ISME

ARTICLE   OPEN

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

Julian Damashek [1,3 ✉], Aimee Oyinlade Okotie-Oyekan[1,4], Scott Michael Gifford [2], Alexey Vorobev[1,5], Mary Ann Moran [1] and James Timothy Hollibaugh [1]

Marine Group II *Euryarchaeota* (*Candidatus* Poseidoniales), abundant but yet-uncultivated members of marine microbial communities, are thought to be (photo)heterotrophs that metabolize dissolved organic matter (DOM), such as lipids and peptides. However, little is known about their transcriptional activity. We mapped reads from a metatranscriptomic time series collected at Sapelo Island (GA, USA) 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 that *Ca.* Thalassarchaeaceae are important photoheterotrophs potentially linking DOM and nitrogen cycling in coastal waters.

ISME Communications (2021)1:5 ; https://doi.org/10.1038/s43705-021-00002-6

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 ocean have increased, several patterns have emerged: MGII is often highly abundant in the euphotic zone and in coastal waters, can reach high abundance following phytoplankton blooms, and largely comprised 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).[11–15] 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.[16,17] 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.[16] 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.[18,19] 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 euryarcheal community.[20,21] 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.[15] Finally, a metaproteomics study found abundant euryarcheal transport proteins for L-amino acids, branched-chain amino acids, and peptides throughout the Atlantic Ocean.[22] Despite these advances, little is known about the similarities or differences in gene transcription between *Ca.* Poseidoniaceae and Thalassarchaeaceae.

We report MAG-resolved metatranscriptomic analyses of *Ca.* Poseidoniales in coastal waters near Sapelo Island (GA, USA). Prior work suggested that *Ca.* Poseidoniales are sporadically active at Sapelo Island[23] and may comprise the majority of archea 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 of the clades were transcriptionally active and identifying highly or differentially transcribed genes. We used two Sapelo Island MAGs[25] combined with recent *Ca*. Poseidoniales MAG collections[16,17] to competitively recruit reads from a metatranscriptomic time series[26] and an incubation experiment[23] to determine which of the 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
The phylogenomic analysis compared SIMO Bins 19-2 and 31-1 (ref. [25]) to previously reported *Ca*. Poseidoniales MAGs, including those binned from TARA Oceans, Mediterranean Sea, Red Sea, Gulf of Mexico, Guaymas Basin, and Puget Sound metagenomes[11,13,15–17,27–32] (Supplementary Table S1). Average nucleotide identity (ANI) was calculated using fastANI[33] to compare nonredundant MAGs from Tully[17] to 15 Port Hacking MAGs[16] and the two SIMO MAGs; any with ANI < 98.5% were added to the nonredundant set. The phylogenomic analysis used sixteen ribosomal proteins[34] within anvi'o v4 (ref. [35]). All genomes were converted to contig databases, and proteins were identified using HMMER,[36] concatenated, aligned using MUSCLE,[37] and used to build a phylogenomic tree using FastTree[38] within anvi'o.

Competitive read mapping
We used competitive read mapping to determine which *Ca*. Poseidoniales genera were transcriptionally active in Sapelo Island metatranscriptomes. Surface water for all metatranscriptomes was collected seasonally in 2008, 2009, and 2014 from the dock at Marsh Landing, Sapelo Island (31°25′4.08 N, 81°17′43.26 W) as previously described.[23,26] Briefly, 3–8 L of surface water was pumped through a 3-μm pore-size prefilter followed by a 0.2-μm pore-size collection filter, which was frozen on liquid nitrogen. RNA was extracted from the collection filter as previously described,[23,26] including the addition of internal RNA standards to calculate volumetric transcript abundances.[39] 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$_0$, which were processed immediately after collection, and T$_{24}$, which were processed following 24 h of in situ incubation in dark containers, 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 July 16, 2020).

Contigs from all MAGs used for phylogenomics (Supplementary Table S1) were used as a read mapping database using Bowtie2 v.2.2.9 (ref. [40]) with the "very-sensitive" flag. Samtools v.1.3.1 (ref. [41]) was used to index BAM files, which were 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 refs. [23,26]) as previously described.[39] Seasonal transcript abundances were compared using a one-way ANOVA test in R[42] with data log-transformed to improve normality. Significant groupings were defined post hoc with Tukey's Honest Significant Difference (HSD) test using the agricolae R package.[43]

Nonmetric multidimensional scaling (NMDS) analysis of metatranscriptome hits was conducted using the vegan R package.[44] NMDS input was a distance matrix constructed by Hellinger transforming the table of transcript hits and calculating Euclidean distance between samples.[45] Genus vectors were calculated using the *envfit* command. The significance of groupings was tested by permutational multivariate analysis of variance (PERMANOVA; *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 previously binned from Red Sea metagenomes (RS440, PBUZ00000000;[27]). 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. [46]) using DIAMOND[47] to search against all orthologous groups in eggNOG-mapper v.1 (refs. [45,46]) and using the BlastKOALA portal (https://www.kegg.jp/blastkoala/, last accessed March 6, 2019).[48] Putative genes for carbohydrate-active enzymes, peptidases, and membrane transport proteins were identified using HMMER searches of dbCAN2 (HMMdb v.7),[49,50] MEROPS v.12.0 (ref. [51]), and the Transporter Classification Database,[52] 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 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. [53]) 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.

Quantitative PCR
DNA samples from SAB field campaigns in 2014 and 2017[24,54] 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] Supplementary Fig. S1) and multiple depths when possible. DNA from the entire microbial community (no prefiltration) was extracted as previously described.[54] Primers were GII-554-f[55] and Eury806-r[56] with cycling conditions as previously reported[57] (Supplementary Table S2). 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). The abundance of *Ca*. Poseidoniales 16S rRNA genes was compared to published bacterial 16S rRNA gene abundance from the same samples.[24,54] 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[58] as previously described.[54] All plots were constructed with anvi'o or the ggplot2 R package.[59]

RESULTS
Euryarcheotal 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; Supplementary Fig. S2). Phylogenomic groupings were generally consistent with previous findings.[16,17]

Both SIMO MAGs contained a proteorhodopsin gene. The presence of a methionine residue at position 315 suggested absorption of green light,[60] and both proteorhodopsin genes grouped in "Archaea Clade B"[11,17,61] (Supplementary Fig. S3). Both MAGs included partial or complete pathways indicating aerobic heterotrophic growth, such as glycolysis, the TCA cycle, and electron transport chain components (Supplementary

Fig. 1 Abundance of transcriptionally active *Ca*. Poseidoniales genera. A *Ca*. Poseidoniales genera in Sapelo Island metatranscriptomes ($n = 24$). The dendrogram (top) shows grouping by similarity. The season is indicated by the 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 the analysis of differential transcription. **B** NMDS of *Ca*. Poseidoniales metatranscriptome reads using a Hellinger distance matrix of genera relative abundances. Each point represents a sample, with color and shape denoting season and day/night, respectively. Shading indicates sampling year and original study. Arrows represent vectors for each genus. NMDS stress and the results of a PERMANOVA analyzing sums of squares by season are indicated. **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.

Table S3). Pathways for the 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}, more than other seasons (Fig. 1A, C and Supplementary Table S4). 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; Supplementary Fig. S4; and Supplementary Table S4). HT (though 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, B and Supplementary Table S4).

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), as expected by the fundamental roles of these pathways in cellular processes. Genes encoding proteins involved in energy production or conservation (ATPases, a membrane-bound pyrophosphatase, and proteorhodopsin) were also highly transcribed. Notably, the *aapJ* and *livK* genes, encoding ligand-binding receptors for L-amino acid and branch-chain amino acid transporters, respectively, were among the most highly transcribed genes in both *Ca*. Thalassarchaeaceae MAGs (Fig. 2 and Supplementary Table S3).

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 glycosyltransferase, 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 and Supplementary Table S3). While both *Ca*. Thalassarchaeaceae MAGs also contained the carbamoyl phosphate synthetase genes, neither transcribed *carA* at high levels (Supplementary Table S3).

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 and Supplementary Table S3). 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, membrane-bound pyrophosphatase (*hppA*), and proteorhodopsin (Fig. 2).

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 and Supplementary Table S5), 16 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 an 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 (Supplementary Fig. S5 and Supplementary Table S5). 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 glycosyltransferase, which all had significantly fewer transcripts in M-rich samples (Supplementary Fig. S6 and Supplementary Table S5). One gene of unknown function was enriched compared to other samples.

Dark-incubation metatranscriptomes

Incubation had little effect on transcription by the dominant genera in LT samples (Fig. 4 and Supplementary Table S4). 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 and Supplementary Table S4). 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 a ligand-binding receptor for a general amino acid transporter (*aapJ*, Fig. 5 and Supplementary Table S6). The 36 O1 genes transcribed at lower levels encoded proteins involved in the 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 O3 or M MAGs were transcribed at significantly different levels between HT incubation time points ($P > 0.1$ for all genes; Supplementary Table S6).

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} (Supplementary Table S7). 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%; Supplementary Table S2). 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 two to three orders of magnitude higher (Fig. 6b).

DISCUSSION

The abundance of *Ca*. Poseidoniales genes in the South Atlantic Bight

Given typical *Ca*. Poseidoniales abundance of 10^6–10^7 genes or cells L^{-1} in oligotrophic waters[62–64] and 10^7–10^8 genes or cells L^{-1} in coastal waters,[9,55,57,65–67] the abundance of their 16S rRNA genes in the coastal SAB (~10^9 genes L^{-1} in some samples) is among the highest measured in the ocean. Greater gene abundance in inshore, nearshore, and mid-shelf waters indicates that *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.[9,66] The DOM in productive, turbid SAB coastal waters supports highly active heterotrophic microbial populations.[68,69] Our data suggest that large populations of *Ca*. Poseidoniales are part of this heterotrophic community and the correlation between *Ca*. Poseidoniales and bacterial abundance (Fig. 6b) suggests that common factors influence the abundance of these two communities.

### Diversity of transcriptionally active *Ca*. Poseidoniales

While numerous studies have demonstrated a high abundance of *Ca*. Poseidoniales in the coastal ocean (see refs. [57,67]), 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 levels of aggregate *Ca*. Poseidoniales transcripts (Fig. 1), indicates that *Ca*. Thalassarchaeaceae are most 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

SIMO Bin 19–2 (O1; *Ca.* Thalassarchaeaceae)
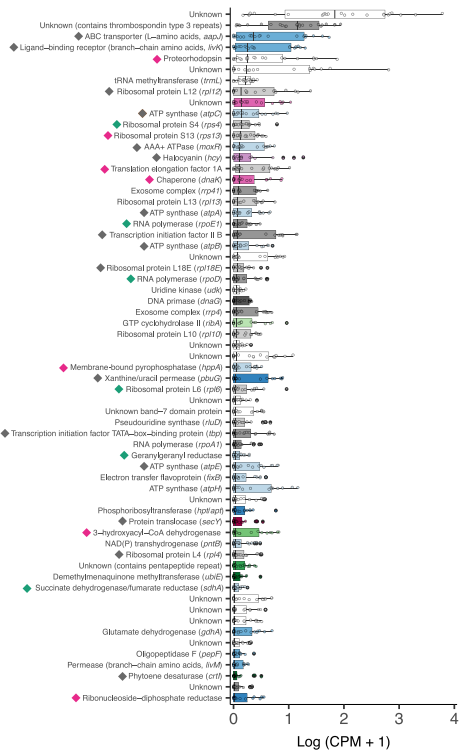
SIMO Bin 31–1 (O3; *Ca.* Thalassarchaeaceae)
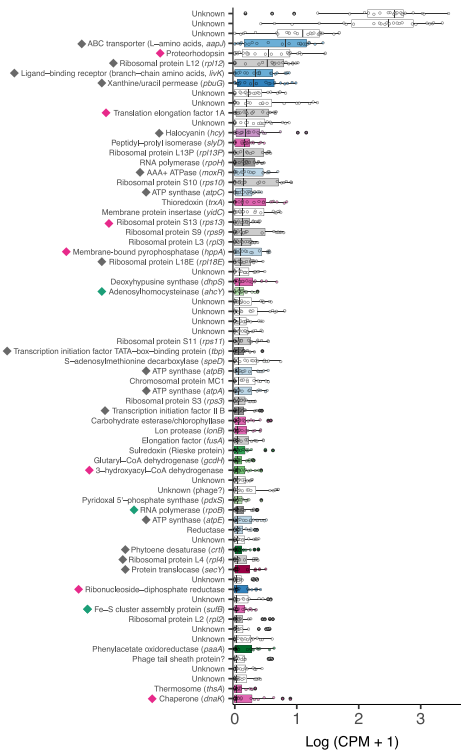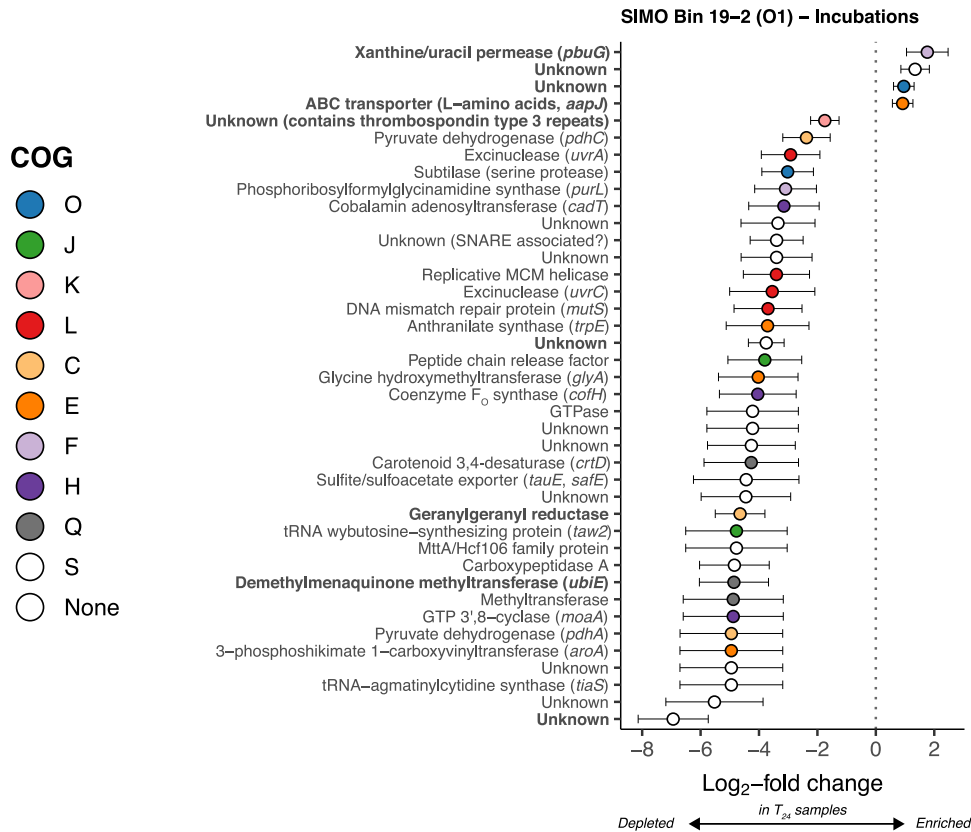
RS440 (M; *Ca.* Poseidoniaceae)

Highly transcribed in:
◆ All 3 MAGs
◆ *Ca.* Thalassarchaeaceae (MGIIb) MAGs only
◆ One *Ca.* Thalassarchaeaceae (MGIIb) and one *Ca.* Poseidoniaceae (MGIIa) MAG

COG
■ D: cell cycle control
■ M: cell wall/membrane biogenesis
■ O: post-translational/protein modification/chaperones
■ T: signal transduction
■ U: intracellular trafficking/secretion
■ J: translation
■ K: transcription
■ L: cell replication/repair
■ C: energy production/conservation
■ E: amino acid metabolism/transport
■ F: nucleotide metabolism/transport
■ G: carbohydrate metabolism/transport
■ H: coenzyme metabolism
■ I: lipid metabolism
■ P: inorganic ion transport/metabolism
■ Q: secondary structure
□ S: unknown function
□ None assigned

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).

J. Damashek et al.

Fig. 3 Log$_2$-fold change of SIMO Bin 19-2 (genus O1) genes differentially transcribed in field metatranscriptomes with high transcriptional activity of genus O1. Poseidoniales were dominated by genus O1 Calculated with DESeq2 (see Fig. 1 and Supplementary Table S4). 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).

Fig. 4 Comparisons of competitive read mapping to *Ca*. Poseidoniales genera at the beginning and end of 24-h incubations of Sapelo Island water conducted by Vorobev et al.[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.

salinity and DOM enriched in marine-origin molecules over the mid-shelf SAB during July 2014.[70] Our data suggest that *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 euryarcheal family in many coastal ecosystems, particularly in summer (see refs. [9,14,57,71,72]), 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 (see refs. [12,73–75]). 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.[17] 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

SIMO Bin 19–2 (O1) – Incubations

COG

- 🔵 O
- 🟢 J
- 🩷 K
- 🔴 L
- 🟡 C
- 🟠 E
- 🟣 F
- 🟪 H
- ⚫ Q
- ⚪ S
- ⚪ None

Genes (top to bottom):
Xanthine/uracil permease (*pbuG*); **Unknown**; **Unknown**; ABC transporter (L–amino acids, *aapJ*); Unknown (contains thrombospondin type 3 repeats); Pyruvate dehydrogenase (*pdhC*); Excinuclease (*uvrA*); Subtilase (serine protease); Phosphoribosylformylglycinamidine synthase (*purL*); Cobalamin adenosyltransferase (*cadT*); Unknown; Unknown (SNARE associated?); Unknown; Replicative MCM helicase; Excinuclease (*uvrC*); DNA mismatch repair protein (*mutS*); Anthranilate synthase (*trpE*); **Unknown**; Peptide chain release factor; Glycine hydroxymethyltransferase (*glyA*); Coenzyme F$_O$ synthase (*cofH*); GTPase; Unknown; Unknown; Carotenoid 3,4-desaturase (*crtD*); Sulfite/sulfoacetate exporter (*tauE, safE*); Unknown; **Geranylgeranyl reductase**; tRNA wybutosine–synthesizing protein (*taw2*); MttA/Hcf106 family protein; Carboxypeptidase A; **Demethylmenaquinone methyltransferase (*ubiE*)**; Methyltransferase; GTP 3',8–cyclase (*moaA*); Pyruvate dehydrogenase (*pdhA*); 3–phosphoshikimate 1–carboxyvinyltransferase (*aroA*); Unknown; tRNA–agmatinylcytidine synthase (*tiaS*); Unknown; **Unknown**

Log$_2$–fold change

Depleted ← *in T$_{24}$ samples* → *Enriched*

Fig. 5 Log$_2$-fold change of differentially abundant SIMO Bin 19-2 (genus O1) genes 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. The color indicates COG functional category (see Fig. 2). Bold indicates genes in the top 5% median transcript coverage across field metatranscriptomes (Fig. 2).

temperature (~23–30 °C[17]) may explain the unusual pattern found in SAB waters: these genera peaked in summer at Sapelo Island, when water temperatures were 29–30 °C (Supplementary Table S4), 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 and Supplementary Table S4).

(3) We identified genes mapping to *Ca.* Poseidoniales MAGs that were differentially transcribed at the beginning versus the end of dark incubations. Since dark incubation separates indigenous microbes from light and sources of short-lived substrates, transcription of the 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, indicating 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 DOM metabolism, transport/metabolism of amino acids and nucleotides, and basic energetic processes. Though the lack of a cultivated representative limits the analysis to computationally inferred functions, these data provide hypotheses regarding the activity of *Ca.* Poseidoniales 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 and 3). Proteorhodopsins consist of a retinal chromophore linked to a transmembrane protein and use light energy to pump protons across the cell membrane.[76] 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 coastal waters (see refs. [77–79]) and our data indicate coastal *Ca.* Poseidoniales conform to this pattern, consistent with recent evidence from other regions.[21,80,81] High transcription of proteorhodopsin supports the photoheterotrophic lifestyle hypothesized for *Ca.* Poseidoniales (see refs. [9,11,12,16,17]).

Since O1 proteorhodopsin transcript abundance did not differ between the beginning and end of dark incubations (Fig. 5), light may not regulate *Ca.* Thalassarchaeaceae proteorhodopsin transcription. However, depletion of O1 *crtD* (carotenoid 3,4-desaturase) transcripts during dark incubation (Fig. 5) suggests
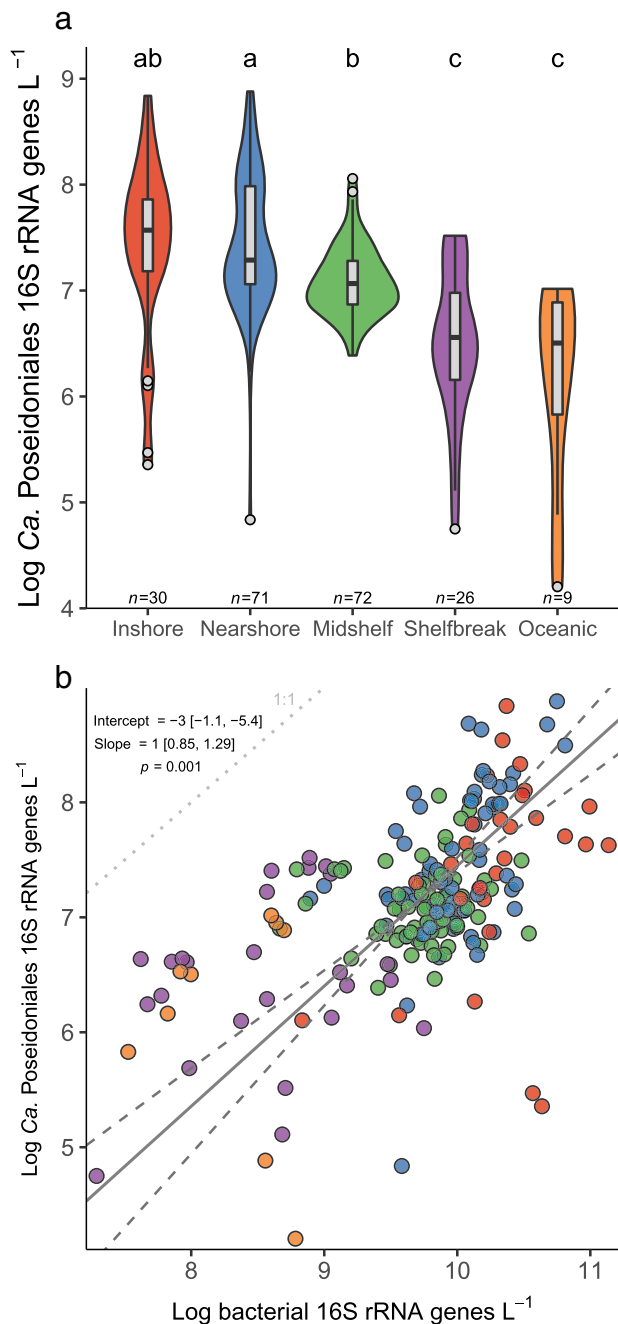
Fig. 6 Ca. Poseidoniales 16S rRNA gene abundance in the SAB. a Violin plots of log-transformed gene abundances across regions in the SAB, with overlain boxplots. The width of the violin plot corresponds to a data probability density. Color denotes sampling region. Letters above boxes denote post hoc 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.

light may regulate retinal synthesis. Whether proteorhodopsin transcription responds to light varies among marine bacteria (see refs. [82,83]) and the function of constitutive transcription is not straightforward: while some bacteria use proteorhodopsin to produce ATP when carbon-limited,[84] high amounts of proteorhodopsin in other bacteria can physically stabilize membranes even when inactive.[85] The high proteorhodopsin transcription in our

data emphasizes, but provides little mechanistic clarification of, the physiological role for proteorhodopsin in *Ca.* Poseidoniales (Table 1).

Numerous genes encoding putative electron transport proteins were highly transcribed by 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 suggest that 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 and Table 1). This gene (which is found across many domains of life,[86] including a diverse array of marine microbes[87]) putatively encodes a membrane-bound pyrophosphatase, which generates a proton or sodium ion gradient via hydrolysis of pyrophosphate, a by-product of numerous cellular processes.[86] In metatranscriptomes from a phytoplankton bloom, enrichment of *hppA* transcripts suggested high pyrophosphate-based energy conservation in oligotrophic waters,[87] and *hppA* transcripts were similarly abundant at night in an oligotrophic lake.[88] Although widespread in MAGs from *Ca.* Poseidoniales,[16] *hppA* has not been previously recognized as a potentially important part of their metabolism. Our data suggest that *Ca.* Poseidoniales may be capable of using pyrophosphatase (along with proteorhodopsin) to generate a protonmotive force (Table 1), conceivably providing an additional source of ATP.

The 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-h 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.[9,10,89] The HT Sapelo Island DOM pool is primarily of marine origin, while LT DOM is more riverine- and marsh-derived,[70] 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 that *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 may have been metabolizing proteins and fatty acids. High transcription of genes encoding proteases or peptidases from all *Ca.* Poseidoniales MAGs (Fig. 2) suggests the likely 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 may have been 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

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

Putative function	Relevant gene(s)	Distribution	Evidence
Proteorhodopsin	Proteorhodopsin gene	Both families	Highly transcribed (Fig. 2); enriched when genus 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 genus 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 genus 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 genus O1 active (Fig. 3); depletion during incubations (Fig. 5)
Amino acid/nucleotide synthesis	*carA*	*Ca.* Poseidoniaceae only	Highly transcribed (Fig. 2)
Carbohydrate synthesis	Family 2 glycosyltransferase	*Ca.* Poseidoniaceae only	Highly transcribed (Fig. 2)

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

high transcription of *Ca.* Poseidoniales peptidase genes in other marine regions.[15,21] 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 likely importance of fatty acid metabolism for both *Ca.* Poseidoniales families (Table 1), matching widespread ß-oxidation genes in *Ca.* Poseidoniales MAGs[16,17] and transcriptional data from the deep ocean.[15]

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.[90] Both are commonly present in *Ca.* Thalassarchaeaceae (MGIIb) but not *Ca.* Poseidoniaceae (MGIIa)[17] 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,[20,21] at the Mid-Cayman Rise,[15] 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 (Supplementary Table S3). 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[90], while LIV transporters can be highly specific for leucine, specific for branched-chain amino acids, or transport diverse amino acids.[90–92] In soil bacteria grown under inorganic nitrogen limitation, elevated transcription of *aapJ* is linked to organic nitrogen use,[93] but it is unclear whether this mechanism translates to *Ca.* Thalassarchaeaceae since amino acids could be used for numerous anabolic or catabolic processes. In addition to these binding proteins, the depletion of transcripts from O1 genes putatively involved in the shikimate pathway of aromatic amino acid synthesis (3-phosphoshikimate 1-carboxyvinyltransferase and anthranilate synthase) during incubations (Fig. 5) suggests that *Ca.* Thalassarchaeaceae in this genus may have been synthesizing aromatic amino acids in situ.

The combination of high *pbuG* transcription by *Ca.* Thalassarchaeaceae (MGIIb) with the high numbers of O1 *pbuG* transcripts in O1-active samples and dark-incubation endpoints (Figs. 2, 3, and 5) suggests an important role for xanthine/uracil permease (the putative product of *pbuG*) in *Ca.* Thalassarchaeaceae metabolism. In some phytoplanktons, *pbuG* is transcribed during nitrogen-stressed growth,[94–96] potentially allowing access to DON. However, *pbuG* and xanthine dehydrogenase (*xdh*) are also transcribed when xanthine is catabolized by marine bacteria.[97] Both *Ca.* Thalassarchaeaceae MAGs contain putative xanthine dehydrogenase genes (*xdhC* and *yagS*; Supplementary Table S3), suggesting the ability to catabolize xanthine (Table 1).

Transcription levels of *carA*, putatively encoding part of carbamoyl phosphate synthetase, appear to be a distinct trait of *Ca.* Poseidoniaceae (MGIIa): while all three MAGs contained this gene (Supplementary Table S3), only *Ca.* Poseidoniaceae *carA* transcription was high. Since carbamoyl phosphate synthetase is a key enzyme for arginine and pyrimidine synthesis from bicarbonate,[98] high *carA* transcription suggests these pathways may be important for *Ca.* Poseidoniaceae growth or survival, though it is not clear whether this transcription is related to the synthesis of amino acids, nucleotides, or both.

CONCLUSIONS

Our metatranscriptomic data and associated experiments provide a novel window into the activity of *Ca.* Poseidoniales families (formerly "MGIIa" and "MGIIb"). They indicate an 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.

REFERENCES

1. Fuhrman, J. A., McCallum, K. & Davis, A. A. Novel major archaebacterial group from marine plankton. *Nature* **356**, 148–149 (1992).
2. DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. USA* **89**, 5685–5689 (1992).
3. Zhang, C. L., Xie, W., Martín-Cuadrado, A.-B. & Rodriguez-Valera, F. Marine group II archaea, potentially important players in the global ocean carbon cycle. *Front. Microbiol.* **6**, 1108 (2015).
4. Santoro, A. E., Richter, R. A. & Dupont, C. L. Planktonic marine archaea. *Annu. Rev. Marine Sci.* **11**, 131–158 (2019).
5. Béjà, O. et al. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* **2**, 516–529 (2000).
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. *Environ. Microbiol.* **6**, 959–969 (2004).
7. Frigaard, N.-U., Martinez, A., Mincer, T. J. & DeLong, E. F. Proteorhodopsin lateral gene transfer between marine planktonic bacteria and archaea. *Nature* **439**, 847–850 (2006).
8. Alonso, C. & Pernthaler, J. Concentration-dependent patterns of leucine incorporation by coastal picoplankton. *Appl. Environ. Microbiol.* **72**, 2141–2147 (2006).
9. Orsi, W. D. et al. Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. *ISME J.* **9**, 1747–1763 (2015).
10. Orsi, W. D. et al. Diverse, uncultivated bacteria and archaea underlying the cycling of dissolved protein in the ocean. *ISME J.* **10**, 2158–2173 (2016).
11. Iverson, V. et al. Untangling genomes from metagenomes: revealing an uncultured class of marine *Euryarchaeota*. *Science* **335**, 587–590 (2012).
12. Martín-Cuadrado, A.-B. et al. A new class of marine Euryarchaeota group II from the Mediterranean deep chlorophyll maximum. *ISME J.* **9**, 1619–1634 (2015).
13. Thrash, J. C. et al. Metabolic roles of uncultivated bacterioplankton lineages in the northern Gulf of Mexico "dead zone." *mBio* **8**, e01017–17 (2017).
14. Orellana, L. H. et al. Niche differentiation among annually recurrent coastal Marine Group II Euryarchaeota. *ISME J.* **13**, 3024–3036 (2019).
15. Li, M. et al. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat. Commun.* **6**, 8933 (2015).
16. Rinke, C. et al. A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (*Ca.* Poseidoniales ord. nov.). *ISME J.* **13**, 663–675 (2019).
17. Tully, B. J. Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. *Nat. Commun.* **10**, 271 (2019).
18. Baker, B. J. et al. Community transcriptomic assembly reveals microbes that contribute to deep-sea carbon and nitrogen cycling. *ISME J.* **7**, 1962–1973 (2013).
19. Ottesen, E. A. et al. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* **345**, 207–212 (2014).
20. Miller, D. et al. Winter mixing impacts gene expression in marine microbial populations in the Gulf of Aqaba. *Aquat. Microb. Ecol.* **80**, 223–242 (2017).
21. Hou, S., Pfreundt, U., Miller, D., Berman-Frank, I. & Hess, W. R. mdRNA-Seq analysis of marine microbial communities from the northern Red Sea. *Sci. Rep.* **6**, 35470 (2016).
22. Bergauer, K. et al. Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proc. Natl. Acad. Sci. USA* **115**, E400–E408 (2018).
23. Vorobev, A. et al. Identifying labile DOM components in a coastal ocean through depleted bacterial transcripts and chemical signals. *Environ. Microbiol.* **20**, 3012–3030 (2018).
24. Liu, Q. et al. Light and temperature control the seasonal distribution of thaumarchaeota in the South Atlantic bight. *ISME J.* **12**, 1473–1485 (2018).
25. Damashek, J. et al. Coastal ocean metagenomes and curated metagenome-assembled genomes from Marsh Landing, Sapelo Island (Georgia, USA). *Microbiol. Resour. Announc.* **8**, e00934–19 (2019).
26. Gifford, S. M., Sharma, S. & Moran, M. A. Linking activity and function to ecosystem dynamics in a coastal bacterioplankton community. *Front. Microbiol.* **5**, 185 (2014).
27. Tully, B. J., Graham, E. D. & Heidelberg, J. F. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci. Data* **5**, 170103 (2018).
28. Tully, B. J., Sachdeva, R., Graham, E. D. & Heidelberg, J. F. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. *PeerJ* **5**, e3558 (2017).
29. Delmont, T. O. et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* **3**, 804–813 (2018).
30. Haroon, M. F., Thompson, L. R., Parks, D. H., Hugenholtz, P. & Stingl, U. A catalogue of 136 microbial draft genomes from Red Sea metagenomes. *Sci. Data* **3**, 160050 (2016).
31. Parks, D. H. et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542 (2017).
32. Haro-Moreno, J. M. et al. Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome* **6**, 128 (2018).
33. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90 K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
34. Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
35. Eren, A. M. et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
36. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
37. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
38. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
39. Satinsky, B. M., Gifford, S. M., Crump, B. C. & Moran, M. A. Use of internal standards for quantitative metatranscriptome and metagenome analysis. in *Methods in Enzymology Microbial Metagenomics, Metatranscriptomics, and Metaproteomics.* (ed. DeLong, E. F.) 237–250 (Elsevier, Inc., 2013).
40. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
41. Li, H. et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
42. R. Core Team. *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/ (R Foundation for Statistical Computing, Vienna, Austria, 2019).
43. de Mendiburu, F. agricolae: statistical procedures for agricultural research. R package version 1.3-1. https://CRAN.R-project.org/package=agricolae (2019).
44. Oksanen, J. et al. vegan: community ecology package. R package version 2.5-5. https://CRAN.R-project.org/package=vegan (2019).
45. Legendre, P. & Legendre, L. *Numerical Ecology*, 3rd edn. (Elsevier, 2012).
46. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
47. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2014).
48. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
49. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
50. Zhang, H. et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
51. Rawlings, N. D. et al. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **46**, D624–D632 (2017).
52. Saier, M. H. Jr et al. The transporter classification database (TCDB): recent advances. *Nucleic Acids Res.* **44**, D372–D379 (2016).
53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
54. Damashek, J. et al. Microbial oxidation of nitrogen supplied as selected organic nitrogen compounds in the South Atlantic Bight. *Limnol. Oceanogr.* **64**, 982–995 (2019).
55. Massana, R., Murray, A. E., Preston, C. M. & DeLong, E. F. Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl. Environ. Microbiol.* **63**, 50–56 (1997).
56. Teira, E., Reinthaler, T., Pernthaler, A., Pernthaler, J. & Herndl, G. J. Combining catalyzed reporter deposition-fluorescence in situ hybridization and micro-autoradiography to detect substrate utilization by bacteria and archaea in the deep ocean. *Appl. Environ. Microbiol.* **70**, 4411–4414 (2004).
57. Galand, P. E., Gutiérrez-Provecho, C., Massana, R., Gasol, J. M. & Casamayor, E. O. Inter-annual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). *Limnol. Oceanogr.* **55**, 2117–2125 (2010).
58. Legendre, P. lmodel2: model II regression. R package version 1.7-3. https://CRAN.R-project.org/package=lmodel2 (2018).
59. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
60. Man, D. et al. Diversification and spectral tuning in marine proteorhodopsins. *EMBO J.* **22**, 1725–1731 (2003).
61. Pinhassi, J., DeLong, E. F., Beja, O., González, J. M. & Pedrós-Alio, C. Marine bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. *Microbiol. Mol. Biol. Rev.* **80**, 929–954 (2016).

62. Herndl, G. J. et al. Contribution of *Archaea* to total prokaryotic production in the deep Atlantic Ocean. *Appl. Environ. Microbiol.* **71**, 2303–2309 (2005).

63. Liu, H. et al. Marine Group II dominates planktonic archaea in water column of the northeastern South China Sea. *Front. Microbiol.* **8**, 1098 (2017).

64. Church, M. J. et al. Abundance and distribution of planktonic Archaea and Bacteria in the waters west of the Antarctic Peninsula. *Limnol. Oceanogr.* **48**, 1893–1902 (2003).

65. Mincer, T. J. et al. Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ. Microbiol.* **9**, 1162–1175 (2007).

66. Stoica, E. & Herndl, G. J. Contribution of *Crenarchaeota* and *Euryarchaeota* to the prokaryotic plankton in the coastal northwestern Black Sea. *J. Plankton Res.* **29**, 699–706 (2007).

67. Xie, W. et al. Localized high abundance of Marine Group II archaea in the subtropical Pearl River Estuary: implications for their niche adaptation. *Environ. Microbiol.* **20**, 734–754 (2018).

68. Wang, Z. A., 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. *Cont. Shelf Res.* **25**, 1917–1941 (2005).

69. Letourneau, M. L. & Medeiros, P. M. Dissolved organic matter composition in a marsh-dominated estuary: response to seasonal forcing and to the passage of a hurricane. *J. Geophys. Res.: Biogeosci.* **124**, 1545–1559 (2019).

70. Medeiros, P. M. et al. Export of terrigenous dissolved organic matter in a broad continental shelf. *Limnol. Oceanogr.* **62**, 1718–1731 (2017).

71. Herfort, L. et al. Variations in spatial and temporal distribution of Archaea in the North Sea in relation to environmental variables. *FEMS Microbiol. Ecol.* **62**, 242–257 (2007).

72. Hugoni, M. et al. Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proc. Natl. Acad. Sci. USA* **110**, 6004–6009 (2013).

73. Bano, N., Ruffin, S., Ransom, B. & Hollibaugh, J. T. Phylogenetic composition of Arctic Ocean archaeal assemblages and comparison with Antarctic assemblages. *Appl. Environ. Microbiol.* **70**, 781–789 (2004).

74. Parada, A. E. & Fuhrman, J. A. Marine archaeal dynamics and interactions with the microbial community over 5 years from surface to seafloor. *ISME J.* **11**, 2510–2525 (2017).

75. Pereira, O., Hochart, C., Auguet, J.-C., Debroas, D. & Galand, P. E. Genomic ecology of Marine Group II, the most common marine planktonic Archaea across the surface ocean. *MicrobiologyOpen* **8**, e852 (2019).

76. Béjà, O. et al. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906 (2000).

77. Frias-Lopez, J. et al. Microbial community gene expression in ocean surface waters. *Proc. Natl. Acad. Sci. USA* **105**, 3805–3810 (2008).

78. Poretsky, R. S. et al. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ. Microbiol.* **11**, 1358–1375 (2009).

79. Satinsky, B. M. et al. Microspatial gene expression patterns in the Amazon River plume. *Proc. Natl. Acad. Sci. USA* **111**, 11085–11090 (2014).

80. Alonso-Sáez, L., Morán, X. A. G. & González, J. M. Transcriptional patterns of biogeochemically relevant marker genes by temperate marine bacteria. *Front. Microbiol.* **11**, 465 (2020).

81. Arandia-Gorostidi, N. et al. Light supports cell-integrity and growth rates of taxonomically diverse coastal photoheterotrophs. *Environ. Microbiol.* **22**, 3823–3837 (2020).

82. Giovannoni, S. J. et al. Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**, 82–85 (2005).

83. Gómez-Consarnau, L. et al. Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445**, 210–213 (2007).

84. Steindler, L., Schwalbach, M. S., Smith, D. P., Chan, F. & Giovannoni, S. J. Energy starved *Candidatus* Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS ONE* **6**, e19725 (2011).

85. Song, Y. et al. Proteorhodopsin overproduction enhances the long-term viability of *Escherichia coli*. *Appl. Environ. Microbiol.* **86**, e02087–19 (2020).

86. Baykov, A. A., Malinen, A. M., Luoto, H. H. & Lahti, R. Pyrophosphate-fueled Na^+ and H^+ transport in prokaryotes. *Microbiol. Mol. Biol. Rev.* **77**, 267–276 (2013).

87. Rinta-Kanto, J. M., Sun, S., Sharma, S., Kiene, R. P. & Moran, M. A. Bacterial community transcription patterns during a marine phytoplankton bloom. *Environ. Microbiol.* **14**, 228–239 (2012).

88. Vila-Costa, M., Sharma, S., Moran, M. A. & Casamayor, E. O. Diel gene expression profiles of a phosphorus limited mountain lake using metatranscriptomics. *Environ. Microbiol.* **15**, 1190–1203 (2013).

89. Needham, D. M. & Fuhrman, J. A. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat. Microbiol.* **1**, 16005 (2016).

90. Hosie, A. H. F. & Poole, P. S. Bacterial ABC transporters of amino acids. *Res. Microbiol.* **152**, 259–270 (2001).

91. Anderson, J. J. & Oxender, D. L. *Escherichia coli* transport mutants lacking binding protein and other components of the branched-chain amino acid transport systems. *J. Bacteriol.* **130**, 384–392 (1977).

92. Montesinos, M. L., 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. *J. Bacteriol.* **179**, 853–862 (1997).

93. Walshaw, D. L., Reid, C. J. & Poole, P. S. The general amino acid permease of *Rhizobium leguminosarum* strain 3841 is negatively regulated by the Ntr system. *FEMS Microbiol. Lett.* **152**, 57–64 (1997).

94. Berg, G. M., Shrager, J., Glöckner, G., Arrigo, K. R. & Grossman, A. R. Understanding nitrogen limitation in *Aureococcus anophagefferens* (Pelagophyceae) through cDNA and qRT-PCR analysis. *J. Phycol.* **44**, 1235–1249 (2008).

95. Wurch, L. L., Gobler, C. J. & Dyhrman, S. T. 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. *Environ. Microbiol.* **16**, 2444–2457 (2014).

96. Cooper, J. T., Sinclair, G. A. & Wawrik, B. Transcriptome analysis of *Scrippsiella trochoidea* CCMP 3099 reveals physiological changes related to nitrate depletion. *Front. Microbiol.* **7**, 639 (2016).

97. Cunliffe, M. Purine catabolic pathway revealed by transcriptomics in the model marine bacterium *Ruegeria pomeroyi* DSS-3. *FEMS Microbiol. Ecol.* **92**, fiv150 (2016).

98. Charlier, D., Le Minh, P. N. & Roovers, M. Regulation of carbamoylphosphate synthesis in *Escherichia coli*: an amazing metabolite at the crossroad of arginine and pyrimidine biosynthesis. *Amino Acids* **50**, 1647–1661 (2018).

ACKNOWLEDGEMENTS

COMPETING INTERESTS

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43705-021-00002-6.

Correspondence and requests for materials should be addressed to J.D.

Reprints and permission information is available at http://www.nature.com/reprints

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