Sequencing Data. *Cell*
698 *Systems*, **9**, 286-296.e8.
699 VELLEND, M. (2020) *The theory of ecological communities (MPB-57)*. Princeton University Press.
700 XU, Z., WANG, M., WU, W., LI, Y., LIU, Q., HAN, Y., JIANG, Y., SHAO, H., MCMINN, A., & LIU, H. (2018) Vertical
701 Distribution of Microbial Eukaryotes From Surface to the Hadal Zone of the Mariana Trench.
702 *Frontiers in Microbiology*, **9**, 2023.
703 YAVEROĞLU, Ö.N., MALOD-DOGNIN, N., DAVIS, D., LEVNAJIC, Z., JANJIC, V., KARAPANDZA, R., STOJMIROVIC, A., &
704 PRŽULJ, N. (2014) Revealing the Hidden Language of Complex Networks. *Scientific Reports*, **4**,
705 4547.
706

FIGURES

Figure 1: Sampling scheme. Location, number, and depth range of samples from the epipelagic zone including surface and DCM layer, the mesopelagic zone, and the bathypelagic zone from the global tropical and subtropical ocean and the Mediterranean Sea.

Figure 2: Spatial recurrence. A) Association prevalence showing the fraction of subnetworks in which an association appeared considering all depth layers across the global tropical and subtropical ocean and the Mediterranean Sea. Associations that occurred more often (black) appeared in the middle of the single static network visualization. Most edges had a low prevalence (blue) <20%. B) The sample-specific subnetworks of the four ocean layers (rows): surface (SRF), DCM, mesopelagic (MES), and bathypelagic (BAT), and the six regions (columns). The histograms show the association prevalence within each depth layer and region (excluding absent associations, i.e., 0% prevalence). The number of samples appears in the upper left corner, the number of edges with a prevalence >0% in the upper right corner, and the depth range in the lower right corner (in m below surface). Note that the prevalence goes up to 100% in B) vs. 66.5% in A).

Figure 3: Highly prevalent associations for each region and depth layer. If an association appears in more than 70% of subnetworks it is classified as highly prevalent. The four ocean layers (rows) are surface (SRF), DCM, mesopelagic (MES), and bathypelagic (BAT). The number of samples appears in the upper left corner, the number of edges in the upper right corner, and the depth range in the lower right corner (in m below surface).

Figure 4: Classification of associations. An association can be classified into global (>70% prevalence, not considering the MS), prevalent (>50%, not considering the MS), low-frequency (>20%, not considering the MS), regional, and other. Regional associations are assigned to one of six ocean basins. The number A) and fraction B) of each type of association are shown for each depth layer: surface (SRF) and DCM (epipelagic), mesopelagic (MES) and bathypelagic (BAT). Color indicates the type of classification. The associations have been classified into the five types based on their prevalence in each region. The prevalence of associations is shown in C). For instance, global associations have a prevalence above 70% in each region (not considering the MS). Regional associations are present in one region (indicated with yellow with mainly low prevalence >0%) and absent in all other regions (0% prevalence not shown in graph).

741 **Figure 5:** Microbial associations across depth layers. For each region and taxonomic domain, we color
742 associations based on when they first appeared: surface (S, yellow), DCM (D, orange), mesopelagic
743 (M, red), and bathypelagic (B, black). Absent ASVs are grouped in the white box. Columns show
744 associations between archaea (Arc), bacteria (Bac), and eukaryotes (Euk).
745
746 **Figure 6:** Minimal Spanning Tree. Each subnetwork is a node in the MST and represents a sample.
747 Nodes are colored according to A) the sample's depth layer, B) the samples ocean region, C) the
748 subnetworks cluster, and D) selected environmental factors. In C), the barplots indicate the different
749 layers within each cluster colored as in A).
750

TABLES

Table 1: Number of classified associations per depth layer. The sum of classified associations (including Other) is the number of present associations. Absent associations appear in other layers but in no subnetwork of a given layer. Global, prevalent, and low-frequency associations have been computed with and without considering the MS. The proportion of regional associations increased with depth (gray row).

Depth layer	Epipelagic (Surface)	Epipelagic (DCM)	Mesopelagic	Bathypelagic
Global	26 (0.14%)	23 (0.31%)	21 (0.20%)	-
Prevalent	22 (0.12%)	47 (0.64%)	10 (0.10%)	7 (0.07%)
Low-frequency	105 (0.58%)	160 (2.17%)	212 (2.05%)	51 (0.51%)
Global (no MS)	86 (0.47%)	52 (0.70%)	28 (0.27%)	9 (0.09%)
Prevalent (no MS)	207 (1.14%)	76 (1.03%)	27 (0.26%)	28 (0.28%)
Low-frequency (no MS)	1361 (7.46%)	219 (2.97%)	342 (3.30%)	489 (4.84%)
Regional	2014 (11.05%)	2290 (31.03%)	3420 (33.00%)	3669 (36.33%)
MS	596 (3.27%)	1295 (17.55%)	2254 (21.75%)	1217 (12.05%)
NAO	577 (3.16%)	306 (4.15%)	422 (4.07%)	1522 (15.07%)
SAO	162 (0.89%)	304 (4.12%)	301 (2.90%)	143 (1.42%)
SPO	152 (0.83%)	105 (1.42%)	40 (0.39%)	109 (1.08%)
NPO	298 (1.63%)	133 (1.80%)	204 (1.97%)	516 (5.11%)
IO	229 (1.26%)	147 (1.99%)	199 (1.92%)	162 (1.60%)
Other*	16067 (88.12%)	4860 (65.85%)	6701 (64.66%)	6372 (63.10%)
Other (no MS)*	14566 (79.88%)	4743 (64.27%)	6547 (62.17%)	55904 (58.46%)
Present	18234 (100%)	7380 (100%)	10364 (100%)	10099 (100%)
Absent	10884	21738	18754	19019

*The number of unclassified (Other) associations is computed from present, regional, global, prevalent, and low-frequency associations. The last three classifications have been done with and without the MS, and subsequently the number of unclassified (other) associations varies.

760 **Table 2:** Dataset compilation. Our data was a compilation of different datasets. We required that each location had to provide data for
761 both eukaryotes and prokaryotes, which resulted in 397 samples. This condition allowed only 13 MalaDeep samples.

Dataset	Samples used for analysis	Stations	Depth range (m)	Water samples	Size Fraction (µm)	16S	18S	Reference	ENA accession number
Malaspina									
MalaSurf	122	120	3	122	0.2-3	122	124	(Ruiz-González *et al.*, 2019; Logares *et al.*, 2020)	PRJEB23913 [18S rRNA genes], PRJEB25224 [16S rRNA genes]
MalaVP	83	13	3-4000	91	0.2-3	91	83	(Giner *et al.*, 2020) & This study	PRJEB23771 [18S rRNA genes], PRJEB45015 [16S rRNA genes]
MalaDeep (Prok)	13	30	~4000	60	0.2-0.8	41	-	(Sanz-Sáez, 2021)	PRJEB45011
MalaDeep (Euk)	13	27	2400-4000	27	0.8-20	-	82	This study	PRJEB45014
Hotmix	179	29	3-4539	188	0.2-3	188	179	(Sebastián, Ortega-Retuerta, *et al.*, 2021)	PRJEB44683 [18S rRNA genes], PRJEB44474 [16S rRNA genes]

762 *16S and 18S refer to sequenced samples; Prok - prokaryotes; Euk - eukaryotes*
763

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY FIGURES

Supplementary Figure 1: Robustness of the third condition for generating sample-specific subnetworks for each region and depth with sufficient samples (DCM layer from the SPO was removed because it contained only one sample). Within each region and depth, the set of samples was randomly subsampled containing between 10% to 90% of the original set using all samples. The y-axis shows the fraction of edges that were kept in the subsampled set compared to the original set. We considered A) only the number of kept edges and B) which edges were kept.

Supplementary Figure 2: Associations occurring in each region and depth layer. If an association appears in more than 20% of subnetworks in each region, it is classified as low-frequency, >50% prevalent, and >70% global. The number of samples appears in the upper left corner, the number of edges in the upper right corner, and the depth range in the lower right corner (in m below surface). We classified the associations considering all six regions (A-D) and considering the five ocean basins neglecting the MS (E-H).

Supplementary Figure 3: Regional associations occurring in each region and depth layer. Within a particular depth layer, if an association appears in at least one subnetwork (present) in one region and in no subnetwork (absent) in other regions, it is classified as regional. The four ocean layers (rows) are surface (SRF), DCM, mesopelagic (MES), and bathypelagic (BAT). The number of samples appears in the upper left corner, the number of edges in the upper right corner, and the depth range in the lower right corner (in m below surface).

Supplementary Figure 4: ASVs across depth layers. For each region, we color ASVs based on the layer they first appeared: surface (S, yellow), DCM (D, orange), mesopelagic (M, red), and bathypelagic (B, black). Absent ASVs are grouped in box “a”. An ASV only appearing in the bathypelagic, is assigned to box “a” in above layers. That is, an ASV detected in the surface and present in the DCM but absent in lower layers, appears in the box (S) in the surface and DCM layer, but in box “a” in the meso- and bathypelagic layer. An ASV cannot be assigned to two layers. Note that most ASVs in the bathypelagic zone have been already detected in upper layers because most ASVs are assigned to the boxes “S”, “D”, and “M” instead of “B”.

Supplementary Figure 5: Global network metrics grouped by region and depth layer.

SUPPLEMENTARY TABLES

Supplementary Table 1: Number of environmentally-driven edges detected by EnDED. We removed environmentally-driven edges (indirect) from the preliminary network, which contained 31966 edges. Only edges that were not environmentally-driven by any environmental factor (not indirect) remained in the network.

Environmental factor	Number of samples	indirect	Not indirect
Fluorescence	394	4 (0.01%)	31962
NO3	361	1563 (4.9%)	30403
PO4	359	1357 (4.2%)	30609
Salinity	395	67 (0.2%)	31899
SiO4	360	632 (2.0%)	31334
Temperature	395	622 (1.9%)	31344
All		2848 (8.9%) = 1779 removed by 1 + 751 removed by 2 + 308 removed by 3 + 10 removed by 4	29118 (91.1%)

Supplementary Table 2: Fraction of microbial associations across depth layers. For each region and layer (rows), we determined the constitution of associations (in percentage %) classifying them based on their first appearance (columns): surface, DCM, mesopelagic, and bathypelagic. We indicated the fractions above 40% in grey.

Region	Layer	Surface	DCM	Mesopelagic	Bathypelagic
MS	SRF	100.00			
	DCM	45.14	54.86		
	Mesopelagic	10.35	18.42	71.24	
	Bathypelagic	2.73	5.12	69.71	22.44
NAO	SRF	100.00			
	DCM	68.30	31.70		
	Mesopelagic	11.64	6.59	81.77	
	Bathypelagic	11.62	1.35	43.49	43.54
SAO	SRF	100.00			
	DCM	45.08	54.92		
	Mesopelagic	6.15	8.50	85.35	
	Bathypelagic	12.22	6.30	26.97	54.61
SPO	SRF	100.00			
	DCM	50.07	49.93		
	Mesopelagic	6.44	2.66	90.90	
	Bathypelagic	9.81	3.32	14.15	72.71
NPO	SRF	100.00			
	DCM	54.23	45.77		
	Mesopelagic	8.33	6.06	85.61	
	Bathypelagic	17.46	5.34	19.92	57.28
IO	SRF	100.00			
	DCM	39.23	60.77		
	Mesopelagic	5.92	7.87	86.21	
	Bathypelagic	11.00	3.84	29.61	55.56

Supplementary Table 3 Subnetwork cluster. Clusters dominated, i.e. over 50%, by one layer or one region are indicated in grey. The last row shows unassigned subnetworks.

cluster ID	Dominated by	Size	Fraction of depth layers					Number of regions (if no number if indicated, it is 1x)				
			Epipelagic			Meso-pelagic	Bathy-pelagic	Epipelagic			Meso-MES	Bathy-BAT
			SRF	EPI	DCM			pelagic	EPI	DCM		
1	MS	5	20.00	20.00	20.00	20.00	20.00	SAO	MS	NAO	MS	MS
2	MS	10	10.00	-	20.00	20.00	50.00	MS	-	2xMS	2xMS	5xMS
3	MS	8	12.50	-	-	25.00	62.50	SRF	-	-	2xMS	5xMS
4	MS, MES	8	-	12.50	-	75.00	12	-	MS	-	6xMS	MS
5	MS, MES	12	16.67	-	-	66.67	16.67	IO, NAO	-	-	7xMS, NAO	2xNAO
6		8	12.50	25.00	12.50	25.00	25.00	IO	MS, NAO	NPO	MS, NAO	2xMS
7	BAT	15	13.33	-	-	26.67	60.00	IO, SPO	-	-	IO, MS, SAO, SPO	IO, MS, NAO, 2xNPO, 2xSAO, 2xSPO
8	DCM	10	10.00	-	90.00	-	-	NPO	-	5xMS, NPO, 3xSAO	-	-
9	DCM	11	36.36	-	63.64	-	-	2xNAO, NPO, SAO	-	3xIO, 2xMS, NPO, SAO	-	-
10		12	-	-	8.33	50.00	41.67	-	-	NAO	IO, MS, NAO, 2xNPO, SAO	IO, 2xNAO, NPO, SAO
11	MES	6	-	-	-	83.33	16.67	-	-	-	IO, MS, NPO, 2xSAO	IO
12	NAO, MES	6	16.67	-	-	83.33	-	NAO	-	-	2xMS, 3xNAO	-
13	SRF	11	54.55	9.09	-	27.27	9.09	IO, MS, NPO, 3xSAO	MS	-	2xMS, NAO	MS
14	BAT	16	12.50	6.25	6.25	6.25	68.75	MS, NAO	MS	MS	MS	5xNAO, 3xNPO, 2xSAO, SPO
15	SRF	8	100.00	-	-	-	-	3xIO, 4xNAO, NPO	-	-	-	-
16	MS, SRF	7	71.43	14.29	-	14.29	-	4xMS, NPO	MS	-	MS	-
17	MS	9	-	11.11	33.33	22.22	33.33	-	MS	MS, NAO, SPO	2xMS	3xMS
18	MS, BAT	8	12.50	25.00	-	-	62.50	IO	2xMS	-	-	3xMS, 2xNAO
19	SRF	7	85.72	14.29	-	-	-	2xIO, NAO, NPO, 2xSAO	MS	-	-	-
20	SRF	15	73.33	-	6.67	6.67	13.33	2xIO, 2xNAO, NPO, 5xSAO, SPO	-	MS	IO	IO, NPO
21		8	25.00	-	12.50	25.00	37.50	IO, SPO	-	MS	MS, SAO	IO, 2xNAO
22		17	23.53	-	5.88	35.29	35.29	3xSAO, SPO	-	MS	NAO, 2xNPO, SAO, 2xSPO	IO, MS, NAO, 3xSAO
23	SRF	8	75.00	12.50	-	12.50	-	IO, 2xMS, NAO, NPO, SPO	MS	-	MS	-
24	MS, MES	13	15.38	7.69	-	61.54	15.38	2xMS	MS	-	IO, 4xMS, 3xNAO	NAO, NPO
25		14	28.57	7.14	14.29	7.14	42.86	2xMS, 2xNAO	MS	2xMS	NAO	MS, 3xNPO, 2xSAO
26	SRF	7	85.72	14.29	-	-	-	2xIO-SRF, MS-EPI, 2xNAO-SRF, 2xNPO-SRF	2xIO-SRF, MS-EPI, 2xNAO-SRF, 2xNPO-SRF	-	-	-
27	SRF	11	100.00	-	-	-	-	2xIO, NAO, 4xNP, 4xSPO	-	-	-	-
28	MS	11	9.09	27.27	-	36.36	27.27	MS	3xNAO	-	4xMS	3xMS
29		12	50.00	-	16.67	16.67	16.67	IO, MS, 3xNAO, SAO	-	MS, NAO	2xMS	2xMS
30		6	50.00	-	16.67	16.67	16.67	IO, NAO, SPO	-	MS	NPO	IO-BAT
31	MS	28	25.00	10.71	7.14	35.71	21.43	4xIO, 2xMS, SAO	3xMS	2xMS	6xMS, 2xNAO, 2xNPO	IO, 2xMS, 3xNAO
32	SRF	6	100.00	-	-	-	-	IO, 2xNA, NPO, 2xSAO	-	-	-	-
33	SRF	6	100.00	-	-	-	-	NAO, 3xNPO, SAO, SPO	-	-	-	-
34	SRF	14	100.00	-	-	-	-	IO, 4xNAO, 5xNPO, 2xSAO, 2xSPO	-	-	-	-
35	SRF	13	69.23	7.69	-	-	23.08	4xIO, 3xNAO, SAO, SPO	MS	-	-	3xMS
36	SRF	7	100.00	-	-	-	-	3xIO, 3xNPO, SAO	-	-	-	-
-		24	41.67	-	12.50	29.17	16.67	2xIO, MS, 2xNAO, 3xNPO, 2xSAO	-	MS, 2xNAO	2xIO, 4xMS, NPO	MS, NAO, NPO, SAO

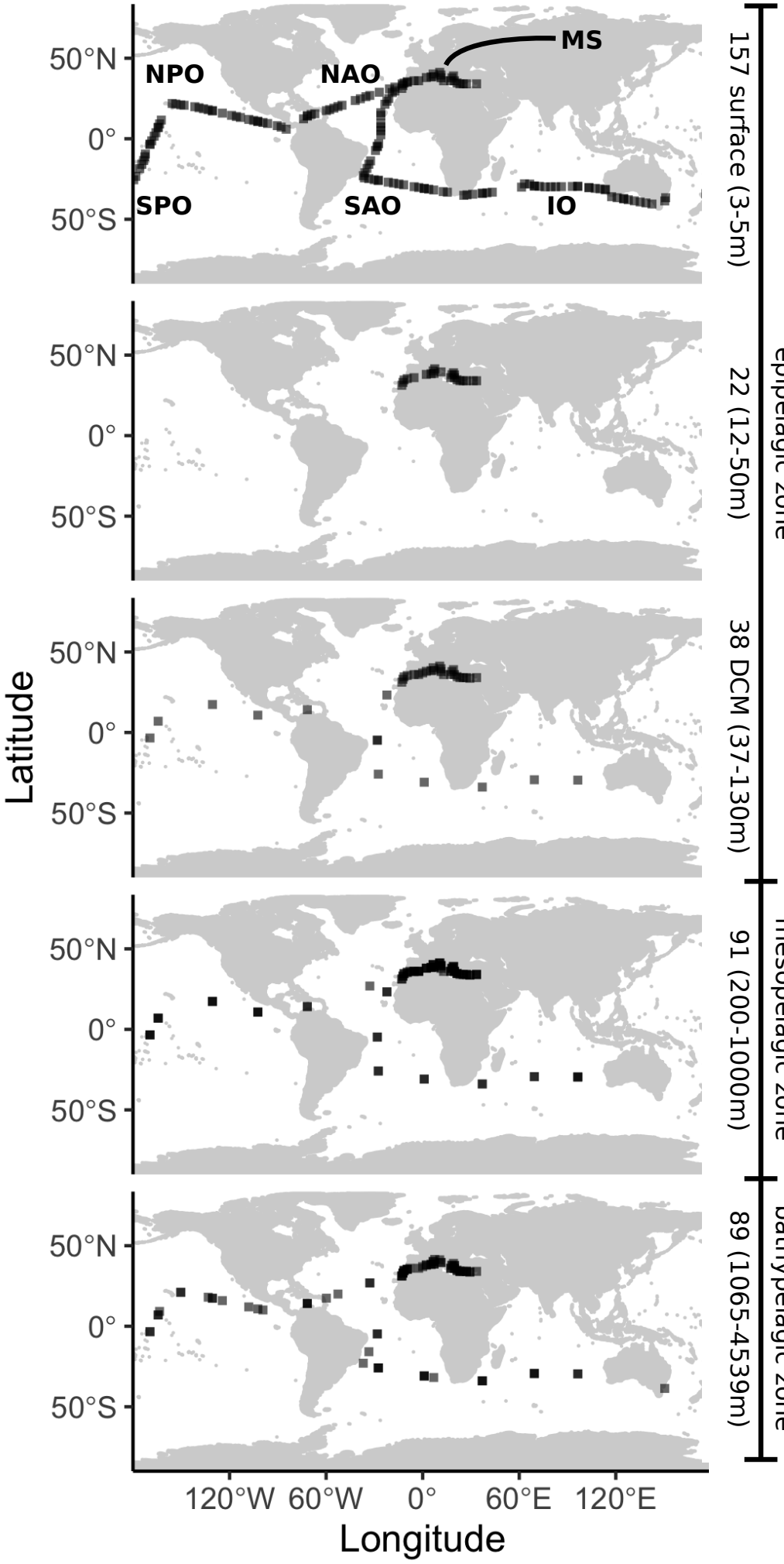
MS – Mediterranean Sea, NAO – North Atlantic Ocean, SAO – South Atlantic Ocean, SPO – South Pacific Ocean, NPO – North Pacific Ocean, IO – Indian Ocean, EPI – epipelagic layer, SRF – surface, DCM – Deep Chlorophyll Maximum, MES – mesopelagic layer, BAT – bathypelagic layer

Supplementary Table 4 Number of edges within each region and depth layer before (J>0%) and after filtering edges with low Jaccard index measuring how often the association partners appeared together in the region and depth layer. The DCM layer in the South Pacific Ocean (SPO) contained only one subnetwork, which resulted in the edge prevalence being 100% for all edges.

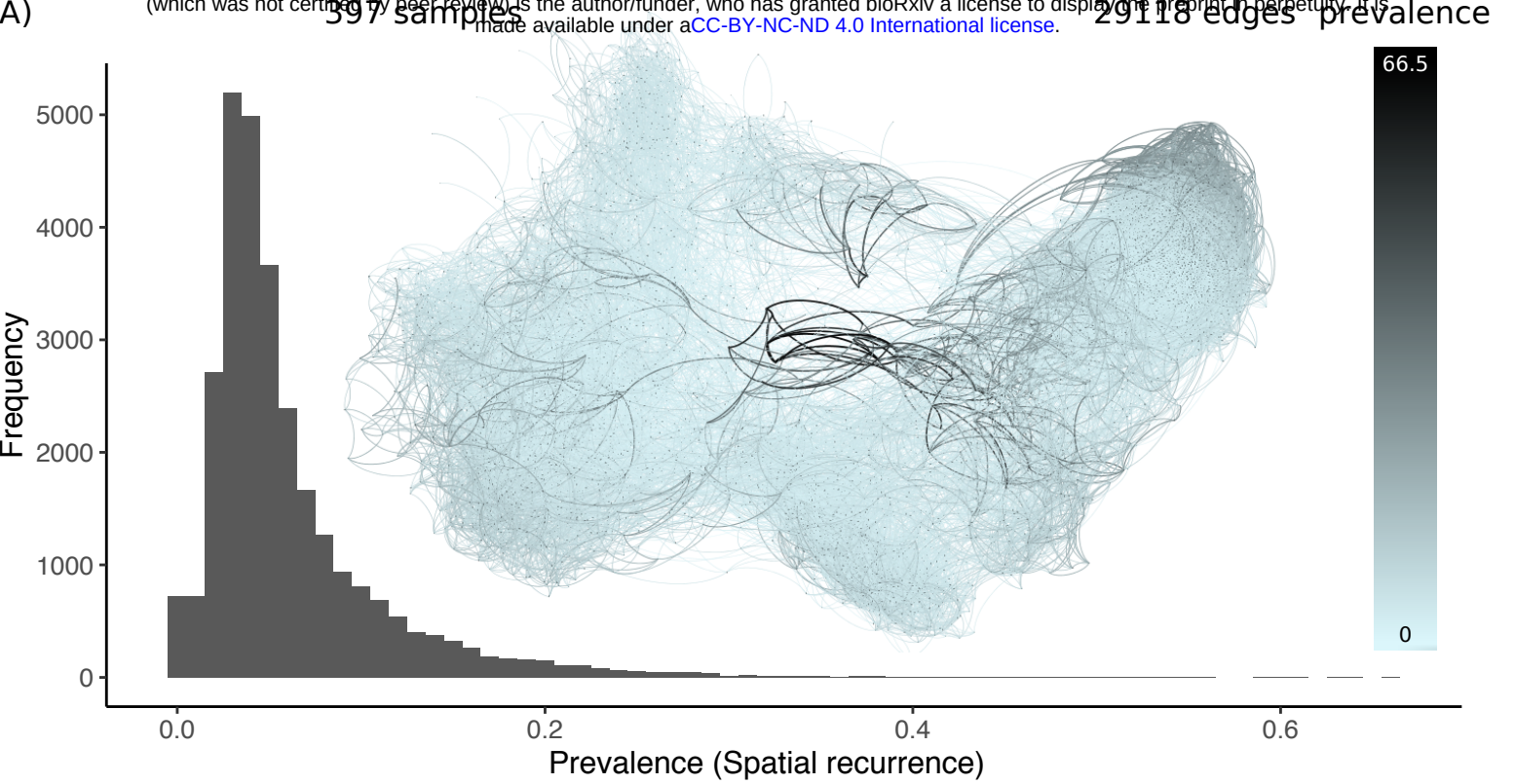
Region	Layer	Samples	Depth (m)	J>0%	J>10%	J>20%	J>30%	J>40%	J>50%
MS	EPI - SRF	19	3	3710	3631	3263	2881	2375	1797
	EPI	18	12-50	4763	4682	4196	3731	3064	2189
	EPI - DCM	21	40-130	5545	5417	4736	4030	3062	2027
	MES	52	200-1000	8756	8403	7336	6179	4629	3088
	BAT	35	1100-3300	4497	4263	3694	3171	2506	1830
NAO	EPI - SRF	34	3	15862	15255	13478	11449	8487	5331
	EPI	4	50	3027	3027	3027	2778	2529	2091
	EPI - DCM	6	70-106	3865	3865	3738	3480	2973	2212
	MES	14	200-800	6325	6289	5689	5109	4169	2978
	BAT	20	1200-4539	7490	7419	6831	6206	5211	3857
SAO	EPI - SRF	26	3	13118	12768	11026	9269	6842	4353
	EPI - DCM	4	80-130	4199	4199	4199	3941	3443	2468
	MES	6	450-850	3937	3937	3740	3440	2687	1614
	BAT	11	1290-4000	4143	4130	3886	3605	3049	2254
NPO	EPI - SRF	29	3	14376	13778	11919	9907	7323	4736
	EPI - DCM	3	37-110	3100	3100	3100	3100	2568	1968
	MES	9	200-780	4197	4197	3781	3343	2583	1625
	BAT	12	2000-4000	5198	5185	4834	4510	4009	3372
SPO	EPI - SRF	14	3-5	12007	11927	10420	8990	6728	4480
	EPI - DCM	1	65	1530	1530	1530	1530	1530	1530
	MES	3	450-650	2066	2066	2066	2066	1756	1318
	BAT	3	1500-4000	3159	3159	3159	3159	2906	2128
IO	EPI - SRF	35	3	14307	13646	11736	9602	6912	4396
	EPI - DCM	3	86-130	3411	3411	3411	3411	2855	2310
	MES	7	400-950	4654	4654	4344	3961	3083	2082
	BAT	8	1065-4000	2928	2928	2790	2563	2101	1290

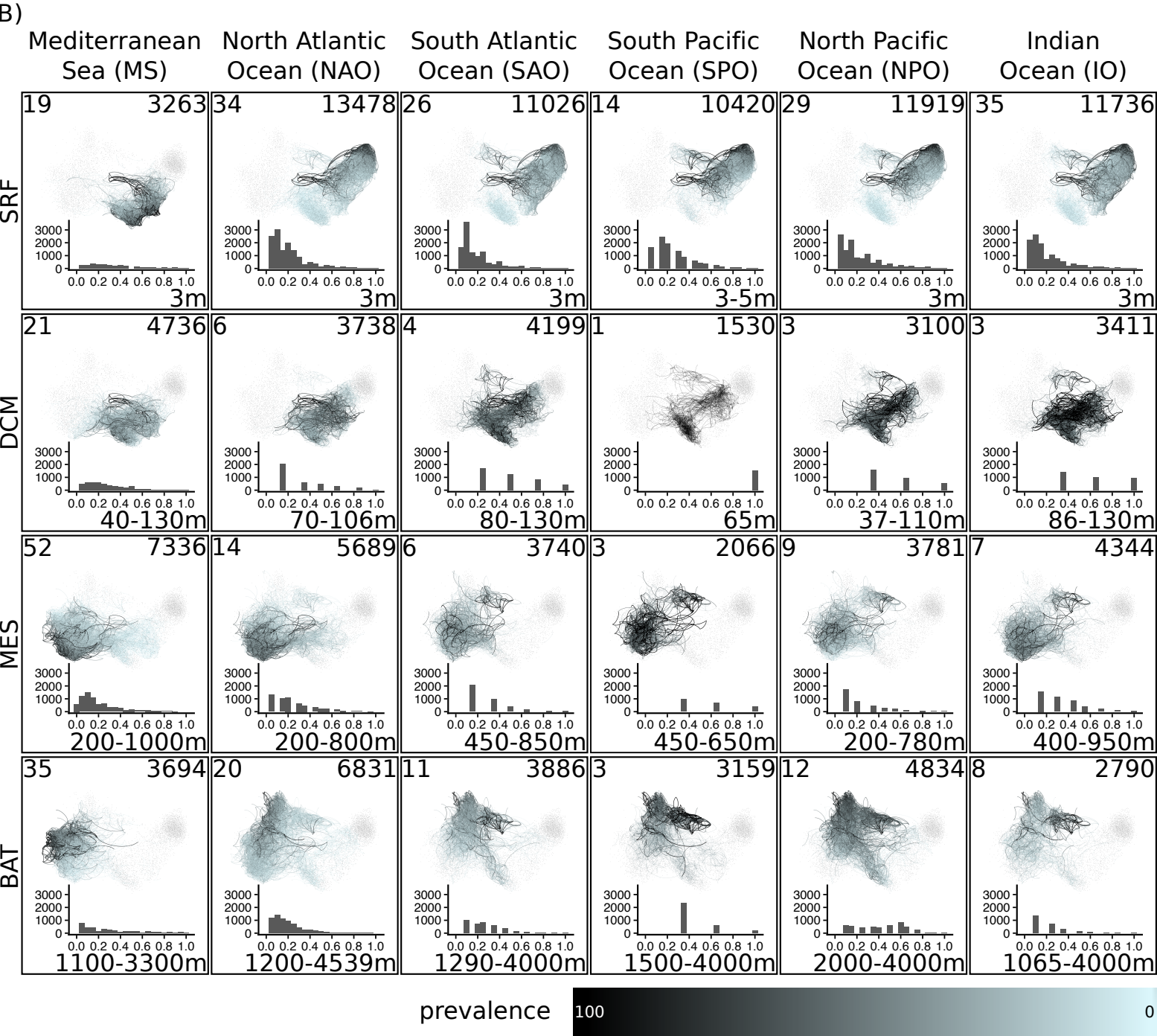
MS – Mediterranean Sea, NAO – North Atlantic Ocean, SAO – South Atlantic Ocean, SPO – South Pacific Ocean, NPO – North Pacific Ocean, IO – Indian Ocean, EPI – epipelagic layer, SRF – surface, DCM – Deep Chlorophyll Maximum, MES – mesopelagic layer, BAT – bathypelagic layer

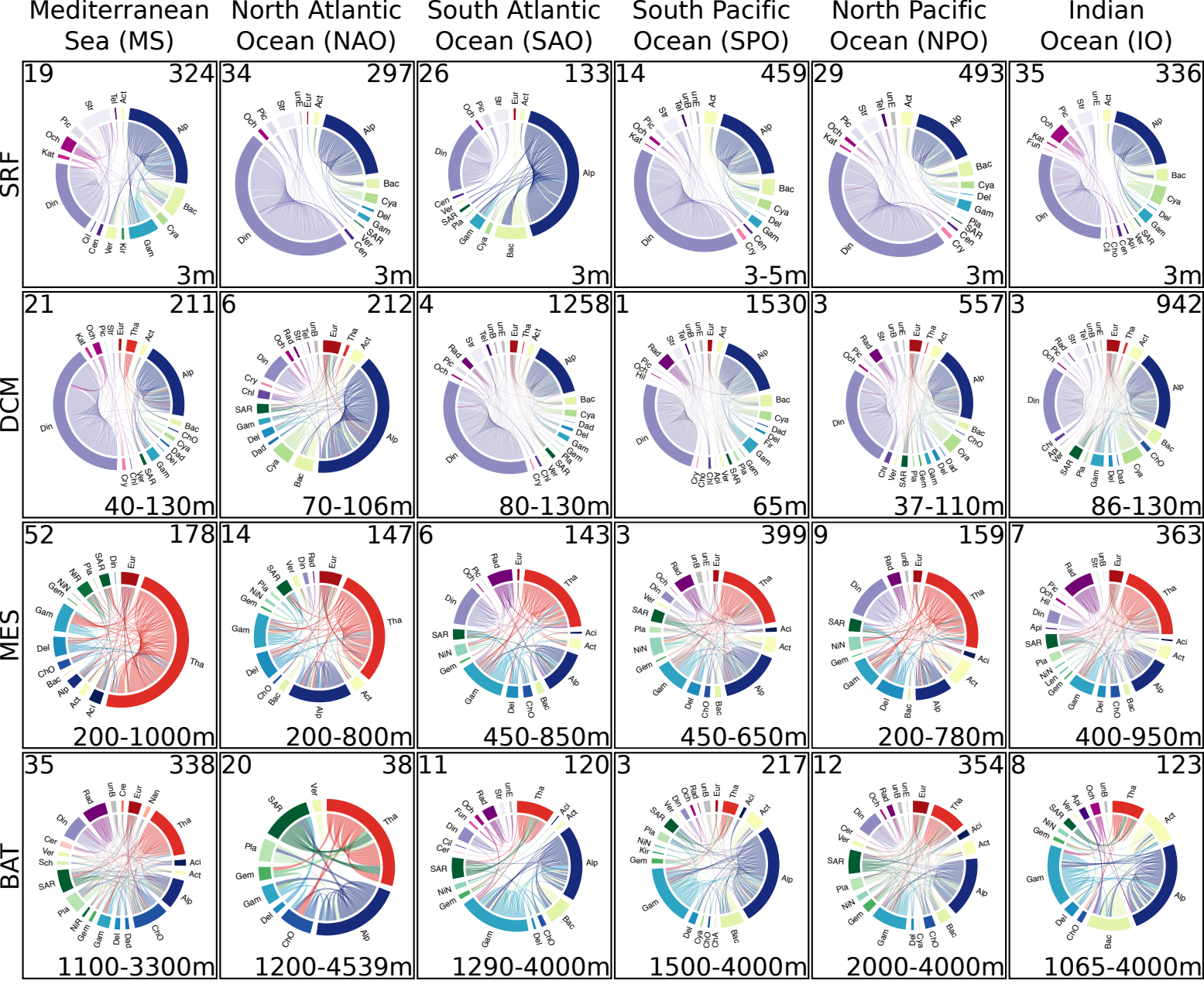
819 **SUPPLEMENTARY MATERIAL**
820
821 **Supplementary Material 1:** Highly prevalent (>70%) regional associations. For each association
822 between two ASVs (first and second column) we list: region (third column), depth layer (fourth
823 column), prevalence in that region and depth layer (fifth column), type: eukaryotic (Euk_Euk),
824 prokaryotic (Prok_Prok), and association between domains (Euk_Prok) (sixth column), and the phyla
825 (seventh and eight column).
826
827 **Supplementary Material 2:** Associations appearing in all layers in at least one region. For each
828 association between two ASVs (first and second column) we list: the classification in each layer (3-6
829 column), overall prevalence (8. column), prevalence in each region and depth layer (9- 34. column),
830 the number of regions in which the association appeared in all layers (AllLayers, 35. column), the
831 number of layers an association appears in a region (36-41. column), type: eukaryotic (Euk_Euk),
832 prokaryotic (Prok_Prok), and association between domains (Euk_Prok) (42. column), and the phyla
833 (43-44. column).



MS - Mediterranean Sea
IO - Indian Ocean
NAO - North Atlantic Ocean
NPO - North Pacific Ocean
SAO - South Atlantic Ocean
SPO - South Pacific Ocean





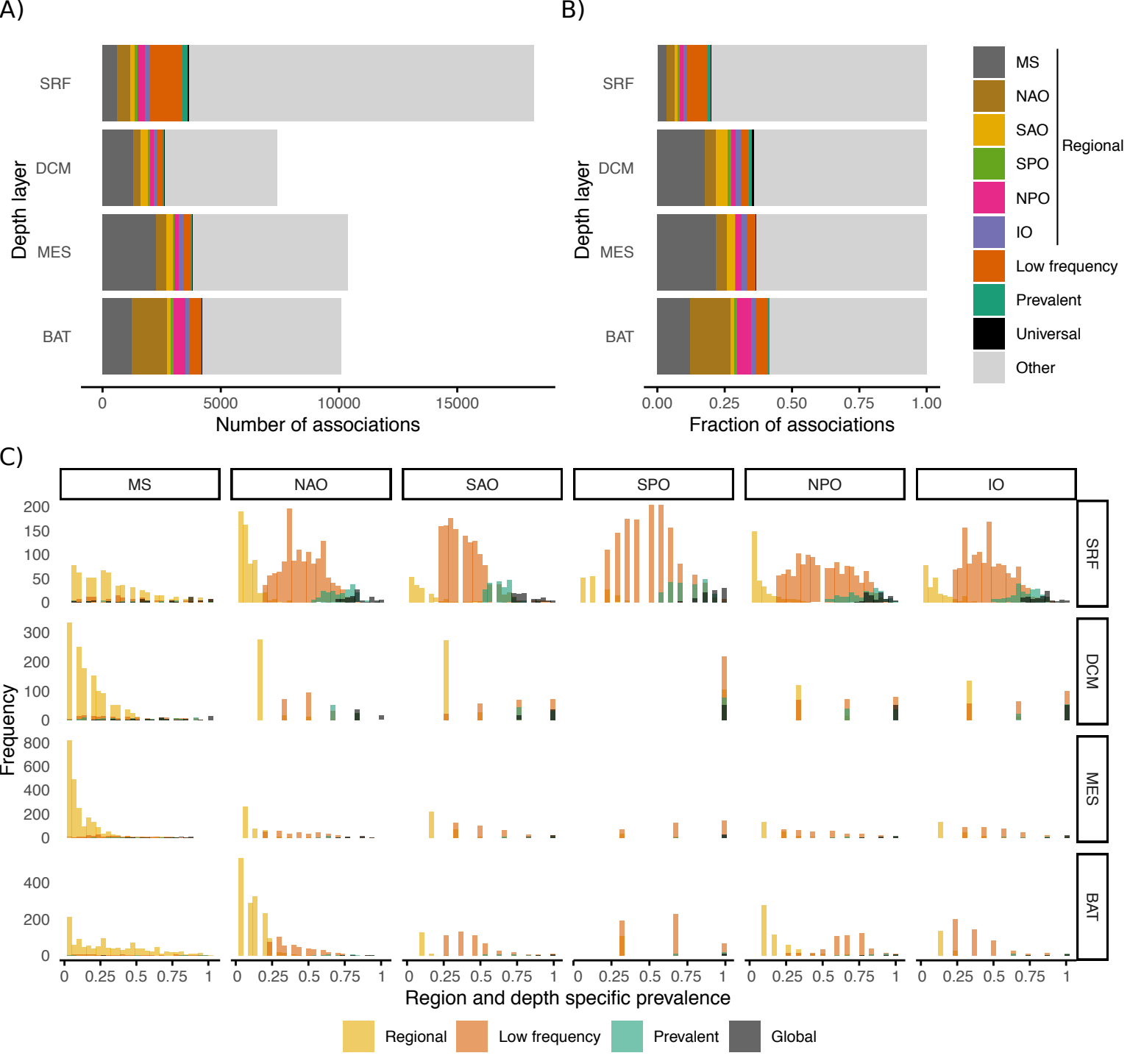


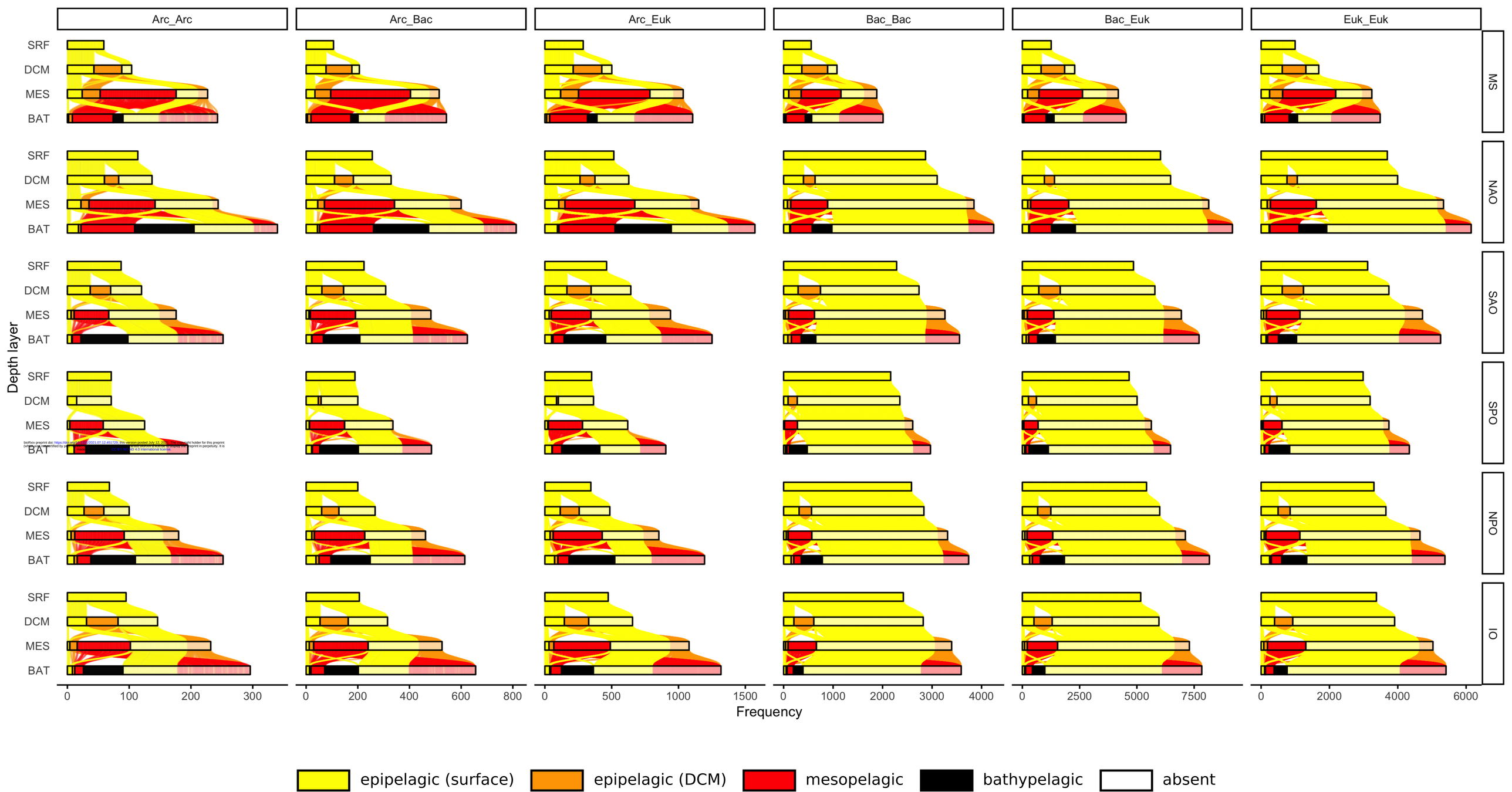
Archaea
- Crenarchaeota
- Euryarchaeota
- Nanoarchaeaeota
- Thaumarchaeota

Bacteria
- Acidobacteria
- Actinobacteria
- Alphaproteobacteria
- Bacteroidetes
- Calditrichaeota
- ChA - Chlamydiae
- ChO - Chloroflexi
- Cyanobacteria
- Dadabacteria
- Deinococcus-Thermus
- Deltaproteobacteria
- Firmicutes
- Gammaproteobacteria
- Gemmatimonadetes
- Hydrogenedentes
- Kiritimatiellaeota
- Lentisphaerae
- Margulisbacteria
- NiN - Nitrospinae
- NiR - Nitrospirae
- Planctomycetes
- SAR406 clade (Marinimicrobia)
- Schekmanbacteria
- Spirochaetes
- Verrucomicrobia

Eukaryota
- Apicomplexa
- Apusomonadidae
- Centroheliozoa
- Cercozoa
- Chlorophyta
- Choanoflagellida
- Ciliophora
- Cryptophyta
- Dinoflagellata
- Fungi
- Hilomonadea (Apusozoa)
- Katablepharidophyta
- Mesomycetozoa (Ichthyosporera)
- Ochrophyta
- Picozoa
- Radiolaria
- Stramenopiles (other)
- Telonemia

Unknown
- unBac
- noProk
- unEuk


A)
B)
C)
SRF
DCM
MES
BAT
Depth layer
0
5000
10000
15000
Number of associations
0.00
0.25
0.50
0.75
1.00
Fraction of associations
MS
NAO
SAO
SPO
NPO
IO
Regional
Low frequency
Prevalent
Universal
Other
Frequency
Region and depth specific prevalence
Regional
Low frequency
Prevalent
Global


Arc_Arc
Arc_Bac
Arc_Euk
Bac_Bac
Bac_Euk
Euk_Euk
MS
NAO
SAO
SPO
NPO
IO
SRF
DCM
MES
BAT
Depth layer
0
100
200
300
0
200
400
600
800
0
500
1000
1500
0
1000
2000
3000
4000
0
2500
5000
7500
0
2000
4000
6000
Frequency
epipelagic (surface)
epipelagic (DCM)
mesopelagic
bathypelagic
absent