Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the Mediterranean Sea in relation to environmental variables

Anne-Catherine Lehours, Matthew T. Cottrell, Océane Dahan, David L. Kirchman, Christian Jeanthon

Submitted to FEMS Microbiol. Ecol.

Title: Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the Mediterranean Sea in relation to environmental variables

Authors: Anne-Catherine Lehours[1,2*], Matthew T. Cottrell[3], Océane Dahan[1,2], David L. Kirchman[3] and Christian Jeanthon[1,2]

Address :

[1]UPMC Univ Paris 06, UMR 7144, Adaptation et Diversité en Milieu Marin, Station Biologique de Roscoff, 29682 Roscoff, France

[2]CNRS, UMR 7144, Adaptation et Diversité en Milieu Marin, Station Biologique de Roscoff, 29682 Roscoff, France

[3]University of Delaware, School of Marine Science and Policy, Lewes, Delaware 19958, USA

*Present address: Clermont Université, Université Blaise Pascal, UMR CNRS 6023, Laboratoire Microorganismes : Génome et Environnement (LMGE), BP80026, F-63171 Aubiere

For correspondence. Email: jeanthon@sb-roscoff.fr. Tel. (+33) 298 292 379; Fax (+33) 298 292 324.

Abstract

Aerobic anoxygenic phototrophic bacteria (AAP) represent an important fraction of bacterioplankton assemblages in various oceanic regimes and have probably a great impact on organic carbon production and cycling in the upper ocean. Although their abundance and distribution have been recently explored in diverse oceanic regions, the environmental factors controlling the population structure and diversity of these photoheterotrophic bacteria remain poorly understood. Here, we investigate the horizontal and vertical distributions and the genetic diversity of AAP populations collected in late summer throughout the Mediterranean Sea using *pufM*-temporal temperature gradient electrophoresis (TTGE) and clone library analyses. The TTGE profiles and clone libraries analyzed using multivariate statistical methods demonstrated a horizontal and vertical zonation of AAP assemblages. Physico-chemical parameters such as pH, inorganic nitrogen compounds, photosynthetically active radiation, total organic carbon and to a lesser extend particulate organic nitrogen and phosphorus, and biogenic activities (e.g. bacterial production, cell densities) acted in synergy to explain population changes with depth. About half of the *pufM* sequences were less than 94% identical to known sequences. The AAP populations were predominantly (~ 80%) composed of *Gammaproteobacteria*, unlike previously explored marine systems. Our results suggest that genetically distinct ecotypes inhabiting different niches may exist in natural AAP populations of the Mediterranean Sea whose genetic diversity is typical of oligotrophic environments.

Introduction

Aerobic anoxygenic phototrophic bacteria (AAP) represent a functional group that was recently found to account for a significant fraction of the bacterioplankton in marine illuminated environments (Kolber *et al*., 2001; Cottrell *et al*., 2006; Koblížek *et al*., 2007; Mašín *et al*., 2006; Sieracki *et al*., 2006; Lami *et al*., 2007; Jiao *et al*., 2007). These bacteriochlorophyll (BChl *a*)-containing prokaryotes, which can use both light and organic matter for energy production, require oxygen and can use reduced organic compounds as electron donors (Yurkov & Csotonyi, 2009). Although they are not primary producers, their higher growth rates and efficiency in organic carbon utilization over strict heterotrophs are likely to make them dynamic and significant contributors to the organic carbon production and cycling in the upper ocean (Koblížek *et al*., 2007). While physiological evidences suggest that they would have a competitive advantage over strict heterotrophs in low-nutrient conditions (Yurkov & van Gemerden 1993; Suyama *et al*., 2002), the emerging findings indicate that these bacteria may be adapted to a broad range of trophic conditions and are abundant in eutrophic and oligotrophic environments (Cottrell *et al*., 2006; Sieracki *et al*., 2006; Waidner & Kirchman, 2007).

Molecular analyses based on the *pufM* gene encoding the M-subunit of the photosynthetic reaction center have revealed that AAP bacteria belong to different groups of *Alpha*-, *Beta*-, and *Gammaproteobacteria* (Béjà *et al*., 2002; Yutin *et al*., 2007). Their abundance and distribution have been explored in diverse oceanic regions and have been shown to vary greatly among oceanic regimes (Cottrell *et al*., 2006; 2008; Jiao *et al*., 2007, Imhoff, 2001; Yutin *et al*., 2007; Waidner & Kirchman, 2008). However, AAP bacteria remain clearly undersampled in several areas, particularly in oligotrophic environments, that

represent 60% of the oceans. The environmental factors controlling the population structure and diversity of these photoheterotrophic bacteria remain yet poorly understood (Eiler, 2006).

The Mediterranean Sea is an ideal environment for these ecological studies as it offers a range of trophic conditions including extreme oligotrophy, particularly in summer when the water column is strongly stratified (Berman *et al*., 1985). While the N:P ratio is close to the Redfield ratio (16:1) in most oceanic waters, Mediterranean waters have a higher ratio, especially in the eastern Basin, leading to strong phosphorus limitation (Moutin & Raimbault, 2002). A complex thermohaline circulation coupled with regional hydrodynamic features also contributes to the establishment of many different oceanic regions throughout the Mediterranean Sea (Manca *et al*., 2004). As an example, the exchange of the Atlantic and Mediterranean water masses at the Strait of Gibraltar induces marked salinity and temperature gradients (Gascard & Richez, 1985). Although the different trophic conditions available in the Mediterranean Sea provide a unique context to link nutrient availability, trophic status and functioning of the food web to the dynamics of photoheterotrophic populations, the distribution and diversity patterns of AAP bacteria have been only partially explored (Oz *et al*., 2005; Yutin *et al*., 2005; 2008).

In the present study, we analyzed the biogeography patterns of AAP populations collected in late summer along two transect during the PROSOPE (PROductivité des Systèmes Océaniques PElagiques) cruise. To explore what environmental factors control structure and diversity of AAP populations in the Mediterranean Sea, we monitored their longitudinal and vertical changes using a *pufM*-based PCR-Temporal Temperature Gel Electrophoresis (TTGE) survey and analyzed the *pufM* clone libraries from selected stations and depths.

Material and Methods

Sampling and nucleic acid extraction

Seawater samples were collected from seven stations along two transects in the Mediterranean Sea (from Gibraltar to the Ionian Sea off the north-east coast of Lybia and north-west to the French coast through the Tyrrhenian and Ligurian Seas) in September and October 1999 during the PROSOPE cruise aboard the R.V. *La Thalassa* (Fig. 1). For molecular diversity studies of AAP populations, 1.45-5 l water samples were retrieved using 12 l Niskin bottles fitted on a Rosette sampler equipped with conductivity, temperature and depth (CTD) sensors. Seawater was prefiltered through 3 µm pore-size, 47 mm diameter, Nuclepore filters using moderate vacuum in order to separate picoplankton from larger organisms. Picoplanktonic cells were collected by filtration as previously described (Marie *et al*., 2006; Garczarek *et al*., 2007). The filtered biomass was transferred into a cryovial containing 3.5 ml of DNA lysis buffer (0.75 M sucrose, 50 mM Tris-HCl, pH 8) and immediately frozen in liquid nitrogen. DNA extraction was performed as previously described (Marie *et al*., 2006). Ancillary data (nutrients, dissolved oxygen, chlorophyll a, salinity, temperature, *etc*.) and methods used to analyze them are available from the PROSOPE web site (http://www.obs-vlfr.fr/cd_rom_dmtt/pr_main.htm).

Environmental *pufM* gene amplification

Multiple combinations of previously designed primer sets (Achenbach *et al*., 2001; Béjà *et al*., 2002; Yutin *et al*., 2005) were tested (data not shown). On the basis of specificity and efficiency (e. g. yield) results, we selected PufMF forward (5'-TACGGSAACCTGTWCTAC-3', Béjà *et al*., 2002) and PufWAW reverse primers (5'-

AYNGCRAACCACCANGCCCA-3', Yutin *et al*., 2005) to amplify partial sequences of the *pufM* gene (245 bp fragments). For TTGE analyses, a 5'-Cy5-labeled PufMF primer and a primer PufWAW with a 40 bp GC-clamp (CGCCCGCCGCGCCCCGCGCCCGTCCCGCCGCCCCCGCCCG) added at the 5'-end were used. Reaction mixture (50 µL) contained the following components: 5X buffer (10µl), 2 mM $MgCl_2$, 10 pmoles of each deoxyribonucleotide triphosphate (dATP, dCTP, dGTP, dTTP; Eurogenetec), 10 pmoles of each oligonucleotide primer, 2.5 U of GoTaq Flexi DNA polymerase (Promega) and 50 to 100 ng of template DNA. Amplifications were carried out in a GeneAmp PCR system 9700 (Applied Biosystems) with the following parameters: 95°C for 5 min, followed by 35 cycles of 95°C for 30 s, 58°C for 30 s, and 72°C for 30 s, with a final extension step at 72°C for 10 min. Amplified products were checked by electrophoresis in 1.5% agarose in 0.5× Tris-Acetate-EDTA (TAE) buffer and further quantified with a DNA quantitation fluorescence assay kit (Sigma-Aldrich).

Temporal Temperature gel Gradient Electrophoresis (TTGE) profiling and analyses

One hundred nanograms of each amplified product were electrophoresed along an 8% (wt/vol) polyacrylamide gel (ratio acrylamide to bis-acrylamide 37.5:1) containing 7M urea, 1.25X TAE, 0.06% of N,N,N',N'-Tetramethylethylenediamine (Temed) and 0.06% ammonium persulfate using the DCode Universal Mutation Detection System (BioRad, Hercules, CA). Runs were performed in 1.25x TAE at 68 V for 17 h with a temperature range of 66 to 69.7°C and a ramp rate of 0.2°C h^{-1}. Standard markers were generated with a mixture of *pufM* PCR products amplified from *Erythrobacter longus* strain OCh 101^T, *E. litoralis* strain $T4^T$, *Roseobacter denitrificans* strain OCh 114^T (CIP104266), *Dinoroseobacter shibae* strain $DFL12^T$. Gel images were obtained at 100-µm resolution using a Typhoon Trio variable mode imager (Amersham Biosciences, Piscataway, NJ). Typhoon scans were acquired using

the 633 nm excitation laser and the 670 BP 30 emission filter as recommended by the manufacturer for the detection of Cy5-labeled molecules. All gels were scanned with photomultiplier tube voltages to maximize signal without saturating fingerprint bands. Band patterns were analyzed with GelCompare 4.6 software package (Applied Maths, Kortrijk, Belgium). In band assignment, a 1% band position tolerance (relative to total length of the gel) was applied, which indicates the maximal shift allowed for two bands in different TTGE patterns to be considered as identical. The number of bands in a profile was expressed as the phylotype richness and the Shannon-Weiner index (H') was calculated as previously described (Hill *et al*., 2003). The Smith and Wilson evenness index (E_{var}, Smith & Wilson, 1996) was calculated using the Ecological Evenness Calculator software (http://www.nateko.lu.se/personal/benjamin.smith/software).

Clone library construction and analyses

Fresh PCR products were cloned using the TOPO TA cloning kit according to the manufacturer's instructions (Invitrogen Corporation, Carlsbad, CA, USA). Recombinant clones were screened for insert-containing plasmids by direct PCR amplification with M13 forward and reverse primers. Clones were sequenced using an ABI 3130 POP7 sequencer (Applied Biosystems, Foster City, CA) at the Biogenouest Sequencing-Genotyping Platform (Roscoff site). Clone libraries were screened for chimeric sequences with Chimera_Check program available on the RDP website (Cole *et al*., 2003). The 388 remaining sequences were subjected to BLAST search against publicly available sequences to determine their approximate phylogenetic affiliations. A conservative value of 94 % nucleic acid sequence similarity (Zeng *et al*., 2007) was chosen for grouping sequences into Operational Taxonomic Units (OTUs) using the phylogenetic analysis software Bosque available at http://bosque.udec.cl (Ramírez-Flandes & Ulloa, 2008). Coverage value (C) was calculated as

previously described (Mullins *et al*., 1995). The Shannon-Weiner index (H', Hill *et al*., 2003), the richness-estimator S_{chao1} (Hughes *et al*., 2001) and the Abundance-base Coverage Estimator (ACE, Chazdon *et al*., 1998) were computed using EstimateS software Version 7.5. (K. Colwell, http://purl.oclc.org/estimates).

A *pufM* database containing more than 700 aligned sequences of cultured species and environmental clones retrieved from GenBank database (http://www.ncbi.nlm.nih.gov/Genbank/) and the GOS scaffold nucleotide sequences was constructed using ARB (Ludwig *et al*., 2004). Sequences were translated to protein and the resulting alignment was then used to manually realign nucleotide sequences. A neighbor-joining tree was first constructed with all the sequences longer than 600 bp and the robustness of inferred tree topologies was tested by bootstrap analysis (1000 resamplings) using PHYLIP (Felsenstein, 1993). Shorter sequences were aligned as above and added to the tree using ADD-BY-PARSIMONY algorithm. Phylogenetic tree display and annotation were performed with iTOL software (Letunic & Bork, 2006). The *pufM* sequences obtained in this study are deposited in the GenBank database under accession **No. GQ468944 to GQ468986.**

Statistical analyses

The normality of environmental variables was checked using Shapiro-Wilk (Shapiro, 1965) and Anderson-Darling (Stephens, 1974) tests and variables were transformed when necessary to correct for deviations from normality. Principal component analysis (PCA), performed with XLSTAT version 6.01 (Addinsoft), was used to group samples according to environmental variables using the Pearson correlation coefficient.

For TTGE patterns, pairwise similarity matrices were calculated using the Dice and the Bray-Curtis equations for presence/absence and relative peak height data, respectively. Dendrograms were generated from the Dice matrix using the method described by Ward

(1963). The consistence of a cluster was expressed by the cophenetic correlation which calculates the correlation between the dendrogram-derived similarities and the matrix similarities. A distance of 35 was used to separate clusters in the hierarchical classification. Analysis of similarity (ANOSIM, Clarke and Warwick, 2001) was used to test the hypothesis that communities within each cluster were more similar to each other than to communities in other clusters. Correlations between similarity matrices were calculated using a Mantel test (Mantel, 1967) with 10000 permutations and were performed with XLSTAT version 6.01. Relationships among samples were visualized using the ordination technique multidimensional scaling (MDS) using a standardized stress with 1000 iterations computed with XLSTAT version 6.01. Canonical correspondence analyses (Legendre & Legendre, 1998) used to determine the extent to which selected environmental variables explained patterns of similarity in community composition were performed using PAST 1.81 (Hammer *et al*., 2001) available at http://palaeo-electronica.org/2001_1/past/issue1_01.htm

Results

Oceanographic context

The PROSOPE cruise sampled the highly eutrophic Morocco upwelling off Agadir (St. UPW, Fig. 1) and then proceeded to a first Mediterranean west-east (W-E) transect, through a markedly decreasing gradient of surface chlorophyll *a* concentrations, from the mesotrophic St. 1 at the Strait of Gibraltar to St. MIO located in a highly oligotrophic area in the center of the Ionian Sea. A second transect sampled the Tyrrhenian and Ligurian Seas (St. 9, St. DYF).

Plots of the depth variations of environmental variables analyzed along both these transect revealed the complexity of the studied zones in which several physical and chemical gradients were superimposed (Figs. S1a to S1l). Temperature and salinity transects reflected

the progression of low-salinity, low-temperature Atlantic waters penetrating through the Gibraltar Strait and flowing along the African coast (Figs. S1a-b). Thus a strong halocline was observed between St. 3 and St. 5 identifying a partitioning of surface waters into westwards and typical high salinity eastwards (21.6°C and 36.7‰ for St. 1 to 26°C and 38.9‰ for St. MIO, respectively). The upwelling at St. UPW supplied abundant dissolved nutrients to the surface layer, which supported phytoplankton growth as indicated by the surface maximum in chlorophyll *a* (Fig. S1c). A progressive W-E deepening of the nitracline and phosphacline down to 98 and 124 m depth at St.1 and St. MIO, respectively, was noted (Figs. S1d-e). Along the second transect (from St. MIO to St. DYF), concentrations of nitrates and phosphates increased below the deep chlorophyll maximum (DCM).

The main biological and physico-chemical parameters of the waters were analyzed using a principal component analysis (PCA) including nineteen variables (salinity, temperature, dissolved oxygen, light, pH, depth, phosphate, nitrate, nitrite, particulate organic carbon, nitrogen, phosphorus, total chlorophyll a, accessory pigments, bacterial production, abundances of *Synechococcus*, *Prochlorococcus*, non photosynthetic bacteria and picoeukaryotes) (Fig. S2). This analysis allowed separation of the Mediterranean Sea euphotic zone into two distinct layers above and in/or below the DCM, whose discrimination was mainly explained by biological variables and nutrient levels (NO_3, NO_2, PO_4) respectively. All surface waters except that of the Morocco upwelling and the Gibraltar Strait (St. 1) were nutrient depleted (Figs. S1 and S2). Covariation of biological parameters with particulate organic compounds, oxygen, pH and photosynthetically available radiation (PAR) are consistent with the characteristics of the upper layer where photosynthesis supports the formation of organic matter (Lucea *et al*., 2003) allowing heterotrophic activities as indicated by the inter-correlation of these parameters with bacterial production and cell densities (Fig. S1 and S2). Deep euphotic waters located in or below the DCM were relatively similar as they

were cold, enriched in nutrients and depleted in oxygen (Fig. S1). Although a strong halocline partitioned W-E surface waters, salinity did not explain variability among samples. This might result from the superimposed longitudinal and vertical variation trends of salinity concentrations (Fig. S1b). Discrimination of surface and deep euphotic layers in the Mediterranean Sea mainly according to their nutrient status suggests that the vertical gradient may prevail over the longitudinal gradient during summer stratification.

Spatial variability, diversity and structure of AAP populations

PCR amplified *pufM* genes from different depths (Fig. S3) were analyzed by temporal temperature gradient gel electrophoresis (TTGE) which yielded a total of 79 unique bands with an average of 34± 6 bands in each sample (Fig. 2). A Mantel test also showed that Bray-Curtis and Dice similarity matrices calculated from TTGE profiles were significantly correlated ($r=0.748$, $p<0.05$). The hierarchical clustering analysis (HCA) identified four clusters of AAP bacteria (Fig. 2) confirmed by ANOSIM statistics performed from both presence/absence and relative intensity of TTGE bands (data not shown). Cluster A contained populations from the coastal stations UPW and St. 1, together with those of the surface waters of the Algerian Basin (St. 3-5m). Cluster B represented the deep euphotic zone of the Algerian Basin (St. 3) and the Strait of Sicily (St. 5) whereas Cluster C grouped TTGE profiles of samples collected above the DCM in the Ionian, Tyrrhenian, Ligurian Seas and the Strait of Sicily. AAP bacteria from deep euphotic waters of stations St. 9 and St. DYF clustered together in Cluster D. Richness and diversity (E_{var} and H') values were not significantly different between clusters (data not shown). E_{var}, H' and richness were positively correlated to PAR and biological activity indicators and were negatively correlated with NO_3 and NO_2 concentrations.

To obtain two-dimensional-coordinates for samples and to confirm HCA groupings, ordination of Bray-Curtis similarities among sample profiles was performed by non-metric

multidimensional scaling (MDS). With the exception of samples St1-80m and St. 3-5m, the four clusters recovered by HCA were defined (data not shown). Populations above the DCM were significantly more similar to each other than to those in or below the DCM (R=0.419, p<0.001). Dimension 1 from the TTGE/MDS analyses was negatively correlated to nutrient variables (NO_3 and NO_2). In contrast to dimension 1, dimension 2 co-varied with numerous variables characterizing biological activity (e.g. oxygen, bacterial production, cell densities).

Variables significantly related to dimensions 1 and 2 were integrated in a Canonical Correspondence Analysis (CCA) performed from relative intensity of TTGE bands. CCA revealed that more than 50% of the variability of AAP communities was described by the 11 selected variables (Fig. 3). The four clusters (A to D) previously detected by HCA analysis were retrieved. Cluster D was mainly separated according to NO_3 and NO_2 variables, while clusters A and C were discriminated by pH, PAR, total organic carbon, biogenic activity variables and particulate nitrogen and phosphorus. None of the selected variables clearly explained pattern of similarity within cluster B.

Phylogenetic analyses of pufM genes

Ten out of the 29 samples were selected for phylogenetic analyses on the basis of their location and the diversity of AAP populations. Of a total of 388 clones analyzed, 44 distinct OTUs were identified after grouping the sequences at 94% nucleic acid sequence similarity (Fig. 4, Table S1). Coverage values (Table 1) and rarefaction curves (data not shown) indicated that most of the diversity was detected in most libraries (> 71%). Differences in AAP population diversity were not significant between samples collected above and below the DCM.

MDS for AAP populations from the 10 selected samples were performed from Dice similarity matrices for both TTGE and clone library analyses (Fig. S4). As also expressed by the significant correlation between Dice similarity matrices for TTGE and clone libraries

(R=0.39, p=0.05), both methods gave reliable information. The discrepancy observed for St.9-65m could be due to the low coverage of the clone library (Table 1).

About half of the *pufM* sequences were less than 94% identical to known sequences (Table S1). Most sequences were related to clones retrieved from the Mediterranean Sea (29%) and coastal environments (52%) including the Delaware and Monterey Bays. A unique sequence (PROSOPE-7) was affiliated with a clone from the Atlantic Ocean. The phylogenetic analysis demonstrated that the sequences were distributed into 7 of the 12 phylogroups previously defined by Yutin *et al*. (2007). None of the sequences was affiliated with the A, D, H, J and L groups (Fig. 4, Table S1) that are mainly found in anoxic environments (Yutin *et al*., 2007).

Only two sequences, belonging to PROSOPE-48 and PROSOPE-52 phylotypes, were highly similar to that of cultured representatives (*Methylobacterium radiotolerans* and *Erythrobacter longus*, respectively). Among the 7 AAP groups recovered from this study, only groups B and K were present in all samples (Fig. 5). PROSOPE-6 which shared 99% similarity with a clone from the Monterey Bay (Béja *et al*., 2002) was the main phylotype of group B. Representing 11% of the *pufM* sequences, PROSOPE-6 was prevalent (> 80%) at coastal and mesotrophic stations (Fig. 4, Table S1). Most of our sequences (~80%) fell into Group K which contained *Gammaproteobacteria* representatives including few isolates such as *Congregibacter litoralis* KT71 (Eilers *et al.,* 2001) and strains NOR5-3 and NOR51B and HTCC2080 (Cho *et al*., 2007) and BAC clones EBAC65D09 and EBAC29C02 (Béjà *et al*., 2002), all related to the OM60 clade (Rappé *et al*., 1997). Among the 30 OTUs affiliated to Group K, 10 (i.e. PROSOPE -10, -11, -38) and 9 (i.e. PROSOPE -12, -15, -45) phylotypes were only detected at the coastal Atlantic (St. UPW and St. 1) and at Mediterranean stations, respectively. Among gammaproteobacterial sequences, PROSOPE-34, which represented up to 20% of the total *pufM* sequences, dominated at meso- and eutrophic stations while other

phylotypes (i.e. PROSOPE-14, -42, -45) prevailed (>80%) at oligotrophic stations. The distribution of the other groups was sporadic. A few clones recovered at MIO and UPW stations clustered in the group C. The *Roseobacter*-like (group G) and *Rhodobacter*-like (groups E and F) clones were only distributed in the nutrient rich coastal waters of the Morocco upwelling (st. UPW) and the Strait of Gibraltar (UPW and St. 1) (Fig. 5). The phylotype PROSOPE-7 consisting of a few clones retrieved at station MIO grouped into the group I. The closest cultured relatives of this group are *Betaproteobacteria* widely distributed in freshwater and estuarine environments (Page *et al*., 2004; Yutin *et al*., 2007). This could suggest that closely related organisms also thrive in oceanic surface waters.

Discussion

The Mediterranean Sea is semi-enclosed allowing the study, on a reduced scale, of processes typical of the world's oceans. It constitutes a unique environment for ecological studies as it offers a range of trophic conditions, including extreme oligotrophy, particularly in summer when the water column is strongly stratified. The PROSOPE cruise was organized a few years before the discovery of the ecological significance of AAP bacteria in oceanic waters (Kolber *et al*., 2001). Therefore, the sampling strategy was not intended to allow the examination of their abundance and their contribution to the microbial community structure using infrared microscopy and fluorometry methods. One of the main objectives of the present study was to investigate environmental factors influencing biogeography of AAPs by analyzing the *pufM* gene diversity in archived bacterial community DNA samples. We agree that top-down regulation factors probably contribute to the distribution of AAP populations (Koblížek *et al*., 2007; Zhang & Jiao, 2009), however, this study provides one of the first look on bottom-up factors that contribute to longitudinal and vertical distribution and diversity of AAP populations.

Linking AAP population structure and diversity to environmental parameters

The diversity of AAP populations as evident in *pufM* sequences was relatively constant along both transects and was similar to that reported by Jiao *et al*. (2007) in oligotrophic oceanic provinces. We found a positive correlation between the PAR and diversity (H'/E_{var}) indices, suggesting that AAP populations are more diverse in the upper surface layer and that light affected their community structure. In contrast to Jiao *et al*. (2007), who reported that the AAP diversity decreased with increasing Chl*a* concentrations, we did not find a significant correlation between these variables. Our study found, however, an inverse relationship between AAP diversity and nitrate levels. This result together with the highest richness observed for the highly oligotrophic station MIO also supports the hypothesis that inorganic nitrogen concentrations may affect their distribution (Waidner & Kirchman, 2007) and that AAP diversity decreases with increasing trophic status (Jiao *et al*., 2007). This suggests that the scarcity of inorganic nutrients and dissolved organic matter in oligotrophic environments leads to a higher specialization of AAP assemblages and increases the functional redundancy (Curtis & Sloan, 2004) for this group, as it has been observed elsewhere for other bacterial populations (Wohl *et al*., 2004). This hypothesis is consistent with the contrasting variation of abundance and diversity of AAP assemblages previously suggested by a large-scale survey analysis of their distribution patterns (Jiao *et al*., 2007).

Variation in the composition of AAP assemblages and environmental context

The HCA and MDS analyses suggest that AAP populations follow nutrient status and that oceanic Mediterranean populations differed from those of the Atlantic coast. This clustering deduced from TTGE analyses was consistent with the clone library results with

some phylotypes retrieved only at coastal stations while others dominated in oligotrophic waters.

Our analyses suggest that stratification is a critical factor determining the vertical distribution of AAP species in the Mediterranean Sea. Similar vertical stratification has been reported for the whole bacterial community structure in the Mediterranean Sea (Acinas *et al.*, 1997; Ghiglione *et al.*, 2008) and in other oligotrophic waters, including the Pacific and Atlantic oceans, the Caribbean Sea (Lee & Fuhrman, 1991), and Antarctic environments (Murray *et al.*, 1998). Multivariate analyses of environmental parameters and molecular fingerprinting profiles revealed that the variation in AAP populations with depth is due to synergetic driving forces similar to those involved in the vertical stratification of bacterial communities at the DYF station (Ghiglione *et al.*, 2008). Fifty percent of the variability of AAP composition was explained by the selected environmental parameters. This suggests that the composition of AAP assemblages may be additionally influenced by other factors such as specific and multiple interactions with other organisms of their environment such as their bacterial counterparts, protists, viruses, and metazoans (Fuhrman, 2009) as well as intrinsic photoheterotrophic capabilities and physiological peculiarities of each species. Indeed, AAP bacteria have diverse metabolisms, ranging from generalists (e.g. *Erythrobacter*, *Roseobacter*) to specialized species (e.g. *Citromicrobium*) able to grow on a broad and a narrow spectrum of carbon sources (Yurkov and Csotonyi 2009). Further in-depth studies on the physiology and metabolism of environmentally significant organisms and on their biological interactions with other planktonic counterparts are essential to better understand the ecology of marine AAP.

Phylogenetic composition of populations

Most *pufM* sequences retrieved from our samples belonged to clades previously identified in a global metagenomic survey (Yutin *et al.*, 2007). As previously observed in

environmental surveys of these genes (e.g. Béjà *et al*., 2002), most of the OTUs retrieved in this study were distantly related to known anoxygenic phototrophs. More than 80% of our *pufM* sequences had their best matches with clone sequences obtained from the Mediterranean Sea and from coastal areas suggesting that AAP populations in this sea are different from those in open oceans. We acknowledge, however, that this observation may be biased by the low number of offshore *pufM* sequences available in the databases.

In contrast to previous PCR-based studies (Oz *et al*., 2005; Yutin *et al*., 2005), *Alphaproteobacteria* constitute a minor fraction of the Mediterranean AAP community. Members of the *Roseobacter*-related group (group G), often a key player in diverse marine systems (Buchan *et al*., 2005; Yutin *et al*., 2007), were only detected at Atlantic coastal stations. While this finding supports the notion that *Roseobacter*-like bacteria are favored by nutrient-rich conditions, in agreement with the common association of *Roseobacter* with phytoplankton blooms (González *et al*., 2000; Suzuki *et al*., 2001), it seems to contradict their suspected important role and dominance in the Mediterranean Sea (Oz *et al*., 2005). However, since *Roseobacter* species can be more abundant in winter than in summer, this discrepancy with previous studies (Oz *et al*., 2005; Yutin *et al*., 2005) could be explained by the different sampling period.

Previous studies have revealed a high contribution of *Gammaproteobacteria* to AAP populations (Béjà *et al*., 2002; Hu *et al*., 2006; Yutin *et al*., 2007). However, the large proportions observed in this study (~ 80%) have never been reported from any environments. Possible PCR biases cannot be definitively ruled out to explain the large dominance of *Gammaproteobacteria* in our samples. However, by using the same primer combination to study the diversity of AAP in arctic waters and in coastal systems of the English Channel, we did not find such high percentages of gammaproteobacterial sequences (Lehours *et al*., Dahan *et al*., unpublished results). Consistent with previous observations, gammaproteobacterial

sequences were retrieved in eutrophic and oligotrophic waters (Hu *et al.*, 2006; Jiao *et al.* 2007). Athough some gammaproteobacterial phylotypes were obtained from both nutrient levels, the distribution of most suggests an adaptation to specific trophic conditions. For example, some of our cloned *pufM* sequences from the Atlantic coast were similar to the sequences belonging to the OM60 clade abundant in coastal oceans (Béjà *et al*., 2002; Yutin *et al*., 2007). Although prevalent at coastal stations, phylogroup B also showed a ubiquitous distribution across the different trophic regimes (Yutin *et al*., 2007). Onboard enrichment experiments performed along the PROSOPE transect demonstrated a switch from a phosphorus limitation in the surface layer to organic carbon limitation in the deep chlorophyll maximum (Van Wambeke *et al*., 2002). Specific adaptation capabilities to extreme oligotrophy are likely to explain the success of phylogroups B and K in typical Mediterranean waters. In line with this hypothesis, genes coding for several storage compounds were identified in the genome of the gammaproteobacterium *Congregibacter litoralis* (Fuchs *et al*., 2007). Among them, the storage compounds cyanophycin and polyphosphate probably reflect an adaptation to fluctuating carbon and phosphorus availability.

Conclusion

The molecular analyses of AAP diversity conclusively demonstrated that typical Mediterranean populations varied greatly along depth profiles, with strong and opposite gradients of light and nutrient availability. This variation with depth was much greater than seen across both transects. During stratification, the vertical distribution of AAP bacteria seems to be governed by the same environmental variables as that of the whole bacterial community (Ghiglione *et al*., 2008). This result, however, does not imply that they may act as strict heterotophs *in situ* and that phototrophy has a minor impact on their distribution and lifestyle. Since evidence from other studies indicate that AAP use their phototrophic

capability at nutrient-poor levels (Cho *et al*., 2007; Spring *et al*., 2009), additional variables may influence their populations in the extreme oligotrophic conditions prevailing in the Mediterranean Sea.

This study was the first to reveal such a high abundance of gammaproteobacterial AAP in the environment. An ecotypic differentiation was suggested by both TTGE and cloning approaches. Further culture efforts are therefore needed to expand the diversity of gammaproteobacterial isolates and to delineate the environmental parameters that govern the activity and distribution of gammaproteobacterial trophic ecotypes.

Acknowledgements

We wish to thank H. Claustre for coordinating the PROSOPE cruise and L. Garczarek, D. Marie and F. Partensky for sampling. We are grateful to M. Perennou (Biogenouest Sequencing Platform at the Station Biologique) for help with sequencing. We also thank Milton Da Costa and I. Wagner-Döbler for the gift of the cultures of *Erythrobacter longus, E. litoralis,* and *Dinoroseobacter shibae*, respectively. This work was supported by the programs PROOF and LEFE-CYBER (CNRS-INSU). A-C. Lehours benefited from a post-doctoral fellowship from the CNRS. Kirchman and Cottrell were supported by NSF MCB-0453993.

References

Achenbach LA, Carey J & Madigan MT (2001) Photosynthetic and phylogenetic primers for detection of anoxygenic phototrophs in natural environments. *Appl Environ Microbiol* **67:**2922–2926.

Acinas SG, Rodríguez-Valera F & Pedrós-Alió C (1997) Spatial and temporal variation in marine bacterioplankton diversity as shown by RFLP fingerprinting of PCR amplified 16S rDNA. *FEMS Microbiol Ecol* **24**:27-40.

Béjà O, Suzuki MT, Heidelberg JF, Nelson WC, Preston CM, Hamada T, Eisen JA, Fraser CM & DeLong EF (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415:** 630–633.

Berman T, Walline PD, Schneller A, Rothenberg J & Towsend DW (1985) Secchi disk depth record: a claim for the eastern Mediterranean. *Limnol Oceanogr* **30**:447-448.

Buchan A, González JM & Moran AM (2005) Overview of the marine Roseobacter lineage. *Appl Environ Microbiol* **71**:5665-5677.

Chazdon RL, Colwell RK, Denslow JS & Guariguata MR (1998) Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of NE Costa Rica. Pp. 285-309 in F. Dallmeier and J. A. Comiskey, eds. *Forest biodiversity research, monitoring and modeling: Conceptual background and Old World case studies.* Parthenon Publishing, Paris.

Cho JC, Stapels MD, Morris RM, Vergin KL, Schwalbach MS, Givan SA, Barofsky DF & Giovannoni SJ (2007) Polyphyletic photosynthetic reaction centre genes in oligotrophic marine Gammaproteobacteria. *Environ Microbiol* **9**:1456–1463.

Clarke KR & Warwick RW (2001) Change in marine communities: an approach to statistical analysis and interpretation, 2nd edition, Primer-E, Plymouth.

Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM & Tiedje JM (2003) The ribosomal database project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acid Res* **31**:442-443.

Cottrell MT, Mannino A & Kirchman D (2006) Aerobic anoxygenic phototrophic bacteria in the mid-Atlantic bight and the north Pacific gyre. *Appl Environ Microbiol* **72**:557-564.

Cottrell MT, Michelou VK, Nemcek N, Ditullio G & Kirchman DL (2008) Carbon cycling by microbes influenced by light in the Northeast Atlantic Ocean. Aquat *Microb Ecol* **50**:239-250.

Curtis TP & Sloan WT (2002) Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. *Curr Op Microbiol* **7**:221-226.

Eiler A (2006) Evidence for the ubiquity of mixotrophic bacteria in the upper Ocean: implications and consequences. *Appl Environ Microbiol* **72**:7431-7437.

Eilers H, Pernthaler J, Peplies J, Glöckner FO, Gerdts G & Amann R (2001) Isolation of novel pelagic bacteria from the german bight and their seasonal contributions to surface picoplankton. *Appl Environ Microbiol* **67**:5134-5142.

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Fuchs BM, Spring S, Teeling H, Quast C, Wulf J, Schattenhofer M, Yan S, Ferriera S, Johnson JM, Glöckner FO & Amann R (2007) Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis. *Proc Natl Acad Sci USA* **104:**2891–2896.

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

Garczarek L, Dufresne A, Rousvoal S, West NJ, Mazard S, Marie D, Claustre H, Raimbault P, Post AF, Scanlan D & Partensky F (2007) High vertical and low horizontal diversity of *Prochlorococcus* ecotypes in the Mediterranean Sea in summer. *FEMS Microbiol Ecol* **60**:189-206.

Gascard JC & Richez C (1985) Water masses and circulation in the Western Alboran Sea and in the Straits of Gibraltar. *Prog Oceanog* **15:** 175–216.

Ghiglione JF, Palacios C, Marty JC, Mével G, Labrune C, Conan P, Pujo-Pay M, Garcia N & Goutx M (2008) Role of environmental factors for the vertical distribution (0-1000 m) of marine bacterial communities in the NW Mediterranean Sea. *Biogeosciences Discuss* **5**:2131-2164.

González JM, Simó R, Massana R, Covert JS, Casamayor EO, Pedrós-Alió C & Moran MA (2000) Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. *Appl Environ Microbiol* **66**:4237–4246.

Hammer Ø, Harper DAT & Ryan PD (2001) PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4: 9pp.

Hill TCJ, Walsh KA, Harris JA & Moffett BF (2003) Using ecological diversity measures with bacterial communities. *FEMS Microbiol Ecol* **43**:1-11.

Hu Y, Du H, Jiao N & Zeng Y (2006) Abundant presence of the gamma-like proteobacterial *pufM* gene in oxic seawater. *FEMS Microbiol Lett* **263:** 200–206.

Hughes JB, Hellmann JJ, Ricketts TH & Bohannan BJM (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* **67**:4399-4406.

Imhoff JF (2001) True marine and halophilic anoxygenic phototrophic bacteria. *Arch Microbiol* **176**:243-254.

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

Koblížek M, Mašín M, Ras J, Poulton AJ & Prášil O (2007) Rapid growth rates of aerobic anoxygenic prototrophs in the Ocean. *Envir Microbiol* **9:** 2401–2406.

Kolber ZS, Plumley FG, Lang AS, Beatty JT, Blankenship RE, VanDover CL, Vetriani C, Koblížek M, Rathgeber C & Falkowski PG (2001) Contribution of aerobic photoheterotrophic Bacteria to the carbon cycle in the ocean. *Science* **292**:2492-2495.

Lami R, Cottrell MT, Ras J, Ulloa O, Obernosterer I, Claustre H, Kirchman DL & Lebaron P (2007) High abundances of aerobic anoxygenic photosynthetic bacteria in the south Pacific Ocean. *Appl Environ Microbiol* **73**:4198-4205.

Lee S & Furhman JA (1991) Spatial and temporal variation of natural bacterioplankton assemblages studied by total genomic DNA cross-hybridization. *Limnol Oceanogr* **36**:1277-1287.

Legendre P & Legendre L (1998) Numerical Ecology, 2nd English ed. Elsevier, 853pp.

Letunic I & Bork P (2006) Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**:127-128.

Lucea A, Duarte CM, Agustí S & Søndergaard M (2003) Nutrient (N, P and Si) and carbon partitioning in the stratified NW Mediterranean. *J Sea Res* **49**:157-170.

Ludwig W, Strunk O, Westram R, Richter L, Meier H *et al*. (2004) ARB: a software environment for sequence data. *Nucleic Acid Res* **32**:1363-1371.

Manca B, Burca M, Giorgetti C, Coatanoan C, Garcia MJ & Iona A (2004) Physical and biochemical averaged vertical profiles in the Mediterranean regions: an important tool to trace the climatology of water masses and to validate incoming data from operational oceanography. *J Mar Syst* **48:** 83–116.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**: 209–220.

Marie D, Zhu F, Balagué V, Ras J & Vaulot D (2006) Eukaryotic picoplankton communities of the Mediterranean Sea in summer assessed by molecular approaches (DGGE, TTGE, QPCR). *FEMS Microbiol Ecol* **55:** 403–415.

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

Moutin T & Raimbault P (2002) Primary production, carbon export and nutrients availability in western and eastern Mediterranean Sea in early summer 1996 (MINOS cruise). *J Mar Syst* **33-34**:273-288.

Mullins TD, Britschgi TB, Krest RL & Giovannoni SJ (1995) Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. *Limnol Oceanogr* **40**:148-158.

Murray AE, Preston CM, Massana R, Taylor LT, Blakis A, Wu K & DeLong EF (1998) Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters off Anvers Island, Antarctica. *Appl Environ Microbiol* **64**:2585–2595.

Oz A, Sabehi G, Koblížek M, Massana R & Béjà O (2005) *Roseobacter* like bacteria in Red and Mediterranean Sea aerobic anoxygenic photosynthetic populations. *Appl Environ Microbiol* **71**:344–353.

Page KA, Connon SA & Giovannoni SJ (2004) Representative freshwater bacterioplankton isolated from Crater Lake, Oregon. *Appl Environ Microbiol* **70**: 6542-6550.

Ramírez-Flandes S & Ulloa O (2008) Bosque: Integrated phylogenetic analysis software. *Bioinformatics* **24**:2539-2541.

Rappé MS, Kemp PF & Giovannoni SJ (1997) Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. *Limnol Oceanogr* **42**:811-826.

Shapiro SS & Wilk MB (1965) An analysis of variance test for normality (complete samples). *Biometrika* **52**: 591–611.

Sieracki ME, Gilg IC, Thier EC, Poulton NJ & Goericke R (2006) Distribution of planktonic aerobic anoxygenic photoheterotrophic bacteria in the northwest Atlantic. *Limnol Oceanogr* **51:** 38–46.

Smith B & Wilson JB (1996) A consumers's guide to evenness indices. *Oïkos* **76**:70-82.

Spring S, Lünsdorf H, Fuchs BM & Tindall BJ (2009) The Photosynthetic Apparatus and Its Regulation in the Aerobic Gammaproteobacterium *Congregibacter litoralis* gen. nov., sp. nov. *Plos One* **4**: 1-23.

Stephens MA (1974) EDF Statistics for Goodness of Fit and Some Comparisons. *J Amer Stat Assoc* **69**: 730–737.

Suyama T, Shigematsu T, Suzuki T, Tokiwa Y, Kanagawa T, Nagashima KVP & Hanada S (2002) Photosynthetic apparatus in *Roseateles depolymerans* 61A is transcriptionally induced by carbon limitation. *Appl Environ Microbiol* **68**:1665-1673.

Suzuki MT, Preston CM, Chavez FP & DeLong EF (2001) Quantitative mapping of bacterioplankton populations in seawater: field tests across an upwelling plume in Monterey Bay. *Aquat Microbiol Ecol* **24:**117–127.

Van Wambeke F, Christaki U, Giannakourou A, Moutin T & Souvemerzoglou K (2002) Longitudinal and vertical trends of bacterial limitation by phosphorus and carbon in the Mediterranean Sea. *Microb Ecol* **43**: 119–133.

Waidner LA & Kirchman DL (2007) Aerobic anoxygenic phototrophic bacteria attached to particles in turbid waters of the Delaware and Chesapeake estuaries. *Appl Environ Microbiol* **73:** 3936–3944.

Waidner LA & Kirchman DL (2008) Diversity and distribution of ecotypes of the aerobic anoxygenic phototrophy gene *pufM* in the Delaware estuary. *Appl Environ Microbiol* **74**:4012-4021.

Ward JH (1963) Hierarchical Grouping to optimize an objective function. *J Amer Stat Assoc* **58**: 236-244.

Wohl DL, Arora S & Gladstone JR (2004) Functional redundancy supports biodiversity and ecosystem function in a closed and constant environment. *Ecology* **85**:1534-1540.

Yurkov V & Csotonyi JT (2009) New light on aerobic anoxygenic phototrophs. *Adv Photosynth Res* **28**: 31–55.

Yurkov V & van Gemerden H (1993) Impact of light/dark regime on growth rate, biomass formation and bacteriochlorophyll synthesis in *Erythromicrobium hydrolyticum*. *Arch Microbiol* **159**:84-89.

Yutin N, Béjà O & Suzuki M (2008) The use of denaturing gradient gel electrophoresis with fully degenerate pufM primers to monitor aerobic anoxygenic phototrophic assemblages *Limnol Oceanogr Methods* **6:**427–440.

Yutin N, Suzuki MT & Béjà O (2005) Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. *Appl Environ Microbiol* **71**:8958–8962.

Yutin N, Suzuki MT, Teeling H, Weber M, Venter JC, Rusch DB & Béjà O (2007) Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. *Environ Microbiol* **9:** 1464–1475.

Zeng YH, Chen XH & Jiao NZ (2007) Genetic diversity assessment of anoxygenic photosynthetic bacteria by distance-based grouping analysis of pufM sequences. *Lett Appl Microbiol* **45**:639-645.

Zhang Y & Jiao N (2009) Roseophage RDJL□1, infecting the aerobic anoxygenic phototrophic bacterium *Roseobacter denitrificans* OCh114. *Appl Environ Microbiol* **75**: 1745-1749.

Table 1. Properties of the distribution of phylotypes in clone libraries. H': Shannon-Weiner index, ACE: Abundance-base Coverage Estimator. The richness estimator SChao1 was computed along with log-linear 95% confidence intervals.

Station-Depth	*Number of distinct OTUs*	*ACE*	*Schao1*	*H'*	*Coverage (%)*
UPW-5m	18	33	25 (20-50)	2.45	78
St1-5m	16	28	21 (17-45)	2.5	88
St1-30m	14	17	14 (13-23)	2.2	87
St1-80m	8	17	10 (7-30)	1.3	83
MIO-5m	13	19	19 (14-51)	2.4	90
MIO-50m	12	13	13 (12-22)	2	71
MIO-90m	18	28	22 (19-40)	2.6	86
St9-65m	13	35	35 (18-110)	2	61
DYF-15m	10	14	11 (10-21)	1.9	93
DYF-50m	7	7.5	7 (6-21)	1.4	82

Figure Legends

Figure 1. Track of the PROSOPE cruise superimposed on the composite SeaWiFS image of surface chlorophyll *a* concentrations in September 1999. Arrows indicated the stations analyzed in this study. ST#, station number.

Figure 2. Temporal temperature gradient gel electrophoresis (TTGE) analyses of *pufM* gene fragments amplified from DNA samples from the Mediterranean Sea. Hierarchical cluster analysis was performed as indicated in the Experimental procedures section. The dashed line indicates the distance chosen for cluster separation. The number of distinct TTGE bands per sample is indicated between brackets. DCM: deep Chlorophyll maximum.

Figure 3. Canonical correspondence analysis (CCA) performed using relative intensity peak of the TTGE bands. Samples and environmental parameters are coded as in Fig. S2.

Figure 4. *pufM* phylogenetic tree showing inferred phylogenetic relationships of *pufM* gene sequences cloned from the Mediterranean samples. The tree is based on a Neighbor-Joining (NJ) tree to which short sequences were added by ARB_PARISMONY. Sequences used to perform NJ tree are marked in bold and grey circles on nodes represent confidence values > 50% for branches found in the initial NJ tree. Color ranges highlight the different groups defined by Yutin *et al.* (2007). Sequences retrieved in this study are indicated by filled triangles. The multi-value bar charts represent the relative frequencies of the corresponding OTU in the different clone libraries. Colors used to represent the clone libraries are indicated at the top of the figure. Green and blue crosses indicate that the corresponding OTU was found to be dominant (>80%) at meso-eutrophic and oligotrophic conditions, respectively.

Figure 5. Distribution of the AAP phylogroups along the PROSOPE transects based on their relative proportion in the clone libraries.

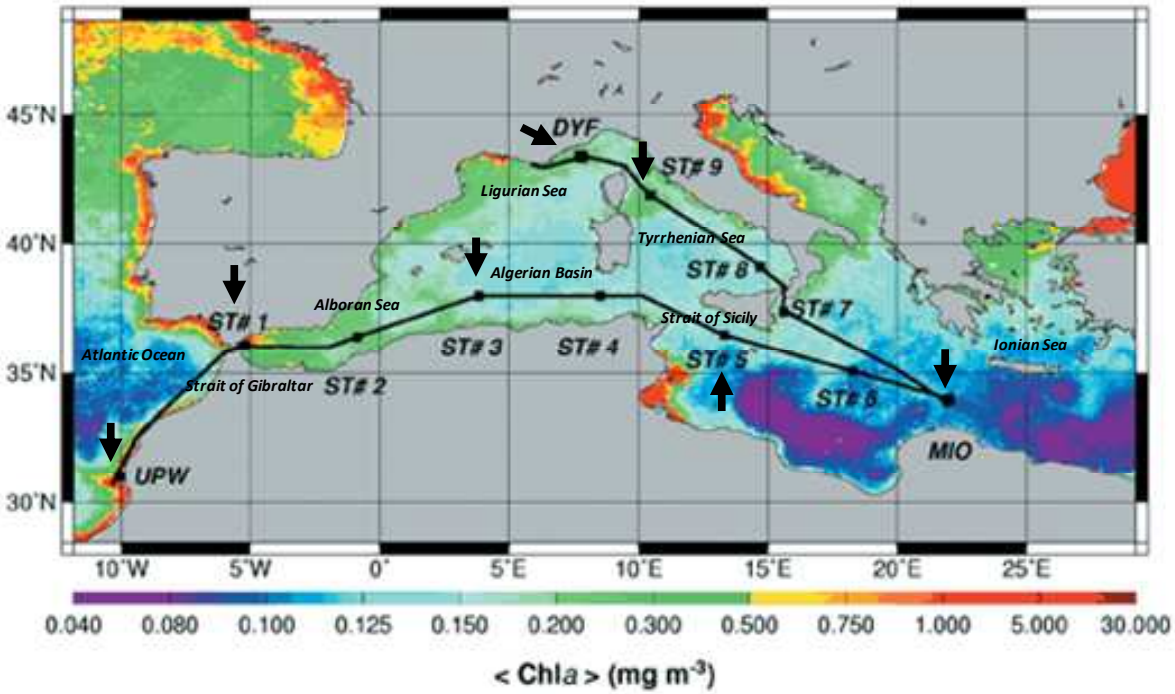
45°N
40°N
35°N
30°N
10°W
5°W
0°
5°E
10°E
15°E
20°E
25°E
DYF
ST# 9
Ligurian Sea
Tyrrhenian Sea
Algerian Basin
ST# 8
ST# 7
Alboran Sea
ST# 1
Strait of Sicily
Atlantic Ocean
ST# 3
ST# 4
ST# 5
Ionian Sea
Strait of Gibraltar
ST# 2
ST# 6
MIO
UPW
0.040
0.080
0.100
0.125
0.150
0.200
0.300
0.500
0.750
1.000
5.000
30.000
< Chla > (mg m-3)

Figure 1

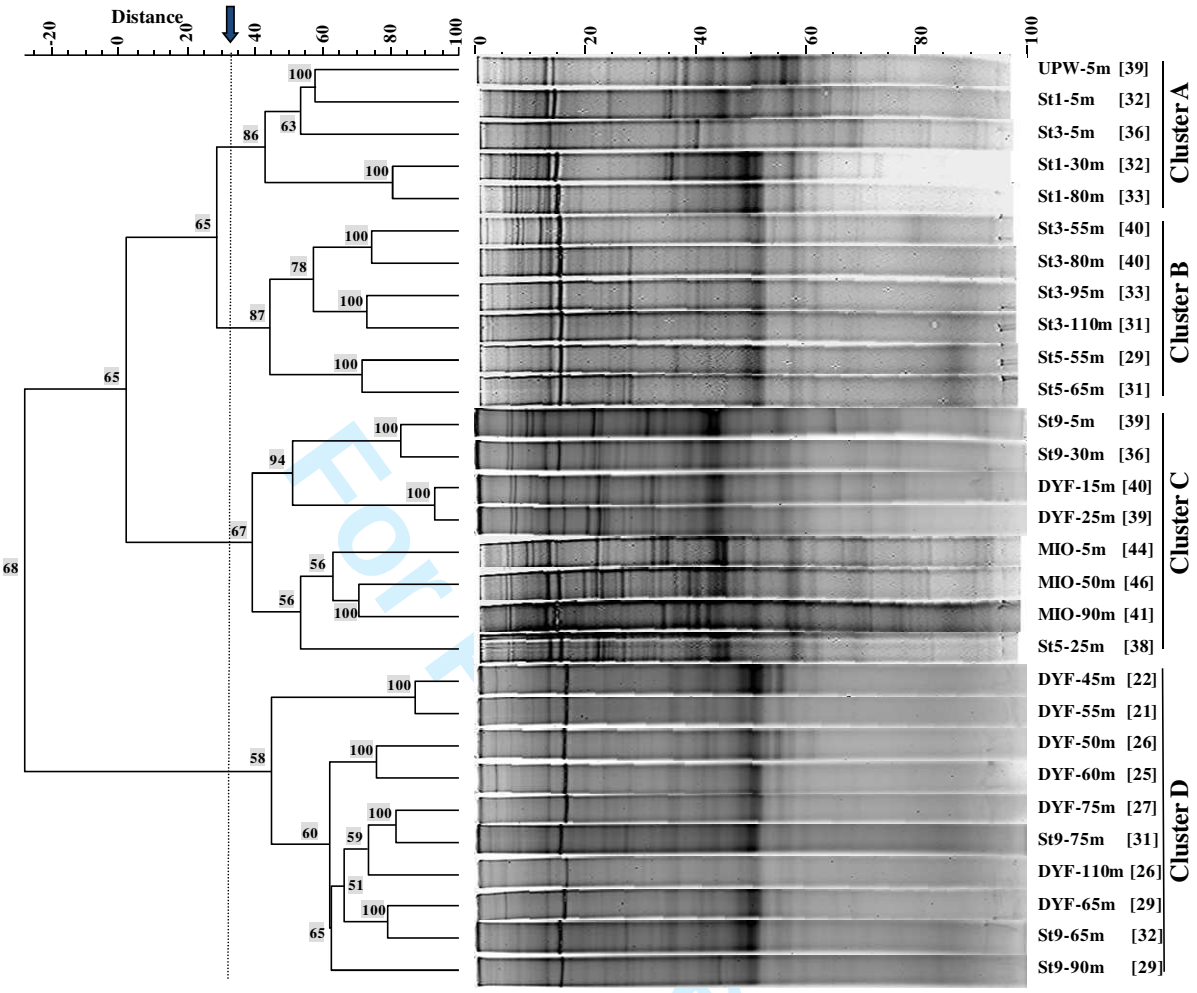

Figure 2

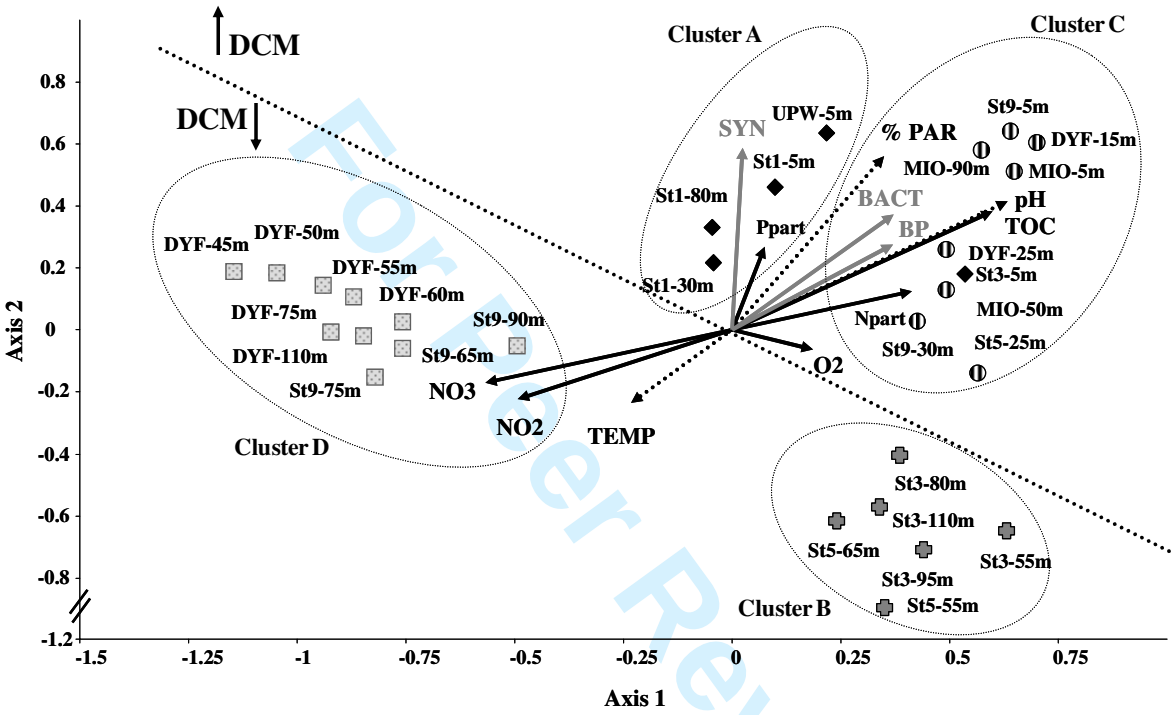

Figure 3

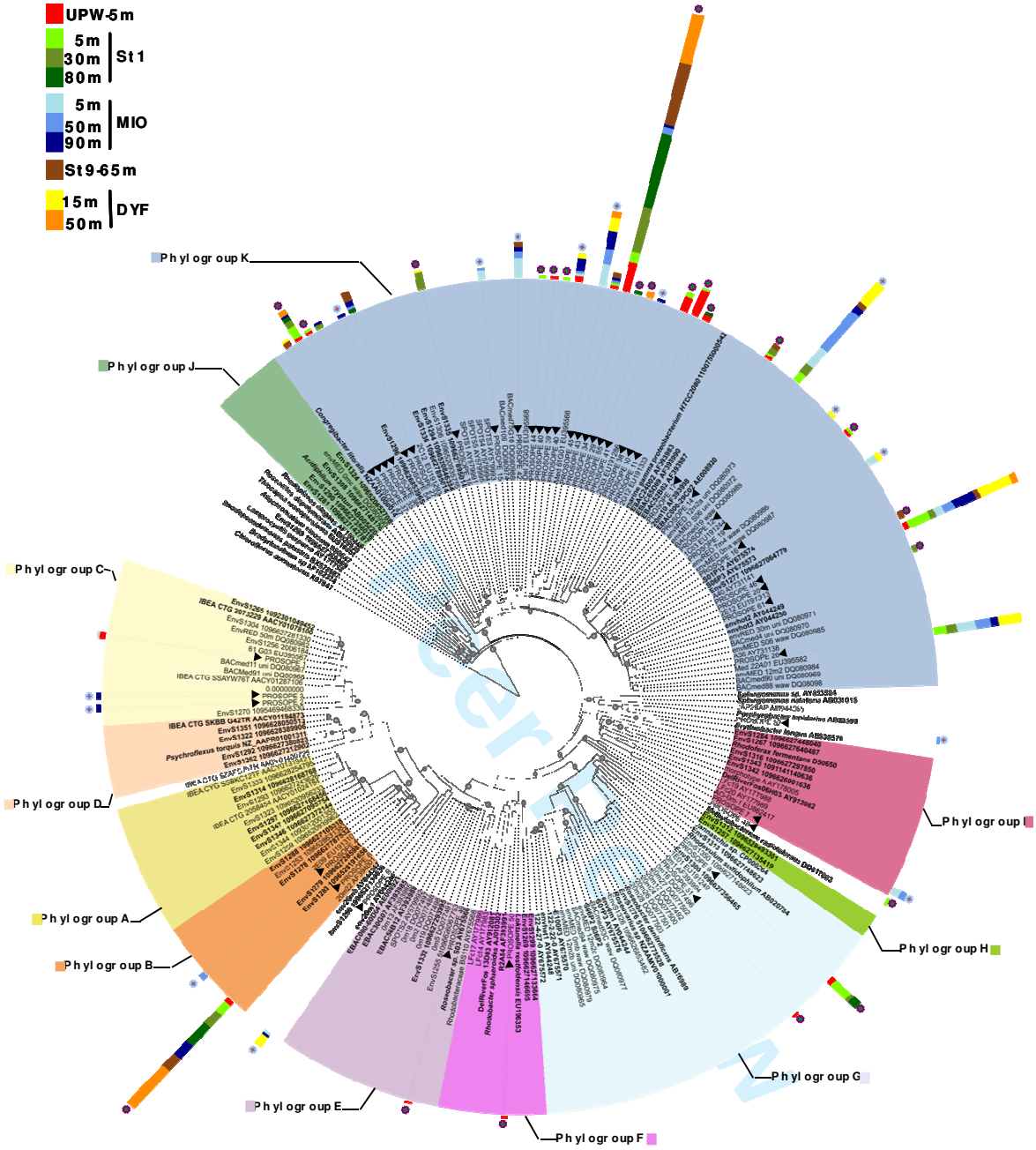
UPW-5m
5m
30m
80m
St 1
5m
50m
90m
MIO
St 9-65m
15m
50m
DYF
Phylogroup K
Phylogroup J
Phylogroup C
Phylogroup D
Phylogroup A
Phylogroup B
Phylogroup E
Phylogroup F
Phylogroup G
Phylogroup H
Phylogroup I

Figure 4

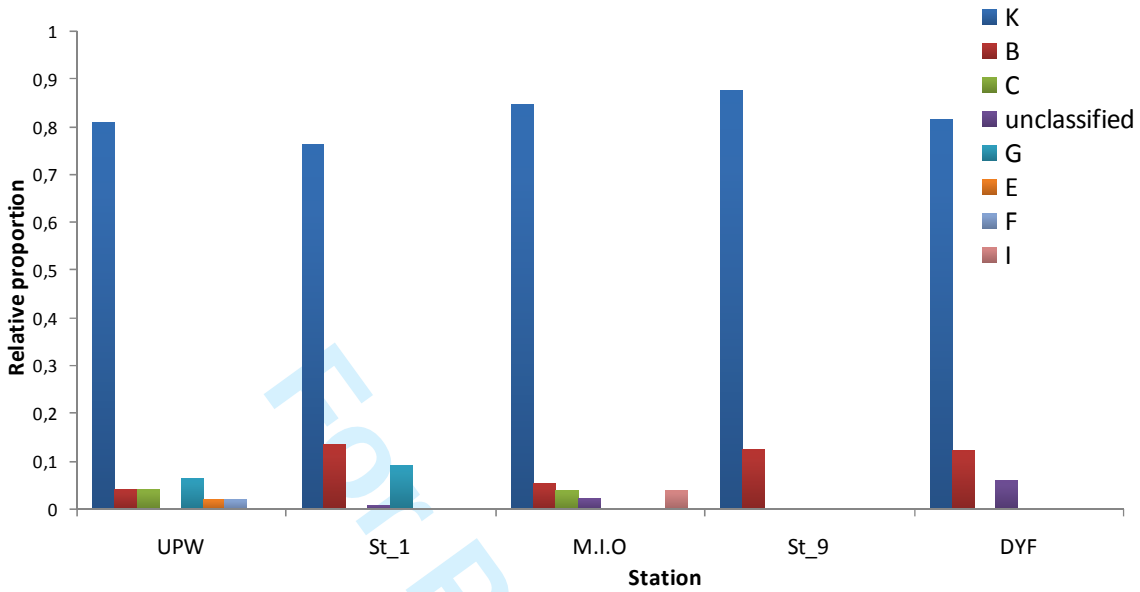

Figure 5

Supplementary material

Table S1. Representative sequence for each OTU, their closest relative by BLAST and percentage identity. Med, Mediterranean Sea; DelBay, Delaware Bay; MontBay, Monterey Bay; AO, Atlantic Ocean; Sarg, Sargasso Sea; Mar. Env, Marine environment.

Figure S1. Hydrological conditions along both transects of the PROSOPE cruise. Both transects (W-E from St. UPW to St. MIO and S-N from St. MIO to St. DYF) were plotted on each graph. (*a*) temperature, (*b*) salinity, (*c*) chlorophyll *a*, (*d*) nitrate, (*e*) phosphate, (*f*) nitrite (*g*) PAR, , (*h*) total organic carbon, (*i*) particulate carbon, (*j*) particulate nitrogen, (*k*) particulate phosphorus, and (*l*) oxygen. Stations and depths (in m) are indicated on the horizontal and vertical axes respectively.

Figure S2. Principal component analysis (PCA) of environmental parameters. The percentage of variability explained by each axis is indicated. Samples (squares) and variables (arrows) are plotted against the first two axes. Samples are labeled as follows: Station-Depth. Stations U, D and M correspond to UPW, DYF and MIO respectively. Black arrows represent variables characterizing the "physical environment": TEMP (temperature, °C), pH, PAR (percentage of the photosynthetically available radiation), O_2 (concentration in dissolved oxygen, $ml.l^{-1}$), SAL (salinity, ‰), Depth (m). Grey arrows represent variables characterizing the "biological activity": PROC (*Prochlorococcus* cell concentration, $cell.ml^{-1}$), SYN (*Synecococcus* cell concentration, $cell.ml^{-1}$), BACT (bacteria cell concentration, $cell.ml^{-1}$), EUK (eukaryotes cell concentration, $cell.ml^{-1}$), BP (bacterial production, ng $C.l^{-1}.h^{-1}$), Chl*a* (chlorophyll *a*, µg. l^{-1}), PIGM (accessory pigments, $\mu g.l^{-1}$). Black dashed arrows represent variables characterizing the "trophic status":TOC (total organic carbon, µM), NO_3 (nitrate,

µM), PO_4 (phosphate, µM), NO_2 (nitrite, µM), N part (particulate N, µM), P part (particulate P, µM). Samples collected above and below the DCM are indicated by grey and black squares respectively. Component 1, 2 and 3 (component 3, not shown) represent 34%, 23% and 12% of the total variance, respectively.

Figure S3. Profiles of chlorophyll *a* (Chl*a*) (solid) and PAR (dashed) with depth. Arrows in the Chl*a* profiles indicate sample depths. Dashed line indicates the limit of the euphotic layer ($Z_{1\%}$).

Figure S4. Multidimensional scaling plot (MDS) of AAP populations based on TTGE and cloning-sequencing of *pufM* genes. This plot corresponds to a two-dimensional visualization of the Dice distance matrix. "(c)" refers to data obtained from clone libraries analyses.

Representative clone	Accession No.	No. of clones in libraries										Closest relative			
		UPW-5m	St1-5m	St1-30m	St1-80m	MIO-5m	MIO-50m	MIO-90m	St9-65m	DYF-15m	DYF-50m	Taxon/Clone	Identity (%)	Accession . No.	Environmental source
PROSOPE 1	GQ468943	2	-	-	-	1	-	-	-	-	-	61G03	99	EU395544	Med
PROSOPE 2	GQ468944	-	-	-	-	-	-	2	-	-	-	46F01	84	EU395550	Med
PROSOPE 3	GQ468945	-	-	-	-	-	-	2	-	-	-	envSargasso-waw	81	DQ080981	Sarg
PROSOPE 4	GQ468946	-	-	-	-	-	2	-	-	-	-	2E09	97	EU191310	DelBay
PROSOPE 6	GQ468947	2	5	5	6	-	-	7	3	-	8	20m22	99	AF394001	MontBay
PROSOPE 7	GQ468948	-	-	-	-	3	2	-	-	-	-	AO-0m-1	98	EU862417	AO
PROSOPE 8	GQ468949	-	-	-	-	1	-	1	-	4	-	1E01	80	EU191249	DelBay
PROSOPE 9	GQ468950	2	8	1	2	-	-	-	-	-	-	1E01	89	EU191249	DelBay
PROSOPE 10	GQ468951	10	1	-	-	-	-	-	-	-	-	JL-XM-C21	81	AY652816	Mar. env
PROSOPE 11	GQ468952	2	-	-	1	-	-	-	-	-	-	2G12	94	EU191333	DelBay
PROSOPE 12	GQ468953	-	-	-	-	-	-	-	-	1	-	Bacmed19_waw	93	DQ080989	Med
PROSOPE 14	GQ468954	-	4	3	-	7	16	3	1	9	-	Bacmed19_waw	98	DQ080989	Med
PROSOPE 15	GQ468955	-	-	-	-	3	1	-	-	-	-	22G04	100	EU395569	Med
PROSOPE 16	GQ468956	-	-	-	-	3	-	-	-	2	-	22G06	98	EU395571	Med
PROSOPE 17	GQ468957	1	1	2	-	-	-	-	1	-	-	EBAC000-29C02	96	AE008920	MontBay
PROSOPE 18	GQ468958	-	-	5	-	-	-	-	-	1	-	2G06	98	EU191328	DelBay
PROSOPE 19	GQ468959	1	1	-	-	-	-	-	-	-	-	1C03	95	EU191244	DelBay
PROSOPE 20	GQ468960	2	9	2	-	3	4	9	1	12	1	2E12	88	EU191313	DelBay
PROSOPE 26	GQ468961	-	5	3	-	6	5	3	-	8	-	Bacmed88_waw	98	DQ80983	Med
PROSOPE 27	GQ468962	-	-	-	1	-	1	2	2	-	-	2B11	99	EU191282	DelBay
PROSOPE 29	GQ468963	-	-	-	1	-	-	-	-	-	-	22A01	79	EU395582	Med
PROSOPE 34	GQ468964	11	4	14	18	-	2	1	12	-	8	2A08	88	EU191269	DelBay
PROSOPE 35	GQ468965	1	1	-	-	-	-	-	-	-	-	2G01	84	EU191324	DelBay
PROSOPE 36	GQ468966	-	-	1	-	-	-	1	-	-	-	2C12	84	EU191293	DelBay
PROSOPE 37	GQ468967	1	5	2	1	-	-	1	-	-	1	1E01	78	EU191249	DelBay
PROSOPE 38	GQ468968	5	2	-	-	-	-	-	-	-	-	2A08	89	EU191269	DelBay
PROSOPE 39	GQ468969	-	1	-	-	-	-	-	-	-	-	61G01	93	EU395566	Med
PROSOPE 40	GQ468970	2	-	-	-	1	1	5	-	2	-	61G01	96	EU395566	Med
PROSOPE 41	GQ468971	-	-	-	-	-	-	1	-	-	-	2A08	88	EU191269	DelBay
PROSOPE 42	GQ468972	-	-	-	-	6	2	2	1	-	-	22G03	98	EU395568	Med
PROSOPE 43	GQ468973	1	1	-	-	-	1	1	1	-		2A08	95	EU191269	DelBay
PROSOPE 44	GQ468974	-	1	-	-	-	-	-	-	-	-	2A08	93	EU191269	DelBay
PROSOPE 45	GQ468975	-	-	-	-	7	5	7	-	5	1	2A08	93	EU191269	DelBay
PROSOPE 46	GQ468976	-	-	-	-	-	-	-	1	-	-	JL-XM-C21	97	AY652816	Mar. env
PROSOPE 47	GQ468977	-	-	-	-	-	-	-	1	1	-	2A08	88	EU191269	DelBay
PROSOPE 48	GQ468978	-	1	-	-	-	-	-	-	-	-	*Methylobacterium radiotolerans*	95	CP001001	
PROSOPE 49	GQ468979	1	-	-	-	-	-	-	-	-	-	1E01	87	EU191249	DelBay
PROSOPE 50	GQ468980	1	-	-	-	-	-	-	-	-	-	Berpuf8	94	AM162694	
PROSOPE 51	GQ468981	1	-	-	-	-	-	-	-	-	-	0m1	98	AF393994	Mar. Env
PROSOPE 52	GQ468982	-	-	-	-	-	1	-	-	-	-	*Erythrobacter longus*	97	D50648	
PROSOPE 54	GQ468983	-	-	-	-	-	-	1	-	-	-	2G03	86	EU191325	DelBay
PROSOPE 55	GQ468984	-	-	-	-	-	-	-	-	-	-	1H02	86	EU191260	DelBay
PROSOPE 60	GQ468985	1	-	-	-	-	-	-	-	-	-	61G01	97	EU395566	Med
PROSOPE 61	GQ468986	-	-	1	-	-	-	-	-	-	-	22A01	91	EU395582	Med
TOTAL		47	50	39	30	41	43	49	24	45	20				

Table S.1

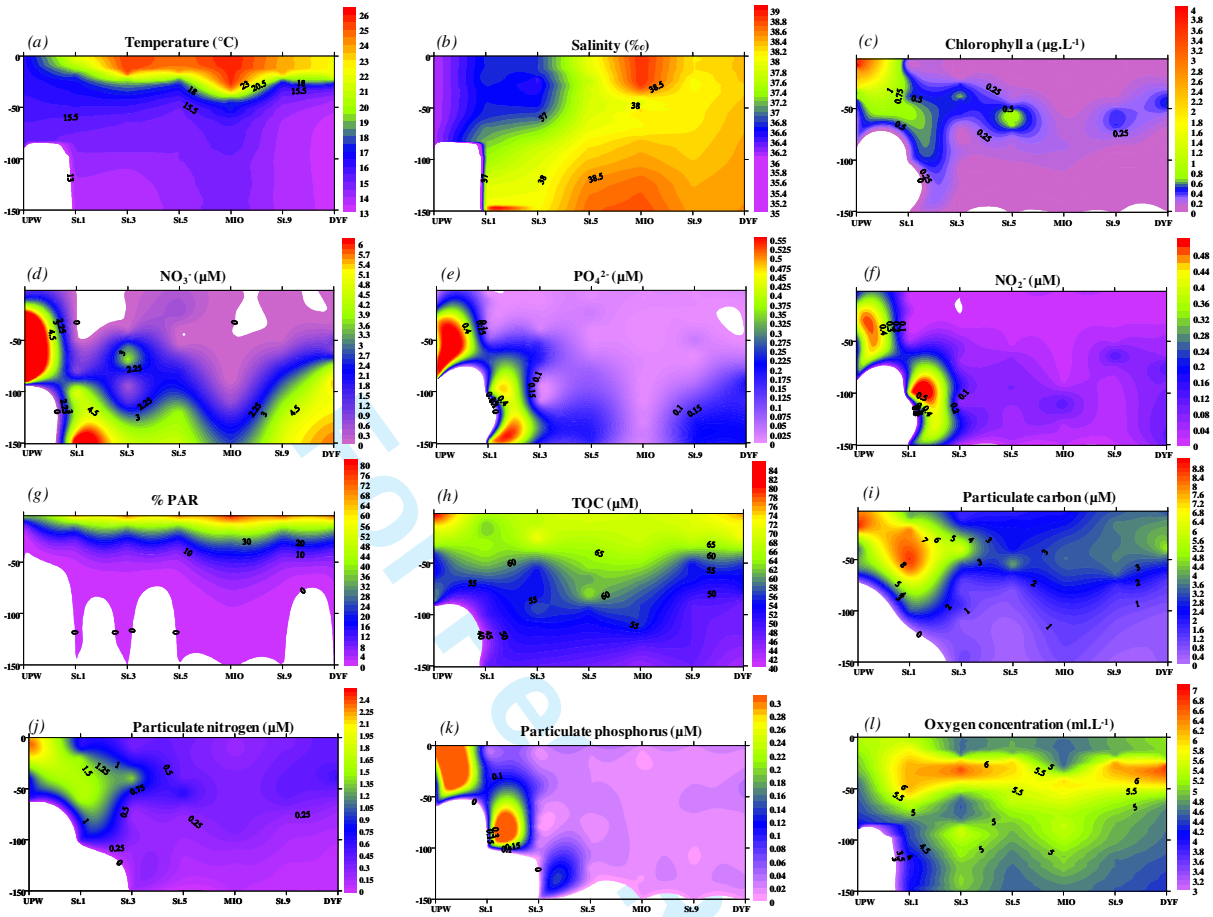

Figure S.1

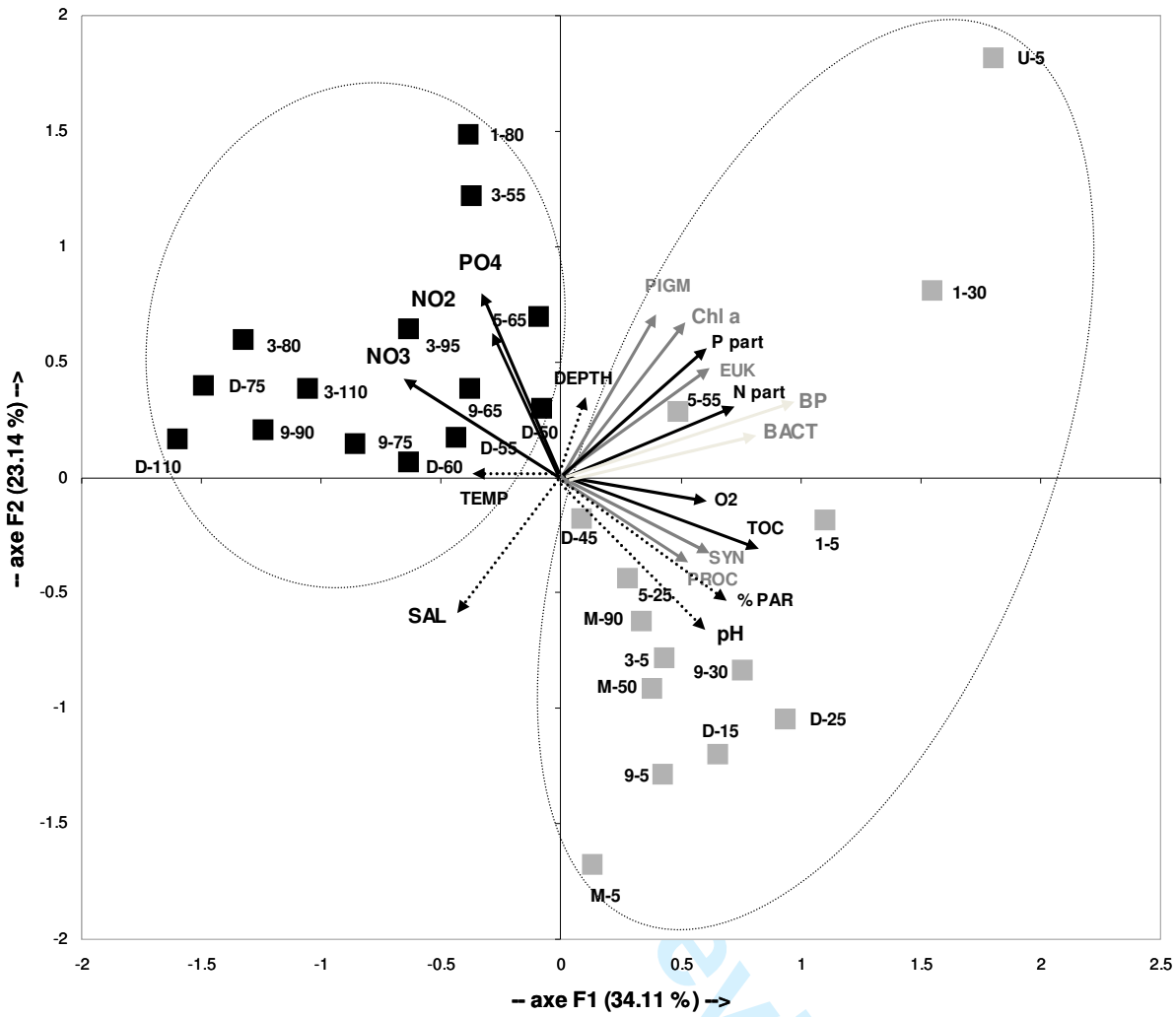
U-5
1-80
3-55
PO4
NO2
5-65
3-80
NO3
3-95
D-75
3-110
9-65
D-50
9-90
9-75
D-55
D-110
D-60
TEMP
SAL
DEPTH
PIGM
Chl a
P part
EUK
N part
5-55
BP
BACT
1-30
O2
TOC
1-5
D-45
SYN
PROC
% PAR
5-25
M-90
pH
3-5
9-30
M-50
D-25
D-15
9-5
M-5
-- axe F1 (34.11 %) -->
-- axe F2 (23.14 %) -->

Figure S.2

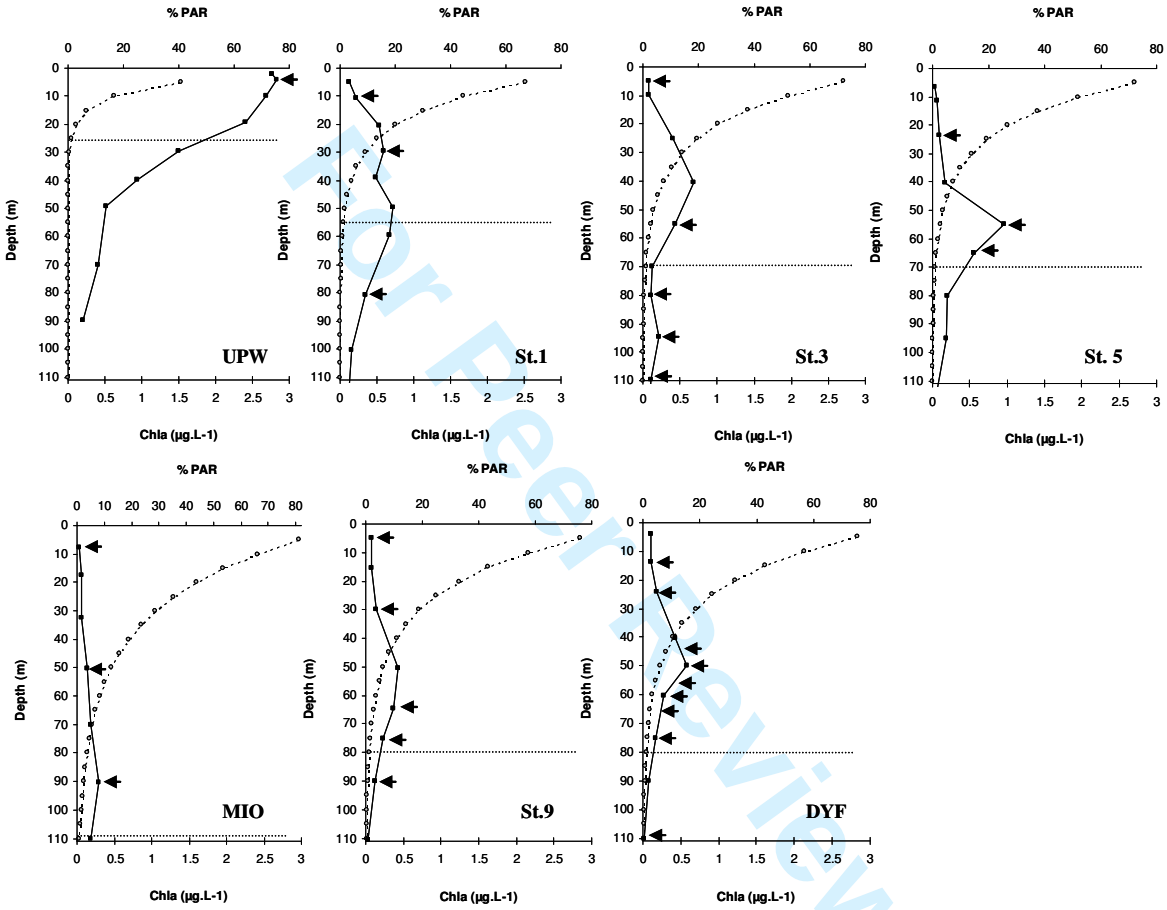

Figure S.3

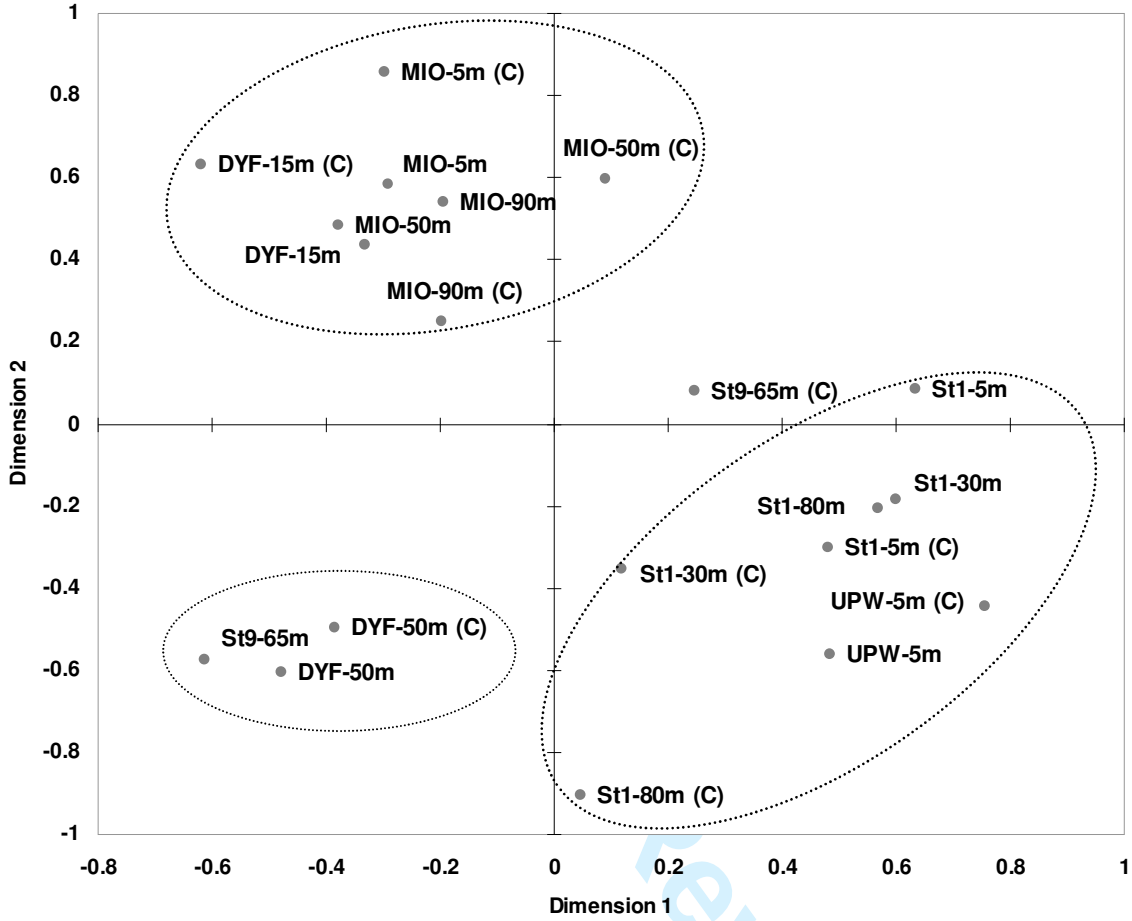

Figure S.4