



**HAL**  
open science

## 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

### ► To cite this version:

Anne-Catherine Lehours, Matthew T. Cottrell, Océane Dahan, David L. Kirchman, Christian Jeanthon. Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the Mediterranean Sea in relation to environmental variables. *FEMS Microbiology Ecology*, 2010, 74 (2), pp.397-409. 10.1111/j.1574-6941.2010.00954.x . hal-00529310

**HAL Id: hal-00529310**

**<https://hal.science/hal-00529310>**

Submitted on 14 Apr 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1  
2  
3 1 **Submitted to FEMS Microbiol. Ecol.**  
4  
5  
6 2  
7  
8 3  
9

10 4 Title: Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the  
11  
12 Mediterranean Sea in relation to environmental variables  
13  
14 5  
15 6

17 7 Authors: Anne-Catherine Lehours<sup>1,2\*</sup>, Matthew T. Cottrell<sup>3</sup>, Océane Dahan<sup>1,2</sup>, David L.  
18  
19 Kirchman<sup>3</sup> and Christian Jeanthon<sup>1,2</sup>  
20  
21 8  
22 9

23  
24  
25 10 Address :

26  
27 11 <sup>1</sup>UPMC Univ Paris 06, UMR 7144, Adaptation et Diversité en Milieu Marin, Station  
28  
29 Biologique de Roscoff, 29682 Roscoff, France  
30  
31 12

32 13 <sup>2</sup>CNRS, UMR 7144, Adaptation et Diversité en Milieu Marin, Station Biologique de Roscoff,  
33  
34 29682 Roscoff, France  
35  
36 15

37 16 <sup>3</sup>University of Delaware, School of Marine Science and Policy, Lewes, Delaware 19958, USA  
38  
39  
40

41 17 \*Present address: Clermont Université, Université Blaise Pascal, UMR CNRS 6023,  
42  
43 Laboratoire Microorganismes : Génome et Environnement (LMGE), BP80026, F-63171  
44  
45 Aubiere  
46  
47  
48 20  
49  
50  
51 21

52  
53 22 For correspondence. Email: [jeanthon@sb-roscoff.fr](mailto:jeanthon@sb-roscoff.fr). Tel. (+33) 298 292 379; Fax (+33)  
54  
55 298 292 324.  
56  
57  
58 24  
59 25  
60 26  
27

1  
2  
3 28  
4 29  
5 30  
6  
7 31  
8  
9  
10 32  
11 33  
12  
13  
14 34  
15  
16 35  
17  
18 36  
19  
20 37  
21  
22 38  
23  
24 39  
25  
26 40  
27  
28 41  
29  
30 42  
31  
32 43  
33  
34 44  
35  
36 45  
37  
38 46  
39  
40 47  
41  
42 48  
43  
44 49  
45  
46 50  
47  
48 51  
49  
50 52  
51  
52 53  
53  
54 54  
55  
56  
57  
58  
59  
60

## 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.

55

56

## Introduction

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80

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

1  
2  
3 79 represent 60% of the oceans. The environmental factors controlling the population structure  
4  
5 80 and diversity of these photoheterotrophic bacteria remain yet poorly understood (Eiler, 2006).  
6  
7

8 81 The Mediterranean Sea is an ideal environment for these ecological studies as it offers  
9  
10 82 a range of trophic conditions including extreme oligotrophy, particularly in summer when the  
11  
12 83 water column is strongly stratified (Berman *et al.*, 1985). While the N:P ratio is close to the  
13  
14 84 Redfield ratio (16:1) in most oceanic waters, Mediterranean waters have a higher ratio,  
15  
16 85 especially in the eastern Basin, leading to strong phosphorus limitation (Moutin & Raimbault,  
17  
18 86 2002). A complex thermohaline circulation coupled with regional hydrodynamic features also  
19  
20 87 contributes to the establishment of many different oceanic regions throughout the  
21  
22 88 Mediterranean Sea (Manca *et al.*, 2004). As an example, the exchange of the Atlantic and  
23  
24 89 Mediterranean water masses at the Strait of Gibraltar induces marked salinity and temperature  
25  
26 90 gradients (Gascard & Richez, 1985). Although the different trophic conditions available in the  
27  
28 91 Mediterranean Sea provide a unique context to link nutrient availability, trophic status and  
29  
30 92 functioning of the food web to the dynamics of photoheterotrophic populations, the  
31  
32 93 distribution and diversity patterns of AAP bacteria have been only partially explored (Oz *et*  
33  
34 94 *al.*, 2005; Yutin *et al.*, 2005; 2008).  
35  
36  
37  
38  
39  
40

41 95 In the present study, we analyzed the biogeography patterns of AAP populations  
42  
43 96 collected in late summer along two transect during the PROSOPE (PROductivité des  
44  
45 97 Systèmes Océaniques PELagiques) cruise. To explore what environmental factors control  
46  
47 98 structure and diversity of AAP populations in the Mediterranean Sea, we monitored their  
48  
49 99 longitudinal and vertical changes using a *pufM*-based PCR-Temporal Temperature Gel  
50  
51 100 Electrophoresis (TTGE) survey and analyzed the *pufM* clone libraries from selected stations  
52  
53 101 and depths.  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 104  
4  
5  
6 105  
7  
8 106

## Material and Methods

### 107 **Sampling and nucleic acid extraction**

108 Seawater samples were collected from seven stations along two transects in the  
109 Mediterranean Sea (from Gibraltar to the Ionian Sea off the north-east coast of Lybia and  
110 north-west to the French coast through the Tyrrhenian and Ligurian Seas) in September and  
111 October 1999 during the PROSOPE cruise aboard the R.V. *La Thalassa* (Fig. 1). For  
112 molecular diversity studies of AAP populations, 1.45-5 l water samples were retrieved using  
113 12 l Niskin bottles fitted on a Rosette sampler equipped with conductivity, temperature and  
114 depth (CTD) sensors. Seawater was prefiltered through 3 µm pore-size, 47 mm diameter,  
115 Nuclepore filters using moderate vacuum in order to separate picoplankton from larger  
116 organisms. Picoplanktonic cells were collected by filtration as previously described (Marie *et*  
117 *al.*, 2006; Garczarek *et al.*, 2007). The filtered biomass was transferred into a cryovial  
118 containing 3.5 ml of DNA lysis buffer (0.75 M sucrose, 50 mM Tris-HCl, pH 8) and  
119 immediately frozen in liquid nitrogen. DNA extraction was performed as previously described  
120 (Marie *et al.*, 2006). Ancillary data (nutrients, dissolved oxygen, chlorophyll a, salinity,  
121 temperature, *etc.*) and methods used to analyze them are available from the PROSOPE web  
122 site ([http://www.obs-vlfr.fr/cd\\_rom\\_dmtt/pr\\_main.htm](http://www.obs-vlfr.fr/cd_rom_dmtt/pr_main.htm)).

123

### 124 **Environmental *pufM* gene amplification**

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

1  
2  
3 129 AYNGCRAACCACCANGCCCA-3', Yutin *et al.*, 2005) to amplify partial sequences of the  
4  
5 130 *pufM* gene (245 bp fragments). For TTGE analyses, a 5'-Cy5-labeled PufMF primer and a  
6  
7  
8 131 primer PufWAW with a 40 bp GC-clamp  
9  
10 132 (CGCCCGCCGCGCCCCGCGCCCGTCCCGCCGCCCCCGCCCG) added at the 5'-end  
11  
12 133 were used. Reaction mixture (50  $\mu$ L) contained the following components: 5X buffer (10 $\mu$ l),  
13  
14 134 2 mM MgCl<sub>2</sub>, 10 pmoles of each deoxyribonucleotide triphosphate (dATP, dCTP, dGTP,  
15  
16 135 dTTP; Eurogenetec), 10 pmoles of each oligonucleotide primer, 2.5 U of GoTaq Flexi DNA  
17  
18 136 polymerase (Promega) and 50 to 100 ng of template DNA. Amplifications were carried out in  
19  
20 137 a GeneAmp PCR system 9700 (Applied Biosystems) with the following parameters: 95°C for  
21  
22 138 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  
23  
24 139 extension step at 72°C for 10 min. Amplified products were checked by electrophoresis in  
25  
26 140 1.5% agarose in 0.5 $\times$  Tris-Acetate-EDTA (TAE) buffer and further quantified with a DNA  
27  
28 141 quantitation fluorescence assay kit (Sigma-Aldrich).  
29  
30  
31  
32  
33  
34  
35

### 36 143 **Temporal Temperature gel Gradient Electrophoresis (TTGE) profiling and analyses**

37  
38 144 One hundred nanograms of each amplified product were electrophoresed along an 8%  
39  
40 145 (wt/vol) polyacrylamide gel (ratio acrylamide to bis-acrylamide 37.5:1) containing 7M urea,  
41  
42 146 1.25X TAE, 0.06% of N,N,N',N'-Tetramethylethylenediamine (Temed) and 0.06%  
43  
44 147 ammonium persulfate using the DCode Universal Mutation Detection System (BioRad,  
45  
46 148 Hercules, CA). Runs were performed in 1.25x TAE at 68 V for 17 h with a temperature range  
47  
48 149 of 66 to 69.7°C and a ramp rate of 0.2°C h<sup>-1</sup>. Standard markers were generated with a mixture  
49  
50 150 of *pufM* PCR products amplified from *Erythrobacter longus* strain OCh 101<sup>T</sup>, *E. litoralis*  
51  
52 151 strain T4<sup>T</sup>, *Roseobacter denitrificans* strain OCh 114<sup>T</sup> (CIP104266), *Dinoroseobacter shibae*  
53  
54 152 strain DFL12<sup>T</sup>. Gel images were obtained at 100- $\mu$ m resolution using a Typhoon Trio variable  
55  
56 153 mode imager (Amersham Biosciences, Piscataway, NJ). Typhoon scans were acquired using  
57  
58  
59  
60

1  
2  
3 154 the 633 nm excitation laser and the 670 BP 30 emission filter as recommended by the  
4  
5 155 manufacturer for the detection of Cy5-labeled molecules. All gels were scanned with  
6  
7  
8 156 photomultiplier tube voltages to maximize signal without saturating fingerprint bands. Band  
9  
10 157 patterns were analyzed with GelCompare 4.6 software package (Applied Maths, Kortrijk,  
11  
12 158 Belgium). In band assignment, a 1% band position tolerance (relative to total length of the  
13  
14  
15 159 gel) was applied, which indicates the maximal shift allowed for two bands in different TTGE  
16  
17 160 patterns to be considered as identical. The number of bands in a profile was expressed as the  
18  
19 161 phylotype richness and the Shannon-Weiner index ( $H'$ ) was calculated as previously  
20  
21 162 described (Hill *et al.*, 2003). The Smith and Wilson evenness index ( $E_{\text{var}}$ , Smith & Wilson,  
22  
23 163 1996) was calculated using the Ecological Evenness Calculator software  
24  
25 164 (<http://www.nateko.lu.se/personal/benjamin.smith/software>).  
26  
27  
28  
29  
30  
31

### 32 166 **Clone library construction and analyses**

33  
34 167 Fresh PCR products were cloned using the TOPO TA cloning kit according to the  
35  
36 168 manufacturer's instructions (Invitrogen Corporation, Carlsbad, CA, USA). Recombinant  
37  
38 169 clones were screened for insert-containing plasmids by direct PCR amplification with M13  
39  
40 170 forward and reverse primers. Clones were sequenced using an ABI 3130 POP7 sequencer  
41  
42 171 (Applied Biosystems, Foster City, CA) at the Biogenouest Sequencing-Genotyping Platform  
43  
44 172 (Roscoff site). Clone libraries were screened for chimeric sequences with Chimera\_Check  
45  
46 173 program available on the RDP website (Cole *et al.*, 2003). The 388 remaining sequences were  
47  
48 174 subjected to BLAST search against publicly available sequences to determine their  
49  
50 175 approximate phylogenetic affiliations. A conservative value of 94 % nucleic acid sequence  
51  
52 176 similarity (Zeng *et al.*, 2007) was chosen for grouping sequences into Operational Taxonomic  
53  
54 177 Units (OTUs) using the phylogenetic analysis software Bosque available at  
55  
56 178 <http://bosque.udec.cl> (Ramírez-Flandes & Ulloa, 2008). Coverage value (C) was calculated as  
57  
58  
59  
60



1  
2  
3 179 previously described (Mullins *et al.*, 1995). The Shannon-Weiner index ( $H'$ , Hill *et al.*, 2003),  
4  
5 180 the richness-estimator  $S_{\text{chao1}}$  (Hughes *et al.*, 2001) and the Abundance-base Coverage  
6  
7  
8 181 Estimator (ACE, Chazdon *et al.*, 1998) were computed using EstimateS software Version 7.5.  
9  
10 182 (K. Colwell, <http://purl.oclc.org/estimates>).

11  
12 183 A *pufM* database containing more than 700 aligned sequences of cultured species and  
13  
14 184 environmental clones retrieved from GenBank database  
15  
16 185 (<http://www.ncbi.nlm.nih.gov/Genbank/>) and the GOS scaffold nucleotide sequences was  
17  
18 186 constructed using ARB (Ludwig *et al.*, 2004). Sequences were translated to protein and the  
19  
20 187 resulting alignment was then used to manually realign nucleotide sequences. A neighbor-  
21  
22 188 joining tree was first constructed with all the sequences longer than 600 bp and the robustness  
23  
24 189 of inferred tree topologies was tested by bootstrap analysis (1000 resamplings) using PHYLIP  
25  
26 190 (Felsenstein, 1993). Shorter sequences were aligned as above and added to the tree using  
27  
28 191 ADD-BY-PARSIMONY algorithm. Phylogenetic tree display and annotation were performed  
29  
30 192 with iTOL software (Letunic & Bork, 2006). The *pufM* sequences obtained in this study are  
31  
32 193 deposited in the GenBank database under accession **No. GQ468944 to GQ468986.**  
33  
34  
35  
36  
37  
38  
39  
40

#### 41 **Statistical analyses**

42  
43 196 The normality of environmental variables was checked using Shapiro-Wilk (Shapiro,  
44  
45 197 1965) and Anderson-Darling (Stephens, 1974) tests and variables were transformed when  
46  
47 198 necessary to correct for deviations from normality. Principal component analysis (PCA),  
48  
49 199 performed with XLSTAT version 6.01 (Addinsoft), was used to group samples according to  
50  
51 200 environmental variables using the Pearson correlation coefficient.  
52

53  
54 201 For TTGE patterns, pairwise similarity matrices were calculated using the Dice and the Bray-  
55  
56 202 Curtis equations for presence/absence and relative peak height data, respectively.  
57  
58 203 Dendrograms were generated from the Dice matrix using the method described by Ward  
59  
60

1  
2  
3 204 (1963). The consistence of a cluster was expressed by the cophenetic correlation which  
4  
5 205 calculates the correlation between the dendrogram-derived similarities and the matrix  
6  
7  
8 206 similarities. A distance of 35 was used to separate clusters in the hierarchical classification.  
9  
10 207 Analysis of similarity (ANOSIM, Clarke and Warwick, 2001) was used to test the hypothesis  
11  
12 208 that communities within each cluster were more similar to each other than to communities in  
13  
14  
15 209 other clusters. Correlations between similarity matrices were calculated using a Mantel test  
16  
17 210 (Mantel, 1967) with 10000 permutations and were performed with XLSTAT version 6.01.  
18  
19  
20 211 Relationships among samples were visualized using the ordination technique  
21  
22 212 multidimensional scaling (MDS) using a standardized stress with 1000 iterations computed  
23  
24 213 with XLSTAT version 6.01. Canonical correspondence analyses (Legendre & Legendre,  
25  
26 214 1998) used to determine the extent to which selected environmental variables explained  
27  
28  
29 215 patterns of similarity in community composition were performed using PAST 1.81 (Hammer  
30  
31 216 *et al.*, 2001) available at [http://palaeo-electronica.org/2001\\_1/past/issue1\\_01.htm](http://palaeo-electronica.org/2001_1/past/issue1_01.htm)  
32  
33

## 34 217 35 36 218 **Results**

### 37 219 **Oceanographic context**

38  
39  
40  
41 220 The PROSOPE cruise sampled the highly eutrophic Morocco upwelling off Agadir  
42  
43 221 (St. UPW, Fig. 1) and then proceeded to a first Mediterranean west-east (W-E) transect,  
44  
45 222 through a markedly decreasing gradient of surface chlorophyll *a* concentrations, from the  
46  
47 223 mesotrophic St. 1 at the Strait of Gibraltar to St. MIO located in a highly oligotrophic area in  
48  
49 224 the center of the Ionian Sea. A second transect sampled the Tyrrhenian and Ligurian Seas (St.  
50  
51 225 9, St. DYF).  
52  
53  
54

55 226 Plots of the depth variations of environmental variables analyzed along both these  
56  
57 227 transect revealed the complexity of the studied zones in which several physical and chemical  
58  
59 228 gradients were superimposed (Figs. S1a to S1l). Temperature and salinity transects reflected  
60

1  
2  
3 229 the progression of low-salinity, low-temperature Atlantic waters penetrating through the  
4  
5 230 Gibraltar Strait and flowing along the African coast (Figs. S1a-b). Thus a strong halocline was  
6  
7  
8 231 observed between St. 3 and St. 5 identifying a partitioning of surface waters into westwards  
9  
10 232 and typical high salinity eastwards (21.6°C and 36.7‰ for St. 1 to 26°C and 38.9‰ for St.  
11  
12 233 MIO, respectively). The upwelling at St. UPW supplied abundant dissolved nutrients to the  
13  
14 234 surface layer, which supported phytoplankton growth as indicated by the surface maximum in  
15  
16 235 chlorophyll *a* (Fig. S1c). A progressive W-E deepening of the nitracline and phosphacline  
17  
18 236 down to 98 and 124 m depth at St.1 and St. MIO, respectively, was noted (Figs. S1d-e).  
19  
20 237 Along the second transect (from St. MIO to St. DYF), concentrations of nitrates and  
21  
22 238 phosphates increased below the deep chlorophyll maximum (DCM).  
23  
24  
25

26  
27 239 The main biological and physico-chemical parameters of the waters were analyzed  
28  
29 240 using a principal component analysis (PCA) including nineteen variables (salinity,  
30  
31 241 temperature, dissolved oxygen, light, pH, depth, phosphate, nitrate, nitrite, particulate organic  
32  
33 242 carbon, nitrogen, phosphorus, total chlorophyll *a*, accessory pigments, bacterial production,  
34  
35 243 abundances of *Synechococcus*, *Prochlorococcus*, non photosynthetic bacteria and  
36  
37 244 picoeukaryotes) (Fig. S2). This analysis allowed separation of the Mediterranean Sea euphotic  
38  
39 245 zone into two distinct layers above and in/or below the DCM, whose discrimination was  
40  
41 246 mainly explained by biological variables and nutrient levels (NO<sub>3</sub>, NO<sub>2</sub>, PO<sub>4</sub>) respectively.  
42  
43 247 All surface waters except that of the Morocco upwelling and the Gibraltar Strait (St. 1) were  
44  
45 248 nutrient depleted (Figs. S1 and S2). Covariation of biological parameters with particulate  
46  
47 249 organic compounds, oxygen, pH and photosynthetically available radiation (PAR) are  
48  
49 250 consistent with the characteristics of the upper layer where photosynthesis supports the  
50  
51 251 formation of organic matter (Lucea *et al.*, 2003) allowing heterotrophic activities as indicated  
52  
53 252 by the inter-correlation of these parameters with bacterial production and cell densities (Fig.  
54  
55 253 S1 and S2). Deep euphotic waters located in or below the DCM were relatively similar as they  
56  
57  
58  
59  
60

1  
2  
3 254 were cold, enriched in nutrients and depleted in oxygen (Fig. S1). Although a strong halocline  
4  
5 255 partitioned W-E surface waters, salinity did not explain variability among samples. This  
6  
7  
8 256 might result from the superimposed longitudinal and vertical variation trends of salinity  
9  
10 257 concentrations (Fig. S1b). Discrimination of surface and deep euphotic layers in the  
11  
12 258 Mediterranean Sea mainly according to their nutrient status suggests that the vertical gradient  
13  
14  
15 259 may prevail over the longitudinal gradient during summer stratification.

### 17 260 **Spatial variability, diversity and structure of AAP populations**

18  
19  
20 261 PCR amplified *pufM* genes from different depths (Fig. S3) were analyzed by temporal  
21  
22 262 temperature gradient gel electrophoresis (TTGE) which yielded a total of 79 unique bands  
23  
24 263 with an average of  $34 \pm 6$  bands in each sample (Fig. 2). A Mantel test also showed that Bray-  
25  
26  
27 264 Curtis and Dice similarity matrices calculated from TTGE profiles were significantly  
28  
29 265 correlated ( $r=0.748$ ,  $p<0.05$ ). The hierarchical clustering analysis (HCA) identified four  
30  
31 266 clusters of AAP bacteria (Fig. 2) confirmed by ANOSIM statistics performed from both  
32  
33  
34 267 presence/absence and relative intensity of TTGE bands (data not shown). Cluster A contained  
35  
36 268 populations from the coastal stations UPW and St. 1, together with those of the surface waters  
37  
38  
39 269 of the Algerian Basin (St. 3-5m). Cluster B represented the deep euphotic zone of the  
40  
41 270 Algerian Basin (St. 3) and the Strait of Sicily (St. 5) whereas Cluster C grouped TTGE  
42  
43 271 profiles of samples collected above the DCM in the Ionian, Tyrrhenian, Ligurian Seas and the  
44  
45  
46 272 Strait of Sicily. AAP bacteria from deep euphotic waters of stations St. 9 and St. DYF  
47  
48 273 clustered together in Cluster D. Richness and diversity ( $E_{var}$  and  $H'$ ) values were not  
49  
50 274 significantly different between clusters (data not shown).  $E_{var}$ ,  $H'$  and richness were positively  
51  
52  
53 275 correlated to PAR and biological activity indicators and were negatively correlated with  $NO_3$   
54  
55 276 and  $NO_2$  concentrations.

56  
57  
58 277 To obtain two-dimensional-coordinates for samples and to confirm HCA groupings,  
59  
60 278 ordination of Bray-Curtis similarities among sample profiles was performed by non-metric

1  
2  
3 279 multidimensional scaling (MDS). With the exception of samples St1-80m and St. 3-5m, the  
4  
5 280 four clusters recovered by HCA were defined (data not shown). Populations above the DCM  
6  
7  
8 281 were significantly more similar to each other than to those in or below the DCM ( $R=0.419$ ,  
9  
10 282  $p<0.001$ ). Dimension 1 from the TTGE/MDS analyses was negatively correlated to nutrient  
11  
12 283 variables ( $\text{NO}_3$  and  $\text{NO}_2$ ). In contrast to dimension 1, dimension 2 co-varied with numerous  
13  
14 284 variables characterizing biological activity (e.g. oxygen, bacterial production, cell densities).

15  
16  
17 285 Variables significantly related to dimensions 1 and 2 were integrated in a Canonical  
18  
19 286 Correspondence Analysis (CCA) performed from relative intensity of TTGE bands. CCA  
20  
21  
22 287 revealed that more than 50% of the variability of AAP communities was described by the 11  
23  
24 288 selected variables (Fig. 3). The four clusters (A to D) previously detected by HCA analysis  
25  
26 289 were retrieved. Cluster D was mainly separated according to  $\text{NO}_3$  and  $\text{NO}_2$  variables, while  
27  
28  
29 290 clusters A and C were discriminated by pH, PAR, total organic carbon, biogenic activity  
30  
31 291 variables and particulate nitrogen and phosphorus. None of the selected variables clearly  
32  
33  
34 292 explained pattern of similarity within cluster B.

### 35 36 293 37 294 **Phylogenetic analyses of pufM genes**

38  
39 295  
40 296 Ten out of the 29 samples were selected for phylogenetic analyses on the basis of their  
41  
42 297 location and the diversity of AAP populations. Of a total of 388 clones analyzed, 44 distinct  
43  
44 298 OTUs were identified after grouping the sequences at 94% nucleic acid sequence similarity  
45  
46 299 (Fig. 4, Table S1). Coverage values (Table 1) and rarefaction curves (data not shown)  
47  
48  
49 300 indicated that most of the diversity was detected in most libraries (> 71%). Differences in  
50  
51 301 AAP population diversity were not significant between samples collected above and below  
52  
53 302 the DCM.

54  
55  
56 303 MDS for AAP populations from the 10 selected samples were performed from Dice  
57  
58 304 similarity matrices for both TTGE and clone library analyses (Fig. S4). As also expressed by  
59  
60 305 the significant correlation between Dice similarity matrices for TTGE and clone libraries

1  
2  
3 306 (R=0.39, p=0.05), both methods gave reliable information. The discrepancy observed for St.9-  
4  
5 307 65m could be due to the low coverage of the clone library (Table 1).  
6  
7

8 308 About half of the *pufM* sequences were less than 94% identical to known sequences  
9  
10 309 (Table S1). Most sequences were related to clones retrieved from the Mediterranean Sea  
11  
12 310 (29%) and coastal environments (52%) including the Delaware and Monterey Bays. A unique  
13  
14 311 sequence (PROSOPE-7) was affiliated with a clone from the Atlantic Ocean. The phylogenetic  
15  
16 312 analysis demonstrated that the sequences were distributed into 7 of the 12 phylogroups  
17  
18 313 previously defined by Yutin *et al.* (2007). None of the sequences was affiliated with the A, D,  
19  
20 314 H, J and L groups (Fig. 4, Table S1) that are mainly found in anoxic environments (Yutin *et*  
21  
22 315 *al.*, 2007).  
23  
24  
25

26 316 Only two sequences, belonging to PROSOPE-48 and PROSOPE-52 phylotypes, were  
27  
28 317 highly similar to that of cultured representatives (*Methylobacterium radiotolerans* and  
29  
30 318 *Erythrobacter longus*, respectively). Among the 7 AAP groups recovered from this study,  
31  
32 319 only groups B and K were present in all samples (Fig. 5). PROSOPE-6 which shared 99%  
33  
34 320 similarity with a clone from the Monterey Bay (Béja *et al.*, 2002) was the main phylotype of  
35  
36 321 group B. Representing 11% of the *pufM* sequences, PROSOPE-6 was prevalent (> 80%) at  
37  
38 322 coastal and mesotrophic stations (Fig. 4, Table S1). Most of our sequences (~80%) fell into  
39  
40 323 Group K which contained *Gammaproteobacteria* representatives including few isolates such  
41  
42 324 as *Congregibacter litoralis* KT71 (Eilers *et al.*, 2001) and strains NOR5-3 and NOR51B and  
43  
44 325 HTCC2080 (Cho *et al.*, 2007) and BAC clones EBAC65D09 and EBAC29C02 (Béjà *et al.*,  
45  
46 326 2002), all related to the OM60 clade (Rappé *et al.*, 1997). Among the 30 OTUs affiliated to  
47  
48 327 Group K, 10 (i.e. PROSOPE -10, -11, -38) and 9 (i.e. PROSOPE -12, -15, -45) phylotypes were  
49  
50 328 only detected at the coastal Atlantic (St. UPW and St. 1) and at Mediterranean stations,  
51  
52 329 respectively. Among gammaproteobacterial sequences, PROSOPE-34, which represented up to  
53  
54 330 20% of the total *pufM* sequences, dominated at meso- and eutrophic stations while other  
55  
56  
57  
58  
59  
60

1  
2  
3 331 phylotypes (i.e. PROSOPE-14, -42, -45) prevailed (>80%) at oligotrophic stations. The  
4  
5 332 distribution of the other groups was sporadic. A few clones recovered at MIO and UPW  
6  
7 333 stations clustered in the group C. The *Roseobacter*-like (group G) and *Rhodobacter*-like  
8  
9 334 (groups E and F) clones were only distributed in the nutrient rich coastal waters of the  
10  
11 335 Morocco upwelling (st. UPW) and the Strait of Gibraltar (UPW and St. 1) (Fig. 5). The  
12  
13 336 phylotype PROSOPE-7 consisting of a few clones retrieved at station MIO grouped into the  
14  
15 337 group I. The closest cultured relatives of this group are *Betaproteobacteria* widely distributed  
16  
17 338 in freshwater and estuarine environments (Page *et al.*, 2004; Yutin *et al.*, 2007). This could  
18  
19 339 suggest that closely related organisms also thrive in oceanic surface waters.  
20  
21  
22  
23  
24

## 25 340 Discussion

26  
27 341 The Mediterranean Sea is semi-enclosed allowing the study, on a reduced scale, of  
28  
29 342 processes typical of the world's oceans. It constitutes a unique environment for ecological  
30  
31 343 studies as it offers a range of trophic conditions, including extreme oligotrophy, particularly in  
32  
33 344 summer when the water column is strongly stratified. The PROSOPE cruise was organized a  
34  
35 345 few years before the discovery of the ecological significance of AAP bacteria in oceanic  
36  
37 346 waters (Kolber *et al.*, 2001). Therefore, the sampling strategy was not intended to allow the  
38  
39 347 examination of their abundance and their contribution to the microbial community structure  
40  
41 348 using infrared microscopy and fluorometry methods. One of the main objectives of the  
42  
43 349 present study was to investigate environmental factors influencing biogeography of AAPs by  
44  
45 350 analyzing the *pufM* gene diversity in archived bacterial community DNA samples. We agree  
46  
47 351 that top-down regulation factors probably contribute to the distribution of AAP populations  
48  
49 352 (Koblížek *et al.*, 2007; Zhang & Jiao, 2009), however, this study provides one of the first look  
50  
51 353 on bottom-up factors that contribute to longitudinal and vertical distribution and diversity of  
52  
53 354 AAP populations.  
54  
55  
56  
57  
58  
59  
60

356

357

**358 Linking AAP population structure and diversity to environmental parameters**

359 The diversity of AAP populations as evident in *pufM* sequences was relatively  
360 constant along both transects and was similar to that reported by Jiao *et al.* (2007) in  
361 oligotrophic oceanic provinces. We found a positive correlation between the PAR and  
362 diversity ( $H'/E_{\text{var}}$ ) indices, suggesting that AAP populations are more diverse in the upper  
363 surface layer and that light affected their community structure. In contrast to Jiao *et al.*  
364 (2007), who reported that the AAP diversity decreased with increasing *Chla* concentrations,  
365 we did not find a significant correlation between these variables. Our study found, however,  
366 an inverse relationship between AAP diversity and nitrate levels. This result together with the  
367 highest richness observed for the highly oligotrophic station MIO also supports the hypothesis  
368 that inorganic nitrogen concentrations may affect their distribution (Waidner & Kirchman,  
369 2007) and that AAP diversity decreases with increasing trophic status (Jiao *et al.*, 2007). This  
370 suggests that the scarcity of inorganic nutrients and dissolved organic matter in oligotrophic  
371 environments leads to a higher specialization of AAP assemblages and increases the  
372 functional redundancy (Curtis & Sloan, 2004) for this group, as it has been observed  
373 elsewhere for other bacterial populations (Wohl *et al.*, 2004). This hypothesis is consistent  
374 with the contrasting variation of abundance and diversity of AAP assemblages previously  
375 suggested by a large-scale survey analysis of their distribution patterns (Jiao *et al.*, 2007).

**376 Variation in the composition of AAP assemblages and environmental context**

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



1  
2  
3 380 some phylotypes retrieved only at coastal stations while others dominated in oligotrophic  
4  
5 381 waters.

6  
7  
8 382 Our analyses suggest that stratification is a critical factor determining the vertical  
9  
10 383 distribution of AAP species in the Mediterranean Sea. Similar vertical stratification has been  
11  
12 384 reported for the whole bacterial community structure in the Mediterranean Sea (Acinas *et al.*,  
13  
14 385 1997; Ghiglione *et al.*, 2008) and in other oligotrophic waters, including the Pacific and  
15  
16 386 Atlantic oceans, the Caribbean Sea (Lee & Fuhrman, 1991), and Antarctic environments  
17  
18 387 (Murray *et al.*, 1998). Multivariate analyses of environmental parameters and molecular  
19  
20 388 fingerprinting profiles revealed that the variation in AAP populations with depth is due to  
21  
22 389 synergetic driving forces similar to those involved in the vertical stratification of bacterial  
23  
24 390 communities at the DYF station (Ghiglione *et al.*, 2008). Fifty percent of the variability of  
25  
26 391 AAP composition was explained by the selected environmental parameters. This suggests that  
27  
28 392 the composition of AAP assemblages may be additionally influenced by other factors such as  
29  
30 393 specific and multiple interactions with other organisms of their environment such as their  
31  
32 394 bacterial counterparts, protists, viruses, and metazoans (Fuhrman, 2009) as well as intrinsic  
33  
34 395 photoheterotrophic capabilities and physiological peculiarities of each species. Indeed, AAP  
35  
36 396 bacteria have diverse metabolisms, ranging from generalists (e.g. *Erythrobacter*, *Roseobacter*)  
37  
38 397 to specialized species (e.g. *Citromicrobium*) able to grow on a broad and a narrow spectrum  
39  
40 398 of carbon sources (Yurkov and Csotonyi 2009). Further in-depth studies on the physiology  
41  
42 399 and metabolism of environmentally significant organisms and on their biological interactions  
43  
44 400 with other planktonic counterparts are essential to better understand the ecology of marine  
45  
46 401 AAP.

#### 47 48 49 50 51 52 53 54 55 402 **Phylogenetic composition of populations**

56  
57 403 Most *pufM* sequences retrieved from our samples belonged to clades previously  
58  
59 404 identified in a global metagenomic survey (Yutin *et al.*, 2007). As previously observed in

1  
2  
3 405 environmental surveys of these genes (e.g. Bèjà *et al.*, 2002), most of the OTUs retrieved in  
4  
5 406 this study were distantly related to known anoxygenic phototrophs. More than 80% of our  
6  
7  
8 407 *pufM* sequences had their best matches with clone sequences obtained from the Mediterranean  
9  
10 408 Sea and from coastal areas suggesting that AAP populations in this sea are different from  
11  
12 409 those in open oceans. We acknowledge, however, that this observation may be biased by the  
13  
14 410 low number of offshore *pufM* sequences available in the databases.

15  
16  
17 411 In contrast to previous PCR-based studies (Oz *et al.*, 2005; Yutin *et al.*, 2005),  
18  
19 412 *Alphaproteobacteria* constitute a minor fraction of the Mediterranean AAP community.  
20  
21  
22 413 Members of the *Roseobacter*-related group (group G), often a key player in diverse marine  
23  
24 414 systems (Buchan *et al.*, 2005; Yutin *et al.*, 2007), were only detected at Atlantic coastal  
25  
26 415 stations. While this finding supports the notion that *Roseobacter*-like bacteria are favored by  
27  
28 416 nutrient-rich conditions, in agreement with the common association of *Roseobacter* with  
29  
30 417 phytoplankton blooms (González *et al.*, 2000; Suzuki *et al.*, 2001), it seems to contradict their  
31  
32 418 suspected important role and dominance in the Mediterranean Sea (Oz *et al.*, 2005). However,  
33  
34 419 since *Roseobacter* species can be more abundant in winter than in summer, this discrepancy  
35  
36 420 with previous studies (Oz *et al.*, 2005; Yutin *et al.*, 2005) could be explained by the different  
37  
38 421 sampling period.

39  
40  
41 422 Previous studies have revealed a high contribution of *Gammaproteobacteria* to AAP  
42  
43 423 populations (Bèjà *et al.*, 2002; Hu *et al.*, 2006; Yutin *et al.*, 2007). However, the large  
44  
45 424 proportions observed in this study (~ 80%) have never been reported from any environments.  
46  
47  
48 425 Possible PCR biases cannot be definitively ruled out to explain the large dominance of  
49  
50 426 *Gammaproteobacteria* in our samples. However, by using the same primer combination to  
51  
52 427 study the diversity of AAP in arctic waters and in coastal systems of the English Channel, we  
53  
54 428 did not find such high percentages of gammaproteobacterial sequences (Lehours *et al.*, Dahan  
55  
56 429 *et al.*, unpublished results). Consistent with previous observations, gammaproteobacterial

1  
2  
3 430 sequences were retrieved in eutrophic and oligotrophic waters (Hu *et al.*, 2006; Jiao *et al.*  
4  
5 431 2007). Although some gammaproteobacterial phylotypes were obtained from both nutrient  
6  
7 432 levels, the distribution of most suggests an adaptation to specific trophic conditions. For  
8  
9 433 example, some of our cloned *pufM* sequences from the Atlantic coast were similar to the  
10  
11 434 sequences belonging to the OM60 clade abundant in coastal oceans (Béjà *et al.*, 2002; Yutin  
12  
13 435 *et al.*, 2007). Although prevalent at coastal stations, phylogroup B also showed a ubiquitous  
14  
15 436 distribution across the different trophic regimes (Yutin *et al.*, 2007). Onboard enrichment  
16  
17 437 experiments performed along the PROSOPE transect demonstrated a switch from a  
18  
19 438 phosphorus limitation in the surface layer to organic carbon limitation in the deep chlorophyll  
20  
21 439 maximum (Van Wambeke *et al.*, 2002). Specific adaptation capabilities to extreme  
22  
23 440 oligotrophy are likely to explain the success of phylogroups B and K in typical Mediterranean  
24  
25 441 waters. In line with this hypothesis, genes coding for several storage compounds were  
26  
27 442 identified in the genome of the gammaproteobacterium *Congregibacter litoralis* (Fuchs *et al.*,  
28  
29 443 2007). Among them, the storage compounds cyanophycin and polyphosphate probably reflect  
30  
31 444 an adaptation to fluctuating carbon and phosphorus availability.  
32  
33  
34  
35  
36  
37  
38  
39  
40

## 41 446 **Conclusion**

42  
43 447 The molecular analyses of AAP diversity conclusively demonstrated that typical  
44  
45 448 Mediterranean populations varied greatly along depth profiles, with strong and opposite  
46  
47 449 gradients of light and nutrient availability. This variation with depth was much greater than  
48  
49 450 seen across both transects. During stratification, the vertical distribution of AAP bacteria  
50  
51 451 seems to be governed by the same environmental variables as that of the whole bacterial  
52  
53 452 community (Ghiglione *et al.*, 2008). This result, however, does not imply that they may act as  
54  
55 453 strict heterotrophs *in situ* and that phototrophy has a minor impact on their distribution and  
56  
57 454 lifestyle. Since evidence from other studies indicate that AAP use their phototrophic  
58  
59  
60

1  
2  
3 455 capability at nutrient-poor levels (Cho *et al.*, 2007; Spring *et al.*, 2009), additional variables  
4  
5  
6 456 may influence their populations in the extreme oligotrophic conditions prevailing in the  
7  
8 457 Mediterranean Sea.

9  
10 458 This study was the first to reveal such a high abundance of gammaproteobacterial  
11  
12 459 AAP in the environment. An ecotypic differentiation was suggested by both TTGE and  
13  
14  
15 460 cloning approaches. Further culture efforts are therefore needed to expand the diversity of  
16  
17 461 gammaproteobacterial isolates and to delineate the environmental parameters that govern the  
18  
19 462 activity and distribution of gammaproteobacterial trophic ecotypes.  
20  
21

22 463

23  
24 464

### Acknowledgements

25  
26  
27 465 We wish to thank H. Claustre for coordinating the PROSOPE cruise and L. Garczarek, D.  
28  
29 466 Marie and F. Partensky for sampling. We are grateful to M. Perennou (Biogenouest  
30  
31 467 Sequencing Platform at the Station Biologique) for help with sequencing. We also thank  
32  
33 468 Milton Da Costa and I. Wagner-Döbler for the gift of the cultures of *Erythrobacter longus*, *E.*  
34  
35 469 *litoralis*, and *Dinoroseobacter shibae*, respectively. This work was supported by the programs  
36  
37 470 PROOF and LEFE-CYBER (CNRS-INSU). A-C. Lehours benefited from a post-doctoral  
38  
39 471 fellowship from the CNRS. Kirchman and Cottrell were supported by NSF MCB-0453993.  
40  
41  
42  
43

44 472

45 473

### References

46 474

47 475

48  
49 476 Achenbach LA, Carey J & Madigan MT (2001) Photosynthetic and phylogenetic primers for  
50 477 detection of anoxygenic phototrophs in natural environments. *Appl Environ Microbiol*  
51 478 **67**:2922–2926.

52 479

53 480 Acinas SG, Rodríguez-Valera F & Pedrós-Alió C (1997) Spatial and temporal variation in  
54 481 marine bacterioplankton diversity as shown by RFLP fingerprinting of PCR amplified 16S  
55 482 rDNA. *FEMS Microbiol Ecol* **24**:27-40.

56 483

57 484 Béjà O, Suzuki MT, Heidelberg JF, Nelson WC, Preston CM, Hamada T, Eisen JA, Fraser  
58 485 CM & DeLong EF (2002) Unsuspected diversity among marine aerobic anoxygenic  
59 486 phototrophs. *Nature* **415**: 630–633.

60 487

- 1  
2  
3 488 Berman T, Walline PD, Schneller A, Rothenberg J & Townsend DW (1985) Secchi disk depth  
4 489 record: a claim for the eastern Mediterranean. *Limnol Oceanogr* **30**:447-448.  
5 490  
6  
7 491 Buchan A, González JM & Moran AM (2005) Overview of the marine Roseobacter lineage.  
8 492 *Appl Environ Microbiol* **71**:5665-5677.  
9 493  
10 494 Chazdon RL, Colwell RK, Denslow JS & Guariguata MR (1998) Statistical methods for  
11 495 estimating species richness of woody regeneration in primary and secondary rain forests of  
12 496 NE Costa Rica. Pp. 285-309 in F. Dallmeier and J. A. Comiskey, eds. *Forest biodiversity*  
13 497 *research, monitoring and modeling: Conceptual background and Old World case studies*.  
14 498 Parthenon Publishing, Paris.  
15 499  
16 500 Cho JC, Stapels MD, Morris RM, Vergin KL, Schwalbach MS, Givan SA, Barofsky DF &  
17 501 Giovannoni SJ (2007) Polyphyletic photosynthetic reaction centre genes in oligotrophic  
18 502 marine Gammaproteobacteria. *Environ Microbiol* **9**:1456-1463.  
19 503  
20 504 Clarke KR & Warwick RW (2001) Change in marine communities: an approach to statistical  
21 505 analysis and interpretation, 2<sup>nd</sup> edition, Primer-E, Plymouth.  
22 506  
23 507 Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM,  
24 508 Schmidt TM, Garrity GM & Tiedje JM (2003) The ribosomal database project (RDP-II):  
25 509 previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy.  
26 510 *Nucleic Acid Res* **31**:442-443.  
27 511  
28 512 Cottrell MT, Mannino A & Kirchman D (2006) Aerobic anoxygenic phototrophic bacteria in  
29 513 the mid-Atlantic bight and the north Pacific gyre. *Appl Environ Microbiol* **72**:557-564.  
30 514  
31 515 Cottrell MT, Michelou VK, Nemcek N, Ditullio G & Kirchman DL (2008) Carbon cycling by  
32 516 microbes influenced by light in the Northeast Atlantic Ocean. *Aquat Microb Ecol* **50**:239-250.  
33 517  
34 518 Curtis TP & Sloan WT (2002) Prokaryotic diversity and its limits: microbial community  
35 519 structure in nature and implications for microbial ecology. *Curr Op Microbiol* **7**:221-226.  
36 520  
37 521 Eiler A (2006) Evidence for the ubiquity of mixotrophic bacteria in the upper Ocean:  
38 522 implications and consequences. *Appl Environ Microbiol* **72**:7431-7437.  
39 523  
40 524 Eilers H, Pernthaler J, Peplies J, Glöckner FO, Gerdts G & Amann R (2001) Isolation of  
41 525 novel pelagic bacteria from the German bight and their seasonal contributions to surface  
42 526 picoplankton. *Appl Environ Microbiol* **67**:5134-5142.  
43 527  
44 528 Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the  
45 529 author. Department of Genetics, University of Washington, Seattle.  
46 530  
47 531 Fuchs BM, Spring S, Teeling H, Quast C, Wulf J, Schattenhofer M, Yan S, Ferriera S,  
48 532 Johnson JM, Glöckner FO & Amann R (2007) Characterization of a marine  
49 533 gammaproteobacterium capable of aerobic anoxygenic photosynthesis. *Proc Natl Acad Sci*  
50 534 *USA* **104**:2891-2896.  
51 535  
52 536 Fuhrman JA (2009) Microbial community structure and its functional implications. *Nature*  
53 537 459: 193-199.

- 1  
2  
3 538  
4 539 Garczarek L, Dufresne A, Rousvoal S, West NJ, Mazard S, Marie D, Claustre H, Raimbault  
5 540 P, Post AF, Scanlan D & Partensky F (2007) High vertical and low horizontal diversity of  
6 541 *Prochlorococcus* ecotypes in the Mediterranean Sea in summer. *FEMS Microbiol Ecol*  
7 542 **60**:189-206.  
8 543  
9 544 Gascard JC & Richez C (1985) Water masses and circulation in the Western Alboran Sea and  
10 545 in the Straits of Gibraltar. *Prog Oceanog* **15**: 175–216.  
11 546  
12 547 Ghiglione JF, Palacios C, Marty JC, Mével G, Labrune C, Conan P, Pujo-Pay M, Garcia N &  
13 548 Goutx M (2008) Role of environmental factors for the vertical distribution (0-1000 m) of  
14 549 marine bacterial communities in the NW Mediterranean Sea. *Biogeosciences Discuss* **5**:2131-  
15 550 2164.  
16 551  
17 552 González JM, Simó R, Massana R, Covert JS, Casamayor EO, Pedrós-Alió C & Moran MA  
18 553 (2000) Bacterial community structure associated with a dimethylsulfoniopropionate-  
19 554 producing North Atlantic algal bloom. *Appl Environ Microbiol* **66**:4237–4246.  
20  
21  
22 555 Hammer Ø, Harper DAT & Ryan PD (2001) PAST: Paleontological Statistics Software  
23 556 Package for Education and Data Analysis. *Palaeontologia Electronica* **4**: 9pp.  
24  
25  
26 557 Hill TCJ, Walsh KA, Harris JA & Moffett BF (2003) Using ecological diversity measures  
27 558 with bacterial communities. *FEMS Microbiol Ecol* **43**:1-11.  
28 559  
29 560 Hu Y, Du H, Jiao N & Zeng Y (2006) Abundant presence of the gamma-like proteobacterial  
30 561 *pufM* gene in oxic seawater. *FEMS Microbiol Lett* **263**: 200–206.  
31 562  
32 563 Hughes JB, Hellmann JJ, Ricketts TH & Bohannan BJM (2001) Counting the uncountable:  
33 564 statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* **67**:4399-  
34 565 4406.  
35 566  
36 567 Imhoff JF (2001) True marine and halophilic anoxygenic phototrophic bacteria. *Arch*  
37 568 *Microbiol* **176**:243-254.  
38 569  
39 570 Jiao N, Zhang Y, Zeng Y, Hong N, Liu R, Chen F & Wang P (2007) Distinct distribution  
40 571 pattern of abundance and diversity of aerobic anoxygenic phototrophic bacteria in the global  
41 572 ocean. *Environ Microbiol* **9**: 3091–3099.  
42 573  
43 574 Koblížek M, Mašín M, Ras J, Poulton AJ & Prášil O (2007) Rapid growth rates of aerobic  
44 575 anoxygenic prototrophs in the Ocean. *Envir Microbiol* **9**: 2401–2406.  
45 576  
46 577 Kolber ZS, Plumley FG, Lang AS, Beatty JT, Blankenship RE, VanDover CL, Vetriani C,  
47 578 Koblížek M, Rathgeber C & Falkowski PG (2001) Contribution of aerobic photoheterotrophic  
48 579 Bacteria to the carbon cycle in the ocean. *Science* **292**:2492-2495.  
49 580  
50 581 Lami R, Cottrell MT, Ras J, Ulloa O, Obernosterer I, Claustre H, Kirchman DL & Lebaron P  
51 582 (2007) High abundances of aerobic anoxygenic photosynthetic bacteria in the south Pacific  
52 583 Ocean. *Appl Environ Microbiol* **73**:4198-4205.  
53 584  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 585 Lee S & Furhman JA (1991) Spatial and temporal variation of natural bacterioplankton  
4 586 assemblages studied by total genomic DNA cross-hybridization. *Limnol Oceanogr* **36**:1277-  
5 587 1287.  
6 588  
7 589 Legendre P & Legendre L (1998) Numerical Ecology, 2<sup>nd</sup> English ed. Elsevier, 853pp.  
8 590  
9 591 Letunic I & Bork P (2006) Interactive tree of life (iTOL): an online tool for phylogenetic tree  
10 592 display and annotation. *Bioinformatics* **23**:127-128.  
11 593  
12 594 Lucea A, Duarte CM, Agustí S & Søndergaard M (2003) Nutrient (N, P and Si) and carbon  
13 595 partitioning in the stratified NW Mediterranean. *J Sea Res* **49**:157-170.  
14 596  
15 597 Ludwig W, Strunk O, Westram R, Richter L, Meier H *et al.* (2004) ARB: a software  
16 598 environment for sequence data. *Nucleic Acid Res* **32**:1363-1371.  
17 599  
18 600 Manca B, Burca M, Giorgetti C, Coatanoan C, Garcia MJ & Iona A (2004) Physical and  
19 601 biochemical averaged vertical profiles in the Mediterranean regions: an important tool to trace  
20 602 the climatology of water masses and to validate incoming data from operational  
21 603 oceanography. *J Mar Syst* **48**: 83–116.  
22  
23 604 Mantel N (1967) The detection of disease clustering and a generalized regression approach.  
24 605 *Cancer Research* **27**: 209–220.  
25  
26 606 Marie D, Zhu F, Balagué V, Ras J & Vaulot D (2006) Eukaryotic picoplankton communities  
27 607 of the Mediterranean Sea in summer assessed by molecular approaches (DGGE, TTGE,  
28 608 QPCR). *FEMS Microbiol Ecol* **55**: 403–415.  
29 609  
30 610 Mašín M, Zdun A, Ston-Egiert J, Nausch M, Labrenz M, Moulisová V & Koblížek M (2006)  
31 611 Seasonal changes and diversity of aerobic anoxygenic phototrophs in the Baltic Sea. *Aquat*  
32 612 *Microbiol Ecol* **45**: 247–254.  
33 613  
34 614 Moutin T & Raimbault P (2002) Primary production, carbon export and nutrients availability  
35 615 in western and eastern Mediterranean Sea in early summer 1996 (MINOS cruise). *J Mar Syst*  
36 616 **33-34**:273-288.  
37 617  
38 618 Mullins TD, Britschgi TB, Krest RL & Giovannoni SJ (1995) Genetic comparisons reveal the  
39 619 same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities.  
40 620 *Limnol Oceanogr* **40**:148-158.  
41 621  
42 622 Murray AE, Preston CM, Massana R, Taylor LT, Blakis A, Wu K & DeLong EF (1998)  
43 623 Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters off  
44 624 Anvers Island, Antarctica. *Appl Environ Microbiol* **64**:2585–2595.  
45 625  
46 626 Oz A, Sabehi G, Koblížek M, Massana R & Bèjà O (2005) *Roseobacter* like bacteria in Red  
47 627 and Mediterranean Sea aerobic anoxygenic photosynthetic populations. *Appl Environ*  
48 628 *Microbiol* **71**:344–353.  
49 629  
50 630 Page KA, Connon SA & Giovannoni SJ (2004) Representative freshwater bacterioplankton  
51 631 isolated from Crater Lake, Oregon. *Appl Environ Microbiol* **70**: 6542-6550.  
52 632  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62

- 1  
2  
3 633 Ramírez-Flandes S & Ulloa O (2008) Bosque: Integrated phylogenetic analysis software.  
4 634 *Bioinformatics* **24**:2539-2541.  
5 635
- 6  
7 636 Rappé MS, Kemp PF & Giovannoni SJ (1997) Phylogenetic diversity of marine coastal  
8 637 picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North  
9 638 Carolina. *Limnol Oceanogr* **42**:811-826.  
10 639
- 11 640 Shapiro SS & Wilk MB (1965) An analysis of variance test for normality (complete samples).  
12 641 *Biometrika* **52**: 591–611.  
13 642
- 14 643 Sieracki ME, Gilg IC, Thier EC, Poulton NJ & Goericke R (2006) Distribution of planktonic  
15 644 aerobic anoxygenic photoheterotrophic bacteria in the northwest Atlantic. *Limnol Oceanogr*  
16 645 **51**: 38–46.  
17 646
- 18 647 Smith B & Wilson JB (1996) A consumers's guide to evenness indices. *Oikos* **76**:70-82.  
19 648
- 20 649 Spring S, Lünsdorf H, Fuchs BM & Tindall BJ (2009) The Photosynthetic Apparatus and Its  
21 650 Regulation in the Aerobic Gammaproteobacterium *Congregibacter litoralis* gen. nov., sp.  
22 651 nov. *Plos One* **4**: 1-23.  
23 652
- 24 653 Stephens MA (1974) EDF Statistics for Goodness of Fit and Some Comparisons. *J Amer Stat*  
25 654 *Assoc* **69**: 730–737.  
26 655
- 27 656 Suyama T, Shigematsu T, Suzuki T, Tokiwa Y, Kanagawa T, Nagashima KVP & Hanada S  
28 657 (2002) Photosynthetic apparatus in *Roseateles depolymerans* 61A is transcriptionally induced  
29 658 by carbon limitation. *Appl Environ Microbiol* **68**:1665-1673.  
30 659
- 31 660 Suzuki MT, Preston CM, Chavez FP & DeLong EF (2001) Quantitative mapping of  
32 661 bacterioplankton populations in seawater: field tests across an upwelling plume in Monterey  
33 662 Bay. *Aquat Microbiol Ecol* **24**:117–127.  
34 663
- 35 664 Van Wambeke F, Christaki U, Giannakourou A, Moutin T & Souvemerzoglou K (2002)  
36 665 Longitudinal and vertical trends of bacterial limitation by phosphorus and carbon in the  
37 666 Mediterranean Sea. *Microb Ecol* **43**: 119–133.  
38 667
- 39 668 Waidner LA & Kirchman DL (2007) Aerobic anoxygenic phototrophic bacteria attached to  
40 669 particles in turbid waters of the Delaware and Chesapeake estuaries. *Appl Environ Microbiol*  
41 670 **73**: 3936–3944.  
42 671
- 43 672 Waidner LA & Kirchman DL (2008) Diversity and distribution of ecotypes of the aerobic  
44 673 anoxygenic phototrophy gene *pufM* in the Delaware estuary. *Appl Environ Microbiol*  
45 674 **74**:4012-4021.  
46 675
- 47 676 Ward JH (1963) Hierarchical Grouping to optimize an objective function. *J Amer Stat Assoc*  
48 677 **58**: 236-244.  
49 678
- 50 679 Wohl DL, Arora S & Gladstone JR (2004) Functional redundancy supports biodiversity and  
51 680 ecosystem function in a closed and constant environment. *Ecology* **85**:1534-1540.  
52 681



- 1  
2  
3 682 Yurkov V & Csotonyi JT (2009) New light on aerobic anoxygenic phototrophs. *Adv*  
4 683 *Photosynth Res* **28**: 31–55.  
5 684  
6  
7 685 Yurkov V & van Gemerden H (1993) Impact of light/dark regime on growth rate, biomass  
8 686 formation and bacteriochlorophyll synthesis in *Erythromicrobium hydrolyticum*. *Arch*  
9 687 *Microbiol* **159**:84-89.  
10 688  
11 689 Yutin N, Béjà O & Suzuki M (2008) The use of denaturing gradient gel electrophoresis with  
12 690 fully degenerate pufM primers to monitor aerobic anoxygenic phototrophic assemblages  
13 691 *Limnol Oceanogr Methods* **6**:427–440.  
14 692  
15 693 Yutin N, Suzuki MT & Béjà O (2005) Novel primers reveal wider diversity among marine  
16 694 aerobic anoxygenic phototrophs. *Appl Environ Microbiol* **71**:8958–8962.  
17 695  
18 696 Yutin N, Suzuki MT, Teeling H, Weber M, Venter JC, Rusch DB & Béjà O (2007) Assessing  
19 697 diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of  
20 698 the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes.  
21 699 *Environ Microbiol* **9**: 1464–1475.  
22 700  
23 701 Zeng YH, Chen XH & Jiao NZ (2007) Genetic diversity assessment of anoxygenic  
24 702 photosynthetic bacteria by distance-based grouping analysis of pufM sequences. *Lett Appl*  
25 703 *Microbiol* **45**:639-645.  
26 704  
27 705 Zhang Y & Jiao N (2009) Roseophage RDJL□1, infecting the aerobic anoxygenic  
28 706 phototrophic bacterium *Roseobacter denitrificans* OCh114. *Appl Environ Microbiol* **75**: 1745-  
29 707 1749.  
30 708  
31 709  
32 710  
33 711  
34 712  
35 713  
36 714  
37 715  
38 716  
39 717  
40 718  
41 719  
42 720  
43 721  
44 722  
45 723  
46 724  
47 725  
48 726  
49 727  
50 728  
51 729  
52 730  
53 731  
54  
55  
56  
57  
58  
59  
60

**Table 1.** Properties of the distribution of phylotypes in clone libraries.  $H'$ : Shannon-Weiner index, ACE: Abundance-base Coverage Estimator. The richness estimator  $S_{Chao1}$  was computed along with log-linear 95% confidence intervals.

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

## 782 **Figure Legends**

1  
2  
3  
4  
5  
6 783 **Figure 1.** Track of the PROSOPE cruise superimposed on the composite SeaWiFS image of  
7  
8 784 surface chlorophyll *a* concentrations in September 1999. Arrows indicated the stations  
9  
10 785 analyzed in this study. ST#, station number.

11  
12  
13 786 **Figure 2.** Temporal temperature gradient gel electrophoresis (TTGE) analyses of *pufM* gene  
14  
15 787 fragments amplified from DNA samples from the Mediterranean Sea. Hierarchical cluster  
16  
17 788 analysis was performed as indicated in the Experimental procedures section. The dashed line  
18  
19 789 indicates the distance chosen for cluster separation. The number of distinct TTGE bands per  
20  
21 790 sample is indicated between brackets. DCM: deep Chlorophyll maximum.

22  
23  
24  
25 791 **Figure 3.** Canonical correspondence analysis (CCA) performed using relative intensity peak  
26  
27 792 of the TTGE bands. Samples and environmental parameters are coded as in Fig. S2.

28  
29  
30 793 **Figure 4.** *pufM* phylogenetic tree showing inferred phylogenetic relationships of *pufM* gene  
31  
32 794 sequences cloned from the Mediterranean samples. The tree is based on a Neighbor-Joining  
33  
34 795 (NJ) tree to which short sequences were added by ARB\_PARISMONY. Sequences used to  
35  
36 796 perform NJ tree are marked in bold and grey circles on nodes represent confidence values >  
37  
38 797 50% for branches found in the initial NJ tree. Color ranges highlight the different groups  
39  
40 798 defined by Yutin *et al.* (2007). Sequences retrieved in this study are indicated by filled  
41  
42 799 triangles. The multi-value bar charts represent the relative frequencies of the corresponding  
43  
44 800 OTU in the different clone libraries. Colors used to represent the clone libraries are indicated  
45  
46 801 at the top of the figure. Green and blue crosses indicate that the corresponding OTU was  
47  
48 802 found to be dominant (>80%) at meso-eutrophic and oligotrophic conditions, respectively.

49  
50  
51 803 **Figure 5.** Distribution of the AAP phylogroups along the PROSOPE transects based on their  
52  
53 804 relative proportion in the clone libraries.

54  
55  
56  
57  
58 805

59  
60 806

1  
2  
3 807  
4  
5 808  
6  
7  
8 809  
9  
10 810  
11  
12 811  
13  
14  
15 812  
16  
17 813  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42 814  
43  
44 815  
45  
46 816  
47  
48  
49 817  
50  
51  
52 818  
53  
54 819  
55  
56 820  
57  
58 821  
59  
60 822

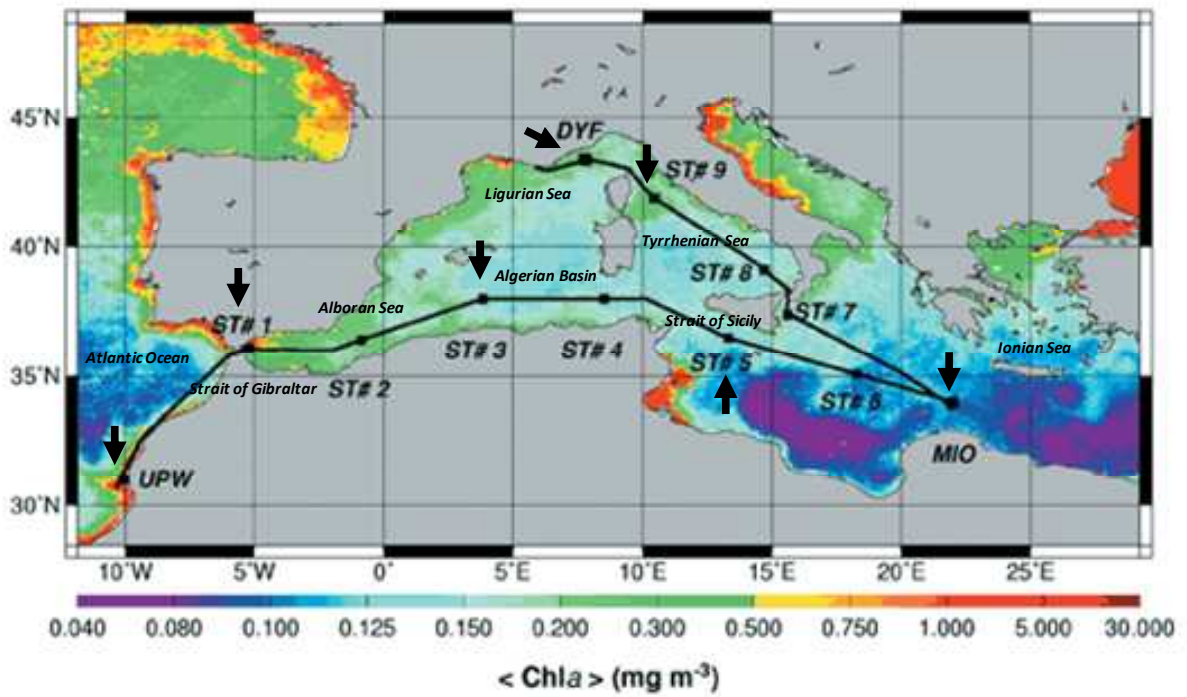


Figure 1

823

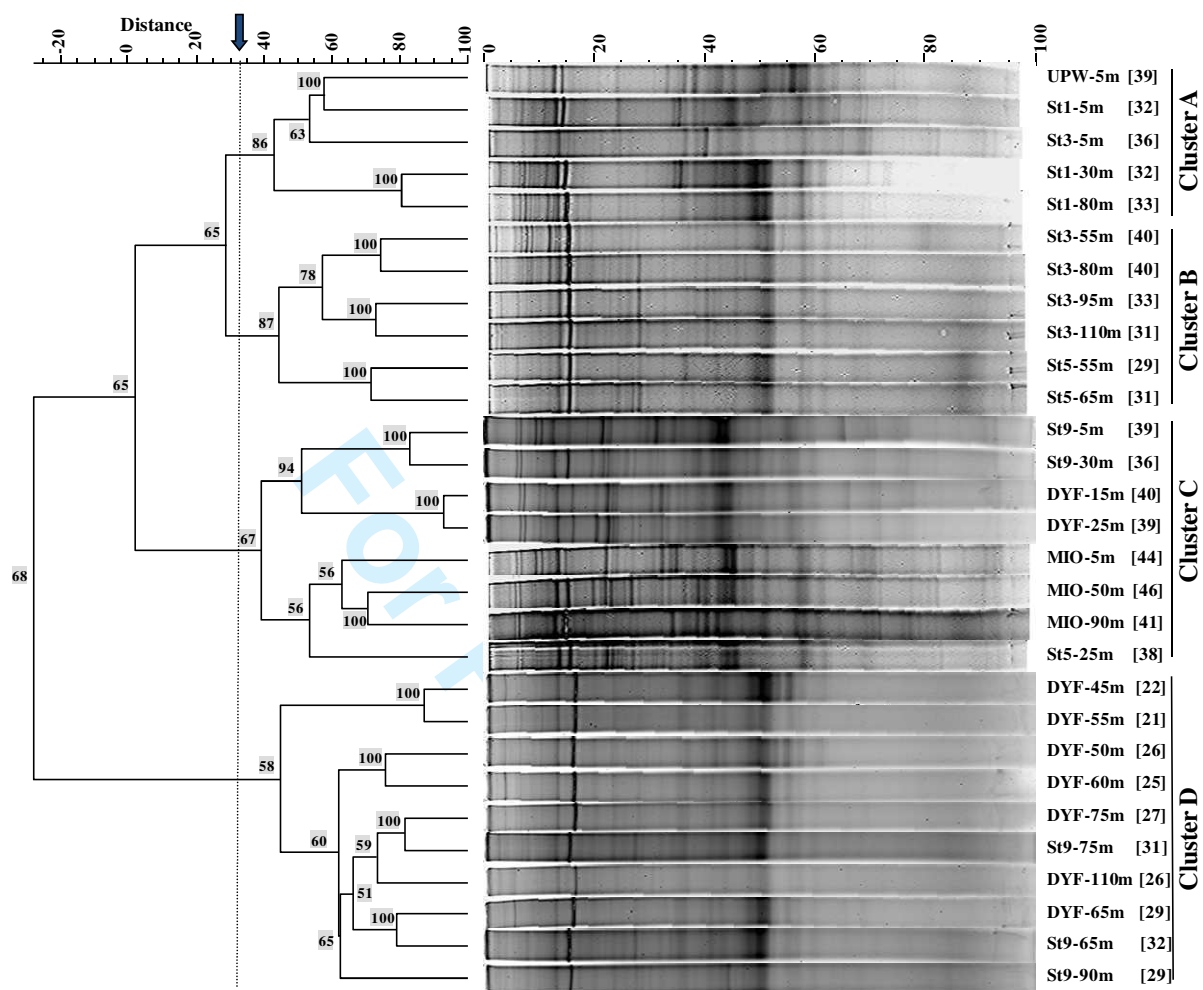


Figure 2

824

825

826

827

828

829

830

831

832

833

834

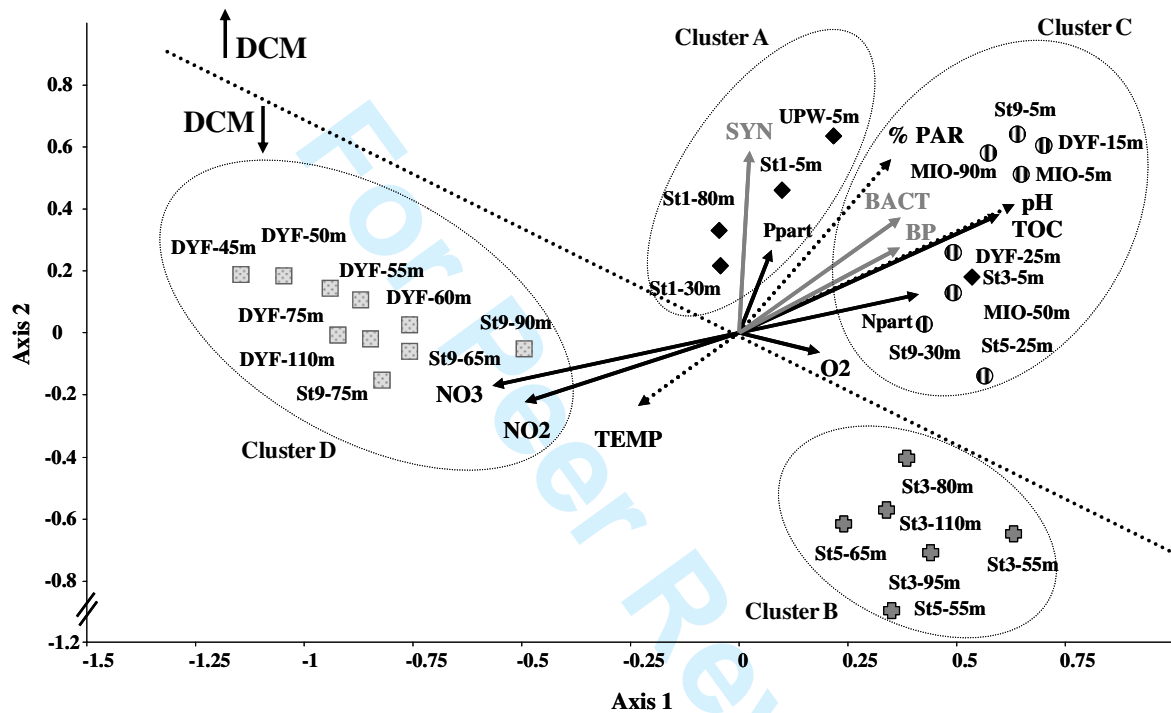
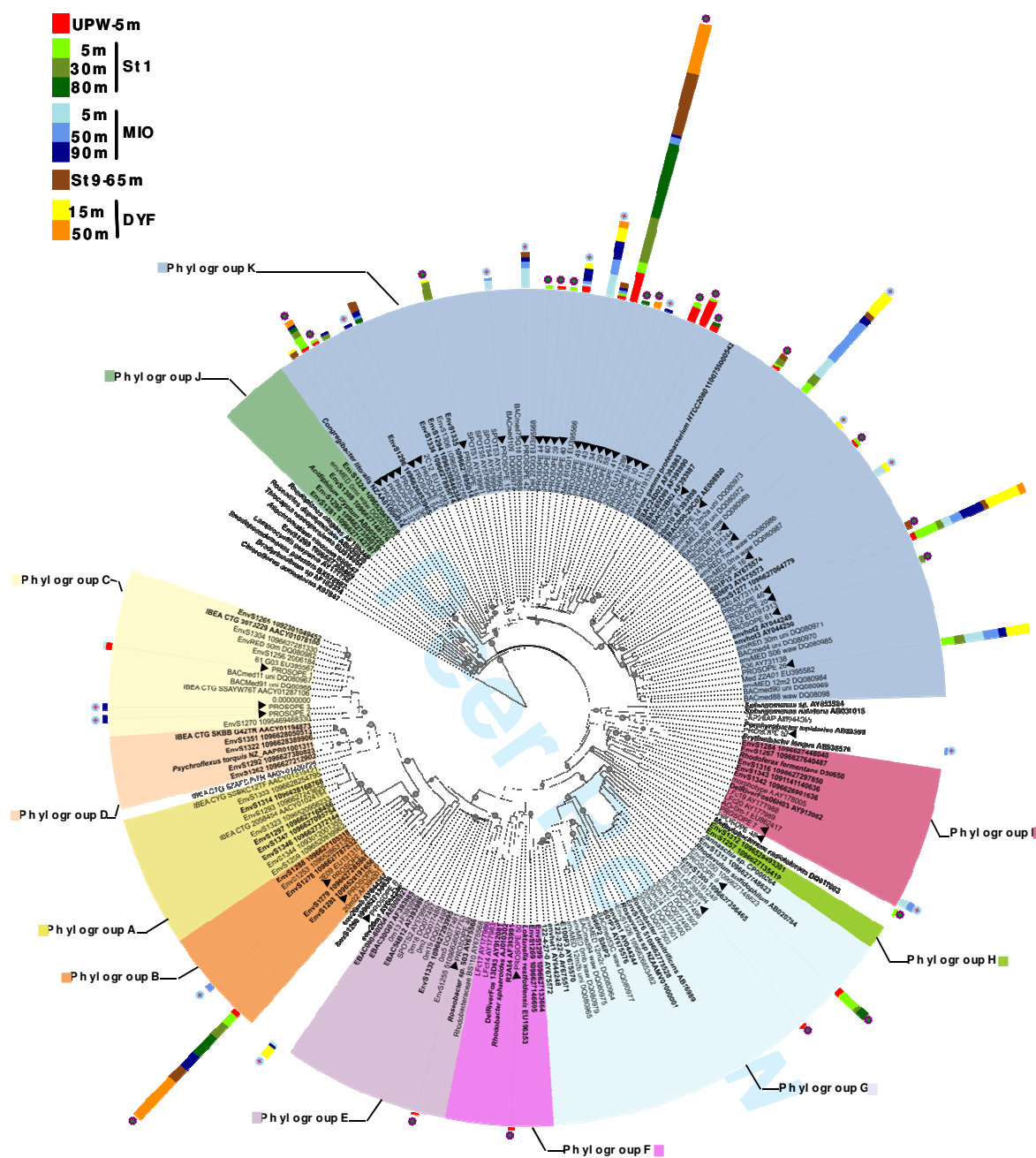


Figure 3

851



852

853

854

Figure 4

855

856

857

858

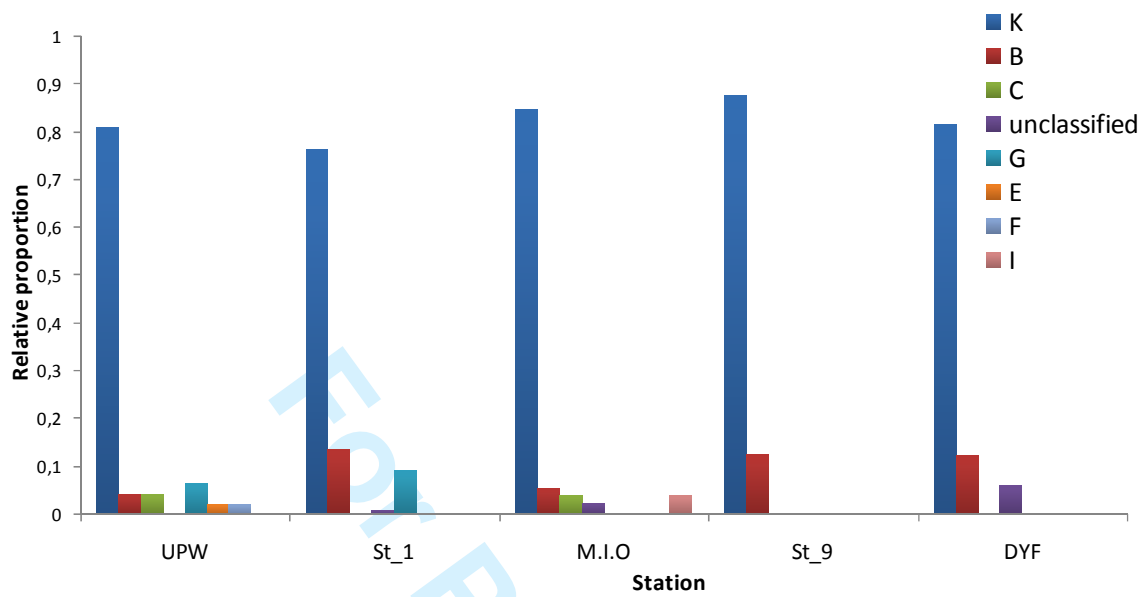


Figure 5



## 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<sub>2</sub> (concentration in dissolved oxygen, ml.l<sup>-1</sup>), SAL (salinity, ‰), Depth (m). Grey arrows represent variables characterizing the “biological activity”: PROC (*Prochlorococcus* cell concentration, cell.ml<sup>-1</sup>), SYN (*Synechococcus* cell concentration, cell.ml<sup>-1</sup>), BACT (bacteria cell concentration, cell.ml<sup>-1</sup>), EUK (eukaryotes cell concentration, cell.ml<sup>-1</sup>), BP (bacterial production, ng C.l<sup>-1</sup>.h<sup>-1</sup>), Chl*a* (chlorophyll *a*, µg. l<sup>-1</sup>), PIGM (accessory pigments, µg.l<sup>-1</sup>). Black dashed arrows represent variables characterizing the “trophic status”: TOC (total organic carbon, µM), NO<sub>3</sub> (nitrate,

1  
2  
3 26  $\mu\text{M}$ ),  $\text{PO}_4$  (phosphate,  $\mu\text{M}$ ),  $\text{NO}_2$  (nitrite,  $\mu\text{M}$ ), N part (particulate N,  $\mu\text{M}$ ), P part (particulate  
4  
5 27 P,  $\mu\text{M}$ ). Samples collected above and below the DCM are indicated by grey and black squares  
6  
7  
8 28 respectively. Component 1, 2 and 3 (component 3, not shown) represent 34%, 23% and 12%  
9  
10 29 of the total variance, respectively.  
11  
12  
13 30  
14

15 31 **Figure S3.** Profiles of chlorophyll *a* (Chl*a*) (solid) and PAR (dashed) with depth. Arrows in  
16  
17 32 the Chl*a* profiles indicate sample depths. Dashed line indicates the limit of the euphotic layer  
18  
19 33 ( $Z_{1\%}$ ).  
20  
21  
22 34

23  
24 35 **Figure S4.** Multidimensional scaling plot (MDS) of AAP populations based on TTGE and  
25  
26 36 cloning-sequencing of *pufM* genes. This plot corresponds to a two-dimensional visualization  
27  
28 37 of the Dice distance matrix. “(c)” refers to data obtained from clone libraries analyses.  
29  
30  
31 38  
32  
33 39  
34  
35 40  
36  
37 41  
38  
39 42  
40  
41 43  
42  
43 44  
44  
45 45  
46  
47 46  
48  
49 47  
50  
51 48  
52  
53 49  
54  
55 50  
56  
57 51  
58  
59 52  
60  
60 53  
54  
55 54  
56  
57 55  
58  
59 56  
60  
60 57  
58  
59 58  
60  
60 59  
60  
60 60

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	-	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	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	-	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	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	
<b>TOTAL</b>		<b>47</b>	<b>50</b>	<b>39</b>	<b>30</b>	<b>41</b>	<b>43</b>	<b>49</b>	<b>24</b>	<b>45</b>	<b>20</b>				

Table S.1

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

111  
112  
113

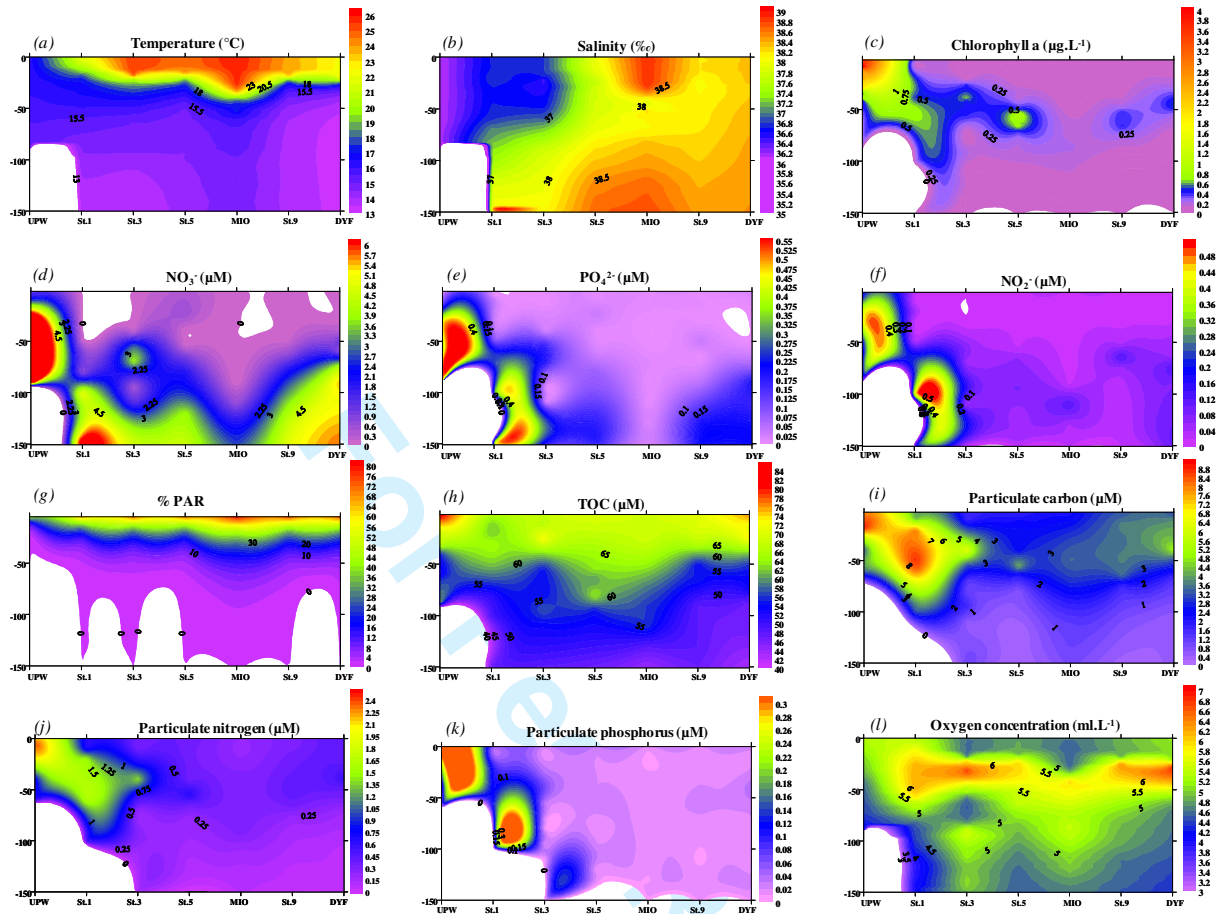


Figure S.1

114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136

137  
138  
139

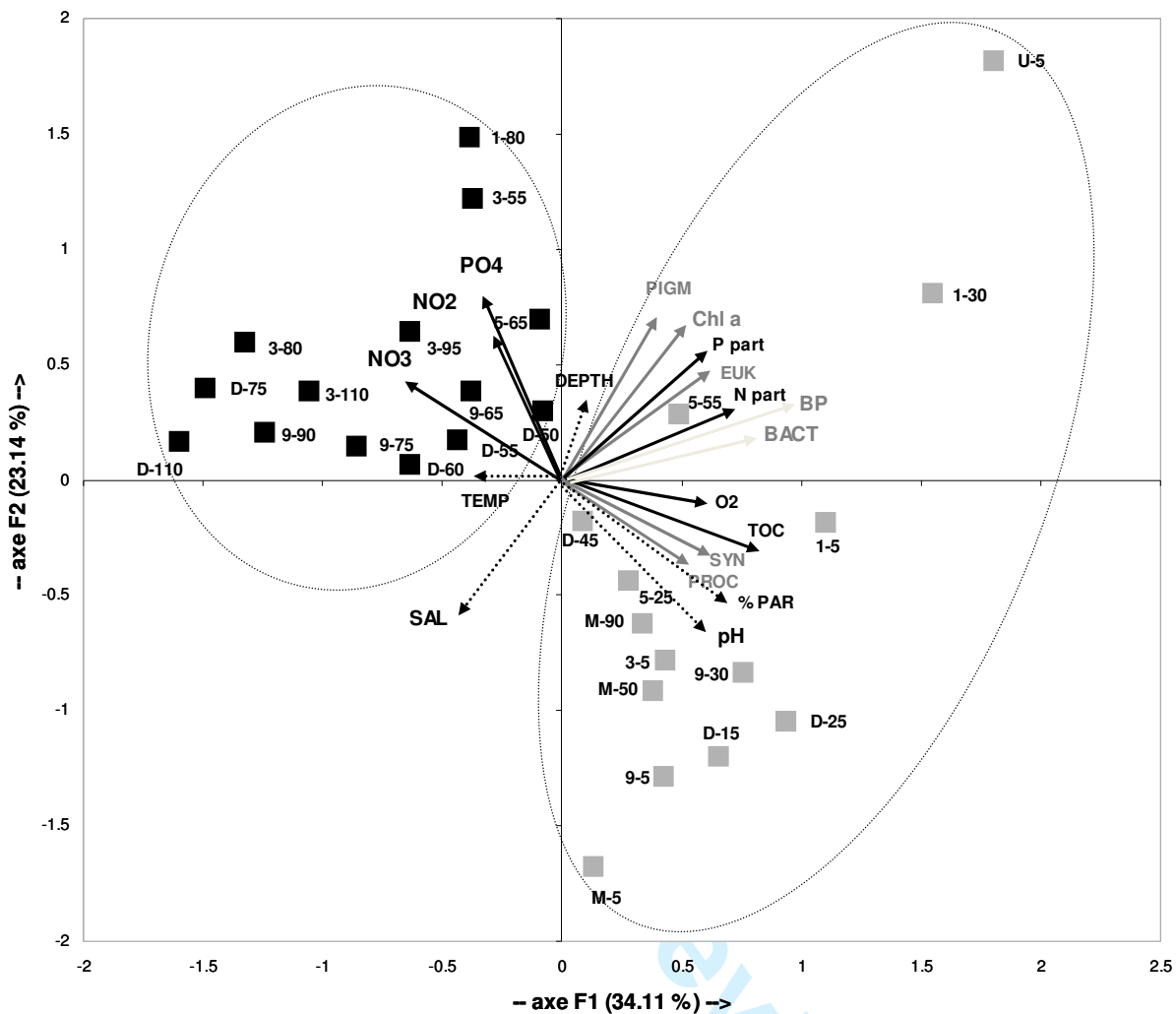


Figure S.2

140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159

1  
2  
3 160  
4 161  
5 162  
6 163  
7 164  
8 165  
9 166  
10 167  
11 168  
12 169

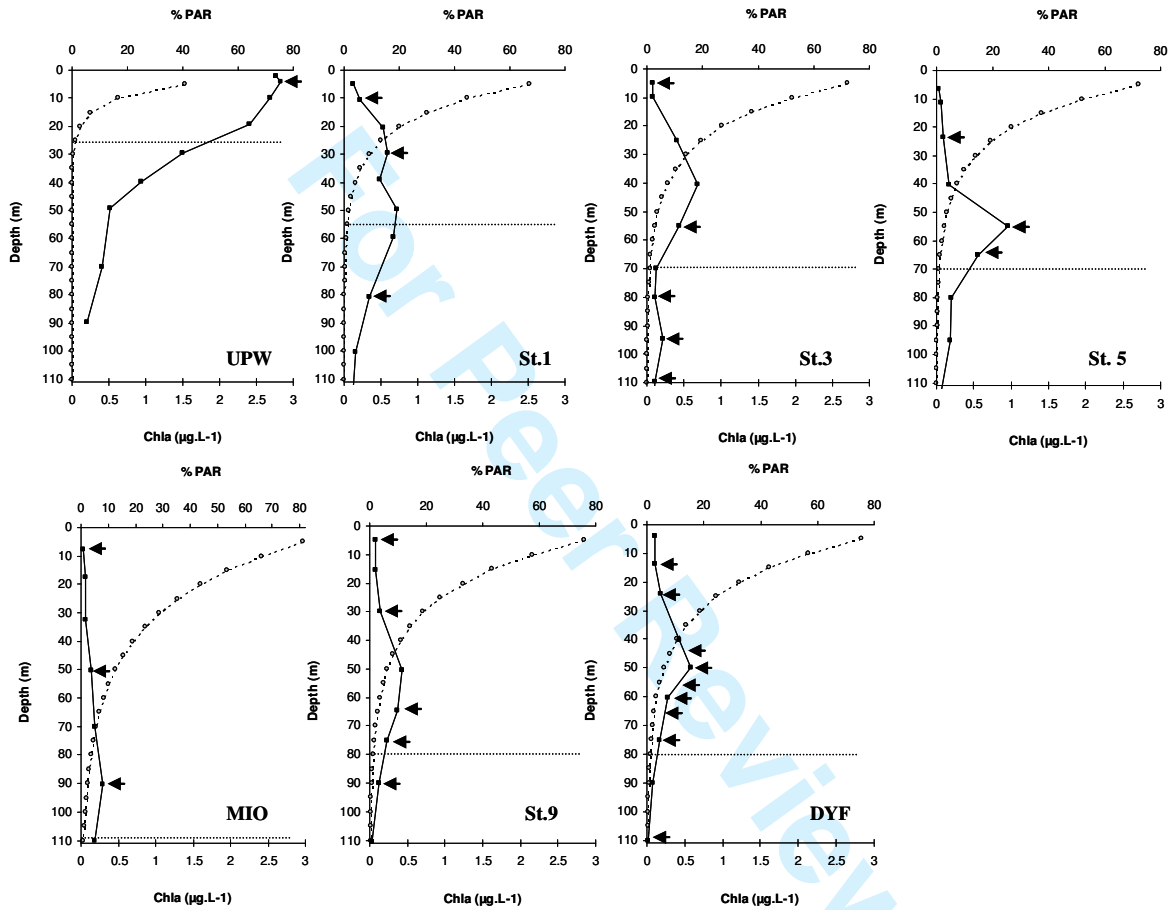


Figure S.3

170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185

186  
187  
188  
189

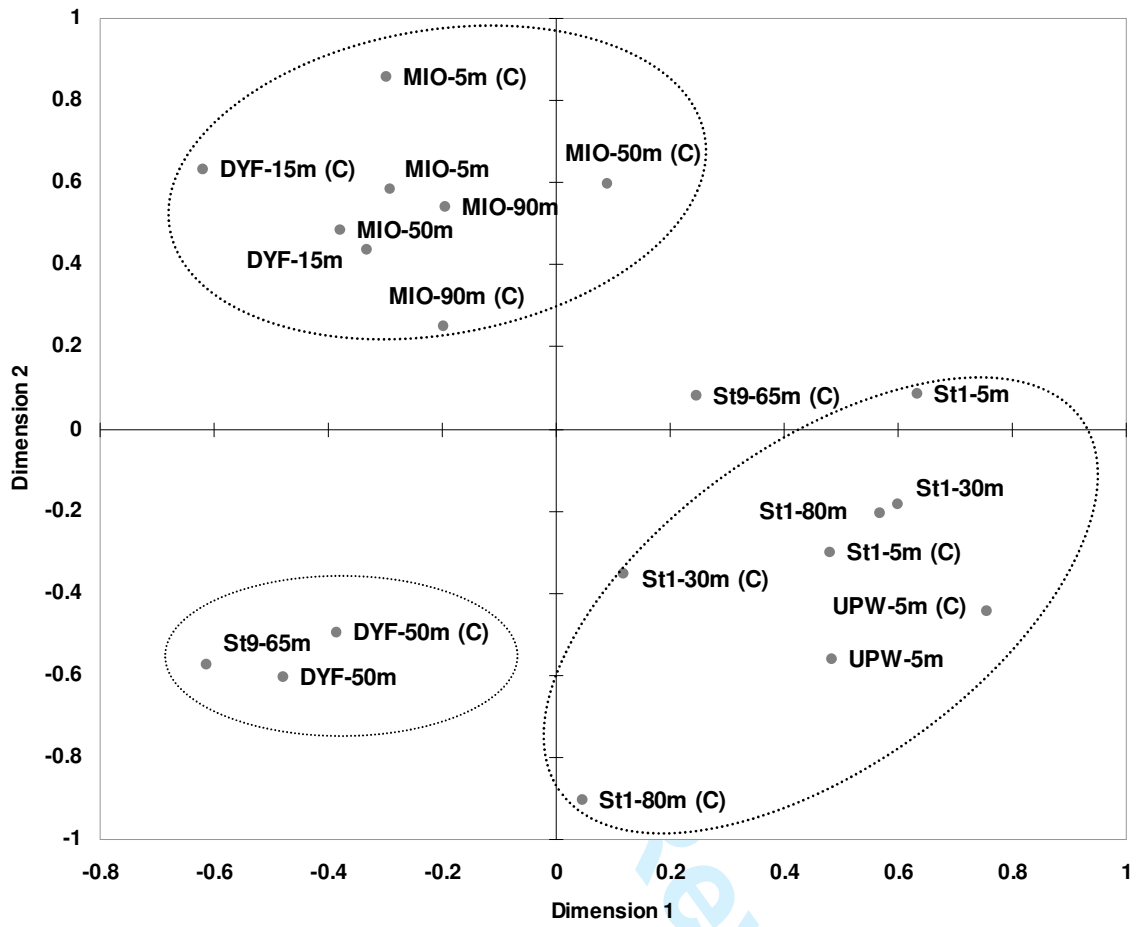


Figure S.4

190  
191  
192  
193  
194  
195  
196  
197  
198

199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210