Universitat Autònoma de Barcelona

Dipòsit digital de documents de la UAB

This is the **accepted version** of the article:

Auladell, Adrià; Sanchez, Pablo; Sánchez Martínez, M. Olga (Universitat Autònoma de Barcelona. Departament de Genètica i de Microbiologia.); [et al.]. «Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria». ISME Journal, Vol. 13 (2019), p. 1975-1987. 13 pàg. DOI 10.1038/s41396-019-0401-4

This version is available at https://ddd.uab.cat/record/254936

under the terms of the IN COPYRIGHT license

A revised manuscript submitted to the ISME Journal, 2019/02/20

Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria

Adrià Auladell[1]*, Pablo Sánchez[1], Olga Sánchez[2], Josep M. Gasol[1, 3] and Isabel Ferrera[1,4]*

[1] Department of Marine Biology and Oceanography, Institut de Ciències del Mar (ICM-CSIC), Barcelona, Catalunya, Spain.

[2] Department of Genetics and Microbiology, Universitat Autònoma de Barcelona (UAB), Bellaterra, Catalunya, Spain.

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

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

Section: Microbial population and community ecology

Running title: Long-term seasonality of photoheterotrophs

Keywords: photoheterotrophy, AAP bacteria, seasonality, Amplicon Sequence Variants (ASVs), metagenome, ecotype, time series.

*Address for correspondence:

Institut de Ciències del Mar, ICM-CSIC
Pg Marítim de la Barceloneta, 37-49
E08003 Barcelona, Catalunya, Spain
Phone: + 34 93 230 95 00
Fax: + 34 93 230 95 55
*aauladell@icm.csic.es, iferrera@icm.csic.es

Competing Interests statement: The authors declare no competing interests.

Statement of authorship: I.F. and J.M.G. conceived the study; I.F. and O.S. designed and performed laboratory analyses; A.A., P.S., O.S. and I.F. analyzed the data; A.A. and I.F. wrote the paper and all authors commented and revised it.

Abstract

We studied the long-term temporal dynamics of the aerobic anoxygenic phototrophic (AAP) bacteria, a relevant functional group in the coastal marine microbial food web, using high-throughput sequencing of the *puf*M gene coupled with multivariate, time series and co-occurrence analyses at the Blanes Bay Microbial Observatory (NW Mediterranean). Additionally, using metagenomics, we tested whether the used primers captured accurately the seasonality of the most relevant AAP groups. Phylogroup K (*Gammaproteobacteria*) was the greatest contributor to community structure over all seasons, with phylogroups E and G (*Alphaproteobacteria*) being prevalent in spring. Diversity indices showed a clear seasonal trend, with maximum values in winter, which was inverse to that of AAP abundance. Multivariate analyses revealed sample clustering by season, with a relevant proportion of the variance explained by day length, temperature, salinity, phototrophic nanoflagellate abundance, chlorophyll *a* and silicate concentration. Time series analysis showed robust rhythmic patterns of co-occurrence, but distinct seasonal behaviors within the same phylogroup, and even within different Amplicon Sequence Variants (ASVs) conforming the same Operational Taxonomic Unit (OTU). Altogether, our results picture the AAP assemblage as highly seasonal and recurrent but containing ecotypes showing distinctive temporal niche partitioning, rather than being a cohesive functional group.

Introduction

Bacteria are extremely abundant and diverse in the ocean where they drive most biogeochemical cycles. Recent developments in sequencing technologies have allowed studying microbial diversity at unprecedented scales. Mapping microbial communities in hundreds of samples from recent global expeditions has resulted in a comprehensive picture of how they vary across space (Yooseph *et al.*, 2007; Salazar *et al.*, 2015; Sunagawa *et al.*, 2015). Likewise, long-term microbial observatories are key to understand microbial variation over time, particularly in temperate zones encompassing contrasting meteorological seasons (Kane, 2004; Buttigieg *et al.*, 2018). To date, different seasonal studies conducted in fixed stations in the Atlantic (Bermuda Atlantic Time Series Study, Western English Channel Time Series) and Pacific (Hawaii Ocean Time Series, San Pedro Ocean Time Series (SPOT)) Oceans, and in the Mediterranean Sea (Service d'Observation du Laboratoire Arago Station) concur that plankton turnover is mostly driven by the environment, and that the seasonal patterns are repeatable over time (Gilbert *et al.*, 2012; Fuhrman *et al.*, 2015; Galand *et al.*, 2018).

Thus far, most seasonal studies have focused on determining the variation of phylogenetic groups based on 16S or 18S rRNA gene sequencing for bacterioplankton and eukaryotic plankton respectively (Kim *et al.*, 2014; Fuhrman *et al.*, 2015; Martin-Platero *et al.*, 2018; Giner *et al.*, 2019). However, these phylogenetic units may include different ecotypes given that closely related or even identical rRNA gene-identified species can possess different functional traits (Martiny *et al.*, 2013) as a result of processes such as horizontal gene transfer (HGT) that can disconnect functional from phylogenetic diversity (Louca *et al.*, 2016). While a considerable amount of information on the seasonality of bulk microbial communities and of some particular phylogroups exists (i.e. Galand *et al.*, 2010), the seasonality of individual functional groups is barely known.

A functional guild of particular interest is the polyphyletic (i.e., derived from more than one common ancestor through HGT) aerobic anoxygenic phototrophic (AAP) bacteria. These organisms have the ability of photoheterotrophy, that is, they are capable of using both organic matter and light as energy sources (Koblížek, 2015). Their discovery challenged previous simplistic views of the structure of ocean microbial food webs (Fenchel, 2001). AAP bacteria are relatively common in the

euphotic zone of the oceans (Schwalbach and Fuhrman, 2005; Jiao *et al.*, 2007; Lami *et al.*, 2007; Cottrell and Kirchman, 2009; Ritchie and Johnson, 2012), exhibit faster growth rates than other bacterioplankton groups (Ferrera *et al.*, 2011; Ferrera, Sánchez, *et al.*, 2017) and their cells are in general larger than most marine heterotrophic bacteria (Sieracki *et al.*, 2006). Altogether, these characteristics make them relevant in the ecosystem by processing a large amount of organic matter (see review by Koblížek, 2015).

Phylogenetically, the AAPs belong to the *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria*. However, since these organisms acquired the ability of photoheterotrophy through HGT, the 16S rRNA gene, typically used for identifying prokaryotes, cannot be used as a genetic marker of AAPs in environmental studies. Alternatively, the *puf*M gene, present in all anoxygenic phototrophs containing type-2 reaction centers, is routinely used in AAP diversity surveys. Based on the phylogeny of this gene and the structure of the *puf* operon, Yutin *et al.* (2007) defined 12 distinct phylogroups (named from A to L) using metagenomic data from the Global Ocean Survey. Currently, the taxonomic assignation of short environmental sequences of the *puf*M gene is commonly done using this 12-phylogroup classification. In recent years, several authors have investigated their diversity and community structure in relation to environmental gradients across spatial scales using the variability of this marker gene (Jiao *et al.*, 2007; Yutin *et al.*, 2007; Ritchie and Johnson, 2012; Boeuf *et al.*, 2013; Bibiloni-Isaksson *et al.*, 2016; Lehours *et al.*, 2018) but much less is known about their temporal dynamics. Two independent studies conducted in the NW Mediterranean (Ferrera *et al.*, 2014) and the East coast of Australia (Bibiloni-Isaksson *et al.*, 2016) examined the variability of AAPs using *puf*M amplicon sequencing and showed that these assemblages seem to be highly dynamic. These two studies analyzed only one year of samples but long-term surveys are necessary to understand their seasonal and interannual patterns of biodiversity, stability, predictability, interactions between species, and responses to environmental changes.

We present here the first long-term exploration of marine AAP assemblages using Illumina sequencing of the amplified *puf*M gene from monthly samples taken over 10 years at the coastal

Blanes Bay Microbial Observatory (BBMO) in the NW Mediterranean Sea. We define the temporal patterns and unveil their recurrence, explore the long-term interactions between the different phylogroups, and identify the main environmental drivers acting upon the observed patterns. Taking advantage of the recent appearance of *threshold-free* algorithms for Amplicon Sequence Variants (ASV) analysis, which surpass the clustering of sequences based on similarity cutoffs (Eren *et al.*, 2015; Callahan *et al.*, 2016), we have gone one step beyond previous studies and explored the seasonality of ASVs potentially representing different AAP ecotypes at a more fine-grained level. These analyses ultimately allow us to explore the level of ecological consistency within the different phylogenetic clades, i.e. whether the different AAP phylotypes are ecologically cohesive or, contrarily, each phylogroup includes organisms presenting temporal niche partitioning. Additionally, by comparing the sequences recovered through amplicon sequencing to those extracted from metagenomes, we test whether the used primers are adequate to evaluate the seasonality of the dominant AAP groups.

Material and Methods

Location and sample collection

Surface water was collected monthly as described elsewhere (Ferrera *et al.*, 2014) from the Blanes Bay Microbial Observatory (41°40'N, 2°48'E), a shallow (~20 m) coastal site about 1 km offshore in the NW Mediterranean coast. A total of 120 samples, from January 2004 to December 2013 were collected and *in situ* prefiltered through a 200-µm mesh. Several environmental parameters were measured alongside sample collection as described in Supplementary Information 1. The measured variables as well as day length were included in an environmental data table containing a total of 23 biotic and abiotic variables that was used for statistical analysis. The environmental data are shown in Figures S1 and S2. The astronomical seasons (based on equinoxes and solstices) were used for establishing spring, summer, autumn and winter periods. Additionally, the mixing layer depth (MLD) was obtained for the first months of 2004 and from 2008 to 2010 as defined in Galí *et al.*, 2013.

*DNA extraction, puf*M *amplification, quantification, sequencing and sequence processing*

About 6 L of 200 µm pre-filtered surface seawater were sequentially filtered through a 20 µm mesh, a 3 µm pore-size polycarbonate filter (Poretics) and a 0.2 µm Sterivex Millipore filter using a peristaltic pump. Sterivex units were filled with 1.8 mL of lysis buffer (50 mM Tris-HCl pH 8.3, 40 mM EDTA pH 8.0 and 0.75 M sucrose), kept at -80°C and extracted using the phenol-chloroform protocol as in Massana *et al.* (1997). Note that AAP bacteria attached to particles larger than 3 µm were not the subject of this study.

Partial amplification of the *puf*M gene (~245 bp fragments) was done in 50 µl reactions using primers pufM forward (5'-TACGGSAACCTGTWCTAC-3', Béjà *et al.*, 2002) and puf_WAW reverse (5'-AYNGCRAACCACCANGCCCA-3', Yutin *et al.*, 2005), each at 0.2 µM final concentration. The final concentration of $MgCl_2$ was 2 mM. PCR conditions were as follows: an initial denaturation step at 95°C for 5 min and 35 cycles at 95°C (30s), 58°C (30s), 72°C (40s) and a final elongation step at 72°C for 10 min. Sequencing was performed in an Illumina MiSeq sequencer (2 *x* 250 bp, Research and Testing Laboratory; http://rtlgenomics.com/). Primers and spurious sequences were trimmed using *cutadapt* v1.14 (Martin, 2011). DADA2 v1.4 was used to differentiate exact sequence variants and remove chimeras (parameters: maxN= 0, maxEE= c(2,4), trunclen= c(200,200)) (Callahan *et al.*, 2016). DADA2 resolves Amplicon Sequence Variants (ASVs) by modeling the errors in Illumina-sequenced amplicon reads. The approach is *threshold-free*, inferring exact variants up to 1 nucleotide of difference using the quality score distribution in a probability model. For comparison purposes, the ASVs were clustered using UCLUST (Edgar, 2010) at 94% of nucleic acid sequence similarity, a threshold typically used for the *puf*M gene (Zeng *et al.*, 2007). After filtering for chimeras and spurious sequences with DADA2, 74% of the initial reads (mean 25692, min 4172, max 135331) were kept for downstream analyses. Sample BL120313 (13 March 2012) was discarded due to low read counts (836 reads). DADA2 read filtering details can be found in Supplementary Table 1. Moreover, in order to determine whether the primers used in this PCR-based approach captured the seasonality patterns accurately, we used 35 metagenomes generated from the same time-series (samples from years 2011 to 2013; see a detailed explanation in Supplementary Information 2) for comparison. Copy numbers of the marker gene *puf*M were estimated by

quantitative polymerase chain reaction (qPCR) as described in Ferrera *et al.*, (2017b) (see Supplementary Information 3 for details).

Phylogenetic classification

A custom-made database was generated combining sequences from previous AAP studies (Cuadrat et al., 2016; Graham et al., 2018; Lehours et al., 2018; Yutin et al., 2007), variants present in the Integrated Microbial Genomes system (Markowitz *et al.*, 2006) and other *puf*M sequences from the GenBank database. Additionally, the predicted sequences from the BBMO metagenomes were included in this database. The nucleotide sequences were aligned with the guidance of amino acid translations using TranslatorX (Abascal *et al.*, 2010), with a posterior manual curation after filtering the sequences by length (>600 bp). From the alignment, a phylogenetic tree was constructed with RAxML v8.2 (Stamatakis, 2014) (GTRGAMMA model, 1000 bootstraps), and the phylogroups were delimited in the resulting tree using iTOL (Letunic and Bork, 2016) (Figure S3). Afterwards, the phylogenetic placement of the amplicon nucleotide sequences was performed with the Evolutionary Placement Algorithm v0.2 (Barbera *et al.*, 2018) to establish their phylogroup classification. Finally, to determine potential primer biases, the forward and reverse primers were contrasted against the nucleotide alignment.

Statistical analyses

All analyses were performed using the R language, with *phyloseq* and *vegan* packages (McMurdie and Holmes, 2013; R Core Team, 2014; Oksanen *et al.*, 2018). Alphadiversity was analyzed using the Chao1 and Shannon indices (Smith and Wilson, 1996). Betadiversity was analyzed using a Bray-Curtis dissimilarity matrix with a previous normalization through rarefying to 4172 reads per sample (Faith *et al.*, 1987; Weiss *et al.*, 2017). We used distance-based Redundancy Analysis (dbRDA, Legendre and Legendre, 1988) to find the environmental predictors (scaled to the mean) that best explained the patterns of community structure and diversity of AAPs over time, with a previous multivariate non-parametric ANOVA for selecting significant variables (p<0.01). A time-decay analysis of the assemblage was computed excluding rare ASVs as recommended elsewhere (Faust

186 *et al.*, 2015). ASVs were considered rare when presenting less than 1% of relative abundance from
187 the rarefied dataset following Alonso-Sáez *et al.* (2015) criterion.

188 *Time series analysis*

189 Fourier time series analysis was performed to study the AAP assemblage dynamics over a decade.
190 An interpolation of the discarded sample (BL120313) was used to maintain equidistant time points.
191 Values were normalized through the Aitchison log-centered ratio transformation (CLR), adequate for
192 compositional data (Gloor *et al.*, 2017). A Fisher G-test with the R package *GeneCycle* was used to
193 determine the significance (p<0.01) of the periodic components (Ahdesmaki *et al.*, 2015). The time
194 series was decomposed in three components - seasonal periodicity (oscillation inside each period),
195 trend (evolution over time) and residuals -through local regression by the *stl* function. Additionally,
196 the autocorrelogram was calculated using the *acf* function.

197 *Network construction*

198 We used Local Similarity Analysis (LSA) (Xia *et al.*, 2011; Durno *et al.*, 2013) with a previous CLR
199 transformation for network construction. Briefly, given a time series data and a delay limit, LSA finds
200 the configuration that yields the highest local similarity (LS) score. Only the ASVs present in >5
201 samples and the environmental variables presenting <5% of missing values were used. The remaining
202 missing values for the variables after filtering were estimated by imputation with the *mice* package
203 (Azur *et al.*, 2012). Only Interactions with LS >=0.5, p<0.001 and 1-month delay were considered.
204 The network was plotted using the *ggraph* package (Pedersen, 2017).

205 *Reproducibility*

206 The code for preprocessing and statistical analyses along with package versions is available in the
207 following repository: https://gitlab.com/aauladell/AAP_time_series. Sequence data has been
208 deposited in Genbank under accession number PRJNA449272.

209 ## Results and discussion

***Patterns of community composition and structure*.** The amplicon sequences retrieved from 10 years of sampling resulted in a total of 820 ASVs whereas the number of OTUs was 406 (94% similarity cutoff). Of the total ASVs, 276 presented only 1 nucleotide variation between sequences. In comparison with previous temporal studies (82 OTUs detected in Ferrera *et al.*, 2014, 89 in Bibiloni-Isaksson *et al.*, 2016), our study presents a more complete picture of *puf*M diversity and is the largest dataset of AAP diversity reported to date. Estimates of richness were higher during winter (mean 51, max 126 observed ASVs), decreasing to minimum values in the spring-summer period, precisely during May-August (mean 35, max 77) (Figure 1). The differences between winter and spring/summer were statistically significant (ANOVA, $p<0.05$). A similar trend was observed when computing the Shannon index (Figure 1). Comparing the amplicon with the metagenomic data from 2011 to 2013, we observed that whereas 188 OTUs and 357 ASVs were present in the amplicons for that period, only a total of 84 different *puf*M sequences were recovered from the metagenomes. However, the Shannon diversity index for the two datasets presented a positive correlation (Pearson R=0.81, p=0.001, N=35) and the followed the same trend of increasing values in winter (Figure 1).

A notable negative correlation between day length and the Shannon index was observed (Pearson R=-0.57, p<0.01, N=119). That relationship of diversity with day length had previously been observed in long-term bulk bacterioplankton community studies (Gilbert *et al.*, 2012), as well as with specific phylogenetic groups such as the SAR11 (Salter *et al.*, 2015). A possible explanation is that the deep winter mixing allows the development of high diversity assemblages in contrast to the selection of specific oligotrophic ecotypes occurring during the stratified summer season (Salter *et al.*, 2015). In fact, mixed layer depth was a significant predictor of the Shannon diversity index (Spearman R=0.56, p<0.001, N=35, Figure S4). Interestingly, this trend of higher alphadiversity in winter is opposed to that of AAP abundance (Figure S5); higher abundances of the *puf*M gene during spring and summer were measured by qPCR as compared to winter and fall (p<0.01). We also found a positive correlation between the qPCR data and the abundance of *puf*M sequences retrieved in the 3 years of metagenomes (Pearson R=0.77, p<0.001, N=12) (Figure S6), in which

237 the higher abundances were found in spring followed by summer. These results support previous
238 observations obtained through various methodological approaches (Ferrera *et al.*, 2014 using
239 microscopy counts; Bibiloni-Isaksson *et al.*, 2016 using qPCR and Galand *et al.*, 2018 using
240 metagenomics) and confirm that there is a clear inverse relationship between AAPs abundance and
241 diversity.

242 Regarding community composition across the decadal period (Figure 2, Figure S7), phylogroup K
243 (*Gammaproteobacteria)* affiliated to the NOR5/OM60 clade) was the most prevalent and dominant over
244 the years (83.8% ± SE 2.3, mean relative abundance), in agreement with previous reports for this
245 station (Ferrera *et al.*, 2014) and for other regions such as the Baltic Sea (Mašín *et al.*, 2006). Yet, a
246 decrease in their contribution was observed during February and March (down to 59.6% and 52% on
247 average respectively). During these months, the contribution of phylogroup E (*Rhodobacter*-like) to
248 community structure was greater, albeit with a high variation over the decade (±26% SD). The previous
249 1-year study of AAP diversity conducted by Ferrera *et al.* (2014) reported a similar observation and
250 moreover, a study of the 16S rRNA gene diversity from the same location also suggested that
251 *Alphaproteobacteria* dominate the bacterial assemblages during the local spring bloom (Alonso-Sáez
252 *et al.*, 2007). Regarding phylogroups D, F (*Rhodobacterales*-like), H (uncultured), J (*Rhodospirillales*-
253 like), I (*Betaproteobacteria*), *Sphingomonas*-like and the unclassified ASVs, these presented a mean
254 relative contribution below 1% in the amplicon dataset. These groups displayed occasional peaks (>1%
255 relative abundance) with no clear periodic trend, for example, *Sphingomonas*-like AAPs showed a
256 contribution of 14% in February 2012 (Figure S7). Overall, these observations are similar to those
257 obtained from the metagenomic distribution (3 years instead of 10) with a good agreement in the
258 relative abundance recovered for the prevailing phylogroups E and K as well as for the less
259 common *Sphingomonas*-like sequences (correlation values: 0.9, 0.69 and 0.91 respectively; Table
260 S2). Contrarily, phylogroup G seems to be underrepresented in the amplicon dataset likely due to
261 the presence of 3 to 5 mismatches in the forward primer with the most abundant metagenomic
262 variants.

Metagenomics is often considered the least biased approach for functional gene analysis since the method is PCR-independent and does not suffer from amplification biases that could result in misrepresentation of the relative abundances of certain populations. Nonetheless, for a given time and money investment, metagenomes retrieve less copies of specific marker genes, offering thus less inquiry potential if the main purpose of our study is the barcoding of a particular group of organisms. We found that richness estimates were higher using amplicon data since more variants of the *puf*M gene were recovered with that approach than from metagenomes, but the seasonal trends in diversity identified by both methodologies were remarkably similar (Figure 1). Likewise, in terms of community structure there was a good agreement for the most prevailing groups with the exception of phylogroup G. Moreover, the seasonal trends observed at the phylogroup and even at the sequence variant level recovered using these two distinct methodological approaches, presented a close resemblance (Figures S8 and S9, see *Seasonality at the fine scale section* below).

Noteworthy, the amplicon approach allowed identifying the seasonal tendencies of many more individual ecotypes that what would have been possible through metagenomics, while metagenomics captured some low abundance groups missing in the amplicon dataset. In particular, phylogroups A, B, C and L, accounting for a total relative abundance of 7.0, 3.1, 3.9 and 0.2% respectively, were only retrieved through metagenomics. Primer coverage analysis revealed that the forward primer contains between 3 and 8 mismatches with the metagenomic sequences from these phylogroups (details not shown), which could explain their absence in the amplicon dataset and why these groups are rarely reported in AAP surveys based on amplicon sequencing (Bibiloni-Isaksson *et al.*, 2016; Lehours *et al.*, 2018). Exceptionally, Ferrera *et al.*, 2014 reported the presence of one single OTU of group C contributing substantially (13% relative abundance) to the community during winter in Blanes Bay, which differs from the present results. To investigate this discrepancy, we carefully compared the sequence of this OTU to our updated database and found that it had been misclassified and it belongs to phylogroup K while does not show any significant similarity to the new phylogroup C sequences retrieved from the metagenomes. These observations highlight the

need to increase the information present in databases to obtain accurate taxonomic assignations. In fact, only a few isolates from phylogroup K exist and none is available for phylogroup C, hampering the classification of these groups as discussed elsewhere (Caliz and Casamayor, 2014). In contrast, phylogroups F, H, I and J (<1% total relative abundance) were recovered only when using amplicons. Their low relative abundance possibly explains their absence in the metagenomic dataset. Overall, these results remark the need to undertake a revision of the primers typically used for high-throughput sequencing of *puf*M in order to increase their phylogenetic recovery but, at the same time, demonstrate that PCR-free metagenomics and amplicon-based approaches perform in a comparable fashion in recovering major AAP groups and, most importantly, that the seasonal patterns observed through amplicon sequencing are robust.

Patterns of betadiversity and recurrence. Non-metric multidimensional scaling (nMDS) using various distance measurements indicated a clear separation of the samples at different temporal scales: by month (Bray Curtis, PERMANOVA R^2=0.51, p<0.001) and by season (Bray Curtis, PERMANOVA R^2=0.31, p<0.001, Figure S10). Spring and winter samples were more dissimilar than those of summer or autumn. The reasons for this pattern are uncertain but could be related to higher date to date environmental variability or to the mixing of the water column that occurs during winter in this station (Gasol *et al.*, 2016).

Community structure was strongly linked to day length, temperature, salinity, phototrophic nanoflagellate abundance, chlorophyll *a* and silicate concentration, as revealed by distance-based redundancy analysis (dbRDA; Figure 3, PERMANOVA p<0.01, Supplementary Information 4), which explained 51.4% of the variation with the first 2 axis explaining 43.6%. In particular, late spring and early summer samples were mostly influenced by day length and temperature, whereas autumn samples were partially influenced by salinity (Figure 3). Day length has previously been shown to explain the seasonal variability of the bulk bacterioplankton (Gilbert *et al.*, 2012) and AAP community structure (Ferrera *et al.*, 2014), but the mechanisms underlying this relationship are unclear. Interestingly, a group of samples from winter and spring appeared to be heavily influenced by the presence of ASVs (ASV8, ASV14 and ASV46) belonging to phylogroup E (*Rhodobacter*-like,

Figure S11), to the abundance of phototrophic nanoflagellate and the concentration of chlorophyll *a* (Figure 3), which could be related to the phytoplankton spring bloom that typically occurs in February-March in Blanes (Nunes *et al.*, 2017). The summer samples were associated to the high contribution of gammaproteobacterial ASV1 and the fall/early-winter cluster to more diverse communities of other gammaproteobacterial ASVs (Figure S11).

Finally, to explore the recurrence of the communities, the Bray-Curtis similarity between samples was plotted against the time lag resulting in the so-called time-decay curve (Figure 4) (Shade *et al.*, 2013; Fuhrman *et al.*, 2015). In our study, the assemblage was maintained over time with a median similarity of 0.45, with 6-month oscillations from the yearly maximum (~0.55) to the minimum (~0.39) values. These results indicate that AAP communities are under strong environmental selection that leads to a high seasonal behavior and translates into yearly repeatable communities. To our knowledge, this is the first time that the recurrence of a functional group of planktonic organisms, defined by a marker gene, has been demonstrated. Comparing the results to the 16S rRNA data from the SPOT and the Western Channel time series results we observe that the seasonal turnover at SPOT is less clear than in our location, and an initial decay of similarity is observed reaching a later plateau over time. In the Western Channel, the seasonality is equally marked but the initial decay is even more pronounced than in SPOT (see Figure 2 in Hatosy *et al.*, 2013). A possible explanation to these differences is that our comparison accounts only for a highly seasonal sub-community (as these organisms are able to use light, and light varies seasonally) while the overall bacterial/prokaryotic community responds to more variables. Further analyses with other functional genes should help understand whether these patterns are robust for distinct groups.

Patterns of co-occurrence. A co-occurrence network was built with 127 ASVs present in >5 samples and 14 environmental variables, presenting 70 nodes and 142 edges after filtering by local similarity and significance (LS >= 0.5, $p<0.001$,) (Figure 5). Noteworthy, most of the ASVs retained in the network were seasonal (46 out of 61) (see below). In terms of topology, the network presents one large cluster and other four minor clusters, being the largest one formed by 54 nodes mainly containing ASVs from phylogroups K and E, displaying multiple interactions with various ecosystem

variables (temperature, day length and the abundance of phototrophic picoeukaryotes and nanoflagelates). Temperature was the variable presenting the largest number of the interactions (14), most of them being delayed one month. Out of these, many were of negative nature with *Gammaproteobacteria*-like ASVs that lower their relative abundance during summer (for example ASV26, 10, 11) while others were positive with ASVs that dominate the AAP community during this season (ASV1). Interestingly, many positive and negative interactions exist between various ASVs of phylogroup K and G and the abundance of phototrophic eukaryotes but none with other phylogroups such as phylogroup E or *Sphingomonas*-like. Strong biotic relationships between AAP species and phytoplankton have been reported, particularly with dinoflagellates (Biebl *et al.*, 2005; Yang *et al.*, 2018) and large fractions of particle-attached AAP bacteria have been observed in various marine environments (Waidner and Kirchman, 2007; Lami *et al.*, 2009). Here, we focused on the free-living fraction of AAPs but further interaction network analyses using both free-living and particle-attached AAP bacteria in combination with phototrophic eukaryotic species data would allow to investigate deeper these biotic relationships.

The majority of correlations occurred within rather than between phylogroups as previously observed (Bibiloni-Isaksson *et al.*, 2016); yet, whereas some groups presented mainly positive intergroup interactions (phylogroup E or *Sphingomonas*-like), phylogroup K showed positive and negative interactions between its ASVs. As an example, a clear negative correlation between ASV1-ASV26 and ASV10-ASV35, all of them being part of phylogroup K, can be observed in Figure 5. We also noticed that various ASVs of phylogroup E, such as ASV14 and ASV8 (*Alphaproteobacteria*-like), were positively related among them while presenting negative associations with ASVs from phylogroup K (*Gammaproteobacteria*-like ASV30 and ASV35). Negative correlations between phylogroup K and phylogroup G (also *Alphaproteobacteria*-like) had previously been reported (Ferrera *et al.*, 2014; Bibiloni-Isaksson *et al.*, 2016). These data thus point towards intergroup competition between members of the *Alpha*- and *Gammaproteobacteria*-like AAPs.

Looking at the interactions within closely related ASVs, i.e. those forming the same OTU, we observed multiple connections between them, for example, among ASV1, ASV2 and ASV15, all

belonging to OTU1 or ASV14, ASV65 and ASV175, all forming OTU14. Nevertheless, network analysis revealed that sometimes there is a dissociation of these closely related ASVs, as seen for ASV17 and ASV27, belonging to OTU1 which do not present connections with other ASVs from the same OTU. These observations support the idea that ASVs may represent individual AAP ecotypes encompassing distinct ecological patterns and reflects the usefulness of breaking apart sequence clusters into variants in order to dig into the ecology of these organisms.

Seasonality at the fine scale. The seasonality of each ASV was measured by evaluating if their relative abundance distribution presented a significant periodicity (Fisher G-test) through the long-term time series, and if so, by comparing them at different levels of resolution: across closely related sequences (ASVs) and across sequence clusters (OTUs and phylogroups). Seasonal patterns (p<0.01) were present in 58 out of 127 ASVs analyzed (those ASVs present in >5 samples), affiliated to phylogroups K (44 ASVs), E (9) and G (2), J (1) and the unclassified group (2) (Table 1, Supplementary Table 3). In order to discard that potential amplification artifacts could influence the observed ASV seasonal trends, we mapped representative ASVs from the prevailing phylogroups E, G and K to the metagenomic sequences and compared the seasonal behavior in both datasets, obtaining a remarkable good concordance (Figure S9).

The seasonal ASVs corresponded to 92% of the total read counts, and 83.4% of the counts corresponded to phylogroup K (*Gammaproteobacteria*). All periodicities found were of 1 year, with the exception of ASV152 (*Gammaproteobacteria*), that presented a periodicity of 2 years. Some of these ASVs always presented relative contributions above 1% regardless of season (all from phylogroup K), some presented values above 1% in a specific season (seasonal contributors), and other ASVs peaked (>1%) only occasionally (herein referred as opportunistic; see examples in Figure S12-S15). In fact, most ASVs presented an opportunistic behavior, with low contribution to total community composition during the decade and peaking occasionally. Various studies of the whole bacterioplankton community have observed this variety of strategies coexisting within a given clade (Shade *et al.*, 2014; Alonso-Sáez *et al.*, 2015; Fuhrman *et al.*, 2015). Our results reveal that this trend is maintained for this specific functional assemblage, with a few prevalent ecotypes and a

larger pool of specialized ASVs, i.e. appearing within specific environmental conditions, within each phylum.

Comparing among the seasonal ASVs, we distinguished different behaviors. For example, ASVs divergent enough to form distinct OTUs but belonging to the same phylogroup did not always follow the same distribution (Figure 6A); e.g. for phylogroup K, the annual maxima of ASV1 occurred during June and July with a minimum in February/March, whereas ASV10 presents the opposite distribution. Contrarily, most ASVs belonging to phylogroups G and E followed a similar trend among them, with their maxima in March, being ASV86 an exception presenting a maximum in September. Looking at a further level of resolution, i.e., comparing the seasonality of closely related ASVs (that would form the same OTU), we observed that these generally displayed similar temporal patterns although some notable exceptions existed. An example is represented in Figure 6B in which the seasonal periodicities of 5 closely related ASVs – all corresponding to OTU1- are plotted together. In this figure, a slight succession of the summer maxima can be observed (ASV2 peaking before ASVs 33 and 1, with ASV57 afterwards), being all these only 1 nucleotide different among them. Yet, ASV128 (presenting a distance of 4 nucleotides to OTU1) displays a different distribution peaking during winter. The existence of divergent distributions of ASVs composing the same OTU demonstrates the need to break apart the clusters of related sequences, since these can hide distinct ecological patterns. Furthermore, while the previous AAP temporal studies provided insights of the inter-annual community structure, this is the first study that identifies the long-term tendencies of individual ecotypes.

At the other end, when we explored the seasonality at the phylogroup level, we found that phylogroup K as a whole did not present a statistically significant seasonal pattern ($p>0.01$) (Figure S16-S17). The disparity of distributions of the various sequences within may be the reason of the loss of a significant signal when computing seasonality at the group level. Contrarily, the autocorrelograms showed phylogroup E presenting a high value (max 0.34 over a year), followed by phylogroup J (Figure S16B). These results could indicate a higher degree of ecotype differentiation in gammaproteobacterial phylogroup K as compared to alphaproteobacterial phylogroups E and G.

A possible explanation is that phylogroup K is phylogenetically broader (based on the 16S rRNA gene sequences) as compared to phylogroups E and G, resulting in more variable tendencies within it. Further analyses including the genomic context with the assignation of sequence variants to Metagenome Assembled Genomes (MAGs) or genome sequencing of new isolates of phylogroup K, would help splitting this phylogroup into smaller phylogenetic clusters, perhaps showing ecological coherence. Lehours *et al*. (2018) recently tested the ecological consistency of the AAP across different oceanic regions and, interestingly, identified clades with good ecological and phylogenetic coherence. Our temporal analyses add a new level of complexity by showing that, despite a certain degree of consistency exists, highly similar ASVs can present very different seasonal distributions that could translate into different ecology.

Concluding remarks. This work shows that the AAPs present a peak of diversity during winter, contrary to their abundance, and that gammaproteobacterial AAPs are the prevalent members of the community in the Mediterranean Sea year-round. Our results also evidence that the AAP assemblages show seasonal patterns repeatable over long periods of time. This study also demonstrates that PCR-free metagenomics and amplicon-based approaches perform in a comparable fashion in recovering major AAP groups and that the seasonal patterns observed through amplicon sequencing are robust. Interestingly, distinct seasonal behaviors were observed within the same phylogroup and even within different ASVs conforming the same OTU. In contrast to the recent spatial study of Lehours *et al.* (2018), in which they reported ecological cohesiveness when comparing contrasting biomes, we found that the different AAP phylotypes do not appear as coherent when studying their seasonal behavior and seem to be rather composed of different ecotypes with distinctive temporal niche partitioning. Overall, these results show that the analysis of long time series allows exploring in-depth patterns of a highly dynamic microbial group and provides a framework for modeling their ecological role in relation to seasonality of marine carbon cycling.

Acknowledgements

We thank the many people involved in maintaining the BBMO and those taking care of sampling, particularly Clara Cardelús and Vanessa Balagué. We would also like to thank Ramon Massana and Irene Forn for providing microscopy counts. We thank the Marbits bioinformatics platform at ICM-CSIC, particularly Ramiro Logares and Célio Dias for providing the metagenomic data and Anders Kristian Krabberød (University of Oslo) for help with network analyses. Access to the MareNostrum Supercomputer was granted to Ramiro Logares under agreement BCV-2017-3-0001. qPCR analysis was done at the Institute of Evolutionary Biology (Barcelona) thanks to José Luís Maestro. This research was funded by grant REMEI (CTM2015-70340-R) from the Spanish Ministry of Economy, Industry and Competitiveness and Grant 2017SGR/1568 from CIRIT-Generalitat de Catalunya to Consolidated Research Groups.

References

Abascal, F., Zardoya, R., and Telford, M.J. (2010) TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**: W7–W13.

Ahdesmaki, M., Fokianos, K., and Strimmer, K. (2015) Package 'GeneCycle'.

Alonso-Sáez, L., Balagué, V., Sà, E.L., Sánchez, O., González, J.M., Pinhassi, J., et al. (2007) Seasonality in bacterial diversity in north-west Mediterranean coastal waters: Assessment through clone libraries, fingerprinting and FISH. *FEMS Microbiol. Ecol.* **60**: 98–112.

Alonso-Sáez, L., Díaz-Pérez, L., and Morán, X.A.G. (2015) The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Environ. Microbiol.* **17**: 3766–3780.

Azur, M.J., Stuart, E.A., Frangakis, C., and Leaf, P.J. (2012) Multiple Imputation by Chained Equations: What is it and how does it work? **20**: 40–49.

Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T., and Stamatakis, A. (2018) EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Syst. Biol.* syy054-syy054.

Béjà, O., Suzuki, M.T., Heidelberg, J.F., Nelson, W.C., Preston, C.M., Hamada, T., et al. (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415**: 630–633.

Bibiloni-Isaksson, J., Seymour, J.R., Ingleton, T., van de Kamp, J., Bodrossy, L., and Brown, M. V.

(2016) Spatial and temporal variability of aerobic anoxygenic photoheterotrophic bacteria along the east coast of Australia. *Environ. Microbiol.* **18**: 4485–4500.

Biebl, H., Allgaier, M., Tindall, B.J., Koblizek, M., Lünsdorf, H., Pukall, R., and Wagner-Döbler, I. (2005) Dinoroseobacter shibae gen. nov., sp. nov., a new aerobic phototrophic bacterium isolated from dinoflagellates. *Int. J. Syst. Evol. Microbiol.* **55**: 1089–1096.

Boeuf, D., Cottrell, M.T., Kirchman, D.L., Lebaron, P., and Jeanthon, C. (2013) Summer community structure of aerobic anoxygenic phototrophic bacteria in the western Arctic Ocean. *FEMS Microbiol. Ecol.* **85**: 417–432.

Buttigieg, P.L., Fadeev, E., Bienhold, C., Hehemann, L., Offre, P., and Boetius, A. (2018) Marine microbes in 4D — using time series observation to assess the dynamics of the ocean microbiome and its links to ocean health. *Curr. Opin. Microbiol.* **43**: 169–185.

Caliz, J. and Casamayor, E.O. (2014) Environmental controls and composition of anoxygenic photoheterotrophs in ultraoligotrophic high-altitude lakes (Central Pyrenees). *Environ. Microbiol. Rep.* **6**: 145–151.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**: 581.

Cottrell, M.T. and Kirchman, D.L. (2009) Photoheterotrophic microbes in the arctic ocean in summer and winter. *Appl. Environ. Microbiol.* **75**: 4958–4966.

Cuadrat, Rafael R. C., Ferrera, Isabel, Grossart Hans-Peter, Dávila, A.M.R. (2016) Picoplankton Bloom in Global South? A High Fraction of Aerobic Anoxygenic Phototrophic Bacteria in Metagenomes from a Coastal Bay (Arraial do Cabo—Brazil). *OMICS*.

Cuadrat, R.R.C., Ferrera, I., Grossart, H.-P., and Dávila, A.M.R. (2016) Picoplankton Bloom in Global South? A High Fraction of Aerobic Anoxygenic Phototrophic Bacteria in Metagenomes from a Coastal Bay (Arraial do Cabo-Brazil). *Omi. A J. Integr. Biol.* **20**:.

Durno, W.E., Hanson, N.W., Konwar, K.M., and Hallam, S.J. (2013) Expanding the boundaries of local similarity analysis. *BMC Genomics* **14**: 1–14.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Eren, A.M., Morrison, H.G., Lescault, P.J., Reveillaud, J., Vineis, J.H., and Sogin, M.L. (2015) Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.* **9**: 968–979.

Faith, D.P., Minchin, P.R., and Belbin, L. (1987) Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* **69**: 57–68.

Faust, K., Lahti, L., Gonze, D., de Vos, W.M., and Raes, J. (2015) Metagenomics meets time series analysis: Unraveling microbial community dynamics. *Curr. Opin. Microbiol.* **25**: 56–66.

Fenchel, T. (2001) Marine bugs and carbon flow. *Science* **292**: 2444 LP-2445.

Ferrera, I., Borrego, C.M., Salazar, G., and Gasol, J.M. (2014) Marked seasonality of aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean Sea as revealed by cell abundance, pigment concentration and pyrosequencing of *pufM* gene. *Environ. Microbiol.* **16**: 2953–2965.

Ferrera, I., Gasol, J.M., Sebastián, M., Hojerová, E., and Kobížek, M. (2011) Comparison of growth rates of aerobic anoxygenic phototrophic bacteria and other bacterioplankton groups in coastal mediterranean waters. *Appl. Environ. Microbiol.* **77**: 7451–7458.

Ferrera, I., Sánchez, O., Kolářová, E., Koblížek, M., and Gasol, J.M. (2017) Light enhances the growth rates of natural populations of aerobic anoxygenic phototrophic bacteria. *ISME J.* **11**: 2391–2393.

Ferrera, I., Sarmento, H., Priscu, J., Chiuchiolo, A., Gonzalez, J.M., and Grossart, H.-P. (2017) Diversity and distribution of freshwater aerobic anoxygenic phototrophic bacteria across a wide latitudinal gradient. *Front. Microbiol.* **8**: 175.

Fuhrman, J.A., Cram, J.A., and Needham, D.M. (2015) Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* **13**: 133–146.

Galand, P.E., Gutiérrez-Provecho, C., Massana, R., Gasol, J.M., and Casamayor, E.O. (2010) Inter-annual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). *Limnol. Oceanogr.* **55**: 2117–2125.

Galand, P.E., Pereira, O., Hochart, C., Auguet, J.C., and Debroas, D. (2018) A strong link between marine microbial community composition and function challenges the idea of functional redundancy. *ISME J.* 1.

Galí, M., Simó, R., Vila-Costa, M., Ruiz-González, C., Gasol, J.M., and Matrai, P. (2013) Diel patterns of oceanic dimethylsulfide (DMS) cycling: Microbial and physical drivers. *Global Biogeochem. Cycles* **27**: 620–636.

Gasol, J.M., Cardelús, C., Morán, X.A.G., Balagué, V., Forn, I., Marrasé, C., et al. (2016) Seasonal patterns in phytoplankton photosynthetic parameters and primary production at a coastal NW Mediterranean site. *Sci. Mar.* **80S1**: 63–77.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., et al. (2012) Defining seasonal marine microbial community dynamics. *ISME J.* **6**: 298–308.

Giner, C.R., Balagué, V., Krabberød, A.K., Ferrera, I., Reñé, A., Garcés, E., et al. (2018)

Quantifying long-term recurrence in planktonic microbial eukaryotes. *Mol. Ecol.* https://doi.org/10.1111/mec.14929

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017) Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* **8**: 1–6.

Graham, E.D., Heidelberg, J.F., and Tully, B.J. (2018) Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J.*

Hatosy, S.M., Martiny, J.B.H., Sachdeva, R., STeele, J., and Fuhrman, J.A. (2013) Beta diversity of marine bacteria depends on temporal scale. *Ecology* **94**: 1898–1904.

Jiao, N., Zhang, Y., Zeng, Y., Hong, N., Liu, R., Chen, F., and Wang, P. (2007) Distinct distribution pattern of abundance and diversity of aerobic anoxygenic phototrophic bacteria in the global ocean. *Environ. Microbiol.* **9**: 3091–3099.

Kane, M.D. (2004) Microbial Observatories: Exploring and Discovering Microbial Diversity in the 21st Century. *Microb. Ecol.* **48**: 447–448.

Kim, D.Y., Countway, P.D., Jones, A.C., Schnetzer, A., Yamashita, W., Tung, C., and Caron, D. a (2014) Monthly to interannual variability of microbial eukaryote assemblages at four depths in the eastern North Pacific. *ISME J.* **8**: 515–30.

Koblížek, M. (2015) Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol. Rev.* **39**: 854–870.

Lami, R., Cottrell, M.T., Ras, J., Ulloa, O., Obernosterer, I., Claustre, H., et al. (2007) High abundances of aerobic anoxygenic photosynthetic bacteria in the South Pacific Ocean. *Appl. Environ. Microbiol.* **73**: 4198–4205.

Lami, R., Ras, J., Lebaron, P., and Koblí, M. (2009) Distribution of free-living and particle-attached aerobic anoxygenic phototrophic bacteria in marine environments. **55**: 31–38.

Legendre, P. and Legendre, L. (1988) Numerical Ecology, Volume 24. *(Developments Environ. Model.* **24**: 870.

Lehours, A.-C., Enault, F., Boeuf, D., and Jeanthon, C. (2018) Biogeographic patterns of aerobic anoxygenic phototrophic bacteria reveal an ecological consistency of phylogenetic clades in different oceanic biomes. *Sci. Rep.* **8**: 4105.

Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**: W242–W245.

Louca, S., Parfrey, L.W., and Doebeli, M. (2016) Decoupling function and taxonomy in the global ocean microbiome. *Science* **353**: 1272 LP-1277.

575 Markowitz, V.M., Korzeniewski, F., Palaniappan, K., Szeto, E., Werner, G., Padki, A., et al. (2006)
576 The integrated microbial genomes (IMG) system. *Nucleic Acids Res.* **34**: D344–D348.

577 Martin-Platero, A.M., Cleary, B., Kauffman, K., Preheim, S.P., McGillicuddy, D.J., Alm, E.J., and
578 Polz, M.F. (2018) High resolution time series reveals cohesive but short-lived communities in
579 coastal plankton. *Nat. Commun.* **9**: 266.

580 Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
581 *EMBnet.journal* **17**: 10.

582 Martiny, A.C., Treseder, K., and Pusch, G. (2013) Phylogenetic conservatism of functional traits in
583 microorganisms. *ISME J.* **7**: 830–838.

584 Mašín, M., Zdun, A., Stoń-Egiert, J., Nausch, M., Labrenz, M., Moulisová, V., and Koblížek, M.
585 (2006) Seasonal changes and diversity of aerobic anoxygenic phototrophs in the Baltic Sea.
586 *Aquat. Microb. Ecol.* **45**: 247–254.

587 Massana, R., Murray, A.E., Preston, C.M., Delong, E.F., Massana, R., Murray, A.E., and Preston,
588 C.M. (1997) Vertical distribution and phylogenetic characterization of marine planktonic
589 Archaea in the Santa Barbara Channel . Vertical Distribution and Phylogenetic
590 Characterization of Marine Planktonic Archaea in the Santa Barbara Channel. **63**: 50–56.

591 McMurdie, P.J. and Holmes, S. (2013) Phyloseq: An R Package for Reproducible Interactive
592 Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**:.

593 Nunes, S., Latasa, M., Gasol, J.M., and Estrada, M. (2017) Seasonal and interannual variability of
594 phytoplankton community structure in a Mediterranean coastal site. *Mar. Ecol. Prog. Ser.* **592**:
595 57–75.

596 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2018) vegan:
597 Community Ecology Package.

598 Pedersen, T.L. (2017) ggraph: An Implementation of Grammar of Graphics for Graphs and
599 Networks.

600 R Core Team (2014) R: A language and environment for statistical computing.

601 Ritchie, A.E. and Johnson, Z.I. (2012) Abundance and genetic diversity of aerobic anoxygenic
602 phototrophic bacteria of coastal regions of the pacific ocean. *Appl. Environ. Microbiol.* **78**:
603 2858–2866.

604 Salazar, G., Cornejo-Castillo, F.M., Benítez-Barrios, V., Fraile-Nuez, E., Álvarez-Salgado, X.A.,
605 Duarte, C.M., et al. (2015) Global diversity and biogeography of deep-sea pelagic prokaryotes.
606 *ISME J.* 1–13.

607 Salter, I., Galand, P.E., Fagervold, S.K., Lebaron, P., Obernosterer, I., Oliver, M.J., et al. (2015)

608 Seasonal dynamics of active SAR11 ecotypes in the oligotrophic Northwest Mediterranean
609 Sea. *Isme J* **9**: 347–360.

610 Schwalbach, M.S. and Fuhrman, J.A. (2005) Wide-ranging abundances of aerobic anoxygenic
611 phototrophic bacteria in the world ocean revealed by epifluorescence microscopy and
612 quantitative PCR. *Limnol. Oceanogr.* **50**: 620–628.

613 Shade, A., Caporaso, J.G., Handelsman, J., Knight, R., and Fierer, N. (2013) A meta-analysis of
614 changes in bacterial and archaeal communities with time. *ISME J.* **754**: 1493–1506.

615 Shade, A., Jones, S.E., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N., and Gilbert, A.
616 (2014) Conditionally Rare Taxa Disproportionately Contribute to Temporal Changes in
617 Microbial Diversity. **5**: 1–9.

618 Sieracki, M.E., Gilg, I.C., Thier, E.C., Poulton, N.J., and Goericke, R. (2006) Distribution of
619 planktonic aerobic anoxygenic photoheterotrophic bacteria in the northwest Atlantic. *Limnol.*
620 *Oceanogr.* **51**: 38–46.

621 Smith, B. and Wilson, J.B. (1996) A Consumer ' s Guide to Evenness Indices. *Oikos* **76**: 70–82.

622 Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
623 phylogenies. *Bioinformatics* **30**: 1312–1313.

624 Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., et al. (2015)
625 Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**:
626 1261359.

627 Waidner, L. a and Kirchman, D.L. (2007) Aerobic anoxygenic phototrophic bacteria attached to
628 particles in turbid waters of the Delaware and Chesapeake estuaries. *Appl. Environ. Microbiol.*
629 **73**: 3936–44.

630 Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017) Normalization
631 and microbial differential abundance strategies depend upon data characteristics. *Microbiome*
632 **5**: 27.

633 Xia, L.C., Steele, J.A., Cram, J.A., Cardon, Z.G., Simmons, S.L., Vallino, J.J., et al. (2011) Extended
634 local similarity analysis (eLSA) of microbial community and other time series data with
635 replicates. *BMC Syst. Biol.* **5**: S15.

636 Yang, Q., Jiang, Z.-W., Huang, C.-H., Zhang, R.-N., Li, L.-Z., Yang, G., et al. (2018) Hoeflea
637 prorocentri sp. nov., isolated from a culture of the marine dinoflagellate Prorocentrum
638 mexicanum PM01. *Antonie Van Leeuwenhoek* **111**: 1845–1853.

639 Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K., et al. (2007)
640 The Sorcerer II global ocean sampling expedition: Expanding the universe of protein families.

641 *PLoS Biol.* **5**: 0432–0466.

642 Yutin, N., Suzuki, M.T., and Be, O. (2005) Novel Primers Reveal Wider Diversity among Marine
643 Aerobic Anoxygenic Phototrophs †. **71**: 8958–8962.

644 Yutin, N., Suzuki, M.T., Teeling, H., Weber, M., Venter, J.C., Rusch, D.B., and Béjà, O. (2007)
645 Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface
646 waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition
647 metagenomes. *Environ. Microbiol.* **9**: 1464–1475.

648 Zeng, Y.H., Chen, X.H., and Jiao, N.Z. (2007) Genetic diversity assessment of anoxygenic
649 photosynthetic bacteria by distance-based grouping analysis of pufM sequences. *Lett. Appl.*
650 *Microbiol.* **45**: 639–645.

651

Tables

Table 1. Summary information for the top 20 ASVs. The following columns are listed: ASV name, OTU correspondence, phylogroup correspondence, taxonomy, occurrence, relative abundance, seasonality behavior, and month of maximum mean relative abundance.

ASV	OTU	Taxonomy	Phylogroup	Occurrence	Relative abundance (%)	Seasonality	Month max. abundance
ASV1	OTU1	*Gammaproteobacteria*	Group K	114	16.37	yes	Jun
ASV2	OTU1	*Gammaproteobacteria*	Group K	112	7.45	yes	Jun
ASV6	OTU6	*Gammaproteobacteria*	Group K	119	6.9	yes	Dec
ASV5	OTU5	*Gammaproteobacteria*	Group K	117	6.75	yes	Nov
ASV8	OTU8	*Alphaproteobacteria*	Group E	81	5.83	yes	Feb
ASV10	OTU10	*Gammaproteobacteria*	Group K	84	4.89	yes	Apr
ASV11	OTU5	*Gammaproteobacteria*	Group K	108	4.58	yes	Jan
ASV14	OTU14	*Gammaproteobacteria*	Group E	57	4.17	yes	Mar
ASV18	OTU18	*Gammaproteobacteria*	Group K	99	3.71	yes	Apr
ASV15	OTU1	*Gammaproteobacteria*	Group K	107	3.28	yes	Nov
ASV17	OTU1	*Gammaproteobacteria*	Group K	78	3.13	yes	Jun
ASV21	OTU5	*Gammaproteobacteria*	Group K	108	2.93	yes	Nov
ASV23	OTU23	*Gammaproteobacteria*	Group K	113	2.41	yes	Dec
ASV13	OTU13	*Gammaproteobacteria*	Group K	97	2.27	yes	Jan
ASV26	OTU10	*Gammaproteobacteria*	Group K	63	1.74	yes	Feb
ASV24	OTU1	*Gammaproteobacteria*	Group K	48	1.42	yes	Jun
ASV27	OTU1	*Alphaproteobacteria*	Group K	71	1.19	yes	Jul
ASV30	OTU30	*Gammaproteobacteria*	Group K	71	1.08	yes	Oct
ASV34	OTU23	*Gammaproteobacteria*	Group K	84	1.06	yes	Jan
ASV37	OTU37	*Gammaproteobacteria*	Group E	81	1.04	no	Apr

Figure Legends

Figure 1. Alphadiversity distribution of the AAP community for each month colored by season. Richness (number of observed ASVs) and Shannon indexes obtained through amplicon sequencing over a decade (2004-2013) are shown in the top and middle panels respectively. Each boxplot presents the median and interquartile range of the distribution of 10 data points shown in grey (with the exception of March, with 9 data points). Whiskers represent 1.5 times the interquartile range. The bottom panel shows the Shannon index values obtained from the metagenomic dataset (2011 to 2013). The colored dots represent the mean monthly values and the bars the standard error of the mean for the 3-year period.

Figure 2. Variation in the relative contributions of AAP phylogroups K (top panel), G, E (middle panel), phylogroups D, F, I, J, *Sphingomonas*-like and the unclassified group (bottom panel) for each month over the studied decade (2004-2013). Each boxplot presents the median and the 25% and 75% limits with the distribution of 10 data points in grey (with the exception of March, with 9 data points), and whiskers represent 1.5 times the interquartile range.

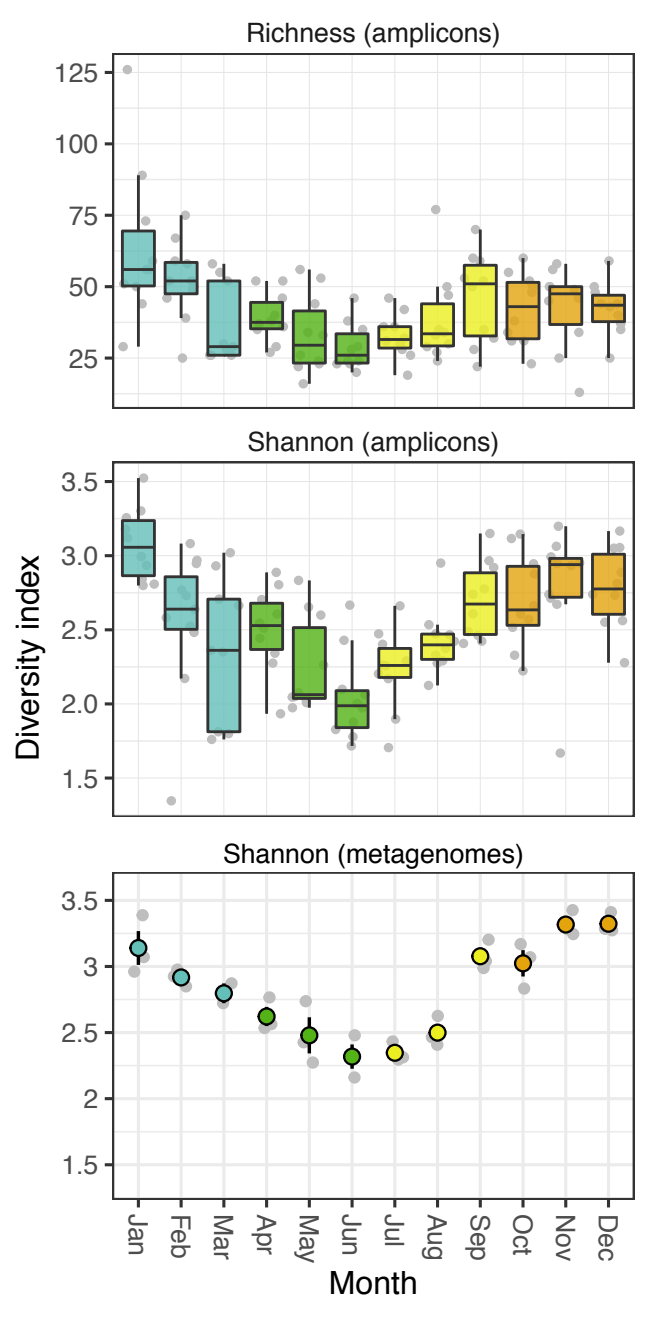
Figure 3. Distance based redundancy analysis of the samples (dots) with the 5 explanatory variables (arrows) influencing the distribution (PERMANOVA $p<0.01$; day length, temperature, salinity, silicate concentration (Si), Chlorophyll *a* (Chl*a*) and phototrophic nanoflagellate abundance (PNF)). The ordination was performed on the Bray-Curtis dissimilarity of $\log_{10}$ transformed data (with a pseudocount of 1) matrix (after rarefying). Samples are colored by season.

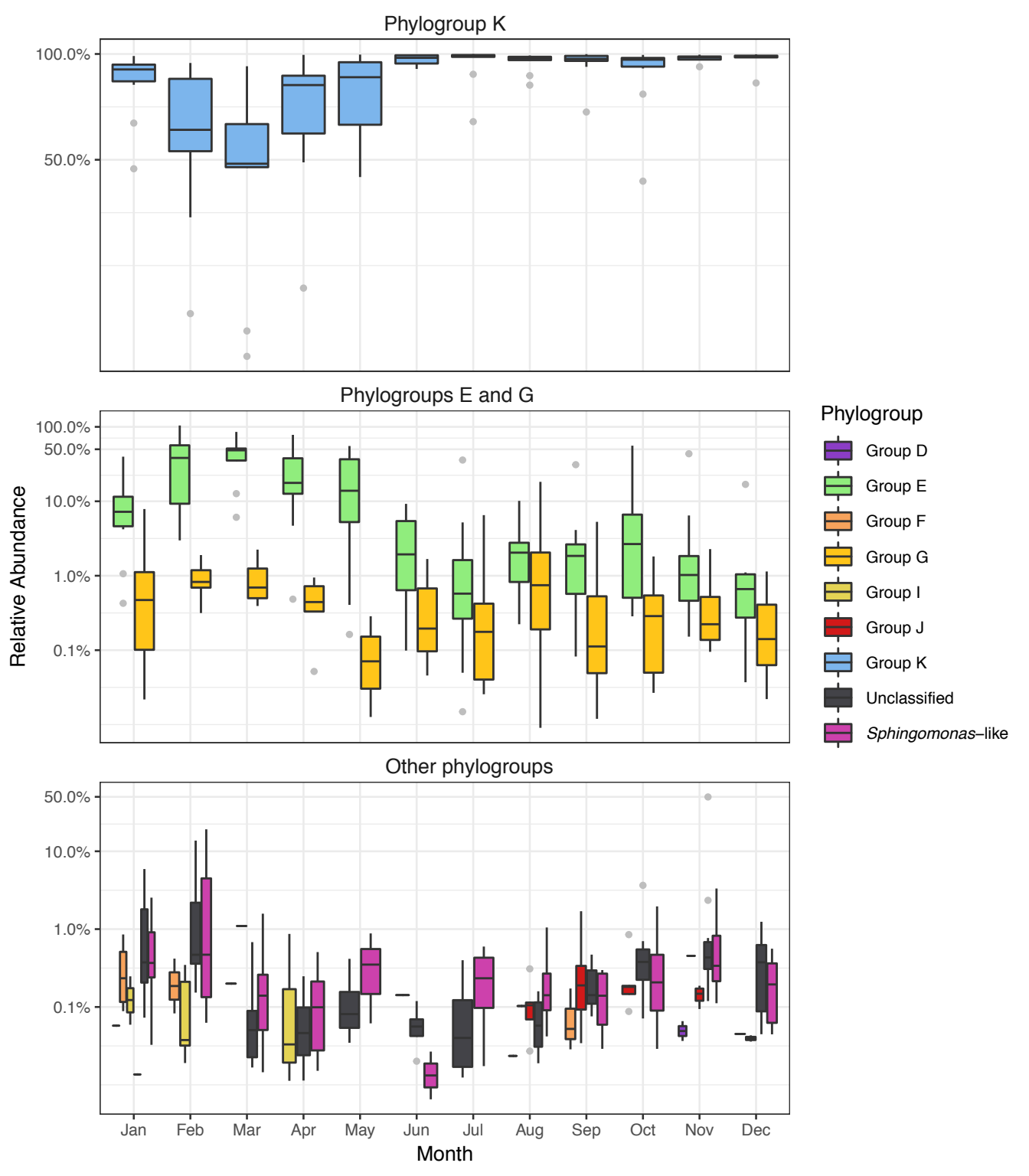
Figure 4. Bray-Curtis similarity between samples plotted against the time lag between each of them (time-decay plot). Mean similarity values for each time lag are plotted in an empty black dot with standard error bars (background grey filled dots show each comparison). A linear regression is plotted with 95% confidence intervals shown.

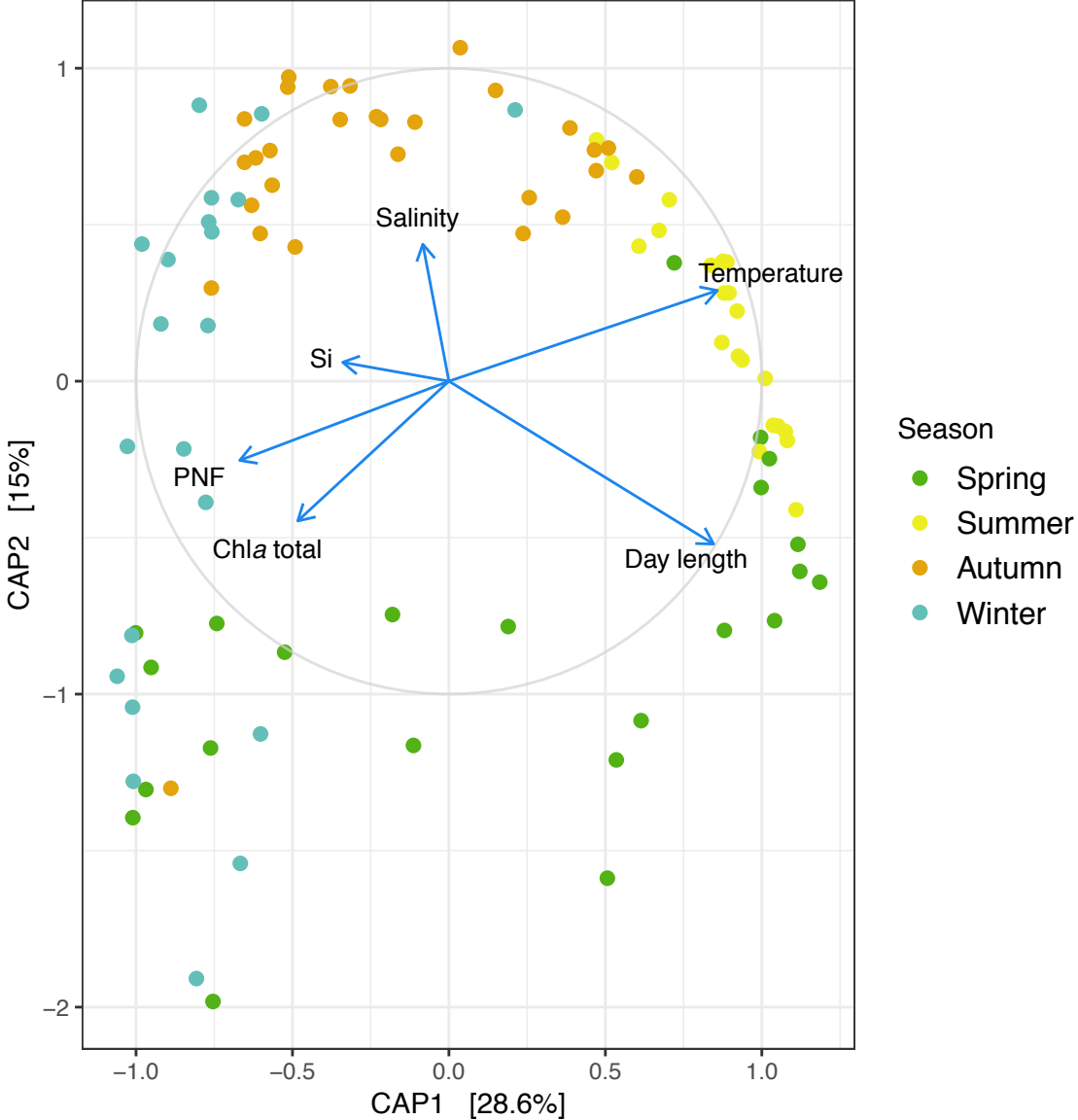
Figure 5. Fast local similarity network showing clusters with >=3 nodes. Node shape designates the type of variable, with the filling specifying the phylogroup, the size the total relative contribution and the stroke color if the ASV displays a seasonal behavior. Edges can be lagged (discontinuous line) or direct and have negative (i.e., anticorrelation) or positive local scores (LS). The label on the nodes indicates the ASV number. Tº: temperature; PNF: abundance of phototrophic nanoflagellates; Peuk1: abundance of picoeukaryotes group I (see Supplementary Information 1 for details).

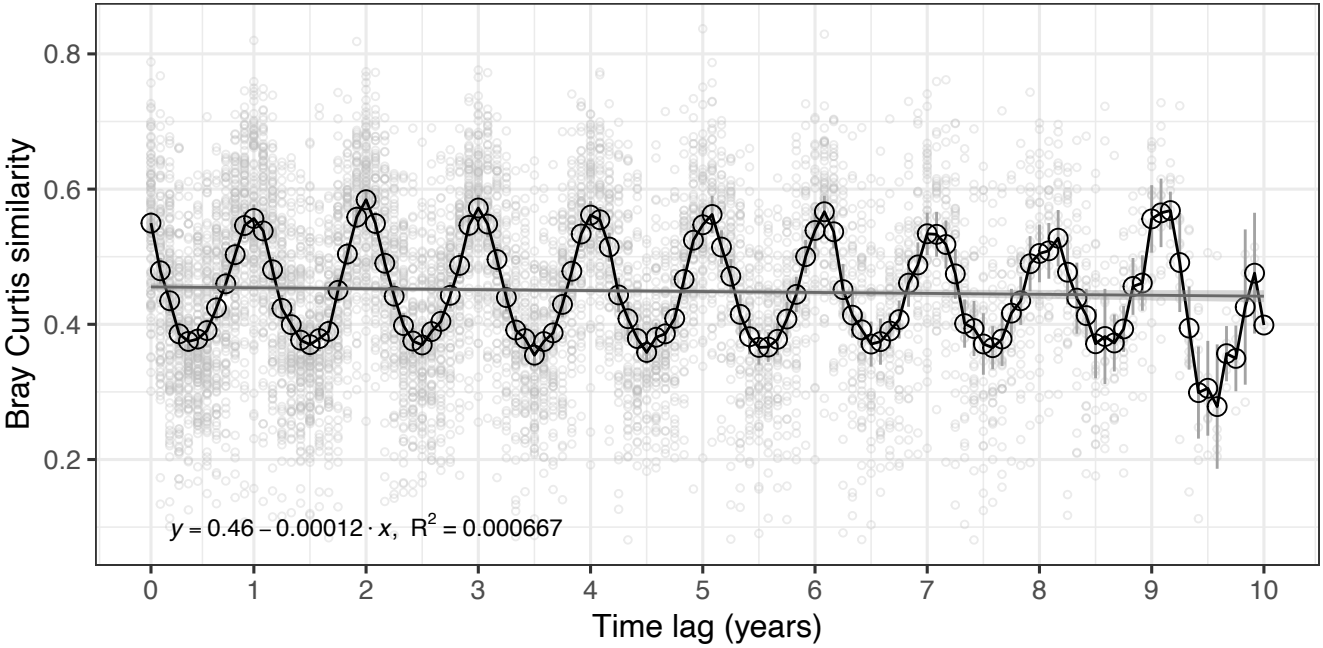
Figure 6. Seasonal component of the relative abundance distribution ($\log_{10}$ +1 transformed) for some remarked ASVs fitted with a polynomial function. (A) Various ASVs with distant nucleotide similarity, colored by phylogroup assignation. (B) Various ASVs belonging to OTU1 (dashed line

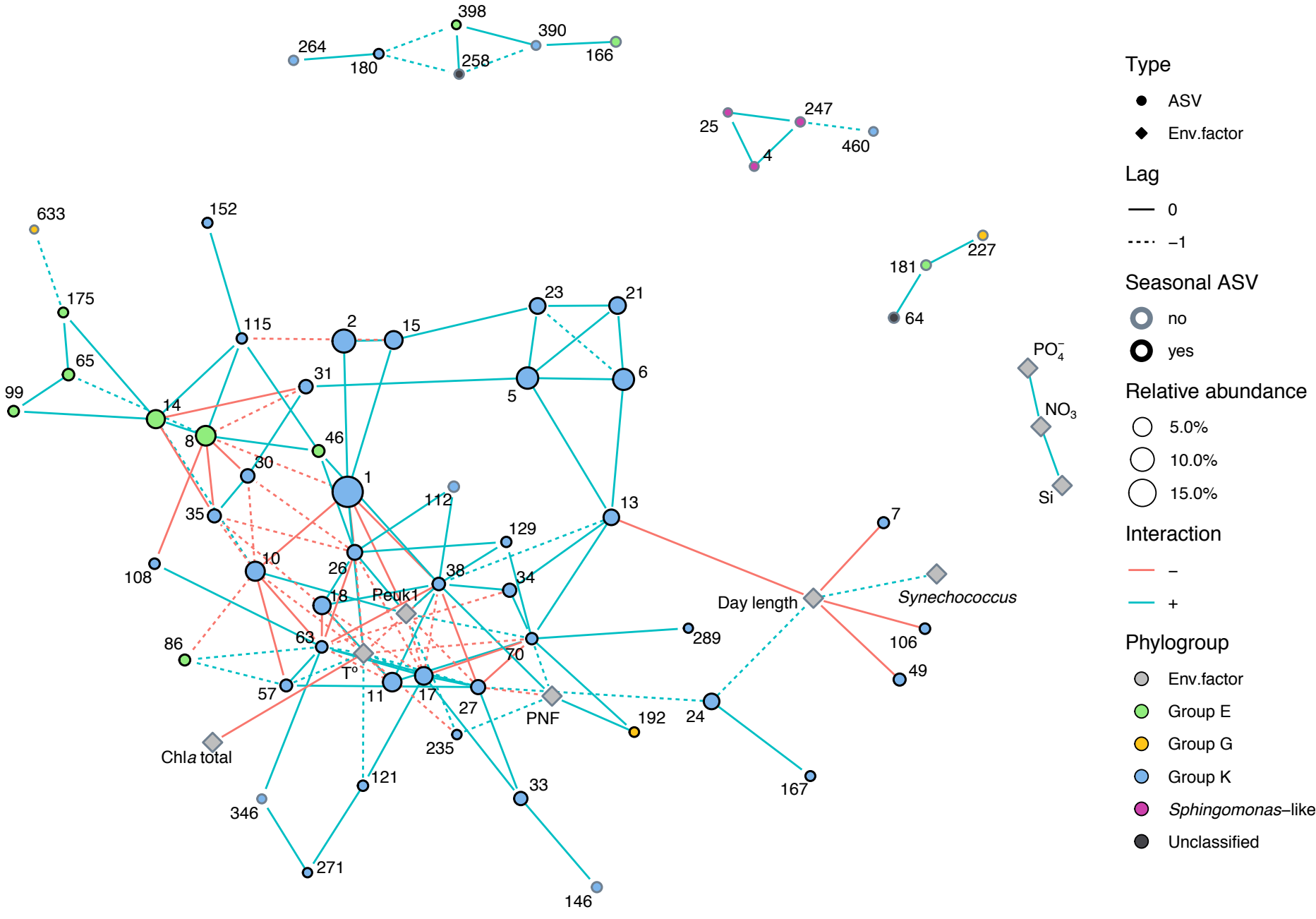
687 corresponds to OTU1). The patterns were defined based on the relative abundance dynamics of 10
688 years by time series analysis.

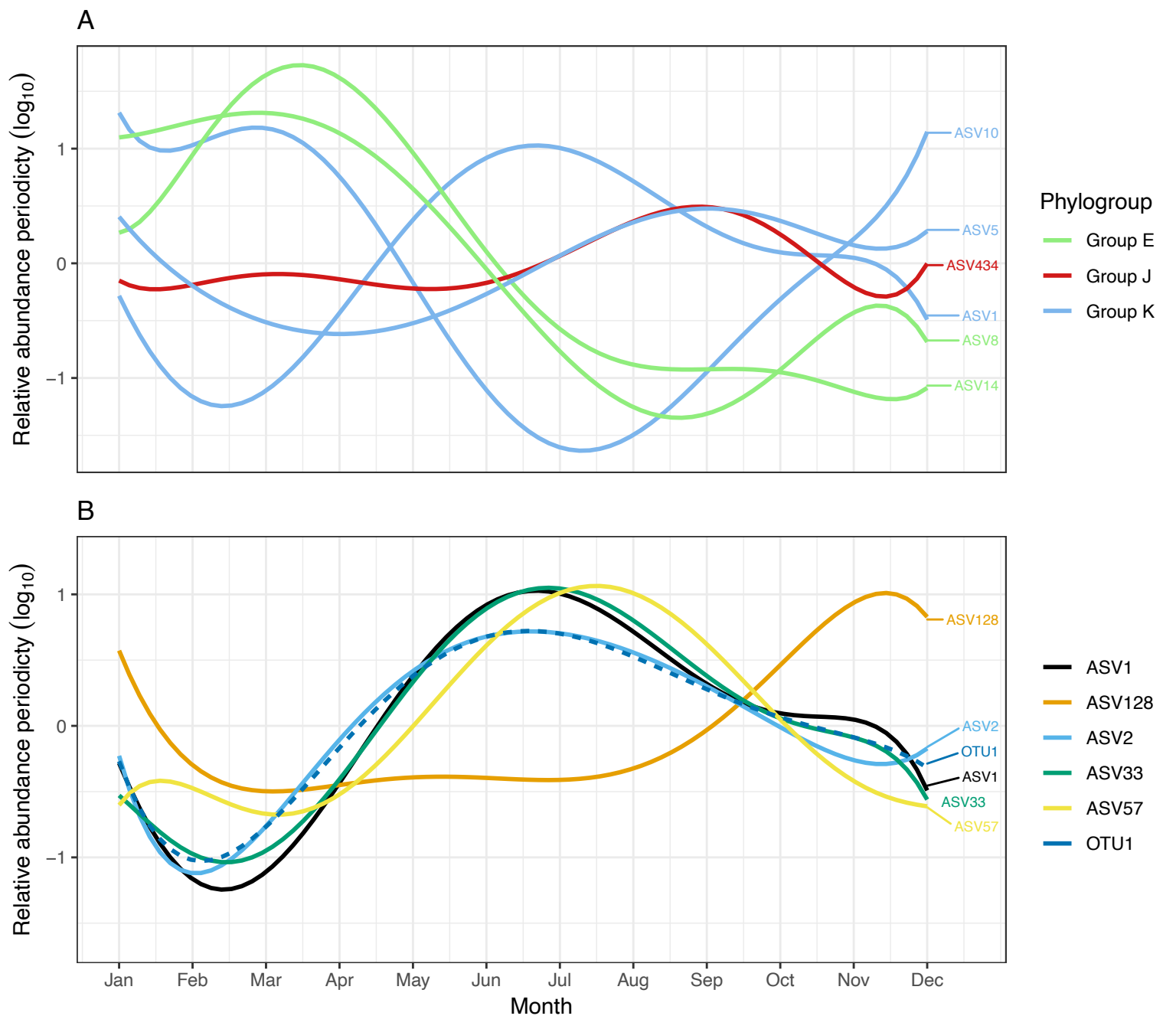

Richness (amplicons)
125
100
75
50
25
Shannon (amplicons)
Diversity index
3.5
3.0
2.5
2.0
1.5
Shannon (metagenomes)
3.5
3
2.5
2
1.5
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month


Phylogroup K
100.0%
50.0%
Phylogroups E and G
Relative Abundance
100.0%
50.0%
10.0%
1.0%
0.1%
Other phylogroups
50.0%
10.0%
1.0%
0.1%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
Phylogroup
Group D
Group E
Group F
Group G
Group I
Group J
Group K
Unclassified
Sphingomonas–like


Salinity
Temperature
Si
PNF
Chl*a* total
Day length
CAP2 [15%]
CAP1 [28.6%]
1
0
−1
−2
−1.0
−0.5
0.0
0.5
1.0
Season
Spring
Summer
Autumn
Winter


Bray Curtis similarity
0.8
0.6
0.4
0.2
$y = 0.46 - 0.00012 \cdot x$, $R^2 = 0.000667$
0
1
2
3
4
5
6
7
8
9
10
Time lag (years)


398
264
390
180
258
166
247
25
460
4
633
152
227
181
175
64
115
23
21
2
15
65
31
6
99
5
14
PO4−
8
46
NO3
30
1
112
13
35
129
7
Si
10
108
26
38
34
18
Peuk1
Day length
Synechococcus
289
86
63
106
T°
70
49
57
11
17
27
PNF
192
24
Chla total
235
167
121
33
346
271
146
Type
ASV
Env.factor
Lag
0
−1
Seasonal ASV
no
yes
Relative abundance
5.0%
10.0%
15.0%
Interaction
−
+
Phylogroup
Env.factor
Group E
Group G
Group K
Sphingomonas–like
Unclassified


A
Relative abundance periodicty (log10)
1
0
−1
ASV10
ASV5
ASV434
ASV1
ASV8
ASV14
Phylogroup
Group E
Group J
Group K
B
Relative abundance periodicty (log10)
1
0
−1
ASV128
ASV2
OTU1
ASV1
ASV33
ASV57
ASV1
ASV128
ASV2
ASV33
ASV57
OTU1
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month