Seasonality of archaeal proteorhodopsin and associated Marine Group IIb ecotypes (*Ca.* Poseidoniales) in the North Western Mediterranean Sea

Olivier Pereira^{1,5} · Corentin Hochart¹ · Dominique Boeuf² · Jean Christophe Auguet³ · Didier Debroas⁴ · Pierre E. Galand¹

Received: 23 July 2020 / Revised: 9 November 2020 / Accepted: 18 November 2020 / Published online: 7 December 2020

© The Author(s), under exclusive licence to International Society for Microbial Ecology 2020

Abstract

The Archaea Marine Group II (MGII) is widespread in the world's ocean where it plays an important role in the carbon cycle. Despite recent discoveries on the group's metabolisms, the ecology of this newly proposed order (*Candidatus* Poseidoniales) remains poorly understood. Here we used a combination of time-series metagenome-assembled genomes (MAGs) and high-frequency 16S rRNA data from the NW Mediterranean Sea to test if the taxonomic diversity within the MGIIb family (*Candidatus* Thalassarchaeaceae) reflects the presence of different ecotypes. The MAGs' seasonality revealed a MGIIb family composed of different subclades that have distinct lifestyles and physiologies. The vitamin metabolisms were notably different between ecotypes with, in some, a possible link to sunlight's energy. Diverse archaeal proteorhodopsin variants, with unusual signature in key amino acid residues, had distinct seasonal patterns corresponding to changing day length. In addition, we show that in summer, archaea, as opposed to bacteria, disappeared completely from surface waters. Our results shed light on the diversity and the distribution of the euryarchaeotal proteorhodopsin, and highlight that MGIIb is a diverse ecological group. The work shows that time-series based studies of the taxonomy, seasonality, and metabolisms of marine prokaryotes is critical to uncover their diverse role in the ocean.

These authors contributed equally: Olivier Pereira, Corentin Hochart, Dominique Boeuf

Supplementary information The online version of this article (<https://doi.org/10.1038/s41396-020-00851-4>) contains supplementary material, which is available to authorized users.

✉ Pierre E. Galand
pierre.galand@obs-banyuls.fr

¹ Sorbonne Universités, CNRS, Laboratoire d'Ecogéochimie des Environnements Benthiques (LECOB), Observatoire Océanologique, Banyuls sur Mer, France

² Daniel K. Inouye Center for Microbial Oceanography, Research and Education, School of Ocean and Earth Science and Technology, University of Hawai'i at Mānoa, Honolulu, HI, United States, Honolulu, HI 96822, USA

³ MARBEC, Université de Montpellier, CNRS, Ifremer, IRD, Montpellier, France, Montpellier, France

⁴ Université Clermont Auvergne, CNRS, Laboratoire Microorganismes: Genome et Environnement, 63000 Clermont-Ferrand, France

⁵ Present address: Shenzhen Key Laboratory of Marine Archaea Geo-Omics, Southern University of Science and Technology, Shenzhen, China

Introduction

Planktonic archaea were first detected in the sea almost 30 years ago [1, 2]. Since then, the extent of the diversity, distribution, and potential function of the third domain of life in the ocean has been extensively studied, but many unknowns remain. Among marine Archaea, the Marine Group II (MGII) is usually more prevalent in surface waters, but not always, and has a global distribution from the poles to tropical seas [3]. Its widespread occurrence suggested that it had an important role in the ocean, but because of the lack of any cultivated representatives, its physiology was largely unknown. The development of metagenomic approaches has, however, allowed a glimpse into the metabolic potential and lifestyle of the MGII [4–10]. Some MGII harbor a proteorhodopsin (PR) gene and could thus be able to use sunlight as energy source [11, 12]. They can potentially degrade large molecules such as lipids, proteins, and polysaccharides [8, 9, 12], and some have genes for motility [6, 7, 12]. The rare experimental tests of the metabolisms of MGII showed the uptake of phytoplankton-derived proteins using stable-

isotope probing [13], and stimulated growth in the presence of whole cells of the picoeukaryote *Micromonas* [9]. The potential metabolisms detected in MGII suggest that the group has an important role for the carbon cycle of the ocean [14].

Among the genes present in MGII, the archaeal PR has some interesting features. Analyses of key amino acid residues show that both green- and blue-light tuned PR types are present within the group [6, 7], with distinct geographic and vertical distributions [10]. The archaeal PR phylogeny is very different from the inferred organismal phylogeny [6], which indicates extensive and ongoing horizontal gene transfers [11, 15]. Although the rhodopsins found in this pelagic archaeal group are predicted to act as a counter-gradient protonic pumps, like canonical PRs, the fact that some genes have a very unusual amino acid sequences [4], and that some MGII can grow in the dark [9], begs additional examination of the genomic diversity and the ecology of the archaeal PRs.

MGII was recently proposed as an order level lineage, *Candidatus* Poseidoniales, separated in two families: MGIIa (*Candidatus* Poseidonaceae) and MGIIb (*Candidatus* Thallasarchaeaceae) [6]. The two families have different spatial and temporal distributions, with MGIIb generally found in deeper waters [4, 10, 16–18], and in surface waters during winter in temperate seas, while MGIIa is predominant in summer [19–21]. These different distributions suggest that the two groups have different ecologic niches corresponding to distinct lifestyles [3]. The specific metabolic traits distinguishing MGIIa and MGIIb were recently described more in details [4, 6, 7, 21, 22]. For example, members of the MGIIa have a larger genome size [7, 21], they are all motile and have a catalase gene possibly protecting the cell from oxidative damage [22], while MGIIb have the potential for assimilatory sulfate reduction [4, 21]. The fact that MGIIa and MGIIb represent different taxonomic and ecological entities is now clear, but the question remains as whether the large phylogenetic diversity seen within each family, and especially within MGIIb, reflects the presence of additional ecotypes.

The main goal of this study was to test for the presence of different ecotypes within the diverse MGIIb family. In addition, we wanted to study if different archaeal PR groups reflect different ecological strategies within MGII archaea. To do so, we used a metagenomic time series gathered over 3 years at the SOLA station in the Bay of Banyuls sur Mer in the North Western Mediterranean Sea [23]. We reconstructed metagenome-assembled genomes (MAGs) and monitored their seasonal dynamics monthly. We also used a high-frequency (twice a week) 16S rRNA amplicon sampling to track the fine succession of ecotypes.

Materials and methods

Sampling and metagenome sequencing

For metagenome data, surface seawater (3 m) was collected monthly from January 2012 to February 2015 (40 samples) at the SOLA station (42°31'N, 03°11'E) in the Bay of Banyuls sur Mer (France) in the northwestern Mediterranean as described previously [23]. Briefly, a volume of 5 L was prefiltered through 3- μ m pore-size polycarbonate filters (Millipore, Billerica, MA, USA), and the microbial biomass was collected on 0.22- μ m pore-size GV Sterivex cartridges (Millipore). The physicochemical parameters were provided by the Service d'Observation en Milieu Littoral (SOMLIT). Metagenomes were sequenced on a HiSeq 2500 “High-Output” paired-end run (2 $\times$ 100 bp) (Illumina). Sequencing produced a total of 2,984,444,036 reads that were archived in the EBI repository under accession number PRJEB26919.

Gene catalog

A gene catalog was built as described earlier [23]. Briefly, for each metagenome, high-quality reads were individually assembled with IDBA-UD [24], and gene prediction was done on contigs $\geq$ 1 kb with MetaGeneAnnotator [25], which generated a total of 6.4 million gene-coding sequences. A catalog of genes was then built by clustering the predicted gene-coding sequences at 95% identity using CD-HIT (v. 4.6, [26]). The resulting catalog contained 1,568,213 non-redundant predicted genes for the SOLA site. An abundance matrix of gene-coding sequences was then built by mapping the metagenomic reads against the predicted SOLA gene catalog using the SOAPaligner [27]. The abundance matrix was normalized to the gene length and per million of reads. The 16S rRNA gene contigs were annotated against the SILVA database (v.128).

Metagenomic co-assemblies and binning

The 40 metagenomes were first compared to each other with the Commet software [28] to assess their pairwise similarity. The method allows an all-against-all comparison of the non-assembled reads based on shared k-mers. The clustering of metagenomes according to their k-mer similarity separated samples in four groups that corresponded to the four seasons: winter, spring, summer, and autumn (Supplementary Fig. 1). We then co-assembled all reads that belonged to a same seasonal group using MEGAHIT (v.1.1.1) with the ‘meta-sensitive’ preset parameter [29]. Contig abundance profiles were produced by mapping back short reads to the co-assemblies using the mem algorithm

from BWA (v.0.7.12) [30]. We then clustered contigs >2.5 kbp with the automatic binning algorithm CONCOCT (v.1.0.0) with a maximal number of clusters set to 100 to minimize the ‘fragmentation error’ [31]. Finally, we manually binned each CONCOCT cluster using the anvi’o interactive interface to produce MAGs [32].

Identification and selection of MAGs

We used CheckM to determine the completeness and degree of contamination of the bins, and all bins with >50% completeness and <10% contamination were defined as MAGs. The taxonomic classification of the MAGs was done with GTDB-Tk (v0.3.2) and GTDB release 82 [33]. 16S rRNA gene were detected using nhmmer (HMMER v.3.2.1) [34] and taxonomically classified against the SILVA database (v.128) using BLAST+ (v.2.2.31) [35]. Only MAGs affiliated as MGII were kept for this study.

Gene prediction was done with prodigal (v.2.6.3, [36] and all the MAG’s predicted proteins were annotated against the KEGG database using GhostKOALA (‘genus_prokaryotes + family_eukaryotes’, v.2.2) [37].

A pangenomic investigation was conducted on the MGIIb subclades O2, O3, O5, and WHARN using anvi’o workflow.

Phylogenetic assessment

We inferred a maximum likelihood tree of MGII archaea based on multiple sequence alignment of 122 single-copy marker proteins as detailed by the Genome Taxonomy Database (GTDB; <http://gtdb.ecogenomic.org/>). Each MAGs were searched for the 122 marker genes using HMMER (v.3.2.1). The gene sequences were aligned with MUSCLE [38], alignments were concatenated and a phylogenomic tree was constructed using FastTreeDbl (v.2.1.10) [39]. The tree included MAGs from this study as well as reference MAGs from earlier studies that had ≥60 single-copy markers [7, 21, 40]. The phylogenetic tree was visualized using the Interactive Tree Of Life webtool (v.5.5.1).

Extraction of euryarchaeal PR sequences

PR homologous sequences were retrieved from the SOLA gene catalog, from the collection of MAGs reconstructed from SOLA metagenomes, and from public databases using the hidden Markov model (HMM) approach [41]. Briefly, an HMM profile was generated from the 7900 manually curated sequences of type-1 rhodopsins stored into the MicRhoDE database [42], and search against the SOLA gene catalog using the HMMER software (v.3.1b2, [34]). Open reading frames from nucleic acid sequences

corresponding to HMM hits were generated using prodigal (v.2.6.3, [36]). The corresponding protein sequences were dereplicated then clustered into PR sequence clusters (PR clusters) at a conservative value of 82% amino acid sequence similarity as used earlier [43]. The clustering was done using CD-HIT (v.4.7, [26]) with parameters set as “-c 1 -n 5 -p 1 -T 6 -g 1 -d 0” and “-c 0.82 -n 5 -p 1 -T 6 -g 1 -d 0”, respectively.

Putative PRs were aligned to the MicRhoDE reference alignment using MAFFT (‘-- addfragments’, v.7.055b, [44]) and were added to a custom version of the MicRhoDE database with ARB software (v.6.0.4, [45]), and protein sequences were placed into the MicRhoDE phylogenetic tree using the add-by-parsimony module of ARB to assess their affiliation. SOLA PR sequences belonging to phylogenetic clusters containing previously identified euryarchaeal PR-types or clustering with a MAG phylogenetically identified as belonging to MGII were extracted. The abundance of each PR sequences was retrieved from the SOLA gene catalog and summed by PR clusters to generate an abundance table.

All aligned sequences were screened for the presence of amino acid residues at position 97, 101, 105, and 108. Partial sequences lacking these sites were not included in the analysis of proton-pumping motif and spectral tuning.

Genes flanking the PR-encoding gene in MAGs were predicted and identified using PROKKA (v.1.14.0, [46]). The synteny analysis was performed by BLASTn homology search and displayed using Easyfig (v.2.2.2, [47]).

16S and 18S rRNA sequencing

A high-frequency amplicon analysis was conducted to target both prokaryotes (16S rRNA) and eukaryotes (18S rRNA). The goal was to infer the fine seasonal dynamics of MGII Archaea and correlate their occurrences with bacterial and eukaryotic taxa. SOLA surface water (3 m) was collected from January 2015 to March 2017: twice a week during the periods of January–March 2015, January–April 2016 and December 2016–March 2017, and roughly once a week otherwise. A total of 5 L of seawater was filtered sequentially through a 3-μm filters and a 0.22-μm Sterivex cartridges. DNA was extracted and 16S rRNA and 18S rRNA were sequenced with the prokaryotic primer (515F-Y [5′-GTGYCAGCMGCCGCGGTAA] and 926R [5′-CCGYCAATYMTTTRAGTTT], and eukaryotic primer (TAREuk_F1 [5′CCAGCASCYGCAGTAATTC] and TAREuk_R [5′ACTTTCGTTCTTGATYRATGA] as described earlier [48]. All sequences were deposited in NCBI under accession number PRJNA579489.

The standard pipeline of the DADA2 package (v1.6) [49] in “R” was used for the analysis of the raw sequences to build an Amplicon Sequence Variant (ASV) table. The

Fig. 1 Seasonal dynamics of archaeal abundance and diversity at the SOLA station over 3 consecutive years. **a** Abundance of 16S rRNA contigs in the metagenomes normalized per million reads. **b** Shannon diversity index for archaeal communities. **c** Polar plot showing the abundance of 16S rRNA contigs belonging to the main

groups of planktonic archaea summed by month. MGI: *Thaumarchaeota* Marine Group I, *Euryarchaeota* MGIIa (Ca. Poseidonaceae), *Euryarchaeota* MGIIb (Ca. Thalassarchaeaceae), and *Euryarchaeota* MGIII. Values are normalized to 1 million reads and averaged per month.

ASVs were compared to the 16S rRNA of the MAGs by BLASTn. ASV abundances were normalized by the median-ratio method implemented in the “DESeq2” package [50].

For the long-term 16S rRNA archaeal amplicon analysis, data were extracted from a 7 years time-series study from the SOLA site [51]. SOLA surface water (3 m) was collected roughly twice a month from October 2007 to January 2015 and archaeal 16S rRNA genes were sequenced and analyzed as described in the original publication [51].

Statistics

Spearman correlations were computed between log-transformed environmental data and normalized ASV abundances, and between normalized prokaryotic ASVs and normalized eukaryotic ASVs abundances. Bray–Curtis similarities were computed on Hellinger transformed MAG and PR abundance data with the vegan package in R [52]. The relative abundance of the archaeal 16S rRNA genes was used to compute the Shannon diversity index with the vegan package in R.

Results

Temporal dynamics of planktonic archaea

Archaeal 16S rRNA contigs obtained from the gene catalog were more abundant in the metagenomes from November to February at the SOLA station (Fig. 1a). They were almost absent in June, July, and August. In contrast, bacterial 16S rRNA contigs were present all year long (Fig. 1a). The diversity of the archaeal communities was highest during the winter months (from October to March), with the exception of a peak of diversity in April 2012 (Fig. 1b).

Among the planktonic archaeal sequences, MGIIa were overall the most abundant (46% of the archaeal sequences), followed by MGIIb (30%), MGI (22%), and MGIII (2%) (Fig. 1c). MGIIb and MGI were present mostly for 3 months from December to February. MGIIa was the most abundant planktonic archaea during the rest of the year and dominated spring and autumn samples (Fig. 1c). MGIII was only present in winter.

MAGs characteristics

MAGs were reconstructed independently from each of the four seasonal group. A total of 40 MGII MAGs satisfying quality standards (>50% complete and <5% contamination) were obtained, of which 12 had a 16S rRNA sequence (Supplementary Table 1). Their mean genome completeness was 79%, and contamination mean 1.4%. Most were obtained from winter samples (21) followed by spring and autumn (9 for each), and summer (only 1).

The SOLA MAGs covered 15 out of the 17 or 21 clades described earlier ([7] and [6] respectively). We did not detect the clades I (1) and O4 (16), and J1-3 and Q (Supplementary Fig. 2). Thirteen MAGs belonged to MGIIb, of which 11 were nonredundant and spread across 9 different clades (Fig. 2). The most abundant MAG belonged to clade O3. A total of 27 MAGs belonged to MGIIa. They separated in 16 nonredundant MAGs that were distributed through 7 different clades (Supplementary Fig. 2). MGIIa MAGs had larger genome sizes than MGIIb (Supplementary Table 1).

MAGs seasonal dynamics

Individual MAGs showed strong seasonal patterns (Fig. 3). Among MGIIa, one MAG showed peaks of abundance from October to November (group 1, Supplementary Fig. 3). It

Fig. 2 Phylogenomic tree for the MGIIb MAGs from SOLA (**bold**) and reference MAGs from the literature. The tree is constructed using concatenated single marker protein gene sequences. Subclades names are indicated following the nomenclature of Galand et al. [20],

and Rinke et al. [6], and Tully [7] in parenthesis. Predicted proteorhodopsin spectral tuning color and amino acid motif are indicated by colors when the gene is present.

was the only MAG that had a positive correlation to ammonium concentrations (Supplementary Table 2). One group of MGIIa was relatively abundant from December to February (group 2, Fig. 3), with an additional peak of abundance in April 2012, and was positively correlated with nitrite and nitrate. A second group of MGIIa was more abundant in February (group 4), positively correlated with nutrients and chlorophyll *a*, but negatively correlated with seawater temperature (Supplementary Table 2). The last

group of MGIIa was more abundant during spring, in April and May, and had strongest correlations to chlorophyll *a* (group 5, Fig. 3).

Among MGIIb, some peaked in late autumn (group 2), others in the middle of the winter (group 3) and one in late winter (group 4) (Fig. 3). All MAGs from group 3 correlated positively with chlorophyll *a*, and with nutrients, nitrite, nitrate and phosphate, and one MAG with silicate (Supplementary Table 2).

Fig. 3 Abundance of the MAGs constructed from metagenomes from the SOLA station sampled from January 2012 to February 2015. MAGs that represented >1% of the total archaeal MAG abundance are represented. Bars are colored according to seawater temperature from coolest (blue) to warmest (yellow). MAG names, MGII

groups (a or b) and clade affiliation as numbers [7] and as letters [6, 20] are indicated at the right. Numbers in bold indicate the groups delineated by Bray–Curtis similarity and MAG ordering follow the clustering order (Supplementary Fig. 3).

Archaeal PR phylogeny and amino acid patterns

PR gene homologs were extracted from both the SOLA gene catalog and reconstructed MAGs, and grouped into PR clusters. The phylogenetic placement of the PR clusters was inferred against the MicRhoDE reference database (Fig. 4). Archaea-related PRs from SOLA were distributed mainly into two clades. A total of 9 PR clusters from the gene catalog and 14 sequences from the MAGs were affiliated with rhodopsin-types previously identified as belonging to the euryarchaeota HF70_39H11/HF70_59C08-like sub-cluster (red branches, Fig. 4). The amino acid residues in position 97, 101, and 108, involved in the ion pumping

mechanism, were aspartate (D), threonine (T), and glutamate (E). This DTE motif is classically associated with proton-pumping PR. The glutamine (Q) residue found in position 105 is characteristic of the blue-light tuned variant of the PR.

A second group of 8 PR clusters from the gene catalog and 29 sequences from MAGs were closely related to the Alphaproteobacterial-like cluster 8/B subcluster (yellow branches, Fig. 4), but constituted a separate sister clade within the supercluster IV, comprising most of the canonical PRs (Supplementary Fig. 4). However, all archaeal sequences from this clade of rhodopsin presented a unique amino acid motif, with a lysine (K) in position 108, in

Fig. 4 Phylogenetic tree of microbial rhodopsins affiliated with *Euryarchaea*. Rhodopsins from SOLA: genes extracted directly from the metagenomes are in bold grey, and from MAGs in bold black. Branches are colored by taxonomy as given in MicRhoDE: Supercluster IV/Alphaproteobacterial-like cluster 8/B subcluster in yellow,

Supercluster IV/Proteobacterial-like cluster 3/HF10_19P19-like subcluster in pink, euryarchaeal-like cluster/HF70_39H11 HF/70_59C08-like subcluster in red, and other microbial rhodopsins/group 2 in orange. Spectral tuning and ion pumping motifs are displayed as internal rings. Uninformative clades have been collapsed.

addition to the usual aspartate (D), threonine (T) in position 97 and 101 (DTK motif) (Supplementary Fig. 5). The methionine (M) at position 105 suggested a green-tuning of this variant.

Both the DTE-Q type from the euryarchaeal clade, and the less common DTK-M variants, presented a conserved putative tertiary structure with seven transmembrane helices, and harbored a lysine in the seventh transmembrane helix, the key residue covalently binding the retinal, the light-absorbing molecule (Supplementary Fig. 5).

Finally, some rhodopsins grouped in clades having no SOLA MAGs representatives. One group was close to the

proteobacterial-like cluster 3/HF10_19P19-like subcluster, and had a DTE-Q motif (pink branches, Fig. 4). Another rhodopsin group was affiliated to an uncharacterized cluster of microbial rhodopsins rooting the proteorhodopsin family with a DTE-L motif (orange branches, Fig. 4).

The DTE-Q and DTK-M types were not homogeneously distributed among MGIIa and MGIIb subclades, which indicates that the archaeal and the rhodopsin phylogenies were not congruent (Supplementary Fig. 6). DTE-Q variants were found among 13 subclades, while DTK-M were found in 11 subclades, and were the only types found in subclades L3, O2, and WHARN.

Archaeal PR seasonal dynamics in the gene catalog

The archaeal PR genes detected in the gene catalog showed seasonal dynamics (Fig. 5). Overall, they were absent from the samples from June to September and most abundant during winter. Four main seasonal dynamics were observed. A group of PR clusters were more abundant in autumn (group 1, Supplementary Fig. 7), another, which contained the largest number of PR clusters, was more abundant in early winter (group 2), and a third was typical for the winter (group 3). The last group showed variable patterns with a peak in April 2014 (group 4). The succession of different archaeal PR clusters corresponded to a succession in the color of the spectral tuning (Fig. 5). Green-tuned types dominated the autumn with DTK-M key amino acids motif, blue-tuned types in early winter with DTE-Q motif and green again in late winter. There were a few exceptions to the color succession, and in particular, the presence of green-tuned PR during earlier winter with the less common DTE-L amino acid pattern.

High-frequency temporal dynamics and long-term occurrence of selected MGII ecotypes based on 16S rRNA amplicon

To monitor the fine temporal dynamics of possible MGII ecotypes at SOLA we focused on MAGs that contained a 16S rRNA sequence, that had distinct seasonal dynamics, and that harbored a PR showing distinct seasonality. We selected one MGIIa that was abundant in autumn (SOLA-AUT-MGII-B2), one MGIIb that was typical for early winter (SOLA-AUT-MGII-B5), and one MGIIb that was more abundant in late winter (SOLA-SPR-MGII-B1). We verified the dynamics of each MAG against the dynamics of the corresponding 16S rRNA (miTAG) in the metagenomes and the corresponding PR clusters for the three metagenome sampling years. The three different markers showed very similar temporal dynamics for each selected MAG (Supplementary Fig. 8).

We used high-frequency sampling data from the SOLA station and tracked the 16S rRNA of the three selected MAGs within the 16S rRNA amplicon dataset (Fig. 6). The dataset sampled twice a week shows the time during which the specific archaea are present, and their temporal dynamics at a fine level of resolution. For MGIIa SOLA_AUT_MGII_B2, the length of the continuous occurrence varied and lasted from 1.5 to 3.5 months. It also showed episodic single re-occurrences after the main period (Fig. 6). For MGIIb SOLA_AUT_MGII_B5, the period of occurrence was longer and more regular lasting for 4 months, and correlated with salinity. Finally, the MGIIb SOLA_SPR_MGII_B1 occurrence varied in intensity from year to year, lasted at least 3 months, and was correlated

with chlorophyll *a* concentration (Supplementary Table 3). For the three archaea, the occurrence was characterized by large variations in relative abundance. In some cases, the occurrence ended abruptly, but not always.

A correlation analysis, based on the high-frequency sampling, between the three archaeal groups and other microorganisms, showed that MGIIa SOLA_AUT_MGII_B2 cooccurred with typical planktonic bacteria such as members of the SAR86 and *Rhodobacteraceae* (Supplementary Table 4), and picoeukaryotes from the Picozoa and Stramenopiles lineages. MGIIb SOLA_AUT_MGII_B5 (O3) cooccurred with prokaryotes usually found in deeper water such as the SAR324 clade or the archaeal Marine Group III, and eukaryotes from the Prymnesiophyceae and Dinophyceae lineages. MGIIb SOLA_SPR_MGII_B1 (WHARN) was associated to a clade of *Roseobacter* and the bloom of the picoeukaryotes *Bathycoccus*, as well as some Spirotrichea and Cryptophyceae (Supplementary Table 5).

The long-term amplicon data confirmed the succession in time of the three selected archaea over 7 years (Fig. 7). It showed that the MGIIa MAG (SOLA-AUT-MGII-B2) had the shortest blooms. The peak of these blooms varied from August to October. For the MGIIb SOLA_AUT_MGII_B5 (O3), blooms lasted for longer periods and they peaked in November and once in December (Fig. 7). The other MGIIb, SOLA_SPR_MGII_B1 (WHARN), also had longer blooms, and showed peaks of abundance from January to March depending on the year (Fig. 7).

Metabolic capabilities of MGIIb O3 and WHARN

We observed a seasonal succession within the abundant MGIIb archaea at the SOLA station and specifically between the group O3 and WHARN. In order to identify potential genotypic differences between these two groups, we focused a pangenomic analysis on the branch of MGIIb tree grouping the O2 to O5 clades and the WHARN clade (Fig. 2). Clade O4, which contains deep-sea representatives, was not included because no representatives were reconstructed from SOLA. The MAGs reconstructed from the SOLA metagenomes were compared with other MGIIb MAGs available in the literature (Supplementary Fig. 9). The analysis shows that the MAGs separated in four main groups according to their gene content. These groups corresponded to the clades delineated by the phylogenetic analysis and showed a clear separation between O2 and O3. The figure also shows that all clades shared a core genome grouping the majority of genes, as illustrated by the black area on the right side of the phylogram (Supplementary Fig. 9). There were also a number of gene clusters that were unique to each clade (left side of the phylogram). Most of these genes were unassigned but among the ones that were annotated we could pinpoint some that were found in O3 and not WHARN: archaeal

Fig. 5 Abundance of the proteorhodopsin clusters extracted from the gene catalog from the SOLA station sampled from January 2012 to February 2015. Bars are colored according to seawater temperature from coolest (blue) to warmest (yellow). Cluster names,

color tuning, and amino acid motives are indicated at the right. Numbers in bold indicate the groups delineated by Bray–Curtis similarity (Supplementary Fig. 7).

Fig. 6 High-frequency temporal dynamics of the 16S rRNA amplicons corresponding to three selected MGII MAGs. Sampling frequency is twice a week from January to March and once a week

otherwise. Bars are colored according to seawater temperature from coolest (blue) to warmest (yellow). Dates indicate the beginning and the end of the MAGs occurrences.

flagellum (COG3354, COG2874, COG0630), beta-glucosidase (COG2723) for the utilization of polysaccharides, or galactokinase (COG0153) for monosaccharides, pyridoxine 5'-phosphate oxidase (COG0259) involved in the vitamin B6 salvage pathway, superoxide dismutase (COG0605) against the reactive oxygen species superoxide, bacterioferritin (COG2193) for iron storage and cysteine synthase (COG0031) for the incorporation of inorganic sulfur into organic compound (cysteine), or threonine synthase (COG0498). For the WHARN clade, we could not identify specific cluster of annotated genes but among vitamin related genes, all O3 and WHARN MAGs had thiamine pyrophosphokinase (COG1564) used for the salvage pathway of vitamin B1 (salvage thiamine to form ThDP), the two genes needed for the de novo B6 synthesis through the DXP-independent pathway were present (COG0311, COG0214) and the catalase peroxiredoxin (COG1225) for the conversion of the reactive oxygen species hydrogen peroxide.

A synteny analysis of the genes surrounding the PR from the WHARN and O3 clades showed that none had genes involved in retinal biosynthesis from fatty acids. There were, however, clear differences between the genomic regions of the two clades (Supplementary Fig. 10). For WHARN, which carries the DTK-M type, the region was very well conserved between MAGs. The PR was juxtaposed with an hypothetical protein close to a phospholipase C (plcA), a fatty acid oxidation complex (fadB), and a GTP cyclohydrolase II (ribA) catalyzing the first step (APy synthesis) of the riboflavin biosynthesis (vitamin B2). In addition, a high number of constitutive genes involved in ribosome constitution and transcription, as well as numerous transfer RNA coding fragments were present (Supplementary Fig. 10). For the O3 clade, the regions containing the DTE-Q type were poorly conserved. Some contigs had genes coding for pyridoxal 5'-phosphate synthase subunit (PdxS) for vitamin B6 synthesis, and other had peroxiredoxins involved in protection against superoxide.

Fig. 7 Temporal dynamics over 7 years of the 16S rRNA corresponding to three selected MGII MAGs. Sampling frequency is roughly twice a month. Bars are colored according to seawater temperature from coolest (blue) to warmest (yellow).

Discussion

The metagenomic time-series combined with high-frequency amplicon data demonstrated the presence of distinct ecotypes within the MGIIb group (*Ca. Thalassarchaeaceae*). Earlier work on MGII *Euryarchaeota* have shown seasonal differences at a broad phylogenetic scale (family) between MGIIa and MGIIb [6, 19–21], but the existence of niche differentiation between clades of MGIIb had never been reported. Here we show a close succession in time between members of the MGIIb clade O3 and WHARN. The clade O3, which bloomed first during winter, had genotype characteristics that draw a picture of microorganisms able to swim, degrade polysaccharides, obtain B6 vitamins through salvages pathways using pyridoxamine and pyridoxal, and protect itself against both the reactive oxygen species superoxide and hydrogen peroxide. In addition, members of O3 had a blue-tuned PR with a DTE-Q amino acid motif. They appear best fitted to occupy the niche defined by the darkest time of the year, and may have a lifestyle similar to the one of SAR86, i.e. photoheterotrophs that live on phytoplankton-derived organic matter, with whom they cooccurred at SOLA. They may thus be competitors but could also co-exist by having

different substrate preferences. Members of the O3 clade could move toward sources of food and protect themselves against the reactive oxygen possibly produced by photosynthetic organisms. Alternatively, peroxiredoxins could also protect the archaea against peroxide produced during the vitamin B6 biosynthesis. Interestingly, the genes for vitamin B6 (PdxS), peroxiredoxin, and PR were located in synteny, which suggests that they could be coexpressed, and that their activation could be potentially enhanced or driven by sunlight.

Many of the genomic features found in the O3 clade were not seen in other MGIIb, but are common to some members of the MGIIa group, including motility and ability to degrade algal polysaccharides [6, 7, 21, 22]. However, members of the MGIIb O3 and MGIIa are distantly related in the phylogenetic tree, which raises the question as whether these common traits were acquired through horizontal gene transfer. Alternatively, the trait may have been common to all MGII before being lost in some clades.

Members of the WHARN clade, in turn, bloomed later, when the water temperature was at its yearly low and day-length increased. WHARN members only have the catalase peroxiredoxin gene, so that they can remove superoxide but not hydrogen peroxide. They can only salvage B6 from

pyridoxal, but can, together with the O3 clades, obtain vitamin B1 through a salvage pathway involving thiamine pyrophosphokinase. WHARN archaea were highly correlated with the abundance of the picoeukaryote *Bathycoccus*. Earlier, the picoeukaryote *Micromonas* was showed to enhance the growth of some MGII, which may be able to take up certain phytoplankton-derived labile organic compounds [9]. It remains to be shown whether the WHARN archaea take up organic material from *Bathycoccus*, if they need each other, or if they simply thrive under similar environmental conditions. WHARN have a green-tuned PR with the unusual DTK-M motif. The conserved genomic region around the PR was rich in tRNAs and genes involved in transcription and regulation processes, which suggest that this PR could be constitutively expressed as previously reported for bacteria [53], and thus important for the metabolism of the group. Our observation of a very conserved genetic environment suggests a strong environmental selection for the maintenance of this structure.

Our data demonstrated the widespread presence within archaea at SOLA of the uncommon DTK PR, DTE being the most common motif defining the capability of establishing a proton motive force, i.e. function as PRs [54]. The DTK motif is very unusual among marine microbial rhodopsins and has only been noted in members of the MGII group [4, 6]. Among bacteria, it is only seen in nonmarine microorganisms from the *Exiguobacterium* subcluster (supercluster I/Firmicutes cluster, ESR) originally detected in *Exiguobacterium sibiricum*, isolated from permafrost soil [55], and an evolutionary relative from a lake (*Exiguobacterium* sp. JL-3 [56]). This newly named ESR retinal protein (after *Exiguobacterium sibiricum* rhodopsin) functions as a light driven proton pump with a unique structural feature and pumping mechanism [57, 58]. The specificity of the marine archaeal PR variant that we describe here is that it is phylogenetically distant from terrestrial ESR, and the presence at position 105 of a methionine (M) instead of a leucine (L). Both, methionine and leucine tune the spectral absorbance to the green. The euryarchaeal DTK-M variant has otherwise characteristics that are similar to the bacteria ESR, with a lysine instead of the typical carboxyl residue at the position that corresponds to the proton donor [57]. In addition, both the bacteria ESR and the euryarchaeal DTK-M variant contain a histidine residue interacting with the proton acceptor aspartic acid, which provides the ability to function over a wider pH range compared to canonical PR [57]. The observation of the DTK motif in both the marine Euryarchaeota and terrestrial microorganisms is intriguing. The role of this rhodopsin and its origin in the sea definitely requires further investigation.

Genes involved in the retinal biosynthesis from fatty acids, notably β -carotene 15,15'-dioxygenase (Blh), were not found in the SOLA O3 nor WHARN MAGs, nor in

most other MGIIb clades [7]. MGIIb thus probably obtain the retinal, a vitamin A aldehyde, through environmental scavenging of compounds produced by other prokaryotes or microalgae. We cannot, however, exclude the possibility that the genes could have been present on a region not well covered in certain MAGs. It should be noted that we detected Blh in numerous MGIIa MAGs, including the L3 subclade that only harbor DTK-M variants, which could pinpoint different PR-based ecological strategies between MGIIa and MGIIb.

The MGII PRs showed clear seasonal patterns with a succession of green-tuned rhodopsins in autumn followed by blue in the middle of the winter, and green again in the spring. The spectral tuning of the prokaryotic PR is thought to be adapted to the dominant light field in deeper (blue) and shallower (green) oceanic waters [41, 59], and to have biogeographic patterns that follow a coastal to open ocean gradient [60]. An earlier winter vs summer comparison in the eastern Mediterranean also suggested seasonality in PR [61]. Our 3 years survey adds to the current knowledge by showing that seasonality is a major driver for the succession of different PR spectral tuning types. The fact that the blue-tuned proteins were present during the middle of the winter could suggest that they give an advantage to archaea in surface waters when days are shorter and light intensity lower. When day length increases, as well as the level of photosynthetically active solar radiation, the archaea fitted with green-tuned PR take over.

The use of microbial metagenomic time-series demonstrated that planktonic archaea can at time disappear from surface marine waters or become very rare. In the NW Mediterranean, they were almost absent in surface waters during the summer month. It strongly contrasts with the pattern observed with the domain Bacteria that is present all year round. It suggests that surface archaea are adapted to the times during which the water column is mixed and the primary production is highest. The oligotrophic conditions found during summer stratification do not seem to appeal to planktonic archaea. The absence of archaea in summer surface waters may not be a specific feature of the Mediterranean Sea. Also limited in time scale, earlier studies based on hybridization have reported that archaea were absent during summer in the Southern Ocean waters [62]. Inversely, in the North Sea, FISH counts showed that archaea were more abundant in summer [21, 63], and especially *Euryarchaeota* [64], after the phytoplankton bloom, and almost absent in winter. They were also present in some summer samples of the Atlantic Ocean [65]. Our and early results thus suggest that MGII archaea thrive under nutrient-rich conditions when primary productivity is high. Inversely, they disappear from surface waters under oligotrophic conditions when phytoplankton is absent.

Our metagenomic data also showed that the seasonality of the two different MGII families is very different from the classical summer versus winter rhythm described in the literature [19–21]. In the NW Mediterranean Sea, MGIIa and MGIIb were both more abundant in winter. The difference between families being that MGIIa was present over a longer period than MGIIb. MGIIb had a narrow peak of yearly abundance from December to February while MGIIa had a longer presence that ranged from November to April. This result contradicts earlier seasonal studies from the Mediterranean Sea that showed a dominance of MGIIa in summer and MGIIb in winter [19, 20], or no seasonal pattern between MGII groups of the San Pedro Channel [18]. The present metagenomic approach, by removing the PCR bias, probably shows a picture that is closer to reality. Our results, however, contradict a metagenome based results from the North Sea depicting summer MGIIa versus winter MGIIb [21]. The temporal range of that survey was however shorter, which could have hidden the full seasonality of the MGII clades.

Finally, our data confirms the phylogeny by Tully [7] separating the O/WHARN branch in six subclades. Rinke et al. [6] delineated only five subclades within that branch. These two earlier studies also had different way of naming the clades: clade WHARN in Tully [7] was named O2 in Rinke et al. [6], and the clade O3 in Rinke et al. [6] grouped together the clades O2 and O3 from Tully [7]. With the thought of avoiding future naming confusion, we propose to adopt the 6 clades defined in Tully [7] and named O1, O2, O3, O4, O5, and WHARN. The letter O refers to the first delineation based on 16S rRNA [20], and we find it important to keep the name WHARN that historically refers to the first published MGII planktonic archaeal sequence (WHAR-N from Woods Hole, [2]).

In conclusion, the combination of time-series and reconstructed genomes data has proven to be a powerful tool for elucidating the ecology of marine microorganisms. We dived into the diversity of the archaea MGIIb family (*Candidatus* Thalassarchaeaceae) and discovered the presence of different ecotypes that closely succeeded one another in a short period of time during the winter season. The different subclades had distinct metabolic capabilities and PR variants that were in adequacy with their ecological niches.

Acknowledgements The work of PEG was supported by the Agence Nationale de la Recherche (ANR) through the projects EUREKA (ANR-14-CE02-0004-01). We are grateful to the captain and the crew of the RV 'Nereis II' for their help in acquiring the samples. We thank the "Service d'Observation", particularly Eric Maria and Paul Labatut, for their help in obtaining and processing of the samples. We extend our acknowledgments to all the researchers that were involved in working with the time series over the years.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Fuhrman JA, McCallum K, Davis AA. Novel major archaeobacterial group from marine plankton. *Nature*. 1992;356:148–9.
- DeLong EF. Archaea in coastal marine environments. *Proc Natl Acad Sci USA*. 1992;89:5685–9.
- Santoro AE, Richter RA, Dupont CL. Planktonic marine archaea. *Annu Rev Mar Sci*. 2019;11:131–58.
- Martin-Cuadrado A-B, García-Heredia I, Molto AG, Lopez-Ubeda R, Kimes N, López-García P, et al. A new class of marine Euryarchaeota group II from the mediterranean deep chlorophyll maximum. *ISME J*. 2015;9:1619–34.
- Martin-Cuadrado AB, Rodriguez-Valera F, Moreira D, Alba JC, Ivars-Martinez E, Henn MR, et al. Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *ISME J*. 2008;2:865–86.
- Rinke C, Rubino F, Messer LF, Youssef N, Parks DH, Chuvpochina M, et al. A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (Ca. Poseidoniales ord. nov.). *ISME J*. 2019;13:663–75.
- Tully BJ. Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. *Nat Commun*. 2019;10:1–12.
- Li M, Baker BJ, Anantharaman K, Jain S, Breier JA, Dick GJ. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat Commun*. 2015;6:8933.
- Orsi WD, Smith JM, Wilcox HM, Swallow JE, Carini P, Worden AZ, et al. Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. *ISME J*. 2015;9:1747–63.
- Pereira O, Hochart C, Auguet JC, Debroas D, Galand PE. Genomic ecology of Marine Group II, the most common marine planktonic Archaea across the surface ocean. *MicrobiologyOpen*. 2019;8:e00852.
- Frigaard NU, Martinez A, Mincer TJ, DeLong EF. Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature*. 2006;439:847–50.
- Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. Untangling genomes from metagenomes: revealing an uncultured class of marine euryarchaeota. *Science*. 2012;335:587–90.
- Orsi WD, Smith JM, Liu S, Liu Z, Sakamoto CM, Wilken S, et al. Diverse, uncultivated bacteria and archaea underlying the cycling of dissolved protein in the ocean. *ISME J*. 2016;10:2158–73.
- Zhang CL, Xie W, Martin-Cuadrado A-B, Rodriguez-Valera F. Marine Group II Archaea, potentially important players in the global ocean carbon cycle. *Front Microbiol*. 2015;6:1108.
- Deschamps P, Zivanovic Y, Moreira D, Rodriguez-Valera F, López-García P. Pangenome evidence for extensive interdomain horizontal transfer affecting lineage core and shell genes in uncultured planktonic thaumarchaeota and euryarchaeota. *Genome Biol Evol*. 2014;6:1549–63.

16. Galand PE, Casamayor EO, Kirchman DL, Potvin M, Lovejoy C. Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing. *ISME J.* 2009;3:860–9.
17. Massana R, DeLong EF, Pedros-Alio C. A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl Environ Microbiol.* 2000;66:1777–87.
18. Parada AE, Fuhrman JA. Marine archaeal dynamics and interactions with the microbial community over 5 years from surface to seafloor. *ISME J.* 2017;11:2510–25.
19. Hugoni M, Taib N, Debroas D, Domaizon I, Jouan Dufournel I, Bronner G, et al. Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proc Natl Acad Sci USA.* 2013;110:6004–9.
20. Galand PE, Gutiérrez-Provecho C, Massana R, Gasol JM, Casamayor EO. Inter-annual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). *Limnol Oceanogr.* 2010;55:2171–2125.
21. Orellana LH, Francis TB, Krüger K, Teeling H, Müller M-C, Fuchs BM, et al. Niche differentiation among annually recurrent coastal Marine Group II Euryarchaeota. *ISME J.* 2019;13:3024–36.
22. Xie W, Luo H, Murugapiran SK, Dodsworth JA, Chen S, Sun Y, et al. Localized high abundance of Marine Group II archaea in the subtropical Pearl River Estuary: implications for their niche adaptation. *Environ Microbiol.* 2018;20:734–54.
23. Galand PE, Pereira O, Hochart C, Auguet JC, Debroas D. A strong link between marine microbial community composition and function challenges the idea of functional redundancy. *ISME J.* 2018;12:2470.
24. Peng Y, Leung HC, Yiu S-M, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics.* 2012;28:1420–8.
25. Noguchi H, Taniguchi T, Itoh T. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* 2008;15:387–96.
26. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006;22:1658–9.
27. Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. *Bioinformatics.* 2008;24:713–4.
28. Maillet N, Collet G, Vannier T, Lavenier D, Peterlongo P. COMMET: comparing and combining multiple metagenomic datasets. *IEEE Int Conf Bioinform Biomed.* 2014:94–8.
29. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics.* 2015;31:1674–6.
30. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009;25:1754–60.
31. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods.* 2014;11:1144–6.
32. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ.* 2015;3:e1319.
33. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics.* 2019;36:1925–7.
34. Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R. Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res.* 1998;26:320–2.
35. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402.
36. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma.* 2010;11:119.
37. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol.* 2016;428:726–31.
38. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinforma.* 2004;5:113.
39. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PloS one.* 2010;5:e9490.
40. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature.* 2013;499:431–7.
41. Olson DK, Yoshizawa S, Boeuf D, Iwasaki W, DeLong EF. Proteorhodopsin variability and distribution in the North Pacific Subtropical Gyre. *ISME J.* 2018;12:1047–60.
42. Boeuf D, Audic S, Brillet-Guéguen L, Caron C, Jeanthon C. MicRhoDE: a curated database for the analysis of microbial rhodopsin diversity and evolution. *Database.* 2015;2015:bav080.
43. Boeuf D, Lami R, Cunningham E, Jeanthon C. Summer abundance and distribution of proteorhodopsin genes in the Western Arctic Ocean. *Front Microbiol.* 2016;7:1584.
44. Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30:3059–66.
45. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, et al. ARB: a software environment for sequence data. *Nucleic Acids Res.* 2004;32:1363–71.
46. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30:2068–9.
47. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics.* 2011;27:1009–10.
48. Lambert S, Lozano J-C, Bouget F-Y, Galand PE. Seasonal marine microorganisms change neighbors under remarkable weather conditions. *Environ Microbiol.* 2021; in press.
49. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016;13:581.
50. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
51. Lambert S, Tragin M, Lozano J-C, Ghiglione J-F, Vaultot D, Bouget F-Y, et al. Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular environmental perturbations. *ISME J.* 2019;13:388–401.
52. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara R, et al. Package 'vegan'. *Community Ecol Package, Version 2019;2:5–6.*
53. Riedel T, Tomasch J, Buchholz I, Jacobs J, Kollenberg M, Gerdt G, et al. Constitutive expression of the proteorhodopsin gene by a flavobacterium strain representative of the proteorhodopsin-producing microbial community in the North Sea. *Appl Environ Microbiol.* 2010;76:3187–97.
54. Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, et al. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science.* 2000;289:1902–6.
55. Rodrigues DF, Ivanova N, He Z, Huebner M, Zhou J, Tiedje JM. Architecture of thermal adaptation in an *Exiguobacterium sibiricum* strain isolated from 3 million year old permafrost: a genome and transcriptome approach. *BMC genomics.* 2008;9:547.
56. Zhu W, Lan Y, Lou X, Han N, Ran T, Xu L, et al. Isolation of proteorhodopsin-bearing bacterium JL-3 from fresh water and characterization of the proteorhodopsin. *FEMS Microbiol Lett.* 2013;344:10–7.

57. Petrovskaya L, Balashov S, Lukashev E, Imasheva E, Gushchin IY, Dioumaev A, et al. ESR—A retinal protein with unusual properties from *Exiguobacterium sibiricum*. *Biochemistry*. 2015;80:688–700.
58. Gushchin I, Chervakov P, Kuzmichev P, Popov AN, Round E, Borshchevskiy V, et al. Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria. *Proc Natl Acad Sci USA*. 2013;110:12631–6.
59. Béja O, Spudich EN, Spudich JL, Leclerc M, DeLong EF. Proteorhodopsin phototrophy in the ocean. *Nature*. 2001;411:786–9.
60. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S. The Sorcerer II Global Ocean Sampling expedition: North-west Atlantic through Eastern Tropical Pacific. *PLoS Biol*. 2007;5:e77.
61. Sabehi G, Kirkup BC, Rozenberg M, Stambler N, Polz MF, Béja O. Adaptation and spectral tuning in divergent marine proteorhodopsins from the eastern Mediterranean and the Sargasso Seas. *ISME J*. 2007;1:48–55.
62. Murray AE, Preston CM, Massana R, Taylor LT, Blakis A, Wu K, et al. Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. *Appl Environ Microbiol*. 1998;64:2585–95.
63. Pernthaler A, Preston CM, Pernthaler J, DeLong EF, Amann R. Comparison of fluorescently labeled oligonucleotide and polynucleotide probes for the detection of pelagic marine Bacteria and Archaea. *Appl Environ Microbiol*. 2002;68:661–7.
64. Herfort L, Schouten S, Abbas B, Veldhuis MJ, Coolen MJ, Wuchter C, et al. Variations in spatial and temporal distribution of Archaea in the North Sea in relation to environmental variables. *FEMS Microbiol Ecol*. 2007;62:242–57.
65. Murray AE, Blakis A, Massana R, Strawzewski S, Passow U, Alldredge A. A time series assessment of planktonic archaeal variability in the Santa Barbara Channel. *Aquat Micro Ecol*. 1999;20:129–45.