Long-term patterns of an interconnected core marine microbiota

Anders K. Krabberød[1*], Ina M. Deutschmann[2], Marit F. M. Bjorbækmo[1], Vanessa Balagué[2], Caterina R. Giner[2], Isabel Ferrera[2,3], Esther Garcés[2], Ramon Massana[2], Josep M. Gasol[2, 4], Ramiro Logares[2,1*]

[1] University of Oslo, Department of Biosciences, Section for Genetics and Evolutionary Biology (Evogene), Blindernv. 31, N-0316 Oslo, Norway

[2] Institute of Marine Sciences (ICM), CSIC, Passeig Marítim de la Barceloneta, Barcelona, Spain

[3] Centro Oceanográfico de Málaga, Instituto Español de Oceanografía, 29640 Fuengirola, Málaga, Spain

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

Corresponding authors:

*Anders K. Krabberød, University of Oslo, Department of Biosciences, Section for Genetics and Evolutionary Biology (Evogene), Blindernv. 31, N-0316 Oslo, Norway
Email: a.k.krabberod@ibv.uio.no

*Ramiro Logares, Institute of Marine Sciences (ICM), CSIC, Passeig Marítim de la Barceloneta, Barcelona, Spain
Email: ramiro.logares@icm.csic.es

Running title: Patterns in a core marine microbiota

Manuscript for: Microbiome

ABSTRACT

Background

Ocean microbes constitute ~70% of the marine biomass, are responsible for ~50% of the Earth's primary production, and are crucial for global biogeochemical cycles. Marine microbiotas include core taxa that are usually key for ecosystem function. Despite their importance, core marine microbes are relatively unknown, which reflects the lack of consensus on how to identify them. So far, most core microbiotas have been defined based on species occurrence and abundance. Yet, species interactions are also important to identify core microbes, as communities include interacting species. Here, we investigate interconnected bacteria and small protists of the core pelagic microbiota populating a long-term marine-coastal observatory in the Mediterranean Sea over a decade.

Results

Core microbes were defined as those present in >30% of the monthly samples over 10 years, with the strongest associations. The core microbiota included 259 Operational Taxonomic Units (OTUs) including 182 bacteria, 77 protists, and 1,411 strong and mostly positive (~95%) associations. Core bacteria tended to be associated with other bacteria, while core protists tended to be associated with bacteria. The richness and abundance of core OTUs varied annually, decreasing in stratified warmers waters and increasing in colder mixed waters. Most core OTUs had a preference for one season, mostly winter, which featured subnetworks with the highest connectivity. Groups of highly associated taxa tended to include protists and bacteria with predominance in the same season, particularly winter. A group of 13 highly-connected hub-OTUs, with potentially important ecological roles dominated in winter and spring. Similarly, 18 connector OTUs with a low degree but high centrality were mostly associated with summer or autumn and may represent transitions between seasonal communities.

Conclusions

We found a relatively small and dynamic interconnected core microbiota in a model temperate marine-coastal site, with potential interactions being more deterministic in winter than in other seasons. These core microbes would be essential for the functioning of this ecosystem over the year. Other non-core taxa may also carry out important functions but would be redundant and non-essential. Our work contributes to the understanding of the dynamics and potential interactions of core microbes possibly sustaining ocean ecosystem function.

Keywords: bacteria, protists, ocean, time-series, seasonality, networks, associations

BACKGROUND

Ecosystems are composed of interacting units embedded in and influenced by their physicochemical environment. Ecosystem function can be broadly defined as the biological, geochemical, and physical processes that occur within it. These processes will likely change or halt if specific organisms or gene-functions are removed, driving the ecosystem towards a new state or its collapse. It is hypothesized that ecological redundancy guarantees continuous ecosystem function, as multiple species could carry out the same or similar function [1]. And while the amount of functional redundancy in microbial ecosystems is a matter of debate [2, 3] it has also been observed that microbiotas in comparable habitats tend to share "core" species that are hypothesized to be fundamental for ecosystem function [4]. These core organisms and the functions they carry out might not be easily replaced.

Identifying the core microbiota is not straightforward as there are different ways of defining a core depending on the habitats and the questions being addressed [4]. One often-used approach is to identify species that tend to be recurrently present across spatiotemporal scales. This definition might not be sufficient, however, since communities are made up of interacting species [5]. A more appropriate definition of a core, therefore, needs to incorporate ecological interactions fundamental for the community in the location under study [4, 5]. This is particularly important in studies using DNA to investigate microbial communities, as a fraction of the detected taxa could be dormant, dead, or transient [6-8]. In the interaction-based definition taxa that do not appear to be interacting are excluded from the core [4].

Core microbiotas based on common presence have been widely studied in terrestrial animals, in particular humans [9] or cattle [10], as well in marine animals, in particular corals [11, 12] and sponges [13, 14]. Core microbiotas in non-host-associated

107 systems, such as soils or the ocean, have been investigated to a lesser extent. In soils,
108 for example, a global analysis identified a core group of 241 ubiquitous and dominant
109 bacterial taxa with more or less invariant abundances and unclear habitat preferences
110 [15]. In the tropical and subtropical global-ocean, a total of 68 bacteria and 57
111 picoeukaryotic operational taxonomic units (OTUs) have been identified that could be
112 part of the core surface microbiota, as they were present in >80% of the globally-
113 distributed samples [16].

114 Analyses of ocean time-series have also pointed to the existence of core
115 microbiotas. For example, Gilbert et al. [17] investigated the microbiota of the English
116 Channel for 6 years and found 12 abundant OTUs that were detected throughout the
117 entire dataset (72 time-points), totaling ~35% of the sequence abundance. Potentially
118 core bacterial OTUs were detected in the SPOT time-series (southern California), in a
119 study covering 10 years of monthly samples in the euphotic zone [18]. These
120 potentially-core bacterial OTUs were present in >75% of the months, represented ~7%
121 (25-28 OTUs depending on depth) of the total richness, and had a high (>10%) relative
122 abundance [18].

123 These studies have provided substantial insights on core marine microbiotas,
124 although they typically define them in terms of species occurrence or abundance over
125 spatiotemporal scales, rather than on potential interactions. As in other ecosystems,
126 microbial interactions are essential for the functioning of the ocean ecosystem, where
127 they guarantee the transfer of carbon and energy to upper trophic levels, as well as the
128 recycling of carbon and nutrients [19]. Despite their importance, most microbial
129 interactions in the ocean remain unknown [20]. A recent literature survey spanning the
130 last 150 years indicated that we have documented a minor fraction of protist interactions
131 in the ocean [21] and most likely, the same is true if not worse for bacteria.

132 During the last decade, association networks have been used to bridge this
133 knowledge gap. Association networks are based on correlations between species'
134 abundances and they may reflect microbial interactions [22]. Contemporaneous
135 positive correlations may point to interactions such as symbiosis, or similar niche
136 preferences, while negative correlations may suggest predation, competition, or
137 opposite niche preferences [23]. So far, network analyses have produced hypotheses on
138 microbial interactions at the level of individual species across diverse ecosystems [22,
139 24, 25], a few of which have been experimentally validated [26]. In addition, networks
140 can help detect species that have relatively more associations to other species ("hubs"),
141 or species that connect different subgroups within a network, and which therefore may
142 have important roles in the ecosystem. Groups of highly associated species in the
143 network ("modules") may represent niches [27, 28], and the amount of these modules
144 may increase with increasing environmental selection [22]. Networks can also produce
145 ecological insight at the community level, since their architecture can reflect
146 community processes, such as selection [27].

147 Network analyses have been particularly useful for the investigation of
148 microbial interactions in the ocean [25, 29]. A surface global-ocean network analysis
149 of prokaryotes and single-celled eukaryotes indicated that ~72% of the associations
150 between microbes were positive and that most associations were between single-celled
151 eukaryotes belonging to different organismal size-fractions [26]. Other studies using
152 networks have indicated a limited number of associations between marine microbes and
153 abiotic environmental variables [17, 18, 23, 26, 30-32], suggesting that microbial
154 interactions have an important role in driving community turnover [32]. Despite the
155 important insights these studies have provided, most of them share the limitation that

156 they do not disentangle whether microbial associations may represent ecological
157 interactions or environmental preferences [22].

158 Even though association networks based on long-term species dynamics may
159 allow a more accurate delineation of core marine microbiotas, few studies have
160 identified them in this manner. Consequently, we have a limited understanding of the
161 interconnected set of organisms that may be key for ocean ecosystem function. Here
162 we identify and investigate the core microbiota occurring in the marine-coastal Blanes
163 Bay Microbial Observatory (Northwestern Mediterranean Sea) over 10 years. We
164 delineated the core microbiota stringently, using potential interactions based on species
165 abundances. We also made an effort to disentangle environmental effects in association
166 networks by identifying and removing species associations that are a consequence of
167 shared environmental preference and not interactions between the species [33]. We
168 analyzed bacteria and protists from the pico- (0.2-3 µm) and nanoplankton (3-20 µm)
169 organismal size fractions, which show a strong seasonality in this location [34-36].
170 Taxa relative abundances were estimated by sequencing the 16S and 18S rRNA-gene
171 and delineating OTUs as Amplicon Sequence Variants (ASVs). Specifically, we ask:
172 What taxa constitute the interconnected core microbiota and what are the main patterns
173 of this assemblage over 10 years? Does the core microbiota feature seasonal sub-groups
174 of highly associated species? What degree of association do bacteria and microbial
175 eukaryotes have and do they show comparable connectivity? Can we identify core
176 OTUs with central positions in the network that could have important ecological roles?
177

178 **RESULTS**

179 *Composition and dynamics of the resident microbiota*

180 Based on the data set containing 2,926 OTUs, (1,561 bacteria and 1,365 microbial
181 eukaryotes) we first defined the resident OTUs as the bacteria and microbial eukaryotes
182 present in >30% of the samples, which equals 36 out of 120 months (not necessarily
183 consecutive). This threshold was selected as it includes seasonal OTUs that would be
184 present recurrently in at least one season. The residents consisted of 709 OTUs: 354
185 Bacteria (~54% relative read abundance) and 355 Eukaryotic OTUs (~46% relative
186 read abundance) [**Table 1,** see methods for calculation of relative read abundance]. The
187 most abundant resident bacteria OTUs belonged to Oxyphotobacteria (mostly
188 *Synechococcus*; ~15% of total relative read abundance), Alphaproteobacteria (mostly
189 SAR11 Clade Ia [~9%, and clade II [~4%]), and Gammaproteobacteria (mainly SAR86;
190 ~2%). The most abundant resident protist OTUs belonged to Dinophyceae
191 (predominantly an unclassified dinoflagellate lineage [~7%], Syndiniales Group I
192 Clade 1 [~7%] and *Gyrodinium* [~4%]), Chlorophyta (mostly *Micromonas* [~3%] and
193 *Bathycoccus* [~2%]), Ochrophyta (predominantly Mediophyceae [~2%] and
194 *Chaetoceros* [~1%]) and Cryptophyceae (mainly a Cryptomonadales lineage [~2%])
195 [**Figure 3, Table S1, Additional file 1**].
196
197 **Table 1.** Description of the datasets.

	OTUs	OTUs (%)	Sequence abundance (%)*
All OTUs[1]	2,926	100	*100*
Bacteria	1,561	53.3	*50.7*
Protists	1,365	46.7	*49.3*
Resident microbiota[2]	709	100	100 (*85*)
Bacteria	354	49.9	53.6
Protists	355	50.1	46.4
Core microbiota[3]	**259**	**100**	**64.5 (*54*)**
Bacteria	182	70.3	46.3
Protists	77	29.7	18.2
Picoplankton	109	42.1	32.4
Nanoplankton	150	57.9	32.1
Protists			
Heterotroph	5	1.9	0.3
Photoautotroph	37	14.3	11.8
Parasite	21	8.1	3.5
Mixotroph	3	1.2	0.7

Symbiont	1	0.4	0.1
Unknown	11	4.3	2.0
Bacteria			
Photoautotroph (cyanobacteria)	19	7.3	19.3
Non-photoautotroph[4]	163	62.5	26.8
Seasonal preference core OTUs			
Winter	156	60.2	21.8
Spring	24	9.3	16.4
Summer	44	17.0	8.2
Autumn	30	11.6	13.7
No seasonality	5	1.9	4.5
Seasonal subnetworks			
Winter	156	60.2	21.8
Spring	19	7.3	13.7
Summer	41	15.8	6.6
Autumn	26	10.0	12.9

[1] Number of OTUs in the full dataset that were left after quality control and rarefaction, which were present in at least 10% of the samples (i.e. 12 months, not necessarily consecutive).
[2] OTUs present in at least 30% of the samples (i.e. 36 months, not necessarily consecutive) [=Resident microbiota].
[3] OTUs included in the core network (core microbiota) with significant correlations (p&q <0.001), local similarity scores >|0.7| and Spearman correlations >|0.7|, being present in at least 30% of the samples.
[4] Includes non-photoautotrophic lifestyles (i.e., chemoautotrophs, photoheterotrophs, chemoheterotrophs, etc.).
* In *Italics* the abundances relative to all OTUs are indicated. All other values in normal text indicate abundances relative to OTUs in the resident microbiota.

The resident microbiota, including both protists and bacteria, showed seasonal variation over 10 years, with communities from the same season but different years tending to group (**Figure 1C and D**). The structure of the resident microbiota correlated to specific environmental variables during winter (nutrients, Total photosynthetic nanoflagellates [PNF; 2-5µm size], and small PNF [2µm]), spring (Total Chlorophyll a [Chla]), summer (daylength, temperature, Secchi disk depth and, the cell abundances of *Synechococcus,* Heterotrophic prokaryotes [HP] and Heterotrophic nanoflagellates [HNF, 2-5µm]) and autumn (salinity) [**Figure 1C**]. The environmental variables most relevant for explaining the variance of the resident microbiota were determined by stepwise model selection and distance-based redundancy analyses (dbRDA) [**Figure 1D**], leading to a dbRDA constrained and unconstrained variation of 41% and 59% respectively (**Figure 1D**). The selected variables were predominantly aligned with the axis summer (daylength, temperature, and the cell abundance of *Synechococcus* and HP) - winter (SiO_2, small PNF [**Figure 1D**]. This dbRDA axis had the highest eigenvalue, explaining ~55% of the constrained variation (**Figure 1D**). Even though the measured environmental variables did not explain the majority of the variation of

223 the resident microbiota, they could account for a substantial fraction. This was further
224 supported by Adonis analyses, which indicated that the measured environmental
225 variables could explain ~45% of the resident microbiota variance, with temperature and
226 daylength having a predominant role by accounting for 30% of this variance (15%
227 each).

228 We then investigated whether temperature and daylength could determine the
229 main niches. We found that ~70% and ~68% of the OTUs in the resident microbiota
230 had niche preferences associated with temperature or daylength respectively (**Figure**
231 **1E-F**; Note that several OTUs preferring Spring or Autumn are not expected to be
232 detected with this approach, as their preferred temperature or daylength may not differ
233 significantly from the randomized mean). In total, 371 OTUs from the resident
234 microbiota had both a temperature and a daylength niche preference that departed
235 significantly from the randomization mean (**Figure 1E-F**). These 371 OTUs
236 represented ~52% of all OTUs in the resident microbiota, corresponding to ~90% of
237 the sequence abundance. In particular, 248 OTUs had a weighted mean for both
238 temperature and daylength below the randomization mean (corresponding to
239 winter/autumn), while 116 OTUs had a weighted mean above the randomization mean
240 for both variables (corresponding to summer/spring). Interestingly, 7 OTUs displayed
241 a weighted mean above and below the randomized mean for temperature and daylength
242 respectively (corresponding to autumn or spring).

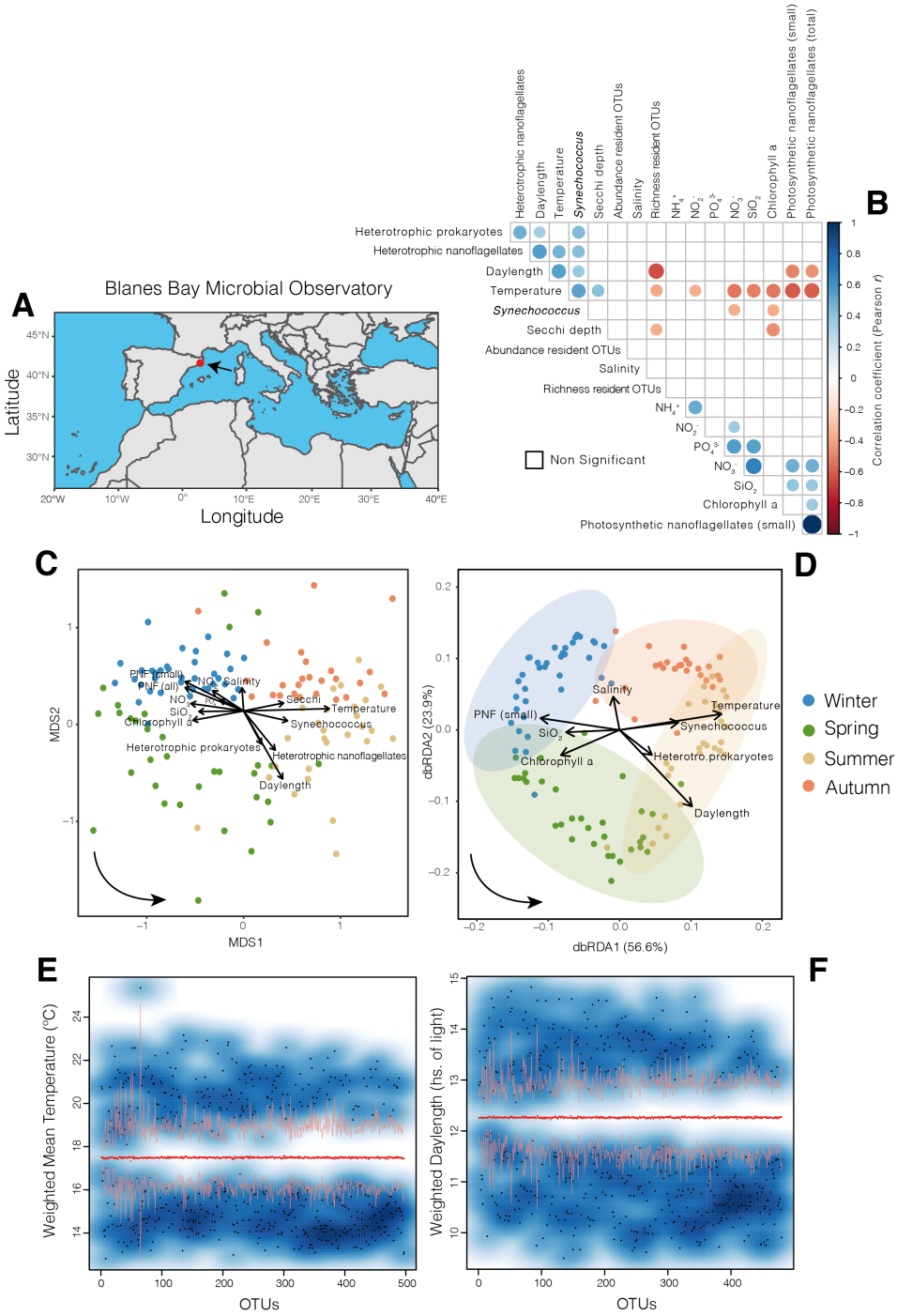


Figure 1. The Blanes Bay Microbial Observatory and the variation of its resident microbiota and measured environmental variables over ten years. A) Location of the Blanes Bay Microbial Observatory. **B)** All possible correlations between the measured environmental variables including the richness and abundance of resident OTUs (NB: only 709 resident OTUs are considered, see **Table1**). Only significant Pearson correlation coefficients are shown ($p<0.01$). The p-values were corrected for multiple inference (Holm's method). **C)** Unconstrained

ordination (NMDS based on Bray Curtis dissimilarities) of communities including resident OTUs only, to which environmental variables were fitted. Only variables with a significant fit are shown (P<0.05). Arrows indicate the direction of the gradient and their length represents the strength of the correlation between resident OTUs and a particular environmental variable. The color of the samples (circles) indicates the season to which they belong. The bottom-left arrow indicates the direction of the seasonal change. PNF = photosynthetic nanoflagellates. **D)** Constrained ordination (Distance-based redundancy analyses, dbRDA, using Bray Curtis dissimilarities) including only the most relevant variables after stepwise model selection using permutation tests. Each axis (i.e., dbRDA1 and dbRDA2) indicates the amount of variance it explains according to the associated eigenvalues. The color of the samples (circles) indicates the season to which they belong. Arrows indicate the direction of the gradient and their length represents the strength of the correlation between resident OTUs and a particular environmental variable. The bottom-left arrow indicates the direction of the seasonal change. E-F) Resident OTUs displaying different niche preferences (blueish areas) in terms of the two most important abiotic variables: Temperature E) and Daylength F). The red dots indicate the randomization mean, and the orange curves represent the confidence limits. Black dots indicate individual OTUs for which temperature or daylength preferences are significantly (p<0.05) higher or lower than a random distribution over 10 years. At least two assemblages with different niches become evident: one preferring higher temperature and longer days (summer/spring), and another one preferring lower temperature and shorter days (winter/autumn). Note that several OTUs associated to Spring or Autumn are not expected to be detected with this approach, as their preferred temperature or daylength may not differ significantly from the randomized mean.

Core network

274 To determine the core microbiota that incorporates possible interactions, we
275 constructed an association network based on the resident OTUs and removed all OTUs
276 that were not involved in strong and significant associations with any other OTUs.
277 Specifically, we kept only the associations (edges in the network) with Local similarity
278 score |LS| > 0.7, a false discovery rate adjusted p-value < 0.001 and Spearman |r| > 0.7.
279 In addition, we removed all associations that seemed to be caused by environmental
280 preferences of OTUs (see Methods). The core network consisted of 1,411 significant
281 and strong correlations (**Figure 2A**) and was substantially smaller than the network
282 based on the resident OTUs without stringent cut-offs (**Figure S1A, Additional file 2**,
283 removed edges in **Figure S1B, Additional file 2**). The core network includes only the
284 strongest microbial associations that are inferred during a decade and, according to our
285 definition, determines the core microbiota. The associations in the core microbiota may
286 represent proxies for species interactions since steps have been taken to remove
287 associations that are driven by environmental factors.

288 In the core network, most associations were positive (~95%), pointing to the
289 dominance of co-existence or symbiotic associations (**Table 2, Figure 2A**). The core
290 network had "small world" properties [37], with a small average path length (i.e.
291 number of nodes between any pair of nodes through the shortest path) and a relatively
292 high clustering coefficient, showing that nodes tend to be connected to other nodes,
293 forming tightly knit groups, more than what it would be expected by chance (**Table 3**).
294 Since node degree was not correlated with OTU abundance (**Figure S2, Additional file**
295 **3**), the associations between OTUs are not caused by a high sequence abundance alone,
296 as the most abundant OTUs did not tend to be the most connected.

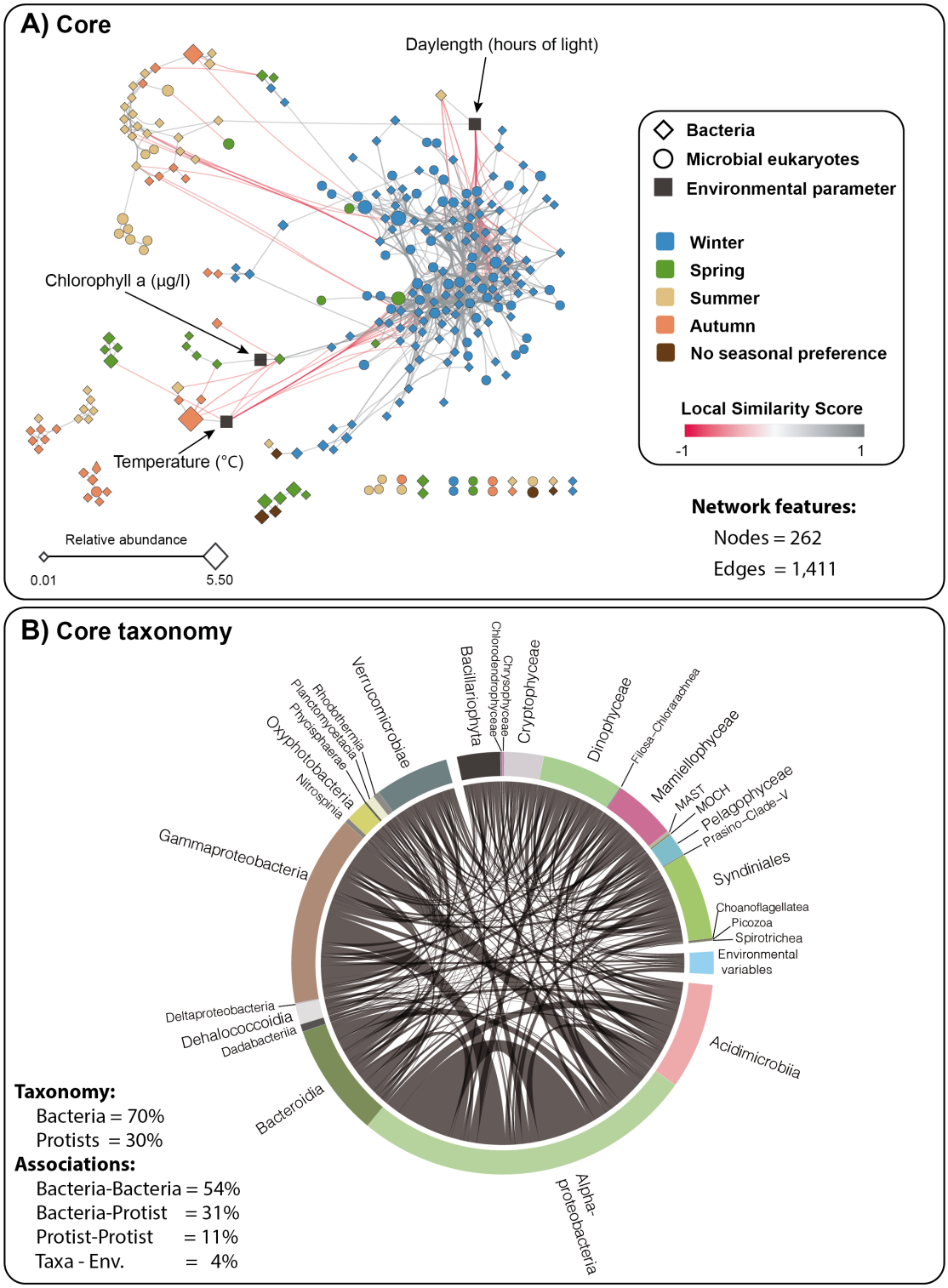


Figure 2. Core microbiota resulting from 10 years of monthly pico- and nanoplankton relative abundances. A) Core network including bacteria and microbial eukaryotic OTUs that occur ≥ 30% of the time during the studied decade (i.e. resident microbiota), with highly

302 significant and strong associations (P<0.001 and Q<0.001, absolute local similarity score |LS| >
303 0.7, Spearman correlation |ρ|>0.7), where detected environmentally-driven edges were
304 removed. The color of the edges (links) indicates whether the association is positive (grey) or
305 negative (red). The shape of nodes indicates bacteria (rhomboid) or microbial eukaryotes (circle),
306 and the color of nodes represents species seasonal preferences, determined using the indicator
307 value (*indval*, p<0.05). Node size indicates OTU relative abundance. B) Core network as a Circos
308 plot, indicating the high-rank taxonomy of the core OTUs. Since 95% of the associations are
309 positive (see Table 2), we do not indicate whether an edge is positive or negative.

310

311 The core network displayed a winter cluster, while no clear clusters could be
312 defined for the other seasons (**Figure 2A**). Of the 15 environmental variables analyzed,
313 only 3 were found to be significantly correlated with core OTUs: *daylength*, showing
314 strong correlations with 33 OTUs, *temperature*, correlated with 14 OTUs, and
315 *Chlorophyll a*, correlated with 1 OTU (**Figure 2A**). Therefore, the analysis of the core
316 network also points to the importance of temperature and daylength in the decade-long
317 seasonal dynamics of the studied microbial ecosystem. It is also coherent with the
318 Adonis and ordination analyses (**Figure 1C-B**). However, the associations between
319 these environmental parameters with taxa represented only 4% of all the associations
320 (**Figure 2B**).

321 **Table 2.** Core associations. See **Figure 2**.

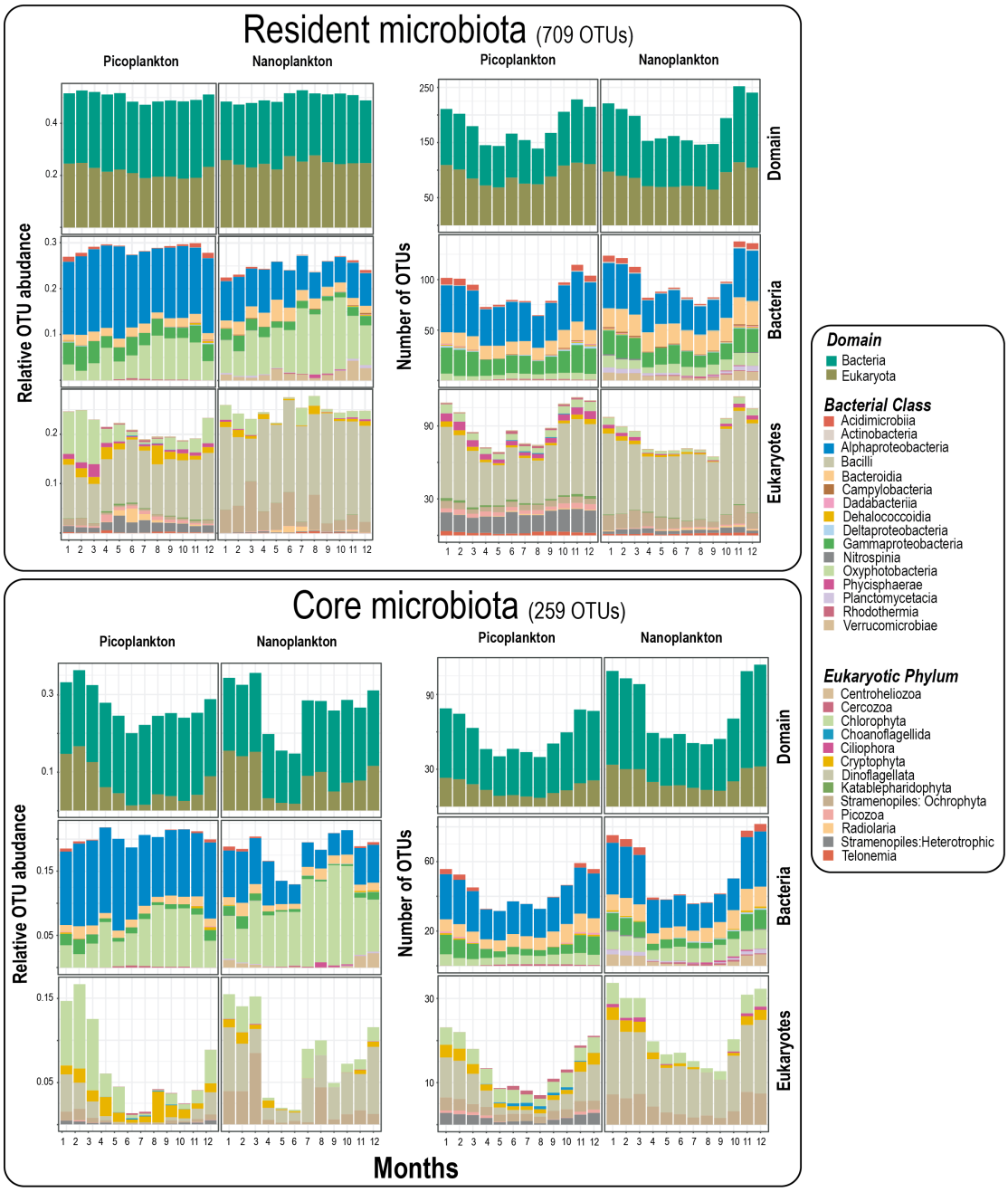
322

	Association # (edges)	Co-occurrences (positive)	Co-exclusions (negative)
All	1,411	1,341 (95.0%)	70 (5.0%)
Within Picoplankton	378	353 (93.3%)	25 (6.6%)
Within Nanoplankton	791	748 (94.6%)	43 (5.4%)
Picoplankton-Nanoplankton	242	240 (99.2%)	2 (0.8%)

323

324 Of the 709 OTUs from the resident microbiota (**Figure 3**), only 259 OTUs
325 (35%) were left in the core network (182 bacteria (~70%) and 77 microbial eukaryotic

326 OTUs (~30%); **Table 1, Figure 2**). The monthly taxonomic composition of the resident
327 microbiota differed from that of the core (**Figure 3**). The core OTUs accounted for
328 ~64% of the relative read abundance of the resident microbiota (**Table 1**). The core
329 OTUs had annual variation in terms of richness and abundance over the 10 years for
330 both the pico- and nanoplankton, with microbial eukaryotes decreasing markedly in
331 OTU richness and relative read abundance in the warmer seasons, and increasing during
332 colder periods (**Figure 3**).



333

334

335 **Figure 3. The monthly variation in the resident and core microbiotas over 10 years.** *Upper*
336 *panels:* The resident microbiota is defined as those eukaryotes and bacteria that occur in at least
337 30% of the samples over 10 years. The relative OTU abundance (left panel) and number of OTUs
338 (right panel) for different domains and taxonomic levels in the resident microbiota are shown.
339 Note that the relative abundance of Bacteria vs. Eukaryotes does not necessarily reflect
340 organismal abundances on the sampling site, but the amplicon relative abundance after PCR.
341 Relative abundances were calculated for each year and aggregated over the corresponding
342 months along the 10 years for the resident microbiota, then split into size fractions (NB: relative
343 abundance for both domains and size fraction sums up to 1 for each month across ten years).
344 *Lower panels:* Core microbiota over 10 years. The relative abundances of core OTUs reflect the
345 remaining proportions after removing all the OTUs that were not strongly associated when
346 building networks. Relative OTU abundance (left panel) and number of OTUs (right panel) for
347 different domains and taxonomic levels among the core OTUs.

348

349 The most abundant bacteria (**Figure 3**; **Table S2, Additional file 1**) among the
350 core OTUs were Oxyphotobacteria (mostly *Synechococcus*), total abundance ~14% of
351 the resident microbiota, followed by Alphaproteobacteria, with SAR11 clades Ia and II
352 representing ~9% and ~2% respectively. The most abundant microbial eukaryotic
353 groups were *Micromonas*, *Bathycoccus*, Dinophyceae, and Cryptomonadales (each
354 ~2%) [**Figure 3**; **Table S3, Additional file 1**]. In terms of diversity and abundance,
355 bacterial non-phototrophs (including chemoautotrophs, photoheterotrophs,
356 chemoheterotrophs) were the most prevalent in the core microbiota, representing ~62%
357 of the OTUs and a quarter of the total relative read abundance (**Table 1**). In turn,
358 protistan heterotrophs represented a minor fraction of the diversity and relative
359 abundance **(Table 1)**. Bacteria photoautotrophs were relatively more abundant than

360 their protistan counterparts but less diverse (**Table 1**). Protistan parasites represented
361 ~8% of the OTUs and ~3% of the abundance, while the remaining protistan lifestyles
362 had a minor relevance in the core microbiota (**Table 1**).
363
364 *Intra- and cross-domain core associations*
365 Bacteria tended to be associated with other bacteria (**Table 3 & 4**; **Figure 2B**), with
366 Bacteria-Bacteria associations making up ~54% of all associations, while Protist-Protist
367 associations accounted for 11% (**Table 4**). The connectivity of the bacterial
368 subnetworks was higher (mean degree ~10) than the protist counterparts (mean degree
369 ~6), regardless of whether these networks included exclusively bacteria, protists, or
370 both (**Table 3**).
371 In particular, there was a substantial number of associations between Alpha-
372 and Gammaproteobacteria, between Alphaproteobacteria and Acidiimicrobia as well as
373 among Alphaproteobacteria OTUs (**Figure 2B**). Eukaryotic OTUs did not show a
374 similar trend with associations between OTUs of the same taxonomic ranks (**Figure**
375 **2B**). In terms of cross-domain associations, Alphaproteobacteria OTUs had several
376 associations with most major protistan groups (i.e. dinoflagellates, diatoms,
377 cryptophytes, Mamiellophyceae, and Syndiniales) [**Figure 2B**].
378
379 *Core associations within the pico- and within the nanoplankton*
380 While the pico- and nano-size fractions indicate different lifestyles in bacteria (free-
381 living or particle-attached), they indicate different cell sizes in protists, and this could
382 be reflected in association networks. Nanoplankton sub-networks were larger and more
383 connected than picoplankton counterparts (**Figure 4**, **Table 3**). This pattern was
384 observed in both sub-networks considering associations from the same or both size

385 fractions (**Table 3**). Nanoplankton sub-networks had a higher average degree (~10)
386 than picoplankton sub-networks (~7; Wilcoxon $p<0.05$), while not differing much in
387 other network statistics (**Table 3**). Most associations in the pico- and nanoplankton
388 were positive (>93%), while the associations between OTUs from different size
389 fractions represented only ~17% of the total, being ~99% positive (**Table 2**).

390 In the pico- or nanoplankton sub-networks that include OTUs from the same
391 size fraction, the number of bacterial core OTUs was higher than the protistan
392 counterparts (103 bacterial vs. 47 protistan OTUs in the nanoplankton, and 79 bacterial
393 vs. 30 protistan OTUs in the picoplankton) (**Figure 4**, **Table 3**). Still, core OTUs in
394 both the pico- and nanoplankton had comparable sequence abundances: ~27% of the
395 resident microbiota in each size fraction. Within the picoplankton, 64% of the
396 associations were between bacteria, 8% between eukaryotes, and 25% between
397 eukaryotes and bacteria (**Table 4**). In turn, in the nanoplankton, 50% of the edges were
398 between bacteria, 14% between eukaryotes, and 31% between eukaryotes and bacteria
399 (**Table 4**). Overall, the BBMO pico- and nanoplankton sub-networks differed in size,
400 connectivity, and taxonomic composition, while they were similar in terms of positive
401 connections and relative sequence abundance.

402

403 **Table 3.** Core network and sub-networks statistics.

Network	Nodes (#OTUs)	Edges	Di.	De.	Average degree	Average path length	Average clustering coefficient	Largest clique (#)	Mod.
Core network	262 (259)	1,411	11	0.04	10.7	3.45	0.52	13 (4)	0.19
Random core network	262	1,411	5	0.04	10.7	2.60	0.03	3(199)	0.13
Picoplankton all [1]	161 (160)*	620*	10	0.05	7.7	3.13	0.55	10(1)	0.22
Picoplankton only [2]	110 (109)	378	9	0.06	6.9	3.15	0.51	9(4)	0.29
Nanoplankton all [3]	197 (194)*	1,033*	10	0.05	10.5	3.18	0.57	13(4)	0.15
Nanoplankton only[4]	153 (150)	791	10	0.07	10.3	3.21	0.56	13(4)	0.17
Bacteria all [5]	233 (230)**	1,236**	10	0.04	10.6	3.34	0.52	11(3)	0.19
Bacteria only [6]	185 (182)	803	10	0.05	8.7	3.50	0.51	10(1)	0.31
Protists all [7]	147 (145)**	608**	5	0.06	8.3	2.40	0.48	8(2)	0.10
Protist only [8]	80 (77)	175	5	0.05	4.4	2.54	0.54	7(1)	0.32

404
405 NB: Networks and sub-networks include OTUs and environmental factors. Di=Network diameter. De=Network density. Largest clique = size of
406 the largest clique(s) in the network, and in brackets, the number of them. Mod = Network modularity inferred using edge betweenness. [1]All
407 associations where picoplankton OTUs are involved (including nanoplankton); [2]Associations between picoplankton OTU only; [3]All associations
408 where nanoplankton OTUs are involved (including picoplankton); [4]Associations between nanoplankton OTU only; [5]All associations where
409 bacterial OTUs are involved (including protists); [6]Associations between bacterial OTU only; [7]All associations where protist OTUs are involved

(including bacteria); [8]Associations between protist OTU only. * Includes nodes and edges shared between pico- and nanoplankton. ** Includes nodes and edges shared between bacteria and protists.

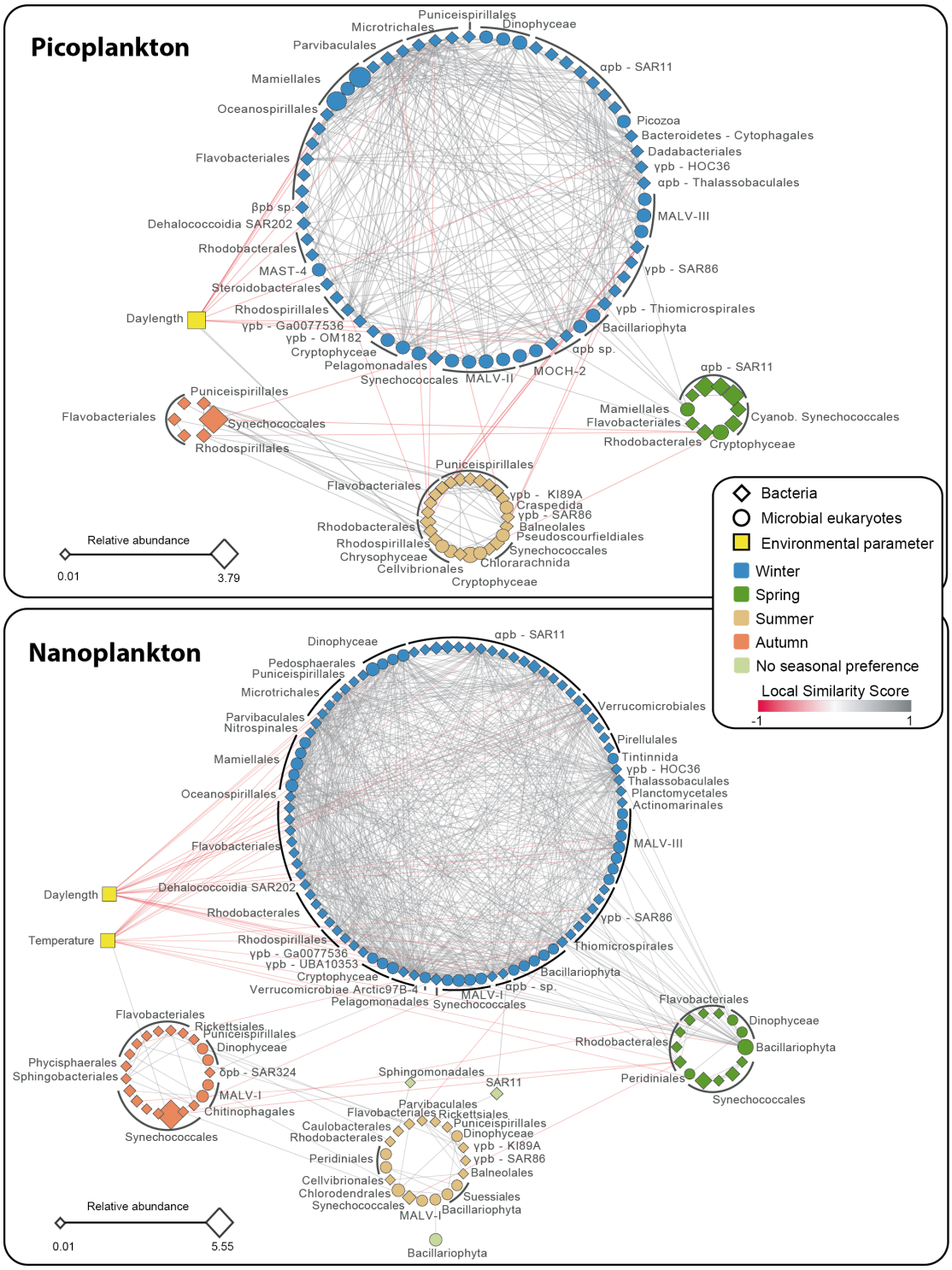


Figure 4. Pico- and nanoplankton core sub-networks. The shape of the nodes indicates bacteria (rhomboid) or microbial eukaryotes (circle), and the color of nodes represents species

417 seasonal preferences, determined using the indicator value ($p<0.05$). The color of the edges
418 indicates if the association is positive (grey) or negative (red). Node size indicates OTU relative
419 abundance from the core microbiota.
420
421 **Table 4.** Core associations within and between taxonomic domains and size fractions.

Network	Association type[1]	# Associations
Core network	Total	1,411
	Bacteria - Bacteria	767 (54%)
	Bacteria - Protist	433 (31%)
	Protist - Protist	161 (11%)
	Environmental factor - Bacteria	36 (3%)
	Environmental factor - Protist	14 (1%)
Picoplankton subnetwork	Total	378
	Bacteria - Bacteria	241 (64%)
	Bacteria - Protist	94 (25%)
	Protist - Protist	31 (8%)
	Environmental factor - Bacteria	12 (3%
	Environmental factor - Protist	0 (0%)
Nanoplankton subnetwork	Total	791
	Bacteria - Bacteria	394 (50%)
	Bacteria - Protist	246 (31%)
	Protist - Protist	113 (14%)
	Environmental factor - Bacteria	24 (3%)
	Environmental factor - Protist	14 (2%)

422 [1] "Bacteria – Bacteria" indicates associations between two bacterial OTUs. "Protist – Protist" are associations between two unicellular eukaryotes
423 and "Bacteria – Protist" are associations between one eukaryote and one bacterial OTU. "Environmental factor – Protist" and "Environmental
424 factor – Bacteria" are associations between an environmental factor and a eukaryotic or bacterial OTU.
425

426 *Network seasonality*

427 The indicator value (IndVal) was used to infer the seasonal preference of core OTUs.
428 Most of the core OTUs (98%; 254 out of 259 OTUs) showed a clear preference for one
429 of the four seasons, pointing to a marked seasonality in the core microbiota (**Figure 4;**
430 **Table 5; Tables S4 & S5, Additional file 1**). Winter had the highest quantity of core
431 OTUs and the highest network connectivity (average degree ~13), compared to the
432 other seasons (average degrees ~2 – ~6) [**Figure 4**; **Table 5**]. The average path length
433 was larger in the core network compared to a random network of the same size (**Table**
434 **3**). Yet, all sub-networks associated with size fractions and seasons (**Table 5**) had
435 shorter path lengths than the random network, indicating that nodes tended to be

436 connected within seasons and size fractions. This was also supported by an increase in
437 network density when comparing the core network (**Table 3**) and the core network
438 subdivided into seasons (**Table 5**), against the core network subdivided into both
439 seasons and size fractions (**Table 5**). The five OTUs that did not show any seasonal
440 preference, among them SAR11 Clades Ia & II, showed high to moderate abundances
441 but had a low number of associations to other OTUs (**Tables S4, S5, S6, Additional**
442 **file 1**). Thus, network connectivity in the BBMO appears to be heterogeneous over
443 time, peaking in winter and remaining low in the other seasons.
444

445 **Table 5:** Subnetworks including core OTUs displaying seasonal preference.
446

	Sub-network	Number of OTUs	Edges	Di.	De.	Average degree	Average path length	Average clustering coefficient	Largest clique (#)	Mod.
All	Winter	156	1,175	7	0.10	15.1	2.62	0.54	13(4)	0.19
	Spring	19	16	4	0.09	1.7	1.56	0.44	4(1)	0.75
	Summer	41	56	7	0.07	2.7	2.90	0.49	6(1)	0.53
	Autumn	26	25	3	0.08	1.9	1.59	0.46	4(2)	0.73
Pico	Winter	63	286	6	0.15	9.1	2.35	0.53	9(4)	0.10
	Spring	8	5	3	0.18	1.2	1.50	0.00	2(5)	0.56
	Summer	25	36	5	0.12	2.9	2.20	0.41	6(1)	0.23
	Autumn	5	3	2	0.30	1.2	1.25	0.00	2(3)	0.44
Nano	Winter	92	658	6	0.16	14.3	2.40	0.61	13(4)	0.04
	Spring	11	11	4	0.20	2.0	1.59	0.57	4(1)	0.56
	Summer	13	17	3	0.22	2.6	1.70	0.65	4(1)	0.50
	Autumn	17	18	3	0.13	2.1	1.35	0.56	4(2)	0.60

447 NB: Subnetworks include OTUs only. Di=Network diameter. De=Network density. Largest clique = size of the largest clique(s) in the network,
448 and in brackets, the number of them. Mod = Network modularity inferred using edge betweenness.
449

450 *Groups of highly associated OTUs*

451 Within the core network, we identified groups that were more connected to each other
452 than to the rest of the network (called modules). These groups of OTUs may indicate
453 recurring associations that are likely important for the stability of ecosystem function.
454 We identified 12 modules in both the pico- and nanoplankton subnetworks (**Table S7,**
455 **Additional file 1**). Modules tended to include OTUs from the same season (**Table S8,**
456 **Additional file 1**), with main modules (i.e. MCODE score >4) including OTUs
457 predominantly associated with winter, summer, and autumn (**Figure 5**). Overall, winter

458 modules prevailed (5 out of 7) among the main modules (**Figure 5**), while modules
459 with scores ≤ 4 did not tend to be associated with a specific season (**Table S8,**
460 **Additional file 1**). Two main winter modules had members that were negatively
461 correlated to temperature and daylength (**Figure 5**; Modules 1 and 4, nanoplankton).
462 The total relative sequence abundance of core OTUs included in modules was
463 ~24% (proportional to the resident microbiota), while the total abundance of individual
464 modules ranged between ~6% and ~0.3% (**Table S7, Additional file 1**). In turn, the
465 relative abundance of core OTUs included in modules ranged between 0.01% and ~2%
466 (**Table S8, Additional file 1**). In most modules, a few OTUs tended to dominate the
467 abundance, although there were exceptions, such as module 4 of the picoplankton,
468 where all SAR11 members featured abundances >1% (**Table S8, Additional file 1**). In
469 addition, several OTUs within modules had relatively low abundances (**Table S8,**
470 **Additional file 1**), supporting modules as a real feature of the network and not just the
471 agglomeration of abundant taxa.
472
473 *Central OTUs*
474 Biological networks typically contain nodes (i.e. OTUs) that hold more "central"
475 positions in the network than others [22]. Even though the ecological role of these hub
476 and connector OTUs is unclear, it is acknowledged that they could reflect taxa with
477 important ecological functions [22]. There is no universal definition for hub or
478 connector OTUs, yet, in this work, we have used stringent thresholds to determine them
479 *ad hoc* (see Methods). We have identified 13 hub-OTUs that were associated with
480 winter or spring (**Table 6**). Hubs did not include highly abundant OTUs, such as
481 *Synechococcus* or SAR11 (**Table 6**), but instead, they included several OTUs with
482 moderate-low abundance (<1%) and high degree (ranging between 26-60) [**Table 6**].

483 For example, the Gammaproteobacteria OTU bn_000226 had a relative abundance of
484 0.04% and a degree of 60 **(Table 6)**. Hubs included other moderately abundant OTUs,
485 such as the eukaryotic picoalgae *Bathycoccus,* which was abundant in winter, as well
486 as an unidentified dinoflagellate (**Table 6**).

487 We identified a total of 18 connector OTUs (featuring relatively low degree and
488 high centrality), which were predominantly associated with summer (5 out of 18) or
489 autumn (6 out of 18), contrasting with hub OTUs, which were associated mostly with
490 winter and spring (**Table 6**). Connectors may be linked to the seasonal transition
491 between main community states (**Figure 1 C & D**) and included several abundant
492 OTUs belonging to *Synechococcus* and SAR11 (**Table 6**). In particular, the SAR11
493 OTU bp_000007 displayed a relatively high abundance (1.4%), but a degree of 3
494 (relatively low) and a betweenness centrality of 0.6 (relatively high). In contrast, two
495 protist OTUs displayed low-moderate abundances (ep_00269, Chrysophyceae,
496 abundance 0.04% and en_00161, Syndiniales, abundance 0.4%), low degree <4, but a
497 high betweenness centrality (>0.8; **Table 6**).
498
499
500
501
502
503
504
505
506
507

508

509 **Table 6.** Central OTUs.

OTU	Class	Lowest rank taxonomy	Relative Abundance (%)[1]	Degree	Betweenness Centrality	Closeness Centrality	Season
Hubs							
en_00092	Mamiellophyceae	*Bathycoccus*	0.51	42	0.04	0.42	Winter
en_00119	Dinophyceae	-	0.41	50	0.03	0.42	Winter
bp_000037	Alphaproteobacteria	Parvibaculales_OCS116	0.31	45	0.08	0.43	Winter
bp_000039	Gammaproteobacteria	SUP05_cluster	0.28	29	0.12	0.41	Spring
bn_000039	Gammaproteobacteria	SUP05_cluster	0.21	42	0.17	0.44	Spring
bn_000037	Alphaproteobacteria	Parvibaculales_OCS116	0.20	40	0.05	0.42	Spring
bp_000059	Gammaproteobacteria	SAR86	0.20	24	0.09	0.40	Spring
ep_00070	Cryptophyceae	Cryptomonadales_X	0.13	40	0.04	0.42	Winter
bn_000059	Gammaproteobacteria	SAR86	0.12	24	0.03	0.40	Spring
bn_000102	Alphaproteobacteria	Nisaeaceae_OM75	0.09	26	0.03	0.38	Winter
bp_000193	Alphaproteobacteria	-	0.06	37	0.03	0.40	Winter
bn_000170	Acidimicrobiia	Sva0996_marine_group	0.06	59	0.06	0.44	Winter
bn_000226	Gammaproteobacteria	HOC36	0.04	60	0.06	0.43	Winter
Connectors							
bp_ 000001	Oxyphotobacteria	*Synechococcus* (CC9902)	3.79	5	0.05	0.30	Autumn
bp_ 000002	Alphaproteobacteria	SAR11 Clade_Ia	2.26	2	0.40	0.56	Spring
bp_ 000004	Alphaproteobacteria	SAR11 Clade_Ia	2.02	3	0.15	0.63	NA
bp_ 000007	Alphaproteobacteria	SAR11 Clade_Ia	1.38	3	0.60	0.71	NA
bp_ 000008	Alphaproteobacteria	SAR11 Clade_Ia	1.15	3	0.15	0.63	NA
bn_ 000008	Alphaproteobacteria	SAR11 Clade_Ia	0.68	5	0.03	0.27	Winter
en_ 00059	Chlorodendrophyceae	*Tetraselmis*	0.66	4	0.05	0.26	Summer
bn_ 000020	Oxyphotobacteria	-	0.56	3	0.60	0.67	Autumn
en_ 00161	Syndiniales	Syndiniales-Group-I-Clade-4_X	0.42	4	0.80	0.75	Autumn
bn_ 000018	Oxyphotobacteria	*Prochlorococcus* MIT9313	0.41	5	0.04	0.24	Winter
bn_ 000054	Alphaproteobacteria	Puniceispirillales_SAR116	0.11	4	0.14	0.40	Autumn
bn_ 000062	Alphaproteobacteria	Puniceispirillales_SAR116	0.08	3	0.55	0.50	Autumn
bn_ 000077	Rhodothermia	*Balneola*	0.07	3	0.17	0.32	Summer
bn_ 000112	Gammaproteobacteria	KI89A	0.06	4	0.53	0.48	Summer
bn_ 000156	Alphaproteobacteria	Parvibaculales_PS1	0.05	4	0.14	0.40	Summer
bn_ 000281	Bacteroidia	Sphingobacteriales_NS11-12	0.05	5	0.16	0.44	Autumn
bn_ 000221	Alphaproteobacteria	Puniceispirillales_SAR116	0.04	5	0.05	0.30	Winter
ep_ 00269	Chrysophyceae	Clade-I_X	0.04	2	1.00	1.00	Summer

510 [1] Proportional to the resident microbiota

511

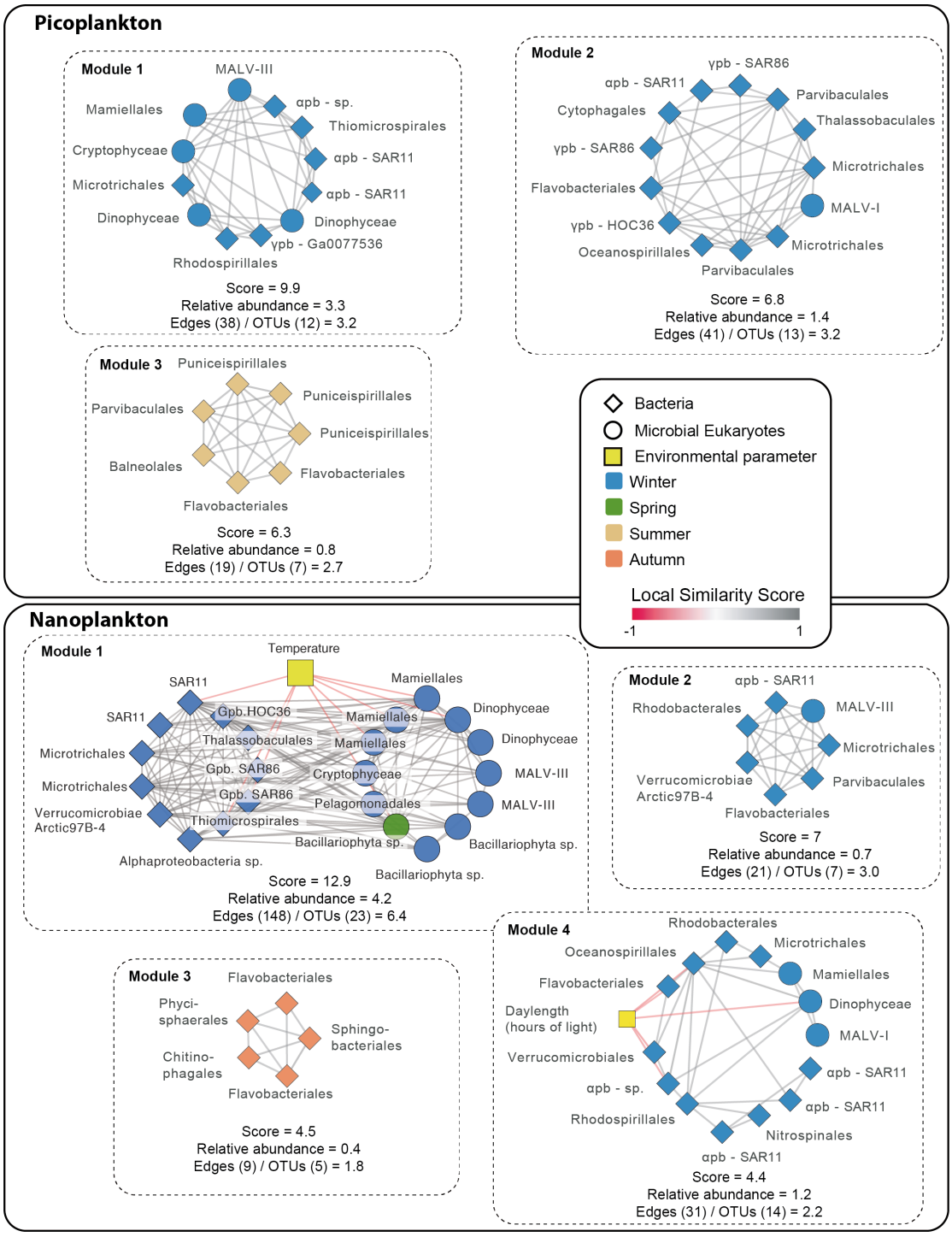


Figure 5. Main modules in the core network. Modules with MCODE score >4 are shown for picoplankton (upper panel) and nanoplankton (lower panel). For each module, the MCODE score and relative amplicon abundance of the taxa included in it (as % of the resident microbiota) are indicated. In addition, the numbers of edges and OTUs within the modules are shown as edges/OTUs; this quotient estimates the average number of edges per OTU within the different

519 modules. The edges represent correlations with |LS| > 0.7, |ρ|>0.7, P<0.001 and Q<0.001. The
520 color of the edges indicates positive (grey) or negative (red) associations. The shape of nodes
521 indicates bacteria (rhomboid) or microbial eukaryotes (circle), and the color of nodes represents
522 species seasonal preferences, determined using the indicator value (p<0.05). pb =
523 Proteobacteria
524
525
526 **DISCUSSION**

527 Identifying the most important microbes for the functioning of the ocean ecosystem is
528 a challenge, which can be addressed by delineating core microbiotas [4]. Recognizing
529 the most abundant and widespread microbes in the ocean is a step towards knowing the
530 core microbiota. However, this does not take into account the importance that both
531 microbial interactions and microbes with moderate or low abundance may have for the
532 functioning of ecosystems [4, 29, 38]. Considering potential interactions when
533 delineating core microbiotas may not only allow identifying moderate/low abundance
534 taxa that may have important roles in the community but could also allow excluding
535 taxa that are present in several locations but that may not have an important role for
536 community function (e.g., dormant cells or cells being dispersed [8]). Here, we have
537 delineated and analyzed the core microbiota of a coastal ecosystem-based on 10 years
538 of occurrence data considering possible interactions.

539 To detect the core microbiota, we first identified the resident OTUs, that is,
540 those that occur >30% of the time (i.e. >36 out of 120 months) over a decade. This
541 threshold was selected as it allows for seasonal OTUs that would be present recurrently
542 in at least one season. Analysis of the resident OTU dynamics indicated a clear
543 seasonality (**Figure 1 C-D**), and that the measured environmental factors could explain
544 ~45% of the resident microbiota variance. The main environmental drivers were

545 temperature and daylength, which is consistent with previous works from the same
546 time-series (BBMO) [34, 39, 40]. These values are lower than what has been reported
547 for bacteria in the English Channel, where daylength explains ~65% of community
548 variance [17], and higher than what has been reported for entire communities in the
549 time-series SPOT (California, 31%) [41] or SOLA (the Mediterranean Sea, ~130 km
550 from BBMO; 7-12%) [42]. Daylength may be more important in the English Channel
551 as it has a more pronounced annual variation than at BBMO, whereas the measured
552 differences could reflect a higher coupling of the resident OTUs with environmental
553 variation in BBMO than in SOLA or SPOT. SOLA is characterized by the occasional
554 winter storms that bring nutrients from the sediments to the water column as well as by
555 the freshwater inputs from nearby rivers during flash floods [43], and this could
556 partially explain the differences with BBMO. The importance of daylength and
557 temperature for community dynamics was reflected by niche analyses, which identified
558 two main niches associated with summer and winter at the BBMO, to which ~50% of
559 the resident OTUs were associated (**Figure 1 E-F)**. Other resident OTUs likely have
560 spring and fall niches as indicated by **Figure 1 C-D**, yet these niches cannot be detected
561 with the used null model analysis, as their preferred temperatures or daylengths will not
562 depart significantly from the randomized mean.

563 Based on the resident OTUs, we built networks to define the core microbiota.
564 We identified a total of 259 core OTUs (182 bacteria and 77 protists) that represented
565 64% of the abundance of the resident microbiota and that showed seasonal variation.
566 We could only find supporting evidence from the literature (PIDA database) [21] for
567 85 associations of the core (6 %), indicating that most of them still need to be validated
568 with direct observation or experimentally. This is not surprising, as the most studied
569 hosts in PIDA are protists from the micro-plankton (>20 μm cell size), which are mostly

570 absent from our pico- and nanoplankton networks. Also, PIDA does not cover Bacteria-
571 Bacteria associations. Nevertheless, the detected core OTUs from BBMO represent a
572 fraction of the core microbiota at this site, since larger microbial size fractions were not
573 sampled. Including these larger size fractions would expand the composition of the core
574 and could unveil additional patterns. For example, in a global ocean network including
575 size fractions >20 µm cell size, protists or small multicellular eukaryotes dominated the
576 interactome [26].

577 Alpha-/Gammaproteobacteria, Bateroidia, Acidimicrobiia were the main
578 bacterial groups in the core, including also common marine taxa, such as
579 *Synechococcus* or SAR11. The main protists in the core included Syndiniales
580 (parasites), Dinoflagellates, Mammiellales (*Micromonas* and *Bathycoccus*), and
581 diatoms. These taxa are likely the most important in sustaining ecosystem function at
582 BBMO, and probably have similar importance in other coastal areas. Other studies have
583 reported important roles in marine association networks for SAR11 and *Synechococcus*
584 [31, 44]. Syndiniales, Haptophytes, and Dinoflagellates dominated networks in terms
585 of the number of nodes and edges at SPOT, while Mamiellales (*Micromonas* &
586 *Bathycoccus*) and diatoms also had relevant roles [41]. Syndiniales, Dinoflagellates,
587 and Diatoms were also predominant in global ocean networks, which is coherent with
588 our results [26].

589 Bacteria-Bacteria associations were the most abundant (54%) in the core
590 BBMO microbiota, followed by Bacteria-Protists (31%) and Protist-Protist (11%)
591 associations. Associations tended to occur among bacteria or protists, rather than
592 between them, in the English Channel time-series [17]. However, the study used
593 microscopy to determine protist community composition, while it used 16S-rRNA gene
594 data for analyzing bacteria communities and this might explain the limited number of

595 connections between protists and bacteria. Most associations occurred among protists
596 in a global-ocean network that included a broad range of microbial size-fractions [26].
597 This suggests that time-series analyses including larger size-fractions may determine a
598 higher proportion of associations among protists, which may turn out to be prevalent.

599 The core network had "small world" properties (that is, high clustering
600 coefficient and relatively short path lengths) [37] when compared to randomized
601 networks (**Table 3**) or particular subnetworks from size fractions or specific seasons
602 (**Table 5**). The small-world topology is characteristic of many different types of
603 networks [45], including marine microbial temporal or spatial networks [23, 26, 30,
604 31]. Some of our network statistics were similar to those obtained at SPOT [23, 30], in
605 particular the averages of degree, clustering coefficient, and path length (**Table 3**).
606 Furthermore, the BBMO network had an average path length similar to a global ocean
607 network [26] and also, similarly to this network, the node degree of the BBMO core
608 members was independent of their relative abundances, showing that the associations
609 between core OTUs were not merely a consequence of high prevalence and abundance.

610 The BBMO core network had a clustering coefficient that was ten times larger
611 than that of an Erdős–Rényi random network of the same size (**Table 3**), which agrees
612 with what was observed at SPOT [23, 30]. The large proportion of positive associations
613 in BBMO networks (~95%) was in agreement with results from other temporal [23, 41]
614 or large-scale spatial [26] microbiota analyses, where positive associations were also
615 predominant (~70-98%), although these values include taxa that are not necessarily part
616 of the core. This suggests that interactions such as syntrophy or symbiotic associations
617 are more important than competition in marine microbial systems and that these types
618 of associations may underpin marine ecosystem function. These findings are also
619 coherent with a recent large-scale literature survey that found that ~47% of the validated

620 associations between protists and bacteria are symbiotic [21]. Nevertheless, it is also
621 possible that common sampling strategies and methodological approaches do not detect
622 a substantial fraction of negative associations. For example, while positive correlations
623 in taxa abundance pointing to positive interactions may be easier to detect, negative
624 associations may be missed due to plummeting species abundances that would prevent
625 establishing significant correlations, or to a delay between the increase and decrease in
626 abundance of interacting taxa that are not synchronized with sampling time. Future
627 studies adapting the sampling scheme to the timing of interactions (e.g., daily or weekly
628 sampling) and the use of other approaches apart from taxa abundances, such as analyses
629 of single-cell genomic data to determine protistan predation, or controlled experiments,
630 will likely generate new insights on negative microbial interactions.

631 The relatively high clustering coefficient of the core network (compared to a
632 random network) and its short path length indicate that most OTUs are connected
633 through < 3 intermediary OTUs. It has been shown that a large proportion of strong
634 positive associations, as in the BBMO core network, may destabilize communities due
635 to positive feedbacks between species [46]. When a species decreases in abundance as
636 a response to environmental variation, it may pull others with it, generating a cascade
637 effect propagated by the many positive associations in the network. Accordingly, the
638 change of abundance in specific OTUs in one section of the network could affect OTUs
639 in other network sections not necessarily affected directly by the environmental
640 variation. This cascade effect may help to explain a paradox: environmental variables
641 affect the structure of marine microbial communities and consequently association
642 networks. Yet, our and others' results [17, 18, 23, 26, 30-32] have reported a limited
643 number of associations between environmental variables and network nodes (OTUs).
644 Environmental heterogeneity might affect network structure by acting on a small subset

645 of nodes (OTUs), which would then influence other nodes through cascading
646 interactions facilitated by the highly interconnected nature of the networks as well as
647 positive feedbacks promoted by the high proportion of positive associations [46].

648 If OTUs susceptible to environmental variation are also highly connected, then
649 their effect on the entire network structure may be larger. In line with this, we found
650 that the connectivity of OTUs associated with environmental variables at BBMO (49
651 OTUs out of 259) had a mean degree of ~25 (SD ~14), while for all the 259 OTUs of
652 the core network, the mean degree was ~11 (SD ~13). The seasonal dynamics of the
653 BBMO microbiota may partially be driven by a subset of OTUs that vary with
654 environmental factors (e.g. temperature, daylength). These may exert a destabilizing
655 influence over the entire community over time, promoting the annual turnover of
656 communities and networks.

657 Most core OTUs (98%) showed a clear preference for one season. Interestingly,
658 the distribution of core OTUs among the seasons was uneven, with 61% of these OTUs
659 showing a winter preference. Network connectivity at BBMO was correspondingly
660 heterogeneous between seasons, peaking in winter and remaining low in the other
661 seasons. Specifically, the winter subnetwork included ~92% of the seasonal edges. This
662 indicates that winter associations are not only specific (i.e. they do not tend to change
663 partners), but they also have a relatively high recurrence (otherwise, winter networks
664 would be smaller). A higher similarity between winter communities when compared to
665 other seasons was also indicated by our ordination analyses of the resident OTUs
666 (**Figure 1**), as well as by studies of the entire protist community at BBMO [34] or whole
667 community analyses at SPOT [23].

668 The structure of communities is determined by the interplay of selection,
669 dispersal, speciation, and ecological drift [47]. Our results indicate that selection, a

670 deterministic process, is stronger in winter, leading to winter sub-communities that tend
671 to be more similar between each other than to communities from other seasons. Given
672 that we have removed edges associated with the measured environmental variables, we
673 do not expect that the identified edges between winter OTUs represent selection
674 associated to these variables (e.g. low temperature). Consequently, winter edges may
675 represent associations linked to unmeasured variables or ecological interactions that
676 may be more likely to develop during winter due to stronger environmental selection.
677 Due to weaker selection in other seasons species occurrence would display less
678 recurrent (or more random) patterns, preventing specific associations to be formed. This
679 also suggests that ecological redundancy changes over time, and is lower in winter
680 compared to the other seasons (even though the number of OTUs is larger in winter).
681 A reduction in redundancy may also promote strong ecological interactions in winter.

682 The existence of subsets of species that interact more often between themselves
683 than with other species (modules), is characteristic of biological networks, and can
684 contribute to overall network stability [48, 49]. Modules can represent divergent
685 selection, niches, the clustering of evolutionary closely related species or co-
686 evolutionary units [50, 51]. Modules in the core BBMO network (total 12) included
687 positive associations between diverse taxa, and could represent divergent selection,
688 driven by unmeasured environmental variables, or examples of syntrophic or symbiotic
689 interactions between microbes from different taxonomic groups.

690 Most BBMO modules included diverse lifestyles (heterotrophs, mixotrophs,
691 phototrophs, parasites), similar to what has been observed at SPOT [41]. Yet, a number
692 of modules appeared to be predominantly heterotrophic or autotrophic (**Table S8,**
693 **Additional file 1**). Some modules included OTUs from the same species, such as
694 Module 4 in the picoplankton, which included several SAR11 Clade I OTUs, and

695 Module 7 of the nanoplankton, which included several *Synechococcus* OTUs. These
696 modules could reflect similar niches, associated with unmeasured variables, or the
697 dependence on metabolites produced by other organisms (auxotrophy). There is
698 evidence of auxotrophy for both SAR11 (e.g. thiamin, glycine)[52-54] and
699 *Synechococcus* (e.g. cobalamin) [55]. Recently it has been observed in co-culture
700 experiments that *Prochlorococcus* may fulfill some metabolic requirements of SAR11,
701 promoting the growth of the latter in a commensal relationship [56]. In our analyses of
702 the BBMO core microbiota, we did not find strong associations between SAR11 and
703 *Prochlorococcus* or the more abundant relative, *Synechococcus*. Yet, SAR11 formed
704 strong associations with a plethora of taxa with which could potentially have
705 commensal relationships.

706 The overall importance of the observed modules was indicated by the total
707 abundance of their constituent OTUs (24% of the reads compared to the resident
708 microbiota). Most of the modules at BBMO were associated with a single season,
709 suggesting that they reflect seasonal niches. Since these modules were inferred over 10
710 years, they represent recurrent network features. Chafee et al. [57] also identified
711 season-specific modules in a 2-year time series in the North Sea (Helgoland), including
712 samples taken weekly or bi-weekly. These modules were much larger than ours, and
713 they may also include environmentally-driven edges. Nevertheless, the Helgoland
714 modules seem to be driven by eutrophic (spring & summer) vs. oligotrophic (autumn
715 & winter) conditions in this location. In contrast, the BBMO modules, displayed weaker
716 correlations with nutrients and seem to be influenced by temperature and daylength
717 (**Figure 5**). Differences in the sampling scheme between Helgoland and BBMO
718 ((bi)weekly vs. monthly) as well as between both locations (different seas and latitudes,
719 affecting temperature and daylength) may explain these differences.

720 Keystone species have a high influence in ecosystems relative to their
721 abundance [58]. Network analyses may help to identify them [24, 59], yet, there is no
722 clear consensus of what network features are the best unequivocal indicator of keystone
723 species [60-62]. Therefore, we focused on identifying central OTUs (hubs or
724 connectors) that may be important for ecosystem function [22, 24] and could represent
725 keystone species. We identified 13 hubs in the BBMO core network with moderate-low
726 abundances (<1%) and high degree (26-60) that were associated with winter or spring.
727 These moderate-low abundance OTUs may affect nutrient cycling directly [63] or
728 indirectly, by affecting other OTUs with higher abundance. The putative stronger
729 selection exerted by low temperatures and short daylengths during winter and early
730 spring, as compared to summer and autumn, may lead to a higher species recurrence
731 [34], larger networks, and possibly, more hubs. An OTU of the abundant picoalgae
732 *Bathycoccus* (en_00092) was identified as a winter hub, which is consistent with
733 reported *Bathycoccus* abundance peaks in late winter (February-March) in both BBMO
734 [64] and the nearby station SOLA [42]. This *Bathycoccus* hub may be associated with
735 diverse taxa, such as prokaryotes that may benefit from algal exudates [65] or even via
736 mixotrophy [66]. In agreement with this, out of the 42 associations of this hub OTU,
737 25 were with bacteria and the rest with protists.

738 In contrast to hubs, connector OTUs were predominantly associated with
739 warmer waters, that is, summer and autumn, and may represent transitions in
740 community states. This was consistent with the associations observed in an abundant
741 *Synechococcus* connector OTU (bp_000001, **Table 6**). This OTU was predominant in
742 summer-autumn, in agreement with previous BBMO reports [36, 67], but it was
743 associated with other OTUs from spring (negative association with bp_000017), winter
744 (negative association with bp_000039), summer (positive association with bp_000087,

745 bp_000012) and autumn (positive association with bp_000022), thus likely holding a
746 central position in the network. Another abundant spring connector OTU (SAR11 Clade
747 Ia, bp_000002), featured only two connections to spring (positive association with
748 bp_000007) and summer (positive association with bp_000046) OTUs.
749

750 **CONCLUSION**

751 Our decade-long analysis of the dynamics of a microbiota populating a time-series in
752 the Mediterranean Sea allowed us to determine the interconnected core microbiota,
753 which likely includes several microbes that are important for the functioning of this
754 coastal ecosystem. We found a relatively small core microbiota that displayed seasonal
755 variation, with a heterogeneous distribution of associations over different seasons,
756 indicating different degrees of recurrence and selection strength over the year. Future
757 analyses of other core marine microbiotas will determine how universal are the patterns
758 found in BBMO. These studies will be crucial to determine potential long-term effects
759 of climate change on the architecture of the interaction networks that underpin the
760 functioning of the ocean ecosystem.
761

762 **METHODS**

763 *Study site and sampling*

764 Surface water (~1 m depth) was sampled monthly from January 2004 to December
765 2013 at the Blanes Bay Microbial Observatory (BBMO) in the Northwestern
766 Mediterranean Sea (41º40'N, 2º48'E) [**Figure 1A**]. The BBMO is an oligotrophic
767 coastal site ~1 km offshore with ~20 m depth and with limited riverine or human
768 influence [36]. Seawater was pre-filtered with a 200 µm nylon mesh and then
769 transported to the laboratory in 20 L plastic carboys and processed within 2 hours.

770 Microbial plankton from about 6 L of the pre-filtered seawater was separated into two
771 size fractions: picoplankton (0.2-3 µm) and nanoplankton fraction (3-20 µm). To
772 achieve this, the seawater was first filtered through a 20 µm nylon mesh using a
773 peristaltic pump. Then the nanoplankton (3-20 µm) was captured on a 3 µm pore-size
774 polycarbonate filter. Subsequently, a 0.2 µm pore-size Sterivex unit (Millipore,
775 Durapore) was used to capture the picoplankton (0.2-3 µm). Sterivex units and 3 µm
776 filters were stored at -80 ºC until further processed. The sequential filtering process
777 aimed to capture free-living bacteria and picoeukaryotes in the 0.2-3 µm size fraction
778 (picoplankton), and particle/protist-attached bacteria or nanoeukaryotes in the 3-20 µm
779 fraction (nanoplankton). The 3µm filter was replaced if clogging was detected; DNA
780 from all 3µm filters from the same sample were extracted together.

781 A total of 15 contextual abiotic and biotic variables were considered for each
782 sampling point: Daylength (hours of light), Temperature (°C), Turbidity (estimated as
783 Secchi disk depth [m]), Salinity, Total Chlorophyll a [Chla] (µg/l), PO_4^{3-} (µM), NH_4^+
784 (µM), NO_2^- (µM), NO_3^- (µM), SiO_2 (µM), abundances of Heterotrophic prokaryotes
785 [HP] (cells/ml), *Synechococcus* (cells/ml), Total photosynthetic nanoflagellates [PNF;
786 2-5µm size] (cells/ml), small PNF (2µm ; cells/ml) and, Heterotrophic nanoflagellates
787 [HNF] (cells/ml) [**Figure 1B**]. Water temperature and salinity were sampled *in situ* with
788 a SAIV A/S SD204 CTD. Inorganic nutrients (NO_3^-, NO_2^-, NH_4^+, PO_4^{3-}, SiO_2) were
789 measured using an Alliance Evolution II autoanalyzer [68]. Cell counts were done by
790 flow cytometry (heterotrophic prokaryotes, *Synechococcus*) or epifluorescence
791 microscopy (PNF, small PNF and HNF). See Gasol *et al.* [36] for specific details on
792 how other variables were measured. Environmental variables were z-score standardized
793 before running statistical analysis.
794

795 *DNA extraction, sequencing, and metabarcoding*

796 DNA was extracted from the filters using a standard phenol-chloroform protocol [69],
797 purified in Amicon Units (Millipore), and quantified and qualitatively checked with a
798 NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific). Eukaryotic PCR
799 amplicons were generated for the V4 region of the 18S rDNA (~380 bp), using the
800 primer pair TAReukFWD1 and TAReukREV3 [70]. The primers Bakt_341F [71] and
801 Bakt_806RB [72] were used to amplify the V4 region of the 16S rDNA. PCR
802 amplification and amplicon sequencing were carried out at the Research and Testing
803 Laboratory (http://rtlgenomics.com/) on the *Illumina* MiSeq platform (2x250 bp paired-
804 end sequencing). DNA sequences and metadata are publicly available at the European
805 Nucleotide Archive (http://www.ebi.ac.uk/ena; accession numbers PRJEB23788 for
806 18S rRNA genes & PRJEB38773 for 16S rRNA genes).

807 A total of 29,952,108 and 16,940,406 paired-end *Illumina* reads were produced
808 for microbial eukaryotes and prokaryotes respectively. Adapters and primers were
809 removed with Cutadapt v1.16 [73]. DADA2 v1.10.1 [74] was used for quality control,
810 trimming, and inference of Operational Taxonomic Units (OTUs) as Amplicon
811 Sequence Variants (ASVs). For both microbial eukaryotes and prokaryotes, the
812 Maximum number of expected errors (MaxEE) was set to 2 and 4 for the forward and
813 reverse reads respectively. No ambiguous bases (Ns) were allowed. Microbial
814 eukaryotic sequences were trimmed to 220 bp (forward) and 190 bp (reverse), while
815 prokaryotic sequences were trimmed to 225 bp (both forward and reverse reads). A
816 total of 28,876 and 19,604 OTUs were inferred for microbial eukaryotes and
817 prokaryotes respectively.

818 OTUs were assigned taxonomy using the naïve Bayesian classifier method [75]
819 together with the SILVA version 132 [76] database as implemented in DADA2.

820 Eukaryotic OTUs were also BLASTed [77] against the Protist Ribosomal Reference
821 database (PR^2, version 4.10.0; [78]). When the taxonomic assignments for the
822 eukaryotes disagreed between SILVA and PR^2, the conflict was resolved manually by
823 inspecting a pairwise alignment of the OTU and the closest hits from the two databases.
824 OTUs assigned to Metazoa, Streptophyta, nucleomorph, chloroplast, and mitochondria
825 were removed before further analysis. Archaea were removed from downstream
826 analyses as the used primers are not optimal for recovering this domain [79].

827 Each sample (corresponding to a specific gene, size fraction, and timepoint) was
828 subsampled with the *rrarefy* function from the R package *Vegan* [80] to 4,907 reads,
829 corresponding to the number of reads in the sample with the lowest sequencing depth,
830 to normalize for different sequencing depth between samples. OTUs present in <10%
831 of the samples were removed. After quality control and rarefaction, the number of
832 OTUs was 2,926 (1,561 bacteria, and 1,365 microeukaryotes; **Table 1**).

833 Due to a suboptimal sequencing of the amplicons, we did not use nanoplankton
834 samples of bacteria and protists from the period May 2010 to July 2012 (27 samples)
835 as well as March 2004 and February 2005. OTU read abundance for samples with
836 missing values were estimated using seasonally aware missing value imputation by
837 weighted moving average for time series as implemented in the R package *imputeTS*
838 [81].

839 Cell/particle dislodging or filter clogging during the sequential filtration process
840 may affect the taxonomic diversity observed in the different size fractions, with
841 nanoplankton DNA leaking into the picoplankton fraction, or picoplankton DNA
842 getting stuck in the nanoplankton fraction. To minimize the effects of cell/particle
843 dislodging or filter clogging on the diversity recovered from the different size fractions,
844 we calculated the sequence-abundance ratio for OTUs appearing in both pico- and

845 nano-plankton fractions. When the ratio exceeded 2:1, we removed the OTU from the
846 size fraction with the lowest number of reads. After subsampling and filtering the OTU
847 tables were joined for each time point, and since the samples had been normalized to
848 the same sequencing depth, we calculated the relative read abundance for the OTUs for
849 each year and aggregated over the corresponding months along the 10 years for the
850 resident microbiota. This means that the relative abundance for both domains and size
851 fractions sums up to 1 for each month across ten years.
852
853 *Resident microbiota*
854 We defined *ad hoc* the resident microbiota as the set of OTUs present in >30% of the
855 samples over 10 years (that is, present in >36 months, not necessarily consecutive).
856 This value was chosen as it allows for seasonal OTUs, which may only be present 3-4
857 months each year, and still be considered as part of the resident microbiota. The
858 residents included 355 eukaryotic and 354 bacteria OTUs (**Table 1**), and excluded a
859 substantial amount of rare OTUs, which can cause spurious correlations during network
860 construction due to sparsity [i.e. too many zeros] [22]. The relative abundance of the
861 taxonomic groups included in the resident microbiota was fairly stable from year to
862 year (**Figure 3**).
863
864 *Environmental variation and resident OTUs*
865 All possible correlations among the measured environmental variables and resident
866 OTU richness and abundance were computed in R and plotted with the package
867 *corrplot*. Only significant Pearson correlation coefficients were considered ($p<0.01$),
868 and the p-values were corrected for multiple inference (Holm's method) using the
869 function *rcorr.adjust* from the R package *RcmdrMisc*. Unconstrained ordination

870 analyses were carried out using NMDS based on Bray Curtis dissimilarities between
871 samples including resident OTUs only. Environmental variables were fitted to the
872 NMDS using the function *envfit* from the R package *Vegan* [80]. Only variables
873 displaying a significant correlation ($p<0.05$) were considered. Constrained ordination
874 was performed using distance-based redundancy analyses (dbRDA) in *Vegan*,
875 considering Bray Curtis dissimilarities between samples including resident OTUs only.
876 The most relevant variables for constrained ordination were selected by stepwise model
877 selection using 200 permutations, as implemented in *ordistep* (*Vegan*). Ordinations
878 were plotted using the R package *ggplot2* and *ggord*. The amount of community
879 variance explained by the different environmental variables was calculated with *Adonis*
880 (Vegan) using 999 permutations. Resident OTUs displaying niche preference in terms
881 of Temperature and Daylength, the most important environmental variables, were
882 determined using the function *niche.val* from the R package *EcolUtils* with 1,000
883 permutations.
884
885 *Delineation of seasons*
886 Seasons were defined following Gasol *et al.* [36] with a small modification: months
887 with water temperature (at the sampling time) >17 °C and daylength >14 h d^{-1} were
888 considered to be summer. Months with water temperature <17 °C and < 11 h d^{-1} of
889 daylength were considered to be winter. Months with water temperature >17°C and
890 daylength <14 h d^{-1} were considered as autumn, while months with water temperature
891 <17°C and > 11 h d^{-1} of daylength were considered to be spring. The indicator value
892 [82] was calculated using the R package *labdsv* [83] to infer OTU seasonal preference.
893
894 *Core microbiota delineated using networks*

895 The OTU table together with the 15 environmental variables were used to construct
896 association networks using extended Local Similarity Analysis (eLSA) [84-86]. eLSA
897 was run on the OTU table with subsampled reads with default normalization: a z-score
898 transformation using the median and median absolute deviation. P-value estimations
899 were run under a mixed model that performs a random permutation test of a co-
900 occurrence only if the theoretical p-values for the comparison are <0.05. Bonferroni
901 false discovery rate (q) was calculated for all edges based on the p-values using the
902 *p.adjust* package in R.

903 To detect environmentally-driven associations between OTUs induced by the
904 measured environmental variables we used the program EnDED [87].
905 Environmentally-driven associations indicate similar or different environmental
906 preferences between OTUs and not ecological interactions. In short, EnDED evaluates
907 associations between two OTUs that are both connected to the same environmental
908 variable based on a combination of four methods: *Sign Pattern*, *Overlap, Interaction*
909 *Information,* and *Data Processing Inequality.* These methods use the sign (positive or
910 negative) and the duration of the association, the relative abundance of OTUs as well
911 as environmental parameters to determine if an association is environmentally-driven.
912 If the four methods agreed that an association was environmentally-driven, then it was
913 removed from the network. The initial number of edges was 199,937, of which 180,345
914 were OTU-OTU edges that were at least in one triplet with an environmental parameter.
915 In total 65,280 (~33%) edges in the network were identified as indirect by EnDED and
916 removed. Afterward, only edges representing the strongest associations (i.e., absolute
917 local similarity score $|LS| > 0.7$, Spearman correlation $|\rho| > 0.7$, $P<0.001$ and $Q<0.001$)
918 and nodes representing the resident OTUs were retained for downstream analysis and
919 are hereafter referred to as "core associations". Those OTUs participating in core

920 associations were defined as core OTUs, although their involvement in ecological
921 interactions need further experimental validation. Both core associations and core
922 OTUs constitute the "core network", which also represents the core microbiota (both
923 "core network" and "core microbiota" are used indistinctively). The core network was
924 randomized using the Erdős–Rényi model [88], using 262 nodes and 1,411 edges.

925 For the core network, we calculated: 1) *Density:* quantifies the proportion of
926 actual network connections out of the total number of possible connections, 2)
927 *Transitivity* or *Clustering coefficient:* measures the probability that nodes connected to
928 a node are also connected, forming tight clusters, 3) *Average path length:* mean number
929 of steps (edges) along the shortest paths for all possible pairs of nodes in the network
930 (a low average path length indicates that most species in the network are connected
931 through a few intermediate species), 4) *Degree:* number of associations per node, 5)
932 *Betweenness centrality:* measures how often an OTU (node) appears on the shortest
933 paths between other OTUs in the network, 6) *Closeness centrality:* indicates how close
934 a node is to all other nodes in a network, 7) *Cliques:* refers to sets of interconnected
935 nodes where all possible connections are realized, 8) *Modularity:* measures the division
936 of a given network into modules (that is, groups of OTUs that are highly interconnected
937 between themselves).

938 The Degree, Betweenness centrality and Closeness centrality were used to
939 identify central OTUs using *ad hoc* definitions. "Hub" OTUs were those with a score
940 above the average for the three statistics and were normally among the top 25% in each
941 score [22, 62, 89]. Specifically, hub OTUs featured a degree >24, Betweenness
942 centrality >0.03 and Closeness centrality >0.3. Similarly, "connector" OTUs were
943 defined as those featuring a relatively low degree and high centrality and could be seen
944 as elements that connect different regions of a network or modules [50]. Connector

OTUs featured a degree <5, Betweenness centrality > 0.03 and Closeness centrality >0.2. Network statistics were calculated with *igraph* in R [90] , Gephi [91] and Cytoscape v3.6.1 [92]. Visualizations were made in Cytoscape v3.6.1. Modules in the core network were identified with MCODE [93].

DECLARATIONS

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

DNA sequences and metadata are publicly available at the European Nucleotide Archive (http://www.ebi.ac.uk/ena; accession numbers PRJEB23788 [18S rRNA genes] & PRJEB38773 [16S rRNA genes]).

Competing interests

The authors declare that they have no competing interests

Funding

RL was supported by a Ramón y Cajal fellowship (RYC-2013-12554, MINECO, Spain). IMD was supported by an ITN-SINGEK fellowship (ESR2-EU-H2020-MSCA-

ITN-2015, Grant Agreement 675752 [ESR2] to RL). This work was supported by the projects INTERACTOMICS (CTM2015-69936-P, MINECO, Spain to RL), MicroEcoSystems (240904, RCN, Norway to RL), MINIME (PID2019-105775RB-I00, AEI, Spain, to RL), ALLFLAGS (CTM2016-75083-R, MINECO to RM), MIAU (RTI2018-101025-B-I00, to JMG) and DEVOTES (grant agreement n° 308392, European Union to EG). It was further supported by Grup Consolidat de Recerca 2017SGR/1568 (Generalitat de Catalunya).

Authors' contributions

AKK & RL designed the study. JMG, RM organized sampling. VB, CRG & IF collected samples, extracted the DNA, and organized its sequencing. AKK, RL & ID analyzed the data, while JMG, RM, IF, CRG & EG, provided contextual ecological or environmental pre-processed data. AKK, MFMB & RL interpreted the results. AKK & RL wrote the manuscript. All authors contributed substantially to manuscript revisions. All authors read and approved the final manuscript.

Acknowledgements

We thank all members of the Blanes Bay Microbial Observatory sampling and analyses team. Bioinformatics analyses were performed at the MARBITS platform of the Institut de Ciències del Mar (ICM; http://marbits.icm.csic.es). We thank the CSIC Open Access Publication Support Initiative through the Unit of Information Resources for Research (URICI) for helping to cover publication fees.

995 **REFERENCES**

996 1. Gitay H, Wilson JB, Lee WG. Species Redundancy: A Redundant Concept?
997 Journal of Ecology. 1996; 84(1):121-124.
998 2. Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the
999 global ocean microbiome. Science. 2016; 353:1272-1277.
1000 3. Galand PE, Pereira O, Hochart C, Auguet JC, Debroas D. A strong link between
1001 marine microbial community composition and function challenges the idea of
1002 functional redundancy. ISME J. 2018; 12(10):2470-2478.
1003 4. Shade A, Handelsman J. Beyond the Venn diagram: the hunt for a core
1004 microbiome. Environ Microbiol. 2012; 14(1):4-12.
1005 5. Little AEF, Robinson CJ, Peterson SB, Raffa KF, Handelsman J. Rules of
1006 Engagement: Interspecies Interactions that Regulate Microbial Communities.
1007 Annual Review of Microbiology. 2008; 62(1):375-401.
1008 6. Lennon JT, Jones SE. Microbial seed banks: the ecological and evolutionary
1009 implications of dormancy. Nat Rev Microbiol. 2011; 9(2):119-130.
1010 7. Singer E, Wagner M, Woyke T. Capturing the genetic makeup of the active
1011 microbiome in situ. ISME J. 2017; 11(9):1949-1963.
1012 8. Mestre M, Höfer J. The Microbial Conveyor Belt: Connecting the Globe
1013 through Dispersion and Dormancy. Trends Microbiol. 2021; in press.
1014 9. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI.
1015 The Human Microbiome Project. Nature. 2007; 449(7164):804-810.
1016 10. Wirth R, Kádár G, Kakuk B, Maróti G, Bagi Z, Szilágyi Á, Rákhely G, Horváth
1017 J, Kovács KL. The Planktonic Core Microbiome and Core Functions in the
1018 Cattle Rumen by Next Generation Sequencing. Front Microbiol. 2018; 9(2285).
1019 11. Rubio-Portillo E, Kersting DK, Linares C, Ramos-Esplá AA, Antón J.
1020 Biogeographic Differences in the Microbiome and Pathobiome of the Coral
1021 *Cladocora caespitosa* in the Western Mediterranean Sea. Front Microbiol.
1022 2018; 9(22).
1023 12. Sweet MJ, Bulling MT. On the Importance of the Microbiome and Pathobiome
1024 in Coral Health and Disease. Frontiers in Marine Science. 2017; 4(9).
1025 13. Lurgi M, Thomas T, Wemheuer B, Webster NS, Montoya JM. Modularity and
1026 predicted functions of the global sponge-microbiome network. Nature
1027 communications. 2019; 10(1):992.
1028 14. Björk JR, O'Hara RB, Ribes M, Coma R, Montoya JM. The dynamic core
1029 microbiome: Structure, dynamics and stability. bioRxiv. 2018.
1030 15. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A,
1031 Eldridge DJ, Bardgett RD, Maestre FT, Singh BK, Fierer N. A global atlas of
1032 the dominant bacteria found in soil. Science. 2018; 359:320-325.
1033 16. Logares R, Deutschmann IM, Junger PC, Giner CR, Krabberød AK, Schmidt
1034 TSB, Rubinat-Ripoll L, Mestre M, Salazar G, Ruiz-González C *et al*.
1035 Disentangling the mechanisms shaping the surface ocean microbiota.
1036 Microbiome. 2020; 8(1):55.
1037 17. Gilbert JA, Steele JA, Caporaso JG, Steinbruck L, Reeder J, Temperton B, Huse
1038 S, McHardy AC, Knight R, Joint I *et al*. Defining seasonal marine microbial
1039 community dynamics. ISME J. 2012; 6(2):298-308.
1040 18. Chow CE, Sachdeva R, Cram JA, Steele JA, Needham DM, Patel A, Parada
1041 AE, Fuhrman JA. Temporal variability and coherence of euphotic zone bacterial
1042 communities over a decade in the Southern California Bight. ISME J. 2013;
1043 7(12):2259-2273.

19. Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. Science. 2015; 347(6223):1257594.
20. Krabberød AK, Bjorbækmo MFM, Shalchian-Tabrizi K, Logares R. Exploring the oceanic microeukaryotic interactome with metaomics approaches. Aquat Microb Ecol. 2017; 79(1):1-12.
21. Bjorbaekmo MFM, Evenstad A, Rosaeg LL, Krabberod AK, Logares R. The planktonic protist interactome: where do we stand after a century of research? ISME J. 2020; 14(2):544-559.
22. Röttjers L, Faust K. From hairballs to hypotheses - biological insights from microbial networks. FEMS Microbiol Rev. 2018; 42(6):761-780.
23. Chow CE, Kim DY, Sachdeva R, Caron DA, Fuhrman JA. Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. ISME J. 2014.
24. Layeghifard M, Hwang DM, Guttman DS. Disentangling Interactions in the Microbiome: A Network Perspective. Trends Microbiol. 2017; 25(3):217-228.
25. Fuhrman JA, Cram JA, Needham DM. Marine microbial community dynamics and their ecological interpretation. Nat Rev Microbiol. 2015; 13(3):133-146.
26. Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, Chaffron S, Ignacio-Espinosa JC, Roux S, Vincent F *et al*. Determinants of community structure in the global plankton interactome. Science. 2015; 348(6237):1262073.
27. Ponisio LC, Valdovinos FS, Allhoff KT, Gaiarsa MP, Barner A, Guimarães PR, Hembry DH, Morrison B, Gillespie R. A Network Perspective for Community Assembly. Frontiers in Ecology and Evolution. 2019; 7(103).
28. Chaffron S, Rehrauer H, Pernthaler J, von Mering C. A global network of coexisting microbes from environmental and whole-genome sequence data. Genome Res. 2010; 20(7):947-959.
29. Krabberod AK, Bjorbaekmo MFM, Shalchian-Tabrizi K, Logares R. Exploring the oceanic microeukaryotic interactome with metaomics approaches. Aquat Microb Ecol. 2017; 79(1):1-12.
30. Cram JA, Xia LC, Needham DM, Sachdeva R, Sun F, Fuhrman JA. Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes. ISME J. 2015; 9(12):2573-2586.
31. Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, Chow CE, Sachdeva R, Jones AC, Schwalbach MS *et al*. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J. 2011; 5(9):1414-1425.
32. Needham DM, Fuhrman JA. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. Nat Microbiol. 2016; 1(4):16005.
33. Deutschmann IM, Lima-Mendez G, Krabberød AK, Raes J, Vallina SM, Faust K, Logares R. Disentangling environmental effects in microbial association networks. preprint. 2021.
34. Giner CR, Balague V, Krabberod AK, Ferrera I, Rene A, Garces E, Gasol JM, Logares R, Massana R. Quantifying long-term recurrence in planktonic microbial eukaryotes. Mol Ecol. 2019; 28(5):923-935.
35. Alonso-Saez L, Balague V, Sa EL, Sanchez O, Gonzalez JM, Pinhassi J, Massana R, Pernthaler J, Pedros-Alio C, Gasol JM. Seasonality in bacterial

diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH. FEMS Microbiol Ecol. 2007; 60(1):98-112.

36. Gasol JM, Cardelus C, Moran XAG, Balague V, Forn I, Marrase C, Massana R, Pedros-Alio C, Sala MM, Simo R *et al*. Seasonal patterns in phytoplankton photosynthetic parameters and primary production at a coastal NW Mediterranean site. Sci Mar. 2016; 80(S1):63-77.

37. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998; 393(6684):440-442.

38. Pedrós-Alió C. The rare bacterial biosphere. Ann Rev Mar Sci. 2012; 4:449-466.

39. Mestre M, Höfer J, Sala MM, Gasol JM. Seasonal Variation of Bacterial Diversity Along the Marine Particulate Matter Continuum. Front Microbiol. 2020; 11(1590).

40. Auladell A, Sánchez P, Sánchez O, Gasol JM, Ferrera I. Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria. ISME J. 2019; 13(8):1975-1987.

41. Berdjeb L, Parada A, Needham DM, Fuhrman JA. Short-term dynamics and interactions of marine protist communities during the spring-summer transition. ISME J. 2018; 12:1907–1917.

42. Lambert S, Tragin M, Lozano JC, Ghiglione JF, Vaulot D, Bouget FY, Galand PE. Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular environmental perturbations. ISME J. 2019; 13(2):388-401.

43. Charles F, Lantoine F, Brugel S, Chrétiennot-Dinet M-J, Quiroga I, Rivière B. Seasonal survey of the phytoplankton biomass, composition and production in a littoral NW Mediterranean site, with special emphasis on the picoplanktonic contribution. Estuarine, Coastal and Shelf Science. 2005; 65(1):199-212.

44. Milici M, Deng Z-L, Tomasch J, Decelle J, Wos-Oxley ML, Wang H, Jáuregui R, Plumeier I, Giebel H-A, Badewien TH *et al*. Co-occurrence Analysis of Microbial Taxa in the Atlantic Ocean Reveals High Connectivity in the Free-Living Bacterioplankton. Front Microbiol. 2016; 7(649).

45. Newman M. Networks: Oxford University Press; 2018.

46. Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: Networks, competition, and stability. Science. 2015; 350:663-666.

47. Vellend M. The theory of ecological communities. Princeton: Princeton University Press; 2016.

48. Stouffer DB, Bascompte J. Compartmentalization increases food-web persistence. Proceedings of the National Academy of Sciences. 2011; 108:3648-3652.

49. Krause AE, Frank KA, Mason DM, Ulanowicz RE, Taylor WW. Compartments revealed in food-web structure. Nature. 2003; 426(6964):282-285.

50. Olesen JM, Bascompte J, Dupont YL, Jordano P. The modularity of pollination networks. Proceedings of the National Academy of Sciences. 2007; 104:19891-19896.

51. Medeiros LP, Garcia G, Thompson JN, Guimarães PR. The geographic mosaic of coevolution in mutualistic networks. Proceedings of the National Academy of Sciences. 2018; 115:12017-12022.

52. Tripp HJ, Schwalbach MS, Meyer MM, Kitner JB, Breaker RR, Giovannoni SJ. Unique glycine-activated riboswitch linked to glycine–serine auxotrophy in SAR11. Environ Microbiol. 2009; 11(1):230-238.

1142 53. Carini P, Steindler L, Beszteri S, Giovannoni SJ. Nutrient requirements for
1143 growth of the extreme oligotroph '*Candidatus Pelagibacter* ubique'
1144 HTCC1062 on a defined medium. ISME J. 2013; 7(3):592-602.
1145 54. Carini P, Campbell EO, Morré J, Sañudo-Wilhelmy SA, Cameron Thrash J,
1146 Bennett SE, Temperton B, Begley T, Giovannoni SJ. Discovery of a SAR11
1147 growth requirement for thiamin's pyrimidine precursor and its distribution in
1148 the Sargasso Sea. ISME J. 2014; 8(8):1727-1738.
1149 55. Włodarczyk A, Selão TT, Norling B, Nixon PJ. Newly discovered
1150 *Synechococcus* sp. PCC 11901 is a robust cyanobacterial strain for high biomass
1151 production. Communications Biology. 2020; 3(1):215.
1152 56. Becker JW, Hogle SL, Rosendo K, Chisholm SW. Co-culture and biogeography
1153 of *Prochlorococcus* and SAR11. ISME J. 2019; 13(6):1506-1519.
1154 57. Chafee M, Fernàndez-Guerra A, Buttigieg PL, Gerdts G, Eren AM, Teeling H,
1155 Amann RI. Recurrent patterns of microdiversity in a temperate coastal marine
1156 environment. ISME J. 2018; 12(1):237-252.
1157 58. Paine RT. A Note on Trophic Complexity and Community Stability. The
1158 American Naturalist. 1969; 103(929):91-93.
1159 59. Banerjee S, Schlaeppi K, van der Heijden MGA. Keystone taxa as drivers of
1160 microbiome structure and functioning. Nat Rev Microbiol. 2018; 16:567-576.
1161 60. Berry D, Widder S. Deciphering microbial interactions and detecting keystone
1162 species with co-occurrence networks. Front Microbiol. 2014; 5(MAY):219.
1163 61. Freilich MA, Wieters E, Broitman BR, Marquet PA, Navarrete SA. Species co-
1164 occurrence networks: Can they reveal trophic and non-trophic interactions in
1165 ecological communities? Ecology. 2018; 99(3):690-699.
1166 62. Banerjee S, Kirkby CA, Schmutter D, Bissett A, Kirkegaard JA, Richardson
1167 AE. Network analysis reveals functional redundancy and keystone taxa
1168 amongst bacterial and fungal communities during organic matter decomposition
1169 in an arable soil. Soil Biology and Biochemistry. 2016; 97:188-198.
1170 63. Pester M, Bittner N, Deevong P, Wagner M, Loy A. A 'rare biosphere'
1171 microorganism contributes to sulfate reduction in a peatland. ISME J. 2010;
1172 4(12):1591-1602.
1173 64. Zhu F, Massana R, Not F, Marie D, Vaulot D. Mapping of picoeucaryotes in
1174 marine ecosystems with quantitative PCR of the 18S rRNA gene. FEMS
1175 Microbiol Ecol. 2005; 52(1):79-92.
1176 65. Seymour JR, Amin SA, Raina JB, Stocker R. Zooming in on the phycosphere:
1177 the ecological interface for phytoplankton-bacteria relationships. Nat
1178 Microbiol. 2017; 2:17065.
1179 66. Farnelid HM, Turk-Kubo KA, Zehr JP. Identification of Associations between
1180 Bacterioplankton and Photosynthetic Picoeukaryotes in Coastal Waters. Front
1181 Microbiol. 2016; 7(339).
1182 67. Auladell A, Barberán A, Logares R, Garcés E, Gasol JM, Ferrera I. Seasonal
1183 niche differentiation between evolutionary closely related marine bacteria.
1184 bioRxiv. 2020.
1185 68. Grasshoff K, Kremling K, Ehrhardt M. Methods of Seawater Analysis: Third,
1186 Completely Revised and Extended Edition; 2007.
1187 69. Massana R, Castresana J, Balague V, Guillou L, Romari K, Groisillier A,
1188 Valentin K, Pedros-Alio C. Phylogenetic and ecological analysis of novel
1189 marine stramenopiles. Appl Environ Microbiol. 2004; 70(6):3528-3534.
1190 70. Stoeck T, Bass D, Nebel M, Christen R, Jones MD, Breiner HW, Richards TA.
1191 Multiple marker parallel tag environmental DNA sequencing reveals a highly

complex eukaryotic community in marine anoxic water. Mol Ecol. 2010; 19 Suppl 1(SUPPL. 1):21-31.
71. Herlemann DPR, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. 2011; 5(10):1571-1579.
72. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol. 2015; 75(2):129-137.
73. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal. 2011; 17(1):10-12.
74. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016; 13(7):581-583.
75. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007; 73(16):5261-5267.
76. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013; 41(Database issue):D590-596.
77. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990; 215(3):403-410.
78. Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J *et al*. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 2013; 41(Database issue):D597-604.
79. McNichol J, Berube PM, Biller SJ, Fuhrman JA. Evaluating and Improving SSU rRNA PCR Primer Coverage via Metagenomes from Global Ocean Surveys. bioRxiv. 2020.
80. Oksanen J, Guillaume Blanchet FFM, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH *et al*. vegan: Community Ecology Package. R package. In.; 2016.
81. Moritz S. imputeTS: Time Series Missing Value Imputation. In.; 2017.
82. Dufrêne M, Legendre P. Species assemblages and indicator species: The need for a flexible asymmetrical approach. Ecological Monographs. 1997.
83. Roberts DW. labdsv: Ordination and Multivariate Analysis for Ecology. R package version 1.8-0. In.; 2016.
84. Ruan Q, Dutta D, Schwalbach MS, Steele JA, Fuhrman JA, Sun F. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. Bioinformatics. 2006; 22(20):2532-2538.
85. Xia LC, Ai D, Cram JA, Liang X, Fuhrman JA, Sun F. Statistical significance approximation in local trend analysis of high-throughput time-series data using the theory of Markov chains. BMC Bioinformatics. 2015; 16(1):301.
86. Xia LC, Ai D, Cram J, Fuhrman JA, Sun F. Efficient statistical significance approximation for local similarity analysis of high-throughput time series data. Bioinformatics. 2013; 29(2):230-237.
87. Deutschmann IM, Lima-Mendez G, Krabberød AK, Raes J, Vallina SM, Faust K, Logares R. Disentangling environmental effects in microbial association networks. ResearchSquare. 2020.

88. Erdős P, Rényi A. On random graphs. Publicationes Mathematicae. 1959; 6:290–297.
89. Banerjee S, Baah-Acheamfour M, Carlyle CN, Bissett A, Richardson AE, Siddique T, Bork EW, Chang SX. Determinants of bacterial communities in Canadian agroforestry systems. Environ Microbiol. 2016; 18(6):1805-1816.
90. Yu G, Chen Y-s, Guo Y-c. Design of integrated system for heterogeneous network query terminal. Journal of Computer Applications. 2009; 29(8):2191-2193.
91. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: *International AAAI Conference on Weblogs and Social Media*. 2009.
92. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics. 2011.
93. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003; 4(1):2.

1258 **FIGURE LEGENDS**
1259
1260

1261 **Figure 1. The Blanes Bay Microbial Observatory and the variation of its resident**
1262 **microbiota and measured environmental variables over ten years. A)** Location of
1263 the Blanes Bay Microbial Observatory. **B)** All possible correlations between the
1264 measured environmental variables including the richness and abundance of resident
1265 OTUs (NB: only 709 resident OTUs are considered, see **Table1**). Only significant
1266 Pearson correlation coefficients are shown ($p<0.01$). The p-values were corrected for
1267 multiple inference (Holm's method). **C)** Unconstrained ordination (NMDS based on
1268 Bray Curtis dissimilarities) of communities including resident OTUs only, to which
1269 environmental variables were fitted. Only variables with a significant fit are shown
1270 ($P<0.05$). Arrows indicate the direction of the gradient and their length represents the
1271 strength of the correlation between resident OTUs and a particular environmental
1272 variable. The color of the samples (circles) indicates the season to which they belong.
1273 The bottom-left arrow indicates the direction of the seasonal change. PNF =
1274 photosynthetic nanoflagellates. **D)** Constrained ordination (Distance-based redundancy
1275 analyses, dbRDA, using Bray Curtis dissimilarities) including only the most relevant
1276 variables after stepwise model selection using permutation tests. Each axis (i.e.,
1277 dbRDA1 and dbRDA2) indicates the amount of variance it explains according to the
1278 associated eigenvalues. The color of the samples (circles) indicates the season to which
1279 they belong. Arrows indicate the direction of the gradient and their length represents
1280 the strength of the correlation between resident OTUs and a particular environmental
1281 variable. The bottom-left arrow indicates the direction of the seasonal change. E-F)
1282 Resident OTUs displaying different niche preferences (blueish areas) in terms of the
1283 two most important abiotic variables: Temperature E) and Daylength F). The red dots

1284 indicate the randomization mean, and the orange curves represent the confidence limits.
1285 Black dots indicate individual OTUs for which temperature or daylength preferences
1286 are significantly ($p<0.05$) higher or lower than a random distribution over 10 years. At
1287 least two assemblages with different niches become evident: one preferring higher
1288 temperature and longer days (summer/spring), and another one preferring lower
1289 temperature and shorter days (winter/autumn). Note that several OTUs associated to
1290 Spring or Autumn are not expected to be detected with this approach, as their preferred
1291 temperature or daylength may not differ significantly from the randomized mean.
1292
1293 **Figure 2. Core microbiota resulting from 10 years of monthly pico- and**
1294 **nanoplankton relative abundances.** A) Core network including bacteria and
1295 microbial eukaryotic OTUs that occur ≥ 30% of the time during the studied decade (i.e.
1296 resident microbiota), with highly significant and strong associations ($P<0.001$ and
1297 $Q<0.001$, absolute local similarity score $|LS| > 0.7$, Spearman correlation $|\rho|>0.7$),
1298 where detected environmentally-driven edges were removed. The color of the edges
1299 (links) indicates whether the association is positive (grey) or negative (red). The shape
1300 of nodes indicates bacteria (rhomboid) or microbial eukaryotes (circle), and the color
1301 of nodes represents species seasonal preferences, determined using the indicator value
1302 (*indval*, $p<0.05$). Node size indicates OTU relative abundance. B) Core network as a
1303 Circos plot, indicating the high-rank taxonomy of the core OTUs. Since 95% of the
1304 associations are positive (see Table 2), we do not indicate whether an edge is positive
1305 or negative.
1306
1307 **Figure 3. The monthly variation in the resident and core microbiotas over 10**
1308 **years.** *Upper panels:* The resident microbiota is defined as those eukaryotes and

1309 bacteria that occur in at least 30% of the samples over 10 years. The relative OTU
1310 abundance (left panel) and number of OTUs (right panel) for different domains and
1311 taxonomic levels in the resident microbiota are shown. Note that the relative abundance
1312 of Bacteria vs. Eukaryotes does not necessarily reflect organismal abundances on the
1313 sampling site, but the amplicon relative abundance after PCR. Relative abundances
1314 were calculated for each year and aggregated over the corresponding months along the
1315 10 years for the resident microbiota, then split into size fractions (NB: relative
1316 abundance for both domains and size fraction sums up to 1 for each month across ten
1317 years). *Lower panels:* Core microbiota over 10 years. The relative abundances of core
1318 OTUs reflect the remaining proportions after removing all the OTUs that were not
1319 strongly associated when building networks. Relative OTU abundance (left panel) and
1320 number of OTUs (right panel) for different domains and taxonomic levels among the
1321 core OTUs.
1322
1323 **Figure 4. Pico- and nanoplankton core sub-networks.** The shape of the nodes
1324 indicates bacteria (rhomboid) or microbial eukaryotes (circle), and the color of nodes
1325 represents species seasonal preferences, determined using the indicator value ($p<0.05$).
1326 The color of the edges indicates if the association is positive (grey) or negative (red).
1327 Node size indicates OTU relative abundance from the core microbiota.
1328
1329 **Figure 5. Main modules in the core network.** Modules with MCODE score >4 are
1330 shown for picoplankton (upper panel) and nanoplankton (lower panel). For each
1331 module, the MCODE score and relative amplicon abundance of the taxa included in it
1332 (as % of the resident microbiota) are indicated. In addition, the numbers of edges and
1333 OTUs within the modules are shown as edges/OTUs; this quotient estimates the average

1334 number of edges per OTU within the different modules. The edges represent
1335 correlations with |LS| > 0.7, |ρ|>0.7, P<0.001 and Q<0.001. The color of the edges
1336 indicates positive (grey) or negative (red) associations. The shape of nodes indicates
1337 bacteria (rhomboid) or microbial eukaryotes (circle), and the color of nodes represents
1338 species seasonal preferences, determined using the indicator value (p<0.05). pb =
1339 Proteobacteria
1340
1341

1342 **TABLE TITLES**

1343

1344 **Table 1.** Description of the datasets.
1345 **Table 2.** Core associations. See **Figure 2**.
1346 **Table 3.** Core network and sub-networks statistics.
1347 **Table 4.** Core associations within and between taxonomic domains and size fractions.
1348 **Table 5:** Subnetworks including core OTUs displaying seasonal preference.
1349 **Table 6.** Central OTUs.
1350
1351

1352 **ADDITIONAL FILES**

1353

1354 **Additional file 1: Table S1**
1355 Relative abundance of bacterial and protistan lineages that are part of the resident and
1356 core microbiotas.
1357
1358

1359 **Additional file 1: Table S2**

1360 Relative abundance of core bacterial taxa.

1361

1362 **Additional file 1: Table S3**

1363 Relative abundance of core eukaryotic taxa.

1364

1365 **Additional file 1: Table S4**

1366 Indicator value for core OTUs in the picoplankton. Sorted by season/kingdom and

1367 relative amplicon abundance.

1368

1369 **Additional file 1: Table S5**

1370 Indicator value for core OTUs in the nanoplankton. Sorted by season/ kingdom and

1371 relative amplicon abundance.

1372

1373 **Additional file 1: Table S6**

1374 Core OTUs without seasonal preference.

1375

1376 **Additional file 1: Table S7**

1377 Module description.

1378

1379 **Additional file 1: Table S8**

1380 OTUs within modules.

1381

1382

1383

1384 **Additional file 2: Figure S1**

1385 **Panel A** shows the full network constructed with the resident microbiota (that is, OTUs
1386 present in >30% of the samples over 10 years; Table 1). **Panel B** displays network
1387 elements that were removed as they did not fulfill the cut-offs (that is, highly significant
1388 correlations (P & Q <0.001), local similarity scores >|0.7| and Spearman correlations
1389 >|0.7|).

1390

1391 **Additional file 3: Figure S2**

1392 OTU relative abundance vs. degree shows no relationship in the core network.
1393

1394

Supplementary Figures

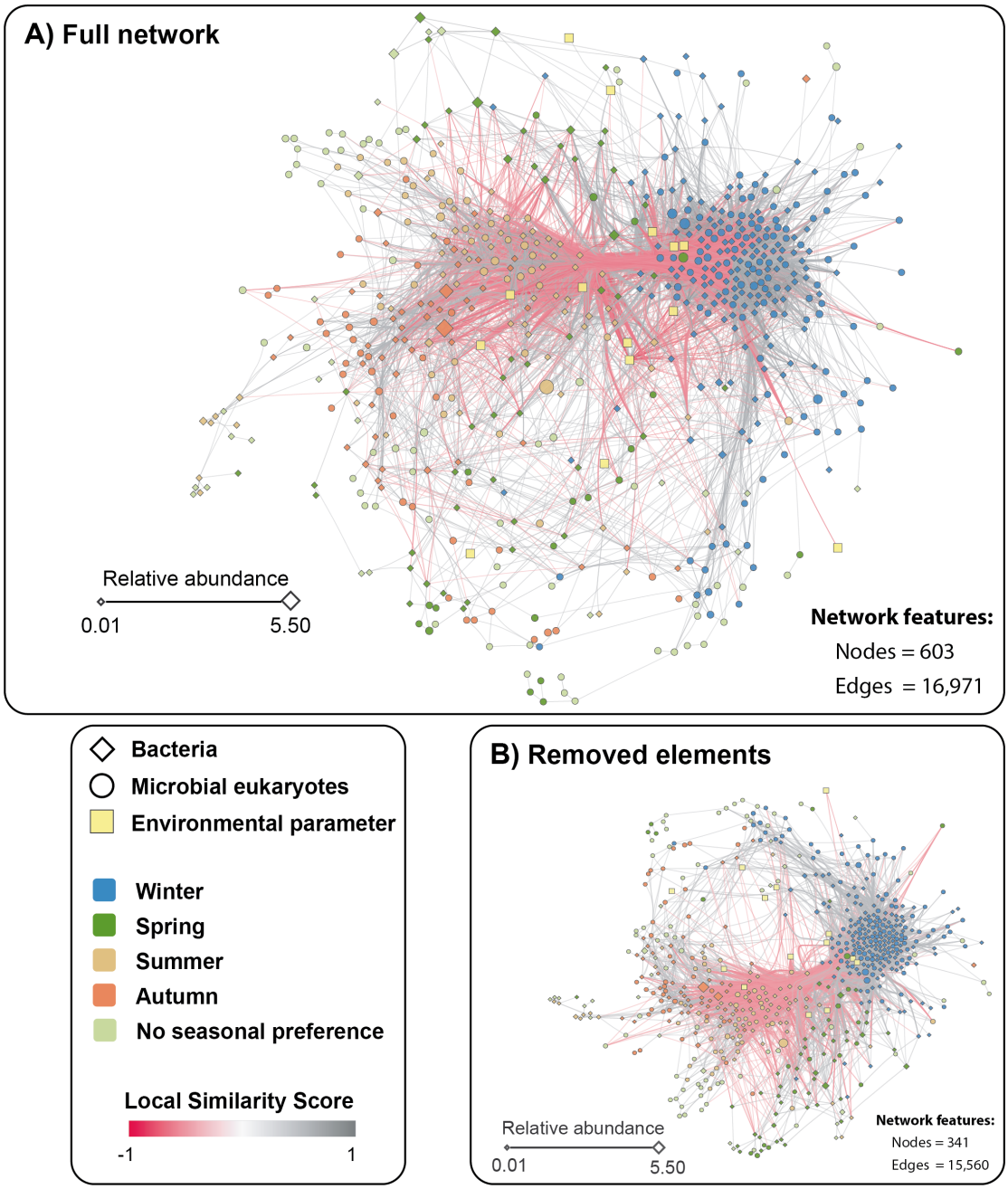


Figure S1. Panel A shows the full network constructed with the resident microbiota (that is, OTUs present in >30% of the samples over 10 years; Table 1). **Panel B** displays network elements that were removed as they did not fulfill the cut-offs (that is, highly significant correlations (P & Q <0.001), local similarity scores >|0.7| and Spearman correlations >|0.7|).

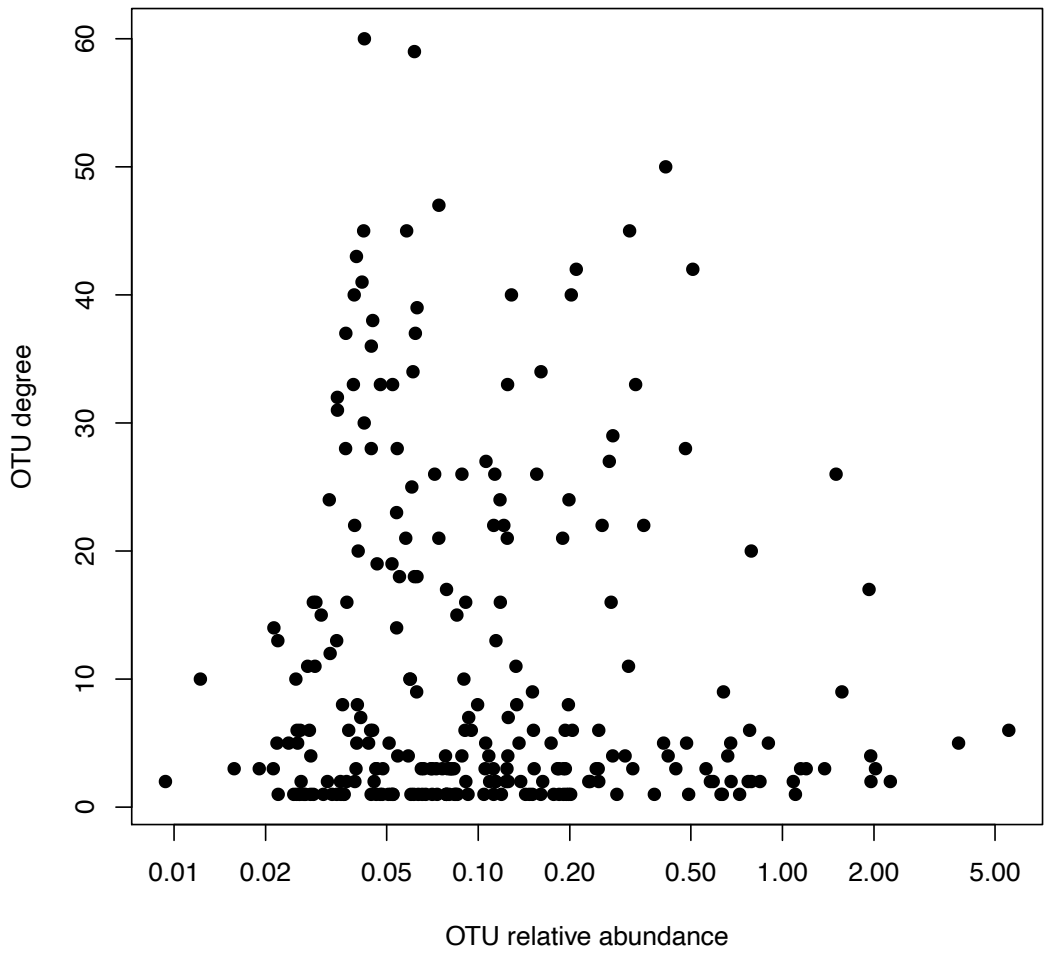


Figure S2. OTU relative abundance vs. degree shows no relationship in the core network.