Transcriptional activity differentiates families of Marine Group II *Euryarchaeota* in the coastal ocean

Julian Damashek[1,3], Aimee O. Okotie-Oyekan[1,4], Scott M. Gifford[2], Alexey Vorobev[1,5], Mary Ann Moran[1], James T. Hollibaugh[1]

[1]Department of Marine Sciences, University of Georgia, Athens, GA, USA
[2]Department of Marine Sciences, University of North Carolina, Chapel Hill, NC, USA
[3]Present address: Department of Biology, Utica College, Utica, NY, USA
[4]Present address: Environmental Studies Program, University of Oregon, Eugene, OR, USA
[5]Present address: INSERM U932, PSL University, Institut Curie, Paris, France

Corresponding author:
Julian Damashek
Utica College
1600 Burrstone Road, Utica, NY 30602
(315) 223-2326; judamash@utica.edu

ABSTRACT

Marine Group II *Euryarchaeota* (*Candidatus* Poseidoniales) are abundant members of marine microbial communities. They are thought to be (photo)heterotrophs that metabolize components of dissolved organic matter (DOM) such as lipids and peptides, but little is known about their transcriptional activity. We mapped reads from metatranscriptomes collected off Sapelo Island, GA to metagenome-assembled genomes to determine the diversity of transcriptionally-active *Ca.* Poseidoniales. Summer metatranscriptomes had the highest abundance of *Ca.* Poseidoniales transcripts, mostly from the O1 and O3 genera within *Ca.* Thalassarchaeaceae (MGIIb). In contrast, transcripts from fall and winter samples were predominantly from *Ca.* Poseidoniaceae (MGIIa). Genes encoding proteorhodopsin, membrane-bound pyrophosphatase, peptidase/proteases, and part of the ß-oxidation pathway were highly transcribed across abundant genera. Highly transcribed genes specific to *Ca.* Thalassarchaeaceae included xanthine/uracil permease and receptors for amino acid transporters. Enrichment of *Ca.* Thalassarchaeaceae transcript reads related to protein/peptide, nucleic acid, and amino acid transport and metabolism, as well as transcript depletion during dark incubations, provided further evidence of heterotrophic metabolism. Quantitative PCR analysis of South Atlantic Bight samples indicated consistently abundant *Ca.* Poseidoniales in nearshore and inshore waters. Together, our data suggest *Ca.* Thalassarchaeaceae are important photoheterotrophs potentially linking DOM and nitrogen cycling in coastal waters.

INTRODUCTION

Since the initial discovery of Marine Group II (MGII) *Euryarchaeota* [1,2], definitive determination of their physiology and ecological roles has remained challenging due to the lack of a cultivated isolate. Nonetheless, as data describing MGII distributions throughout the oceans have increased, several patterns have emerged: MGII are often highly abundant in the euphotic zone and in coastal waters, reach high abundance following phytoplankton blooms, and are largely comprised of two subclades, MGIIa and MGIIb [3,4]. Early metagenomic studies provided the first evidence that MGII may be aerobic (photo)heterotrophs [5-7], a hypothesis supported by incubation experiments [8-10] and by the gene content of diverse metagenome-assembled genomes (MAGs) [11-14]. Two recent studies deepened our understanding of the phylogenomics and metabolic potential of MGII by analyzing hundreds of MAGs, highlighting clade-specific differences in genomic potential for transport and degradation of organic molecules, light harvesting proteorhodopsins, and motility [15,16]. Here, we refer to MGII as the putative order "*Candidatus* Poseidoniales," MGIIa and MGIIb as the putative families "*Ca.* Poseidoniaceae" and "*Ca.* Thalassarchaeaceae," respectively, and putative genera as specified by Rinke et al. [15]. We occasionally use "MGIIa" and "MGIIb" for consistency with previous literature.

Metatranscriptomics is one strategy for gleaning information about microbial activity in the environment. *Ca.* Poseidoniales transcripts can be abundant in marine metatranscriptomes, suggesting transiently high transcriptional activity [17,18]. When metatranscriptome reads from the Gulf of Aqaba were mapped to metagenomic contigs from the Mediterranean Sea, genes involved in amino acid transport, carbon metabolism, and cofactor synthesis were highly transcribed in the aggregate euryarchaeal community [19,20]. In another study, mapping deep-

sea metatranscriptome reads to novel *Ca.* Poseidoniales MAGs indicated transcription of genes related to protein, fatty acid, and carbohydrate transport and metabolism, likely fueling aerobic heterotrophy [21]. Finally, a metaproteomics study found that euryarchaeal transport proteins for L-amino acids, branched-chain amino acids, and peptides were present throughout the Atlantic Ocean [22]. Despite these advances, little is known about similarities or differences in gene transcription between *Ca.* Poseidoniales and Thalassarchaeaceae.

We report MAG-resolved metatranscriptomic analyses of *Ca.* Poseidoniales in coastal waters near Sapelo Island (GA, USA). Prior work suggested *Ca.* Poseidoniales are sporadically active at Sapelo Island [23] and may comprise the majority of archaea in mid-shelf surface waters of the South Atlantic Bight (SAB) [24]. Since other studies thoroughly described the genomic content of *Ca.* Poseidoniales MAGs, our focus instead was determining which clades were transcriptionally active and identifying highly or differentially transcribed genes. We used two Sapelo Island MAGs [25] combined with recent marine MAG collections [15,16] to competitively recruit reads from a metatranscriptomic time series [26] and an incubation experiment [23] to determine which clades were active over time. We then used representative MAGs from highly active genera to determine which *Ca.* Poseidoniales genes were transcribed. Finally, we used quantitative PCR (qPCR) to measure the abundance of *Ca.* Poseidoniales 16S rRNA genes in DNA samples throughout the SAB to assess the prevalence *Ca.* Poseidoniales in this region.

MATERIALS AND METHODS

PHYLOGENOMICS

Phylogenomic analyses compared SIMO Bins 19-2 and 31-1 (ref. 25) to previously-reported *Ca.* Poseidoniales MAGs [15,16]. Average nucleotide identity (ANI) was calculated using fastANI [27] to compare the set of non-redundant MAGs from Tully [16] to 15 Port Hacking MAGs [15] and the two SIMO MAGs; MAGs with ANI <98.5% were added to the non-redundant set. Phylogenomic analysis was conducted using a set of sixteen ribosomal proteins [28] within anvi'o v4 (ref. 29). All genomes were converted to contig databases and ribosomal proteins were identified using HMMER [30]. These proteins were concatenated, aligned using MUSCLE [31], and used to build a phylogenomic tree using FastTree [32] within anvi'o.

COMPETITIVE READ MAPPING

We used competitive read mapping [33] to determine which *Ca.* Poseidoniales genera were transcriptionally active in free-living Sapelo Island metatranscriptomes [23,26]. Analyses of "field" communities included Gifford et al. metatranscriptomes (iMicrobe Accession CAM_P_0000917) [26] and the T_0 metatranscriptomes from Vorobev et al. [23], while dark incubation analyses included only Vorobev et al. samples (T_0 and T_{24}; NCBI BioProject PRJNA419903). Temperature, salinity, dissolved oxygen, pH, and turbidity data corresponding to metatranscriptome sampling times were downloaded from the NOAA National Estuarine Research Reserve System website (http://cdmo.baruch.sc.edu; last accessed 16 July 2020).

Contigs from all MAGs from the phylogenomic analysis were used as a database for read mapping using Bowtie2 v.2.2.9 (ref. 34) with the "very-sensitive" flag. Samtools v.1.3.1 (ref. 35) was used to index resulting BAM files, which were then profiled and summarized in anvi'o. Contig genus identity was imported to the anvi'o contig database as an external collection. The

number of transcripts L^{-1} was calculated by scaling the number of mapped reads by the volume of water filtered and the recovery of internal standards (reported in [23,26]) as previously described [36]. Seasonal transcript abundances were compared using a one-way ANOVA test in R [37] with data log-transformed as necessary to improve normality. When ANOVA results were significant, groupings were defined post-hoc with Tukey's Honest Significant Difference (HSD) test using the agricolae R package [38].

Non-metric multidimensional scaling (NMDS) analysis of metatranscriptome hits was conducted using the vegan R package [39]. NMDS input was a distance matrix constructed by Hellinger-transforming the table of transcript hits and calculating Euclidean distance between samples [40]. Genus vectors were calculated using the *envfit* command. Significance of groupings were tested by permutational multivariate analysis of variance (the *adonis* command) with 999 permutations.

MAG-SPECIFIC ANNOTATION AND TRANSCRIPT ANALYSIS

Gene-specific analyses focused on three MAGs: two from the SIMO collection (SIMO Bin 19-2, Genbank: VMDE00000000; SIMO Bin 31-1, VMBU00000000; [25]) and one (RS440, PBUZ00000000; [41]) binned from TARA Oceans metagenomes [42]. These MAGs represented genera O1, O3, and M, respectively, which were highly abundant in metatranscriptomes (see Fig. 1). RS440 was selected due to a high number of transcripts recruited when genus M was abundant (data not shown).

MAGs were annotated using the archaeal database in Prokka v.1.13 (ref. 43), using DIAMOND [44] to search against all orthologous groups in eggNOG-mapper v.1 (refs. 45, 46), and using the online BlastKOALA portal (https://www.kegg.jp/blastkoala/, last accessed 6

March 2019) [47]. Putative genes for carbohydrate-active enzymes, peptidases, and membrane transport proteins were identified using HMMER searches of dbCAN2 (HMMdb v.7) [48,49], *MEROPS* v.12.0 (ref. 50), and the Transporter Classification Database [51], respectively.

Transcript reads were mapped to MAGs (combined into a single database such that each read mapped to only one MAG) to identify *Ca.* Poseidoniales genes that were highly or differentially transcribed. Coverage was calculated by profiling BAM files in anvi'o and normalized to coverage per million reads (CPM) by dividing by the total number of reads per sample. For each MAG, "highly transcribed" genes were the 5% of putative genes with the highest median CPM across metatranscriptomes (SIMO Bin 19-2: 63 genes, SIMO Bin 31-1: 70 genes, RS440: 77 genes).

DESeq2 v.3.11 (ref 52) was used to identify genes from each MAG that were differentially transcribed when each genus was highly transcriptionally active. For each MAG, "treatment" samples in DESeq2 were those where the respective genus recruited ≥50% of *Ca.* Poseidoniales reads from the metatranscriptome. Thus, positive fold-change values are genes transcribed at higher levels when the genus is highly transcriptionally active (compared to other metatranscriptomes). DESeq2 was also used to identify differentially transcribed genes for each MAG between T_0 and T_{24} samples in high tide (HT) dark incubations [23]. Since T_{24} samples were the "treatment" condition in DESeq2, positive fold-change values here are genes transcribed at higher levels in T_{24} compared to T_0 samples. In all DESeq2 analyses, genes with Benjamini-Hochberg adjusted $p<0.1$ were counted as having significantly different transcription.

16S rRNA QUANTITATIVE PCR

DNA samples from SAB field campaigns in 2014 and 2017 [24,53] were used as templates for qPCR reactions targeting the *Ca.* Poseidoniales 16S rRNA gene. Samples included the variety of shelf habitats (inshore, nearshore, mid-shelf, shelf-break, and oceanic as previously defined [24]; Fig. S1). Primers were GII-554-f [54] and Eury806-r [55] with cycling conditions as previously reported [56] (Table S1). Reactions (25 µL, triplicate) used iTaq Universal Green SYBR Mix (Bio-Rad, Hercules, CA) in a C1000 Touch Thermal Cycler/CFX96 Real-Time System (Bio-Rad, Hercules, CA). Each plate included a no-template control and a standard curve (serial dilutions of a linearized plasmid containing a previously-sequenced, cloned amplicon). Abundance of *Ca.* Poseidoniales 16S rRNA genes was compared to published bacterial 16S rRNA gene abundance from the same samples [24,53]. Regional variability of gene abundance was assessed using a one-way ANOVA and a post-hoc HSD test as described above. Model II regressions of log-transformed qPCR data were estimated using the lmodel2 R package [57] as previously described [53]. All plots were constructed with anvi'o or the ggplot2 R package [58].

RESULTS

EURYARCHAEOTAL MAGs

SIMO Bins 19-2 and 31-1 were estimated as 82.5-92.3% and 77.5-96.2% complete, respectively, with redundancy <0.6% [25]. Phylogenomics placed both in the putative family *Ca.* Thalassarchaeaceae (MGIIb) and genera O1 (SIMO Bin 19-2) and O3 (SIMO Bin 31-1; Fig. S2). Phylogenomic groupings were generally consistent with previous findings [15,16].

Both SIMO MAGs contained a proteorhodopsin gene. Presence of a methionine residue at position 315 suggested absorption of green light [59,60], and both proteorhodopsin genes grouped in "Archaea Clade B" [11,16,61] (Fig. S3). Both MAGs included partial or complete

pathways indicating aerobic heterotrophic growth, such as glycolysis, the TCA cycle, and electron transport chain components (Table S2). Pathways for metabolism of compounds such as fatty acids, peptides, and proteins were also present, as were transport systems and metabolic pathways for amino acids and nucleotides.

DOMINANT GENERA IN FIELD METATRANSCRIPTOMES

There were significant seasonal differences in transcript recruitment by the combined set of *Ca.* Poseidoniales MAGs ($F_{3,28}$=4.9, p=0.007): most summer samples had >10^{10} *Ca.* Poseidoniales transcripts L^{-1}, significantly more than in other seasons (Fig. 1A,C; Table S3). The diversity of transcriptionally-active *Ca.* Poseidoniales also changed seasonally. Genera O1 and O3 accounted for most reads mapped from summer samples (typically 89.5-99.5% of *Ca.* Poseidoniales reads), with most mapping to O1 (Fig. 1A,B). HT (and not low tide; LT) metatranscriptomes from July 2014 also had a moderate fraction of reads (37.5-39.6%) mapped to *Ca.* Poseidoniaceae. In contrast to summer samples, November 2008 and May 2009 transcripts were predominantly O3, while those from February 2009 and October 2014 were mostly *Ca.* Poseidoniaceae genera M, L1, or L2 (Fig. 1A, Table S3). NMDS analysis showed clear seasonal groupings (PERMANOVA: r^2=0.693, p=0.001; Fig. S4).

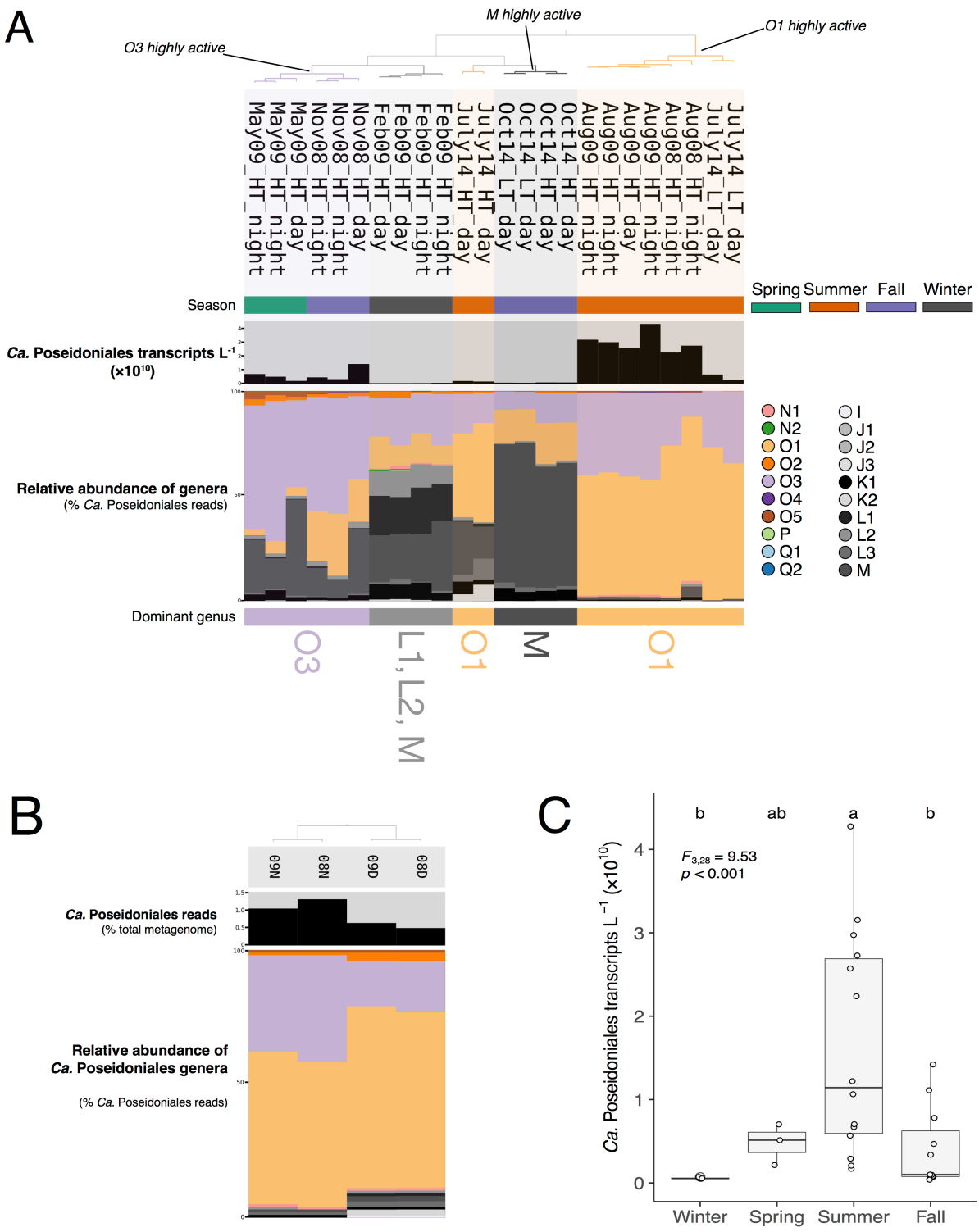

Fig. 1A) Relative abundance of *Ca.* Poseidoniales genera in Sapelo Island metatranscriptomes (*n*=24). The dendrogram (top) shows grouping by similarity. Season is indicated by color beneath sample names. The bar chart shows the abundance of *Ca.* Poseidoniales transcripts L^{-1} and the stacked bar charts show the relative abundance of genera (% total *Ca.* Poseidoniales transcripts), colored by genus. Dominant genera are indicated below the stacked bar chart. "Highly active" samples for each genus are marked and were used for analysis of differential transcription. **B)** Relative abundance of genera in Sapelo Island metagenomes (*n*=4), as described above. Since internal standards were not included in metagenomes, total *Ca.* Poseidoniales reads are shown as a percentage of the total metagenome. **C)** Boxplots of *Ca.* Poseidoniales reads L^{-1}, grouped by season (winter, *n*=4; spring, *n*=3; summer, *n*=10; fall, *n*=7). Values from individual metatranscriptomes are overlain. Results of an ANOVA are indicated; letters at the top indicate post-hoc groups according to Tukey's HSD test.

HIGHLY TRANSCRIBED *CA.* POSEIDONIALES GENES

Many highly transcribed *Ca.* Thalassarchaeaceae (MGIIb) genes were involved in translation, transcription, replication/repair, or post-translation protein modification (Fig. 2). Genes encoding proteins involved in energy production or conservation (ATPases, a pyrophosphate-energized proton pump, and proteorhodopsin) were also highly transcribed. Notably, the *aapJ* and *livK* genes, encoding substrate-binding proteins of L-amino acid and branch-chain amino acid transporters, respectively, were among the most highly transcribed genes in both *Ca.* Thalassarchaeaceae MAGs (Fig. 2, Table S2).

Many of the highly transcribed genes mapping to the *Ca.* Poseidoniaceae (MGIIa) MAG were not highly transcribed by *Ca.* Thalassarchaeaceae, including genes encoding a carbamoyl phosphate synthetase subunit (*carA*), a family 2 glycosyl transferase, chromosomal protein MC1b, and a ftsX-like permease. The *carA* gene had the highest median coverage of *Ca.* Poseidoniaceae genes across coastal metatranscriptomes (Fig. 2, Table S2). While both *Ca.* Thalassarchaeaceae MAGs also contained the carbamoyl phosphate synthetase genes, neither transcribed *carA* at high levels (Table S2).

Twelve genes were highly transcribed by *Ca.* Thalassarchaeaceae and not by *Ca.* Poseidoniaceae, including genes encoding ATP synthase, transcription initiation factor IIB, halocyanin, phytoene desaturase, protein translocase, xanthine/uracil permease, and receptors for amino acid transporters (Fig. 2, Table S2). Other than those encoding ribosomal proteins, only six genes were highly transcribed in all three MAGs: a chaperone protein, a ribonucleoside-diphosphate reductase, translation elongation factor 1A, 3-hydroxyacyl-CoA dehydrogenase, a membrane-bound pyrophosphatase (*hppA*), and proteorhodopsin (Fig. 2).

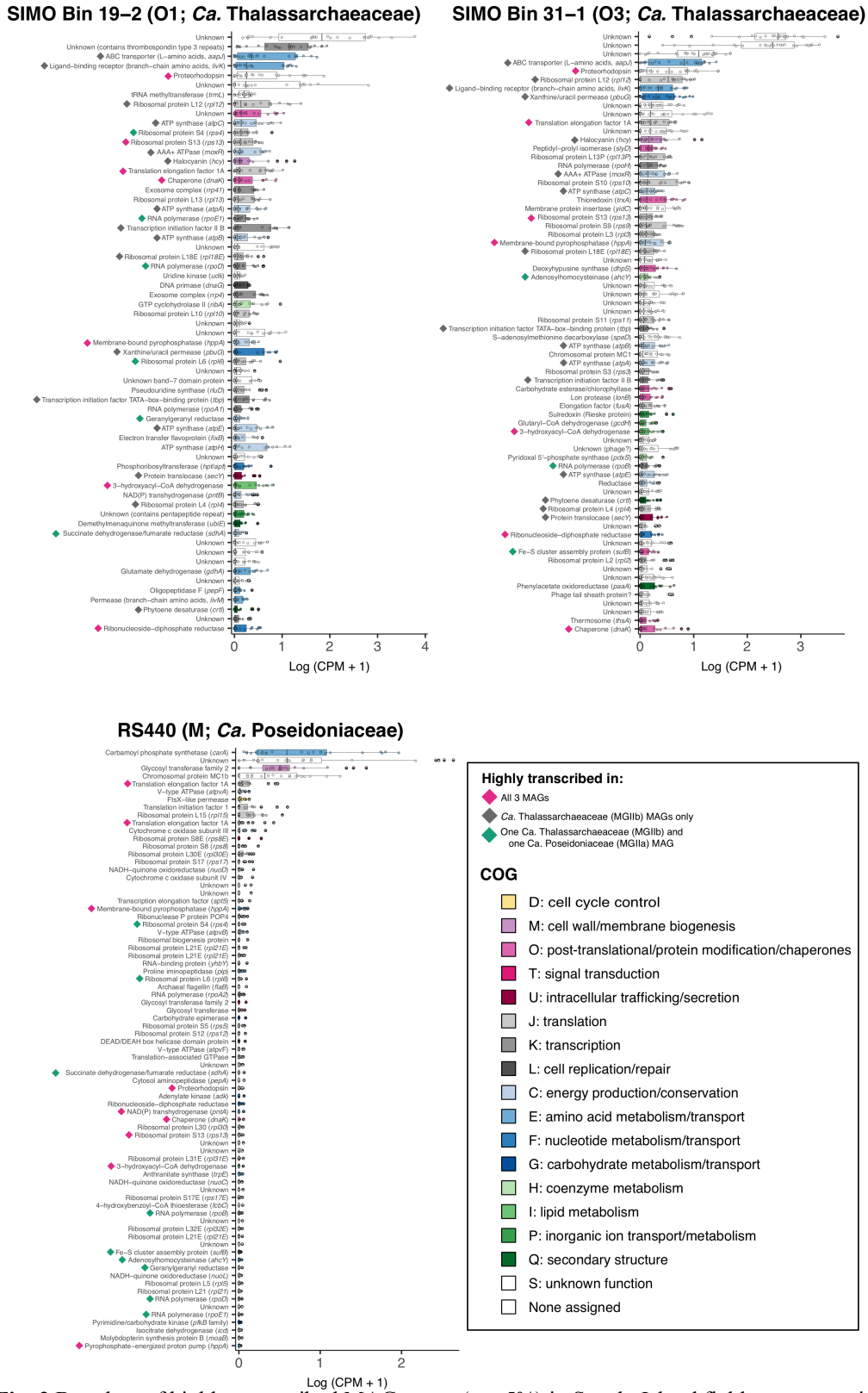

Fig. 2 Boxplots of highly-transcribed MAG genes (top 5%) in Sapelo Island field metatranscriptomes. Overlain points show CPM for individual metatranscriptomes (n=24). Shading indicates COG functional category assigned by eggNOG-mapper (genes assigned to group S were similar to proteins of unknown function in the COG database, while genes with no COG assignment did not match proteins in the COG database). Diamonds indicate genes highly transcribed in all MAGs (pink), in *Ca.* Thalassarchaeaceae (MGIIb) MAGs only (gray), or in the *Ca.* Poseidoniaceae (MGIIa) MAG and one *Ca.* Thalassarchaeaceae MAG (green).

DIFFERENTIAL GENE TRANSCRIPTION

We were interested in identifying genes with variable transcription levels when genera O1, O3, and M were highly transcriptionally active in the ocean. Twenty-three genes were differentially transcribed in O1-active samples (Fig. 3, Table S4), sixteen of which had higher abundance in O1-active metatranscriptomes compared to other metatranscriptomes. These highly transcribed genes encoded proteorhodopsin, two copper-containing redox proteins (halocyanin and plastocyanin), and proteins involved in lipid metabolism (3-hydroxyacyl-CoA dehydrogenase and oligosaccharyltransferase), nucleotide transport/metabolism (ribonucleotide-diphosphate reductase and xanthine/uracil permease), and amino acid transport (ligand-binding receptor for a L-amino acid transporter, *aapJ*). Differentially transcribed genes mapping to the O3 MAG were mostly depleted in O3-active metatranscriptomes and largely encoded proteins of unknown function; only the gene encoding ribosomal protein L12 was enriched in O3-active samples (Fig. S5, Table S4). Only four genes mapping to the M MAG were differentially transcribed in M-active metatranscriptomes. Annotated genes encoded chromosomal protein MC1b, an ATPase subunit, and a glycosyl transferase, which all had significantly fewer transcripts in M-rich samples (Fig. S6, Table S4). One gene of unknown function was enriched compared to other samples.

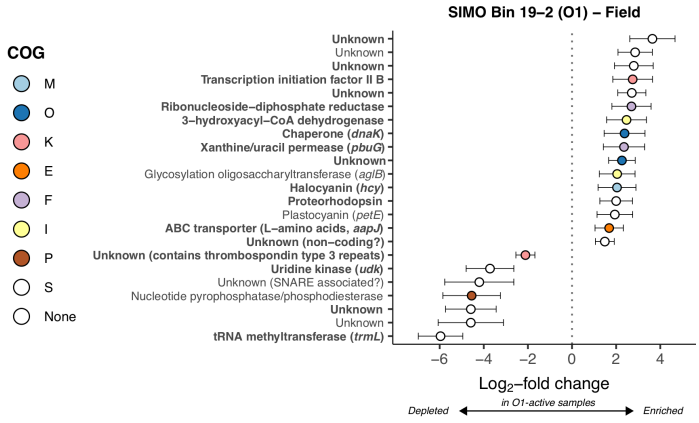

Fig. 3 Log_2-fold change of SIMO Bin 19-2 (genus O1) genes differentially transcribed in field metatranscriptomes where transcriptional activity of *Ca.* Poseidoniales was dominated by genus O1 (see Fig. 1, Table S3), calculated with DESeq2. Error bars show estimated standard error. Only genes with adjusted *p*-values <0.1 are shown. Color indicates COG functional category (see Fig. 2). Bold indicates genes in the top 5% median transcript coverage across field metatranscriptomes (Fig. 2).

DARK INCUBATION METATRANSCRIPTOMES

Incubation had little effect on transcription by the dominant genera in LT samples (Fig. 4, Table S3). In contrast, there were distinct shifts in transcriptionally active populations during incubations of all HT samples. July HT metatranscriptomes initially contained 60.4-62.5% *Ca.* Thalassarchaeaceae (MGIIb) while hits from the corresponding T_{24} samples were 98.1-99.3% *Ca.* Thalassarchaeaceae, due to increased transcript hits to genus O1 (Fig. 4, Table S3). Likewise, October 2014 HT samples initially contained 65.0-66.6% hits to *Ca.* Poseidoniaceae (MGIIa) but changed to 78.3-98.8% hits to *Ca.* Thalassarchaeaceae at T_{24} due to an increase in hits to O1 (Fig. 4).

DESeq2 identified 40 differentially transcribed genes mapping to the O1 MAG between HT T_0 and T_{24} metatranscriptomes. Four O1 genes had higher transcription at T_{24}, including xanthine/uracil permease (*pbuG*) and an amino acid transporter substrate-binding domain (*aapJ*; Fig. 5, Table S5). The 36 O1 genes transcribed at lower levels encoded proteins involved in repair of UV-damaged DNA, amino acid or nucleotide metabolism, coenzyme synthesis, peptidases or proteases, transcription, DNA replication, and lipid biosynthesis, as well as phytoene desaturase (*crtD*) and multiple subunits of pyruvate dehydrogenase (*pdhC*, *pdhA*). None of the genes mapping to the

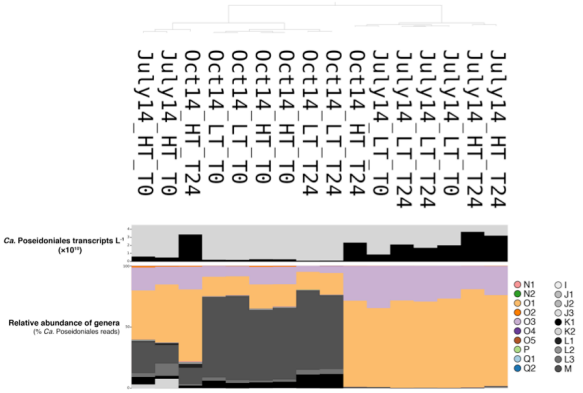

Fig. 4 Comparisons of competitive read mapping to *Ca.* Poseidoniales genera at the beginning and end of 24-hour incubations of Sapelo Island water conducted by [23]. The dendrogram (top) shows grouping by similarity. The bar chart below the dendrogram is the abundance of *Ca.* Poseidoniales transcripts L^{-1}. Stacked bar charts show the relative abundance of genera (% total *Ca.* Poseidoniales transcript reads), colored by genus.

O3 or M MAGs were transcribed at significantly different levels between HT incubation timepoints (p>0.1 for all genes; Table S5).

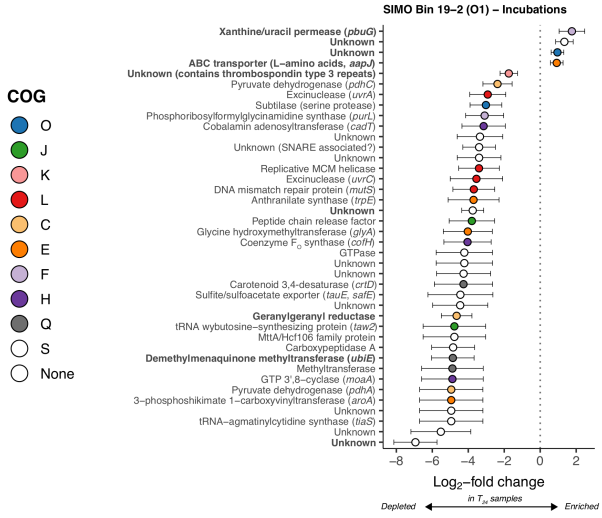

Fig. 5 Log_2-fold change of SIMO Bin 19-2 (genus O1) genes differentially abundant in T_{24} versus T_0 metatranscriptomes from Sapelo Island high tide waters. Error bars show estimated standard error. Only genes with adjusted p-values<0.1 are shown. Color indicates COG functional category (see Fig. 2). Bold indicates genes in the top 5% median transcript coverage across field metatranscriptomes (Fig. 2).

16S rRNA GENE ABUNDANCE

Ca. Poseidoniales 16S rRNA genes were detected in all SAB DNA samples (n=208), with a range from 1.6×10^4 to 7.6×10^8 genes L^{-1} (Table S6). Standard curves for the *Ca.* Poseidoniales assay always had r^2>0.99 (mean ± standard deviation: 0.99±0.001) and the mean efficiency was 93.4% (±2.0%; Table S1). When data from all cruises were combined, *Ca.* Poseidoniales genes were most abundant throughout inshore and nearshore waters and least abundant in shelf-break and oceanic samples ($F_{4,204}$=18.5, p<0.001; Fig. 6A). There was a strong linear relationship between log-transformed bacterial and *Ca.* Poseidoniales 16S rRNA gene abundances, with bacterial abundances 2-3 orders of magnitude higher (Fig. 6B).

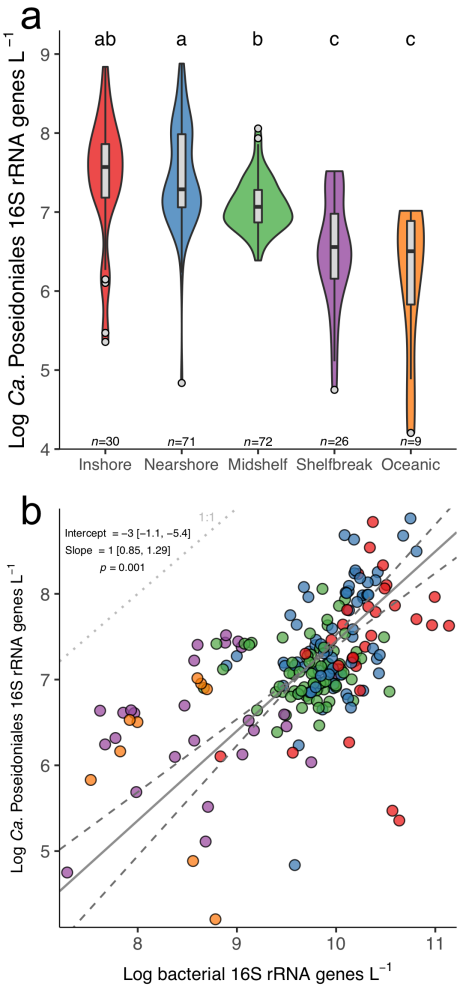

Fig. 6 A) Violin plots of log-transformed *Ca.* Poseidoniales 16S rRNA gene abundances across regions in the SAB, with overlain boxplots. Width of the violin plot corresponds to data probability density. Color denotes sampling region. Letters above boxes denote post-hoc grouping according to Tukey's HSD test. B) Scatterplot of bacterial and *Ca.* Poseidoniales 16S rRNA gene abundances in the SAB. The solid line shows the best fit of a model II (major axis) linear regression, with dashed lines showing a 95% confidence interval of the slope. Regression parameters are shown on the plot.

DISCUSSION

ABUNDANCE OF *CA.* POSEIDONIALES GENES IN THE SOUTH ATLANTIC BIGHT

The abundance of *Ca.* Poseidoniales 16S rRNA genes in the coastal SAB is among the highest measured in the ocean, with nearly 10^9 genes L^{-1} in some samples. Typical *Ca.* Poseidoniales abundance is 10^6-10^7 genes or cells L^{-1} in oligotrophic waters [62-66] and 10^7-10^8 genes or cells L^{-1} in coastal waters [9,54,56,67-69]. Greater gene abundance in inshore, nearshore, and mid-shelf waters indicates *Ca.* Poseidoniales are more abundant over the shallow shelf than further offshore (Fig. 6A), matching clone library data from the SAB [24] and data from the Central California Current and the Black Sea [9,68]. The DOM in productive, turbid

SAB coastal waters supports highly active heterotrophic microbial populations [70,71]. Our data suggest large populations of *Ca.* Poseidoniales are part of this heterotrophic community, and the correlation between *Ca.* Poseidoniales and bacterial abundance (Fig. 6B) suggests common factors influence the abundance of these populations.

TRANSCRIPTIONALLY ACTIVE *CA.* POSEIDONIALES

While numerous studies have demonstrated high abundance of *Ca.* Poseidoniales in the coastal ocean (e.g., [56,69]), little is known about which clades are transcriptionally active in these regions. At Sapelo Island, the striking dominance of the *Ca.* Thalassarchaeaceae (MGIIb) genera O1 and O3 in summer samples, which also contained the highest amount of aggregate *Ca.* Poseidoniales transcripts (Fig. 1), indicates that *Ca.* Thalassarchaeaceae are highly-active during the summertime. Outliers to this pattern were July 2014 HT samples, which contained abundant *Ca.* Poseidoniaceae (MGIIa) transcripts (though *Ca.* Thalassarchaeaceae still comprised the majority of their *Ca.* Poseidoniales reads). A previous study found relatively high salinity and DOM enriched in marine-origin molecules over the mid-shelf SAB during July 2014 [72]. Our data suggest *Ca.* Poseidoniaceae were relatively active over the shelf during this time and were transported inshore during flood tides, leading to shifts in transcriptional diversity between LT waters (dominated by *Ca.* Thalassarchaeaceae) and HT waters (which included a higher number of *Ca*. Poseidoniaceae).

Ca. Poseidoniaceae (MGIIa) are the predominant euryarchaeal family in many coastal ecosystems, particularly in summer (e.g., [9,14,56,73,74]), but it is unclear what general patterns govern *Ca.* Poseidoniales distributions in coastal waters worldwide. Studies of *Ca.* Poseidoniales ecology often focus on distributions with depth, typically finding abundant *Ca.*

Thalassarchaeaceae (MGIIb) in deeper waters and *Ca.* Poseidoniaceae (MGIIa) more prevalent in euphotic waters (e.g., [12,54,75-77]). A recent mapping of global ocean metagenome reads showed that coastal populations of *Ca.* Poseidoniales were primarily *Ca.* Poseidoniaceae, though *Ca.* Thalassarchaeaceae MAGs recruited a substantial number of reads from some coastal metagenomes [16]. Our data match this latter pattern, with *Ca.* Poseidoniales populations in surface waters off Sapelo Island dominated by highly active *Ca.* Thalassarchaeaceae (Fig. 1). The higher abundance of MAGs from genera O1 and O3 (also referred to as MGIIb.12 and MGIIb.14) in some mesopelagic and coastal samples with relatively high temperature (~23–30°C; [16]) may explain the unusual pattern found in SAB waters: these genera peaked in summer at Sapelo Island, when water temperatures were 29–30°C (Table S2), suggesting they may be adapted to growth at relatively low light and high temperature.

CA. POSEIDONIALES GENE TRANSCRIPTION PATTERNS

Sapelo Island metatranscriptome reads that mapped to *Ca.* Poseidoniales were analyzed in three ways:

(1) We determined sets of "highly transcribed" genes mapping to MAGs of transcriptionally-active *Ca* Poseidoniales genera (5% of MAG genes with the highest median transcript coverage);

(2) We identified genes mapping to *Ca* Poseidoniales MAGs that were differentially transcribed when its genus was highly active (≥50% of *Ca*. Poseidoniales transcripts in a sample; Fig. 1, Table S2);

(3) We identified genes mapping to *Ca* Poseidoniales MAGs that were differentially transcribed at the beginning versus end of dark incubations. Since dark incubation

separates indigenous microbes from light and sources of short-lived substrates, transcription of corresponding transporters and metabolic genes ceases during incubations as substrates are consumed. We therefore assume transcript depletion in T_{24} compared to T_0 metatranscriptomes indicates genes that were transcriptionally active in the field [23]. This interpretation was bolstered by significant transcript depletions for genes involved in repairing UV damage to DNA (*uvrA*, *uvrC*, and *cofH*; Fig. 5), an expected result given alleviation of UV stress in dark incubations.

In the following sections, we synthesize these analyses to discuss *Ca.* Poseidoniales transcriptional activity related to metabolism of labile DOM, transport/metabolism of amino acids and nucleotides, and basic energetic processes. Though lack of a cultivated representative limits the analysis to computationally-inferred functions, these data provide hypotheses regarding the activity of *Ca.* Poseideniales families in the coastal ocean.

PROTON GRADIENTS AND ELECTRON TRANSPORT

Our analysis revealed that *Ca* Poseidoniales genes involved in establishing transmembrane proton gradients were highly transcribed in our samples. Genes encoding proteorhodopsin were among the most highly transcribed by both *Ca.* Thalassarchaeaceae MAGs and were highly transcribed in O1-active samples (Figs. 2,3). Proteorhodopsins consist of a retinal chromophore linked to a transmembrane protein and use light energy to pump protons across the cell membrane [78]. The resulting energy can be coupled to ATP production or other chemiosmotic processes and often supports photoheterotrophy, though its function varies widely [61]. Proteorhodopsin genes are highly transcribed in the photic zone of both open ocean and

355 coastal waters (e.g., [79-81]) and our data indicate coastal *Ca.* Poseidoniales conform to this
356 pattern, consistent with recent evidence from other regions [20,82,83]. High transcription of
357 proteorhodopsin supports the photoheterotrophic lifestyle hypothesized for *Ca.* Poseidoniales
358 (e.g., [9,11,12,15,16]).

359 Since O1 proteorhodopsin transcript abundance did not differ between the beginning and
360 end of dark incubations (Fig. 5), light may not regulate *Ca.* Thalassarchaeaceae proteorhodopsin
361 transcription. However, depletion of O1 *crtD* (carotenoid 3,4-desaturase) transcripts during dark
362 incubation (Fig. 5) suggests light may regulate retinal synthesis. Whether proteorhodopsin
363 transcription responds to light varies among marine bacteria [60,84,85], and the function of
364 constitutive transcription is not straightforward: while some bacteria use proteorhodopsin to
365 produce ATP when carbon-limited [86], high amounts of proteorhodopsin in other bacteria can
366 physically stabilize membranes even when inactive [87]. The high proteorhodopsin transcription
367 in our data emphasizes, but provides little mechanistic clarification of, the physiological role for
368 proteorhodopsin in *Ca.* Poseidoniales (Table 1).

Table 1 Transcriptional traits shared or distinct among euryarchaeal families[1]

Putative function	Relevant gene(s)	Distribution	Evidence
Proteorhodopsin	Proteorhodopsin gene	Both families	Highly transcribed (Fig. 2); enriched when O1 active (Fig. 3)
Pyrophosphatase	*hppA*	Both families	Highly transcribed (Fig. 2)
Protease/peptidase	*pepF* , *lonB* , *pepA* , *pip* , carboxypeptidase A, subtilase	Both families	Highly transcribed (Fig. 2); depletion during dark incubation (Fig. 5)
ß-oxidation	3-hydroxyacyl-CoA dehydrogenase	Both families	Highly transcribed (Fig. 2); enriched when O1 active (Fig. 3)
Proteorhodopsin retinal synthesis	*crtI* , *crtD*	*Ca.* Thalassarchaeaceae only	Highly transcribed in SIMO MAGs (Fig. 2); depletion during dark incubation (Fig. 5)
Electron transport	Halocyanin gene	*Ca.* Thalassarchaeaceae only	Highly transcribed in SIMO MAGs (Fig. 2); enriched when O1 active (Fig. 3)
Amino acid transport/metabolism	*aapJ* , *livK* , *aroA* , *trpE*	*Ca.* Thalassarchaeaceae only	Highly transcribed (Fig. 2); depletion during dark incubation (Fig. 5)
Xanthine/uracil permease	*pbuG*	*Ca.* Thalassarchaeaceae only	Highly transcribed (Fig. 2); enriched when O1 active (Fig. 3); depletion during incubations (Fig. 5)
Amino acid synthesis	*carA*	*Ca.* Poseidoniaceae only	Highly transcribed (Fig. 2)
Carbohydrate synthesis	Family 2 glycosyl transferase	*Ca.* Poseidoniaceae only	Highly transcribed (Fig. 2)

[1]Putative *Ca.* Poseidoniales families are *Ca.* Poseidoniaceae (MGIIa; MAG RS440) and *Ca.* Thalassarchaeaceae (MGIIb; SIMO Bins 19-2, 31-1).

369

370 Numerous genes encoding putative electron transport proteins were highly transcribed by
371 at least one *Ca.* Poseidoniales family (Fig. 2). Like proteorhodopsin, a halocyanin gene was

among the most highly transcribed *Ca.* Thalassarchaeaceae genes and was enriched when genus O1 was transcriptionally active (Fig. 3). Halocyanins are involved in the electron transport chain and have been posited to increase the energy yield of aerobic respiration in *Ca.* Thalassarchaeaceae to stimulate rapid growth [12]. The similar transcription patterns of proteorhodopsin and halocyanin suggests proteorhodopsin activity in coastal *Ca.* Thalassarchaeaceae may function to increase growth rates during respiration, aiding rapid population growth when conditions permit.

The *hppa* gene from all *Ca* Poseidoniales MAGs was highly transcribed (Fig. 2; Table 1). The gene putatively encodes a membrane-bound pyrophosphatase, which generates a proton gradient via hydrolysis of pyrophosphate, a byproduct of numerous cellular processes [88]. In metatranscriptomes from a phytoplankton bloom, enrichment of *hppA* transcripts suggested high pyrophosphate-based energy conservation in oligotrophic waters [89]. Similarly, *hppA* was abundant in metatranscriptomes from Lake Llebreta, particularly at night [90]. Although *hppA* is widespread in MAGs from *Ca.* Poseidoniales [15] it has not been recognized as a potentially important part of their metabolism. Our data suggest *Ca.* Poseidoniales may be capable of using pyrophosphatase (along with proteorhodopsin) to generate a protonmotive force (Table 1).

POTENTIAL IMPORTANCE OF MARINE DOM IN *CA.* POSEIDONIALES METABOLISM

Although the T_0 metatranscriptomes from summer versus fall were dominated by transcripts from different *Ca.* Poseidoniales families, a 24-hour dark incubation consistently favored transcription by *Ca.* Thalassarchaeaceae when samples were collected at HT (Fig. 4). This tidal stage-linked increase in *Ca.* Thalassarchaeaceae transcription could relate to differences in DOM availability between HT and LT, consistent with numerous studies

implicating DOM in shaping *Ca.* Poseidoniales populations [9,10,91]. The HT Sapelo Island DOM pool is primarily of marine origin while LT DOM is more riverine- and marsh-derived [72], which may select for growth of different *Ca.* Poseidoniales families in the water masses present at different tidal stages. Furthermore, the depletion of transcripts encoding two pyruvate dehydrogenase subunits (*pdhA* and *pdhC*) during incubations (Fig. 5) suggests *Ca.* Poseidoniales were metabolizing phytoplankton photosynthate *in situ*. Alternatively, these tidal stage-driven transcriptional patterns may relate to differential light adaptation in populations originating in offshore versus nearshore waters. Inshore populations, potentially adapted to life in turbid waters, may increase transcription upon dark enclosure, whereas offshore populations (transported shoreward during flood tide) may be adapted to clearer waters and reduce transcription in dark conditions.

Multiple lines of evidence indicate coastal *Ca.* Poseidoniales were metabolizing proteins and fatty acids. High transcription of genes encoding proteases or peptidases mapping to all *Ca.* Poseidoniales MAGs (Fig. 2) suggests metabolism of proteins or peptides by both families (Table 1). Furthermore, decreased transcription of protease genes mapping to the O1 MAG during dark incubations (Fig. 5) suggests protein metabolism by *Ca.* Thalassarchaeaceae was active *in situ* prior to incubation. While some of these genes could be involved in intracellular recycling (particularly lon protease and cytosol aminopeptidase), active protein metabolism is consistent with previous experiments demonstrating protein assimilation [10] and high transcription of *Ca.* Poseidoniales peptidase genes in other marine regions [20,21].

In addition to genes encoding protein catabolism, high transcription of the 3-hydroxyacyl-CoA dehydrogenase gene from all three MAGs (Fig. 2), and its enrichment in O1-active field samples (Fig. 3), suggests the importance of fatty acid metabolism for both *Ca.*

Poseidoniales families (Table 1). This conclusion fits with the widespread occurrence of ß-oxidation genes in *Ca.* Poseidoniales MAGs [15,16], as well as transcriptional data from the deep ocean [21].

DISTINCT PATTERNS OF AMINO ACID AND NUCLEOTIDE UPTAKE AND METABOLISM

Transcription of *livK* and *aapJ* appears to differentiate *Ca.* Poseidoniales families in the coastal ocean (Table 1). These genes putatively encode ligand-binding receptors for ABC transporters: *aapJ* for a general L-amino acid transporter and *livK* for a branched-chain amino acid transporter [92]. Both are commonly present in *Ca.* Thalassarchaeaceae (MGIIb) but not *Ca.* Poseidoniaceae (MGIIa) [16] and were among the most highly transcribed *Ca.* Thalassarchaeaceae genes (Fig. 2). Previous studies noted high transcription of euryarchaeal *livK* and *aapJ* genes in the water column of the Red Sea [19,20], at the Mid-Cayman Rise [21], and throughout the Atlantic Ocean [22]. Our data suggest this activity was probably associated with *Ca.* Thalassarchaeaceae.

The *aapJ* and *livK* genes were collocated with genes putatively encoding the full transporters in the *Ca.* Thalassarchaeaceae MAGs (Table S1). Unfortunately, it is difficult to guess their substrates from sequence data alone: AAP transporters are typically capable of transporting a range of L-amino acids [92] while LIV transporters can be highly specific for leucine, specific for branched-chain amino acids, or transport diverse amino acids [92-94]. In soil bacteria grown under inorganic nitrogen limitation, elevated transcription of *aapJ* is linked to organic nitrogen use [95], but it is unclear whether this mechanism translates to *Ca.* Thalassarchaeaceae since amino acids could be used for numerous anabolic or catabolic

441 processes. In addition to these binding proteins, the depletion of transcripts from O1 genes
442 putatively involved in the shikimate pathway of aromatic amino acid synthesis (3-
443 phosphoshikimate 1-carboxyvinyltransferase and anthranilate synthase) during incubations (Fig.
444 5) suggests these archaea were synthesizing aromatic amino acids *in situ*.

445 The combination of high *pbuG* transcription by *Ca.* Thalassarchaeaceae (MGIIb) with the
446 high numbers of O1 *pbuG* transcripts in O1-active samples and dark incubation endpoints (Figs.
447 2,3,5) suggests an important role for xanthine/uracil permease (the putative product of *pbuG*) in
448 *Ca.* Thalassarchaeaceae metabolism. In some phytoplankton, *pbuG* is transcribed during
449 nitrogen-stressed growth [96-98], potentially allowing access to DON. However, *pbuG* and
450 xanthine dehydrogenase (*xdh*) are also transcribed when xanthine is catabolized by marine
451 bacteria [99]. Both *Ca.* Thalassarchaeaceae MAGs contain putative xanthine dehydrogenase
452 genes (*xdhC* and *yagS*; Table S1), suggesting the ability to catabolize xanthine (Table 1).

453 Transcription levels of *carA*, putatively encoding part of carbamoyl phosphate synthetase,
454 appears to be a distinct trait of *Ca.* Poseidoniaceae (MGIIa): while all three MAGs contained this
455 gene (Table S1), only *Ca.* Poseidoniaceae *carA* transcription was high. Since carbamoyl
456 phosphate synthetase is a key enzyme for arginine and pyrimidine synthesis from bicarbonate,
457 high *carA* transcription suggests these pathways may be important components of *Ca.*
458 Poseidoniaceae growth or survival.

459

460 **CONCLUSIONS**

461 Our metatranscriptomic data and associated experiments provide a window into the
462 activity of *Ca.* Poseidoniales families (formerly "MGIIa" and "MGIIb"). They indicate an
463 important role for *Ca.* Thalassarchaeaceae (MGIIb) as coastal photoheterotrophs, particularly in

warm waters. High transcription of proteorhodopsin and membrane-bound pyrophosphatase genes suggested common methods for establishing proton gradients. Furthermore, high transcription of genes involved in protein/peptide metabolism and ß-oxidation of fatty acids confirmed peptide and lipid metabolism as a common trait. However, high transcription of *Ca.* Thalassarchaeaceae genes encoding amino acid binding proteins and nucleotide transporters suggests uptake of these substrates may distinguish the two families. These data confirm the importance of DOM metabolism by *Ca.* Poseidoniales and suggest a potential role for organic nitrogen in *Ca.* Thalassarchaeaceae metabolism.

ACKNOWLEDGMENTS

We thank Qian Liu, Bradley Tolar, the Georgia Coastal Ecosystems LTER field team, the University of Georgia Marine Institute at Sapelo Island staff, and the captain and crew of the R/V *Savannah* for assistance with original field sampling. This work was funded in part by NSF OCE grant 1538677 and OPP grant 1643466 (to JTH), NSF Grant IOS1656311 (to MAM), and NSF OCE grant 1832178 (to the Georgia Coastal Ecosystems LTER), as well as by resources and technical expertise from the Georgia Advanced Computing Resource Center, a partnership between the University of Georgia's Office of the Vice President for Research and Office of the Vice President for Information Technology.

The authors declare no competing financial interests.

REFERENCES

1. Fuhrman JA, McCallum K, Davis AA. Novel major archaebacterial group from marine plankton. Nature 1992;356:148–9.
2. DeLong EF. Archaea in coastal marine environments. Proceedings of the National Academy of Sciences 1992;89:5685–9.
3. Zhang CL, Xie W, Martín-Cuadrado A-B, Rodriguez-Valera F. Marine Group II Archaea, potentially important players in the global ocean carbon cycle. Frontiers in Microbiology 2015;6:1108.
4. Santoro AE, Richter RA, Dupont CL. Planktonic marine archaea. Annual Review of Marine Science 2019;11:131–58.

5. Béjà O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP, et al. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. Environmental Microbiology 2000;2:516–29.
6. Moreira D, Rodriguez-Valera F, López-García P. Analysis of a genome fragment of a deep-sea uncultivated Group II euryarchaeote containing 16S rDNA, a spectinomycin-like operon and several energy metabolism genes. Environmental Microbiology 2004;6:959–69.
7. Martín-Cuadrado A-B, Rodriguez-Valera F, Moreira D, Alba JC, Ivars-Martínez 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. The ISME Journal 2008;2:865–86.
8. Alonso C, Pernthaler J. Concentration-dependent patterns of leucine incorporation by coastal picoplankton. Applied and Environmental Microbiology 2006;72:2141–7.
9. Orsi WD, Smith JM, Wilcox HM, Swalwell JE, Carini P, Worden AZ, et al. Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. The ISME Journal 2015;9:1747–63.
10. 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. The ISME Journal 2016;10:2158–73.
11. 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.
12. Martín-Cuadrado A-B, Garcia-Heredia I, Gonzaga Moltó A, López-Ubeda R, Kimes N, López-Garcia P, et al. A new class of marine Euryarchaeota group II from the mediterranean deep chlorophyll maximum. The ISME Journal 2015;9:1619–34.
13. Thrash JC, Seitz KW, Baker BJ, Temperton B, Gillies LE, Rabalais NN, et al. Metabolic roles of uncultivated bacterioplankton lineages in the northern Gulf of Mexico "dead zone." mBio 2017;8:e01017–17.
14. 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. The ISME Journal 2019;13:3024–36.
15. Rinke C, Rubino F, Messer LF, Youssef N, Parks DH, Chuvochina M, et al. A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (*Ca.* Poseidoniales ord. nov.). The ISME Journal 2019;13:663–75.
16. Tully BJ. Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. Nature Communications 2019;10:271.
17. Baker BJ, Sheik CS, Taylor CA, Jain S, Bhasi A, Cavalcoli JD, et al. Community transcriptomic assembly reveals microbes that contribute to deep-sea carbon and nitrogen cycling. The ISME Journal 2013;7:1962–73.
18. Ottesen EA, Young CR, Gifford SM, Eppley JM, Marin R, Schuster SC, et al. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. Science 2014;345:207–12.
19. Miller D, Pfreundt U, Hou S, Lott SC, Hess WR, Berman-Frank I. Winter mixing impacts gene expression in marine microbial populations in the Gulf of Aqaba. Aquatic Microbial Ecology 2017;80:223–42.
20. Hou S, Pfreundt U, Miller D, Berman-Frank I, Hess WR. mdRNA-Seq analysis of marine microbial communities from the northern Red Sea. Scientific Reports 2016;6:35470.
21. 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. Nature Communications 2015;6:8933.
22. Bergauer K, Fernandez-Guerra A, Garcia JAL, Sprenger RR, Stepanauskas R, Pachiadaki MG, et al. Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. Proceedings of the National Academy of Sciences 2018;115:E400–8.
23. Vorobev A, Sharma S, Yu M, Lee J, Washington BJ, Whitman WB, et al. Identifying labile DOM components in a coastal ocean through depleted bacterial transcripts and chemical signals. Environmental Microbiology 2018;20:3012–30.

24. Liu Q, Tolar BB, Ross MJ, Cheek JB, Sweeney CM, Wallsgrove NJ, et al. Light and temperature control the seasonal distribution of thaumarchaeota in the South Atlantic bight. The ISME Journal 2018;12:1473–85.
25. Damashek J, Edwardson CF, Tolar BB, Gifford SM, Moran MA, Hollibaugh JT. Coastal ocean metagenomes and curated metagenome-assembled genomes from Marsh Landing, Sapelo Island (Georgia, USA). Microbiology Resource Announcements 2019;8:e00934–19.
26. Gifford SM, Sharma S, Moran MA. Linking activity and function to ecosystem dynamics in a coastal bacterioplankton community. Frontiers in Microbiology 2014;5:185.
27. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications 2018;9:5114.
28. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. Nature Microbiology 2016;1:16048.
29. 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.
30. Eddy SR. Accelerated profile HMM searches. PLoS Computational Biology 2011;7:e1002195.
31. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 2004;32:1792–7.
32. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS ONE 2010;5:e9490.
33. Santoro AE, Dupont CL, Richter RA, Craig MT, Carini P, McIlvin MR, et al. Genomic and proteomic characterization of "*Candidatus* Nitrosopelagicus brevis": An ammonia-oxidizing archaeon from the open ocean. Proceedings of the National Academy of Sciences 2015;112:1173–8.
34. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods 2012;9:357–9.
35. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–9.
36. Satinsky BM, Gifford SM, Crump BC, Moran MA. Use of Internal Standards for Quantitative Metatranscriptome and Metagenome Analysis. In: DeLong EF, editor. Methods in Enzymology Microbial Metagenomics, Metatranscriptomics, and Metaproteomics. Elsevier Inc; 2013. pages 237–50.
37. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. 2019;
38. de Mendiburu F. agricolae: Statistical procedures for agricultural research. R package version 1.3-1. https://CRAN.R-project.org/package=agricolae. 2019;
39. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community ecology package. R package version 2.5-5. https://CRAN.R-project.org/package=vegan. 2019;
40. Legendre P, Legendre L. Numerical Ecology. 3rd ed. Elsevier; 2012.
41. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Scientific Data 2018;5:170103.
42. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. Science 2015;348:1261359.
43. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–9.
44. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nature Methods 2014;12:59–60.
45. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Research 2016;44:D286–93.
46. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Mering von C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Molecular Biology and Evolution 2017;34:2115–22.
47. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. Journal of Molecular Biology 2016;428:726–31.

48. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research 2014;42:D490–5.
49. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Research 2018;46:W95–W101.
50. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Research 2017;46:D624–32.
51. Saier MH Jr, Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. The Transporter Classification Database (TCDB): recent advances. Nucleic Acids Research 2016;44:D372–9.
52. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 2014;15:550.
53. Damashek J, Tolar BB, Liu Q, Okotie Oyekan AO, Wallsgrove NJ, Popp BN, et al. Microbial oxidation of nitrogen supplied as selected organic nitrogen compounds in the South Atlantic Bight. Limnology and Oceanography 2019;64:982–95.
54. Massana R, Murray AE, Preston CM, DeLong EF. Vertical distribution and phylogenetic characterization of marine planktonic *Archaea* in the Santa Barbara Channel. Applied and Environmental Microbiology 1997;63:50–6.
55. Teira E, Reinthaler T, Pernthaler A, Pernthaler J, Herndl GJ. Combining catalyzed reporter deposition-fluorescence in situ hybridization and microautoradiography to detect substrate utilization by bacteria and archaea in the deep ocean. Applied and Environmental Microbiology 2004;70:4411–4.
56. 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). Limnology and Oceanography 2010;55:2117–25.
57. Legendre P. lmodel2: Model II regression. R package version 1.7-3. https://CRAN.R-project.org/package=lmodel2. 2018;
58. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer; 2016.
59. Man D, Wang W, Sabehi G, Aravind L, Post AF, Massana R, et al. Diversification and spectral tuning in marine proteorhodopsins. The EMBO Journal 2003;22:1725–31.
60. Gómez-Consarnau L, González JM, Coll-Lladó M, Gourdon P, Pascher T, Neutze R, et al. Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. Nature 2007;445:210–3.
61. Pinhassi J, DeLong EF, Beja O, González JM, Pedrós-Alio C. Marine bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. Microbiology and Molecular Biology Reviews 2016;80:929–54.
62. Church MJ, DeLong EF, Ducklow HW, Karner MB, Preston CM, Karl DM. Abundance and distribution of planktonic Archaea and Bacteria in the waters west of the Antarctic Peninsula. Limnology and Oceanography 2003;48:1893–902.
63. Herndl GJ, Reinthaler T, Teira E, van Aken H, Veth C, Pernthaler A, et al. Contribution of *Archaea* to total prokaryotic production in the deep Atlantic Ocean. Applied and Environmental Microbiology 2005;71:2303–9.
64. Galand PE, Lovejoy C, Hamilton AK, Ingram RG, Pedneault E, Carmack EC. Archaeal diversity and a gene for ammonia oxidation are coupled to oceanic circulation. Environmental Microbiology 2009;11:971–80.
65. Lincoln SA, Wai B, Eppley JM, Church MJ, Summons RE, DeLong EF. Planktonic Euryarchaeota are a significant source of archaeal tetraether lipids in the ocean. Proceedings of the National Academy of Sciences 2014;111:9858–63.
66. Liu H, Zhang CL, Yang C, Chen S, Cao Z, Zhang Z, et al. Marine Group II dominates planktonic archaea in water column of the northeastern South China Sea. Frontiers in Microbiology 2017;8:1098.
67. Mincer TJ, Church MJ, Taylor LT, Preston C, Karl DM, DeLong EF. Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. Environmental Microbiology 2007;9:1162–75.

68. Stoica E, Herndl GJ. Contribution of *Crenarchaeota* and *Euryarchaeota* to the prokaryotic plankton in the coastal northwestern Black Sea. Journal of Plankton Research 2007;29:699–706.

69. 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. Environmental Microbiology 2018;20:734–54.

70. Wang ZA, Cai W-J, Wang Y, Ji H. The southeastern continental shelf of the United States as an atmospheric CO_2 source and an exporter of inorganic carbon to the ocean. Continental Shelf Research 2005;25:1917–41.

71. Letourneau ML, Medeiros PM. Dissolved organic matter composition in a marsh-dominated estuary: response to seasonal forcing and to the passage of a hurricane. Journal of Geophysical Research: Biogeosciences 2019;124:1545–59.

72. Medeiros PM, Babcock Adams L, Seidel M, Castelao RM, Di Iorio D, Hollibaugh JT, et al. Export of terrigenous dissolved organic matter in a broad continental shelf. Limnology and Oceanography 2017;62:1718–31.

73. Herfort L, Schouten S, Abbas B, Veldhuis MJW, Coolen MJL, Wuchter C, et al. Variations in spatial and temporal distribution of Archaea in the North Sea in relation to environmental variables. FEMS Microbiology Ecology 2007;62:242–57.

74. 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. Proceedings of the National Academy of Sciences 2013;110:6004–9.

75. Bano N, Ruffin S, Ransom B, Hollibaugh JT. Phylogenetic composition of Arctic Ocean archaeal assemblages and comparison with Antarctic assemblages. Applied and Environmental Microbiology 2004;70:781–9.

76. Parada AE, Fuhrman JA. Marine archaeal dynamics and interactions with the microbial community over 5 years from surface to seafloor. The ISME Journal 2017;11:2510–25.

77. Pereira O, Hochart C, Auguet J-C, Debroas D, Galand PE. Genomic ecology of Marine Group II, the most common marine planktonic Archaea across the surface ocean. MicrobiologyOpen 2019;8:e852.

78. Béjà 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.

79. Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, et al. Microbial community gene expression in ocean surface waters. Proceedings of the National Academy of Sciences 2008;105:3805–10.

80. Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. Environmental Microbiology 2009;11:1358–75.

81. Satinsky BM, Crump BC, Smith CB, Sharma S, Zielinski BL, Doherty M, et al. Microspatial gene expression patterns in the Amazon River Plume. Proceedings of the National Academy of Sciences 2014;111:11085–90.

82. Alonso-Sáez L, Morán XAG, González JM. Transcriptional patterns of biogeochemically relevant marker genes by temperate marine bacteria. Frontiers in Microbiology 2020;11:465.

83. Arandia-Gorostidi N, González JM, Huete-Stauffer T, Ansari MI, Morán XAG, Alonso-Sáez L. Light supports cell-integrity and growth rates of taxonomically diverse coastal photoheterotrophs. Environmental Microbiology 2020;

84. Giovannoni SJ, Bibbs L, Cho J-C, Stapels MD, Desiderio R, Vergin KL, et al. Proteorhodopsin in the ubiquitous marine bacterium SAR11. Nature 2005;438:82–5.

85. Kimura H, Young CR, Martinez A, DeLong EF. Light-induced transcriptional responses associated with proteorhodopsin-enhanced growth in a marine flavobacterium. The ISME Journal 2011;5:1641–51.

86. Steindler L, Schwalbach MS, Smith DP, Chan F, Giovannoni SJ. Energy starved *Candidatus* Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. PLoS ONE 2011;6:e19725.

87. Song Y, Cartron ML, Jackson PJ, Davison PA, Dickman MJ, Zhu D, et al. Proteorhodopsin overproduction enhances the long-term viability of *Escherichia coli*. Applied and Environmental Microbiology 2020;86:e02087–19.

88. Baykov AA, Malinen AM, Luoto HH, Lahti R. Pyrophosphate-fueled Na+ and H+ transport in prokaryotes. Microbiology and Molecular Biology Reviews 2013;77:267–76.

89. Rinta-Kanto JM, Sun S, Sharma S, Kiene RP, Moran MA. Bacterial community transcription patterns during a marine phytoplankton bloom. Environmental Microbiology 2012;14:228–39.

90. Vila-Costa M, Sharma S, Moran MA, Casamayor EO. Diel gene expression profiles of a phosphorus limited mountain lake using metatranscriptomics. Environmental Microbiology 2013;15:1190–203.

91. Needham DM, Fuhrman JA. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. Nature Microbiology 2016;1:16005.

92. Hosie AHF, Poole PS. Bacterial ABC transporters of amino acids. Research in Microbiology 2001;152:259–70.

93. Anderson JJ, Oxender DL. *Escherichia coli* transport mutants lacking binding protein and other components of the branched-chain amino acid transport systems. Journal of Bacteriology 1977;130:384–92.

94. Montesinos ML, Herrero A, Flores E. Amino acid transport in taxonomically diverse cyanobacteria and identification of two genes encoding elements of a neutral amino acid permease putatively involved in recapture of leaked hydrophobic amino acids. Journal of Bacteriology 1997;179:853–62.

95. Walshaw DL, Reid CJ, Poole PS. The general amino acid permease of *Rhizobium leguminosarum* strain 3841 is negatively regulated by the Ntr system. FEMS Microbiology Letters 1997;152:57–64.

96. Berg GM, Shrager J, Glöckner G, Arrigo KR, Grossman AR. Understanding nitrogen limitation in *Aureococcus anophagefferens* (Pelagophyceae) through cDNA and qRT-PCR analysis. Journal of Phycology 2008;44:1235–49.

97. Wurch LL, Gobler CJ, Dyhrman ST. Expression of a xanthine permease and phosphate transporter in cultures and field populations of the harmful alga *Aureococcus anophagefferens*: tracking nutritional deficiency during brown tides. Environmental Microbiology 2014;16:2444–57.

98. Cooper JT, Sinclair GA, Wawrik B. Transcriptome analysis of *Scrippsiella trochoidea* CCMP 3099 reveals physiological changes related to nitrate depletion. Frontiers in Microbiology 2016;7:639.

99. Cunliffe M. Purine catabolic pathway revealed by transcriptomics in the model marine bacterium *Ruegeria pomeroyi* DSS-3. FEMS Microbiology Ecology 2016;92:fiv150.