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

Running title: Euryarchaeal transcriptomes 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], James Timothy 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 13502

(315) 223-2326; judamash@utica.edu

Version 2 for biorXiv

ABSTRACT

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 *Ca.* Thalassarchaeaceae are important photoheterotrophs potentially linking DOM and nitrogen cycling in coastal waters.

42 **INTRODUCTION**

43 Since the initial discovery of Marine Group II (MGII) *Euryarchaeota* [1,2], definitive
44 determination of their physiology and ecological roles has remained challenging due to the lack
45 of a cultivated isolate. Nonetheless, as data describing MGII distributions throughout the oceans
46 have increased, several patterns have emerged: MGII are often highly abundant in the euphotic
47 zone and in coastal waters, can reach high abundance following phytoplankton blooms, and are
48 largely comprised of two subclades, MGIIa and MGIIb [3,4]. Early metagenomic studies
49 provided the first evidence that MGII may be aerobic (photo)heterotrophs [5-7], a hypothesis
50 supported by incubation experiments [8-10] and by the gene content of diverse metagenome-
51 assembled genomes (MAGs) [11-15]. Two recent studies deepened our understanding of the
52 phylogenomics and metabolic potential of MGII by analyzing hundreds of MAGs, highlighting
53 clade-specific differences in genomic potential for transport and degradation of organic
54 molecules, light harvesting proteorhodopsins, and motility [16,17]. Here, we refer to MGII as the
55 putative order "*Candidatus* Poseidoniales," MGIIa and MGIIb as the putative families "*Ca.*
56 Poseidoniaceae" and "*Ca.* Thalassarchaeaceae," respectively, and putative genera as specified by
57 Rinke et al. [16]. We occasionally use "MGIIa" and "MGIIb" for consistency with previous
58 literature.

59 Metatranscriptomics is one strategy for gleaning information about microbial activity in
60 the environment. *Ca.* Poseidoniales transcripts can be abundant in marine metatranscriptomes,
61 suggesting transiently high transcriptional activity [18,19]. When metatranscriptome reads from
62 the Gulf of Aqaba were mapped to metagenomic contigs from the Mediterranean Sea, genes
63 involved in amino acid transport, carbon metabolism, and cofactor synthesis were highly
64 transcribed in the aggregate euryarchaeal community [20,21]. In another study, mapping deep-

65 sea metatranscriptome reads to novel *Ca.* Poseidoniales MAGs indicated transcription of genes
66 related to protein, fatty acid, and carbohydrate transport and metabolism, likely fueling aerobic
67 heterotrophy [15]. Finally, a metaproteomics study found abundant euryarchaeal transport
68 proteins for L-amino acids, branched-chain amino acids, and peptides throughout the Atlantic
69 Ocean [22]. Despite these advances, little is known about similarities or differences in gene
70 transcription between *Ca.* Poseidoniales and Thalassarchaeaceae.

71 We report MAG-resolved metatranscriptomic analyses of *Ca.* Poseidoniales in coastal
72 waters near Sapelo Island (GA, USA). Prior work suggested *Ca.* Poseidoniales are sporadically
73 active at Sapelo Island [23] and may comprise the majority of archaea in mid-shelf surface
74 waters of the South Atlantic Bight (SAB) [24]. Since other studies thoroughly described the
75 genomic content of *Ca.* Poseidoniales MAGs, our focus instead was determining which clades
76 were transcriptionally active and identifying highly or differentially transcribed genes. We used
77 two Sapelo Island MAGs [25] combined with recent *Ca.* Poseidoniales MAG collections [16,17]
78 to competitively recruit reads from a metatranscriptomic time series [26] and an incubation
79 experiment [23] to determine which clades were active over time. We then used representative
80 MAGs from highly active genera to determine which *Ca.* Poseidoniales genes were transcribed.
81 Finally, we used quantitative PCR (qPCR) to measure the abundance of *Ca.* Poseidoniales 16S
82 rRNA genes in DNA samples throughout the SAB to assess the prevalence *Ca.* Poseidoniales in
83 this region.

84

85 **MATERIALS AND METHODS**

86 PHYLOGENOMICS

87 Phylogenomic analyses compared SIMO Bins 19-2 and 31-1 (ref. 25) to previously-
88 reported *Ca.* Poseidoniales MAGs, including those binned from TARA Oceans, Mediterranean
89 Sea, Red Sea, Gulf of Mexico, Guaymas basin, and Puget Sound metagenomes [11,13,15-17,27-
90 32] (Table S1). Average nucleotide identity (ANI) was calculated using fastANI [33] to compare
91 non-redundant MAGs from Tully [17] to 15 Port Hacking MAGs [16] and the two SIMO MAGs;
92 any with ANI <98.5% were added to the non-redundant set. Phylogenomic analysis used sixteen
93 ribosomal proteins [34] within anvi'o v4 (ref. 35). All genomes were converted to contig
94 databases and proteins were identified using HMMER [36], concatenated, aligned using
95 MUSCLE [37], and used to build a phylogenomic tree using FastTree [38] within anvi'o.
96
97 COMPETITIVE READ MAPPING
98 We used competitive read mapping to determine which *Ca.* Poseidoniales genera were
99 transcriptionally active in Sapelo Island metatranscriptomes. Surface water for all
100 metatranscriptomes was collected seasonally in 2008, 2009, and 2014 from the dock at Marsh
101 Landing, Sapelo Island (31°25′4.08 N, 81°17′43.26 W) as previously described [23,26]. Briefly,
102 3-8 L of surface water was pumped through a 3-µm pore size prefilter followed by a 0.2-µm pore
103 size collection filter, which was frozen on liquid nitrogen. RNA from the free was extracted from
104 the collection filter as previously described [23,26], including the addition of internal RNA
105 standards to calculate volumetric transcript abundances [39]. Analyses of "field" communities
106 included Gifford et al. metatranscriptomes (iMicrobe Accession CAM_P_0000917) [26] and the
107 T_0 metatranscriptomes from Vorobev et al. [23], while "dark incubation" analyses included only
108 Vorobev et al. samples (T_0, which were processed immediately after collection, and T_{24}, which
109 were processed following 24 hours of *in situ* incubation in dark containers; NCBI BioProject

110 PRJNA419903). Temperature, salinity, dissolved oxygen, pH, and turbidity data corresponding
111 to metatranscriptome sampling times were downloaded from the NOAA National Estuarine
112 Research Reserve System website (http://cdmo.baruch.sc.edu; last accessed 16 July 2020).
113 Contigs from all MAGs used for phylogenomics (Table S1) were used as a read mapping
114 database using Bowtie2 v.2.2.9 (ref. 40) with the "very-sensitive" flag. Samtools v.1.3.1 (ref. 41)
115 was used to index BAM files, which were profiled and summarized in anvi'o. Contig genus
116 identity was imported to the anvi'o contig database as an external collection. The number of
117 transcripts L^{-1} was calculated by scaling the number of mapped reads by the volume of water
118 filtered and the recovery of internal standards (reported in [23,26]) as previously described [39].
119 Seasonal transcript abundances were compared using a one-way ANOVA test in R [42] with data
120 log-transformed as necessary to improve normality. When ANOVA results were significant,
121 groupings were defined post-hoc with Tukey's Honest Significant Difference (HSD) test using
122 the agricolae R package [43].
123 Non-metric multidimensional scaling (NMDS) analysis of metatranscriptome hits was
124 conducted using the vegan R package [44]. NMDS input was a distance matrix constructed by
125 Hellinger-transforming the table of transcript hits and calculating Euclidean distance between
126 samples [45]. Genus vectors were calculated using the *envfit* command. Significance of
127 groupings were tested by permutational multivariate analysis of variance (*adonis* command) with
128 999 permutations.
129
130 MAG-SPECIFIC ANNOTATION AND TRANSCRIPT ANALYSIS
131 Gene-specific analyses focused on three MAGs: two from the SIMO collection (SIMO
132 Bin 19-2, Genbank: VMDE00000000; SIMO Bin 31-1, VMBU00000000; [25]) and one

133 previously binned from Red Sea metagenomes (RS440, PBUZ00000000; [27]). These MAGs
134 represented genera O1, O3, and M, respectively, which were highly abundant in
135 metatranscriptomes (see Fig. 1). RS440 was selected due to a high number of transcripts
136 recruited when genus M was abundant (data not shown).

137 MAGs were annotated using the archaeal database in Prokka v.1.13 (ref. 46), using
138 DIAMOND [47] to search against all orthologous groups in eggNOG-mapper v.1 (refs. 45,46),
139 and using the BlastKOALA portal (https://www.kegg.jp/blastkoala/, last accessed 6 March 2019)
140 [48]. Putative genes for carbohydrate-active enzymes, peptidases, and membrane transport
141 proteins were identified using HMMER searches of dbCAN2 (HMMdb v.7) [49,50], MEROPS
142 v.12.0 (ref. 51), and the Transporter Classification Database [52], respectively.

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

150 DESeq2 v.3.11 (ref. 53) was used to identify genes from each MAG that were
151 differentially transcribed when each genus was highly transcriptionally active. For each MAG,
152 "treatment" samples in DESeq2 were those where the respective genus recruited ≥50% of *Ca.*
153 Poseidoniales reads from the metatranscriptome. Thus, positive fold-change values are genes
154 transcribed at higher levels when the genus is highly transcriptionally active (compared to other
155 metatranscriptomes). DESeq2 was also used to identify differentially transcribed genes for each

156 MAG between T_0 and T_{24} samples in high tide (HT) dark incubations [23]. Since T_{24} samples
157 were the "treatment" condition in DESeq2, positive fold-change values here are genes
158 transcribed at higher levels in T_{24} compared to T_0 samples. In all DESeq2 analyses, genes with
159 Benjamini-Hochberg adjusted $p<0.1$ were counted as having significantly different transcription.
160
161 QUANTITATIVE PCR
162 DNA samples from SAB field campaigns in 2014 and 2017 [24,54] were used as
163 templates for qPCR reactions targeting the *Ca.* Poseidoniales 16S rRNA gene. Samples included
164 the variety of shelf habitats (inshore, nearshore, mid-shelf, shelf-break, and oceanic as previously
165 defined [24]; Fig. S1) and multiple depths when possible. DNA from the entire microbial
166 community (no prefiltration) was extracted as previously described [54]. Primers were GII-554-f
167 [55] and Eury806-r [56] with cycling conditions as previously reported [57] (Table S2).
168 Reactions (25 µL, triplicate) used iTaq Universal Green SYBR Mix (Bio-Rad, Hercules, CA) in
169 a C1000 Touch Thermal Cycler/CFX96 Real-Time System (Bio-Rad, Hercules, CA). Each plate
170 included a no-template control and a standard curve (serial dilutions of a linearized plasmid
171 containing a previously-sequenced, cloned amplicon). Abundance of *Ca.* Poseidoniales 16S
172 rRNA genes was compared to published bacterial 16S rRNA gene abundance from the same
173 samples [24,54]. Regional variability of gene abundance was assessed using a one-way ANOVA
174 and a post-hoc HSD test as described above. Model II regressions of log-transformed qPCR data
175 were estimated using the lmodel2 R package [58] as previously described [54]. All plots were
176 constructed with anvi'o or the ggplot2 R package [59].
177
178

179 **RESULTS**

180 EURYARCHAEOTAL MAGs

181 SIMO Bins 19-2 and 31-1 were estimated as 82.5-92.3% and 77.5-96.2% complete,
182 respectively, with redundancy <0.6% [25]. Phylogenomics placed both in the putative family *Ca.*
183 Thalassarchaeaceae (MGIIb) and genera O1 (SIMO Bin 19-2) and O3 (SIMO Bin 31-1; Fig. S2).
184 Phylogenomic groupings were generally consistent with previous findings [16,17].

185 Both SIMO MAGs contained a proteorhodopsin gene. Presence of a methionine residue
186 at position 315 suggested absorption of green light [60], and both proteorhodopsin genes grouped
187 in "Archaea Clade B" [11,17,61] (Fig. S3). Both MAGs included partial or complete pathways
188 indicating aerobic heterotrophic growth, such as glycolysis, the TCA cycle, and electron
189 transport chain components (Table S3). Pathways for metabolism of compounds such as fatty
190 acids, peptides, and proteins were also present, as were transport systems and metabolic
191 pathways for amino acids and nucleotides.

192

193 DOMINANT GENERA IN FIELD METATRANSCRIPTOMES

194 There were significant seasonal differences in transcript recruitment by the combined set
195 of *Ca.* Poseidoniales MAGs ($F_{3,28}$=4.9, p=0.007): most summer samples had >10^{10} *Ca.*
196 Poseidoniales transcripts L^{-1}, significantly more than in other seasons (Fig. 1A,C; Table S4). The
197 diversity of transcriptionally-active *Ca.* Poseidoniales also changed seasonally. Genera O1 and
198 O3 accounted for most reads mapped from summer samples (typically 89.5-99.5% of *Ca.*
199 Poseidoniales reads), with most mapping to O1 (Fig. 1A,B; Fig. S4; Table S4). HT (and not low
200 tide; LT) metatranscriptomes from July 2014 also had a moderate fraction of reads (37.5-39.6%)
201 mapped to *Ca.* Poseidoniaceae. In contrast to summer samples, November 2008 and May 2009

202 transcripts were predominantly O3, while those from February 2009 and October 2014 were
203 mostly *Ca.* Poseidoniaceae genera M, L1, or L2 (Fig. 1A,B, Table S4).
204
205 HIGHLY TRANSCRIBED *CA.* POSEIDONIALES GENES
206 Many highly transcribed *Ca.* Thalassarchaeaceae (MGIIb) genes were involved in
207 translation, transcription, replication/repair, or post-translation protein modification (Fig. 2), as
208 expected by their fundamental role in cellular processes. Genes encoding proteins involved in
209 energy production or conservation (ATPases, a pyrophosphate-energized proton pump, and
210 proteorhodopsin) were also highly transcribed. Notably, the *aapJ* and *livK* genes, encoding
211 ligand-binding receptors for L-amino acid and branch-chain amino acid transporters,
212 respectively, were among the most highly transcribed genes in both *Ca.* Thalassarchaeaceae
213 MAGs (Fig. 2, Table S3).
214 Many of the highly transcribed genes mapping to the *Ca.* Poseidoniaceae (MGIIa) MAG
215 were not highly transcribed by *Ca.* Thalassarchaeaceae, including genes encoding a carbamoyl
216 phosphate synthetase subunit (*carA*), a family 2 glycosyl transferase, chromosomal protein
217 MC1b, and a ftsX-like permease. The *carA* gene had the highest median coverage of *Ca.*
218 Poseidoniaceae genes across coastal metatranscriptomes (Fig. 2, Table S3). While both *Ca.*
219 Thalassarchaeaceae MAGs also contained the carbamoyl phosphate synthetase genes, neither
220 transcribed *carA* at high levels (Table S3).
221 Twelve genes were highly transcribed by *Ca.* Thalassarchaeaceae and not by *Ca.*
222 Poseidoniaceae, including genes encoding ATP synthase, transcription initiation factor IIB,
223 halocyanin, phytoene desaturase, protein translocase, xanthine/uracil permease, and receptors for
224 amino acid transporters (Fig. 2, Table S3). Other than those encoding ribosomal proteins, only

225 six genes were highly transcribed in all three MAGs: a chaperone protein, a ribonucleoside-
226 diphosphate reductase, translation elongation factor 1A, 3-hydroxyacyl-CoA dehydrogenase, a
227 membrane-bound pyrophosphatase (*hppA*), and proteorhodopsin (Fig. 2).
228
229 DIFFERENTIAL GENE TRANSCRIPTION
230 We were interested in identifying genes with variable transcription levels when genera
231 O1, O3, and M were highly transcriptionally active in the ocean. Twenty-three genes were
232 differentially transcribed in O1-active samples (Fig. 3, Table S5), sixteen of which had higher
233 abundance in O1-active metatranscriptomes compared to other metatranscriptomes. These highly
234 transcribed genes encoded proteorhodopsin, two copper-containing redox proteins (halocyanin
235 and plastocyanin), and proteins involved in lipid metabolism (3-hydroxyacyl-CoA
236 dehydrogenase and oligosaccharyltransferase), nucleotide transport/metabolism (ribonucleotide-
237 diphosphate reductase and xanthine/uracil permease), and amino acid transport (ligand-binding
238 receptor for a L-amino acid transporter, *aapJ*). Differentially transcribed genes mapping to the
239 O3 MAG were mostly depleted in O3-active metatranscriptomes and largely encoded proteins of
240 unknown function; only the gene encoding ribosomal protein L12 was enriched in O3-active
241 samples (Fig. S5, Table S5). Only four genes mapping to the M MAG were differentially
242 transcribed in M-active metatranscriptomes. Annotated genes encoded chromosomal protein
243 MC1b, an ATPase subunit, and a glycosyl transferase, which all had significantly fewer
244 transcripts in M-rich samples (Fig. S6, Table S5). One gene of unknown function was enriched
245 compared to other samples.
246
247 DARK INCUBATION METATRANSCRIPTOMES

248 Incubation had little effect on transcription by the dominant genera in LT samples (Fig. 4,
249 Table S4). In contrast, there were distinct shifts in transcriptionally active populations during
250 incubations of all HT samples. July HT metatranscriptomes initially contained 60.4-62.5% *Ca.*
251 Thalassarchaeaceae (MGIIb) while hits from the corresponding T_{24} samples were 98.1-99.3%
252 *Ca.* Thalassarchaeaceae, due to increased transcript hits to genus O1 (Fig. 4, Table S4).
253 Likewise, October 2014 HT samples initially contained 65.0-66.6% hits to *Ca.* Poseidoniaceae
254 (MGIIa) but changed to 78.3-98.8% hits to *Ca.* Thalassarchaeaceae at T_{24} due to an increase in
255 hits to O1 (Fig. 4).
256 DESeq2 identified 40 differentially transcribed genes mapping to the O1 MAG between
257 HT T_0 and T_{24} metatranscriptomes. Four O1 genes had higher transcription at T_{24}, including
258 xanthine/uracil permease (*pbuG*) and a ligand-binding receptor for a general amino acid
259 transporter (*aapJ*; Fig. 5, Table S6). The 36 O1 genes transcribed at lower levels encoded
260 proteins involved in repair of UV-damaged DNA, amino acid or nucleotide metabolism,
261 coenzyme synthesis, peptidases or proteases, transcription, DNA replication, and lipid
262 biosynthesis, as well as phytoene desaturase (*crtD*) and multiple subunits of pyruvate
263 dehydrogenase (*pdhC*, *pdhA*). None of the genes mapping to the O3 or M MAGs were
264 transcribed at significantly different levels between HT incubation timepoints (p>0.1 for all
265 genes; Table S6).
266
267 16S rRNA GENE ABUNDANCE
268 *Ca.* Poseidoniales 16S rRNA genes were detected in all SAB DNA samples (n=208),
269 with a range from 1.6×10^4 to 7.6×10^8 genes L^{-1} (Table S7). Standard curves for the *Ca.*
270 Poseidoniales assay always had r^2>0.99 (mean ± standard deviation: 0.99±0.001) and the mean

271 efficiency was 93.4% (±2.0%; Table S2). When data from all cruises were combined, *Ca.*
272 Poseidoniales genes were most abundant throughout inshore and nearshore waters and least
273 abundant in shelf-break and oceanic samples ($F_{4,204}$=18.5, p<0.001; Fig. 6A). There was a strong
274 linear relationship between log-transformed bacterial and *Ca.* Poseidoniales 16S rRNA gene
275 abundances, with bacterial abundances 2-3 orders of magnitude higher (Fig. 6B).
276
277 **DISCUSSION**
278 ABUNDANCE OF *CA.* POSEIDONIALES GENES IN THE SOUTH ATLANTIC BIGHT
279 Given typical *Ca.* Poseidoniales abundance of 10^6-10^7 genes or cells L^{-1} in oligotrophic
280 waters [62-64] and 10^7-10^8 genes or cells L^{-1} in coastal waters [9,55,57,65-67], the abundance of
281 their 16S rRNA genes in the coastal SAB (nearly 10^9 genes L^{-1} in some samples) is among the
282 highest measured in the ocean. Greater gene abundance in inshore, nearshore, and mid-shelf
283 waters indicates *Ca.* Poseidoniales are more abundant over the shallow shelf than further
284 offshore (Fig. 6A), matching clone library data from the SAB [24] and data from the Central
285 California Current and the Black Sea [9,66]. The DOM in productive, turbid SAB coastal waters
286 supports highly active heterotrophic microbial populations [68,69]. Our data suggest large
287 populations of *Ca.* Poseidoniales are part of this heterotrophic community, and the correlation
288 between *Ca.* Poseidoniales and bacterial abundance (Fig. 6B) suggests common factors influence
289 the abundance of these two communities.
290
291 DIVERSITY OF TRANSCRIPTIONALLY ACTIVE *CA.* POSEIDONIALES
292 While numerous studies have demonstrated high abundance of *Ca.* Poseidoniales in the
293 coastal ocean (e.g., [57,67]), little is known about which clades are transcriptionally active in

294 these regions. At Sapelo Island, the striking dominance of the *Ca.* Thalassarchaeaceae (MGIIb)
295 genera O1 and O3 in summer samples, which also contained the highest levels of aggregate *Ca.*
296 Poseidoniales transcripts (Fig. 1), indicates that *Ca.* Thalassarchaeaceae are most active during
297 the summertime. Outliers to this pattern were July 2014 HT samples, which contained abundant
298 *Ca.* Poseidoniaceae (MGIIa) transcripts (though *Ca.* Thalassarchaeaceae still comprised the
299 majority of their *Ca.* Poseidoniales reads). A previous study found relatively high salinity and
300 DOM enriched in marine-origin molecules over the mid-shelf SAB during July 2014 [70]. Our
301 data suggest *Ca.* Poseidoniaceae were relatively active over the shelf during this time and were
302 transported inshore during flood tides, leading to shifts in transcriptional diversity between LT
303 waters (dominated by *Ca.* Thalassarchaeaceae) and HT waters (which included a higher number
304 of *Ca*. Poseidoniaceae).

305 *Ca.* Poseidoniaceae (MGIIa) are the predominant euryarchaeal family in many coastal
306 ecosystems, particularly in summer (e.g., [9,14,57,71,72]), but it is unclear what general patterns
307 govern *Ca.* Poseidoniales distributions in coastal waters worldwide. Studies of *Ca.* Poseidoniales
308 ecology often focus on distributions with depth, typically finding abundant *Ca.*
309 Thalassarchaeaceae (MGIIb) in deeper waters and *Ca.* Poseidoniaceae (MGIIa) more prevalent
310 in euphotic waters (e.g., [12,55,73-75]). A recent mapping of global ocean metagenome reads
311 showed that coastal populations of *Ca.* Poseidoniales were primarily *Ca.* Poseidoniaceae, though
312 *Ca.* Thalassarchaeaceae MAGs recruited a substantial number of reads from some coastal
313 metagenomes [17]. Our data match this latter pattern, with *Ca.* Poseidoniales populations in
314 surface waters off Sapelo Island dominated by highly active *Ca.* Thalassarchaeaceae (Fig. 1).
315 The higher abundance of MAGs from genera O1 and O3 (also referred to as MGIIb.12 and
316 MGIIb.14) in some mesopelagic and coastal samples with relatively high temperature (~23–

317 30°C; [17]) may explain the unusual pattern found in SAB waters: these genera peaked in
318 summer at Sapelo Island, when water temperatures were 29–30°C (Table S4), suggesting they
319 may be adapted to growth at relatively low light and high temperature.
320
321 *CA.* POSEIDONIALES GENE TRANSCRIPTION PATTERNS
322 Sapelo Island metatranscriptome reads that mapped to *Ca.* Poseidoniales were analyzed
323 in three ways:
324 (1) We determined sets of "highly transcribed" genes mapping to MAGs of
325 transcriptionally-active *Ca* Poseidoniales genera (5% of MAG genes with the highest
326 median transcript coverage);
327 (2) We identified genes mapping to *Ca* Poseidoniales MAGs that were differentially
328 transcribed when its genus was highly active (≥50% of *Ca*. Poseidoniales transcripts
329 in a sample; Fig. 1, Table S4);
330 (3) We identified genes mapping to *Ca* Poseidoniales MAGs that were differentially
331 transcribed at the beginning versus end of dark incubations. Since dark incubation
332 separates indigenous microbes from light and sources of short-lived substrates,
333 transcription of corresponding transporters and metabolic genes ceases during
334 incubations as substrates are consumed. We therefore assume transcript depletion in
335 T_{24} compared to T_0 metatranscriptomes indicates genes that were transcriptionally
336 active in the field [23]. This interpretation was bolstered by significant transcript
337 depletions for genes involved in repairing UV damage to DNA (*uvrA*, *uvrC*, and
338 *cofH*; Fig. 5), an expected result given alleviation of UV stress in dark incubations.
339

340 In the following sections, we synthesize these analyses to discuss *Ca.* Poseidoniales
341 transcriptional activity related to DOM metabolism, transport/metabolism of amino acids and
342 nucleotides, and basic energetic processes. Though lack of a cultivated representative limits the
343 analysis to computationally-inferred functions, these data provide hypotheses regarding the
344 activity of *Ca.* Poseidoniales families in the coastal ocean.
345
346 PROTON GRADIENTS AND ELECTRON TRANSPORT
347 Our analysis revealed that *Ca* Poseidoniales genes involved in establishing
348 transmembrane proton gradients were highly transcribed in our samples. Genes encoding
349 proteorhodopsin were among the most highly transcribed by both *Ca.* Thalassarchaeaceae MAGs
350 and were highly transcribed in O1-active samples (Figs. 2,3). Proteorhodopsins consist of a
351 retinal chromophore linked to a transmembrane protein and use light energy to pump protons
352 across the cell membrane [76]. The resulting energy can be coupled to ATP production or other
353 chemiosmotic processes and often supports photoheterotrophy, though its function varies widely
354 [61]. Proteorhodopsin genes are highly transcribed in the photic zone of both open ocean and
355 coastal waters (e.g., [77-79]) and our data indicate coastal *Ca.* Poseidoniales conform to this
356 pattern, consistent with recent evidence from other regions [21,80,81]. High transcription of
357 proteorhodopsin supports the photoheterotrophic lifestyle hypothesized for *Ca.* Poseidoniales
358 (e.g., [9,11,12,16,17]).
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 (e.g., [82,83]), and the function of
364 constitutive transcription is not straightforward: while some bacteria use proteorhodopsin to
365 produce ATP when carbon-limited [84], high amounts of proteorhodopsin in other bacteria can
366 physically stabilize membranes even when inactive [85]. 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).

369 Numerous genes encoding putative electron transport proteins were highly transcribed by
370 at least one *Ca.* Poseidoniales family (Fig. 2). Like proteorhodopsin, a halocyanin gene was
371 among the most highly transcribed *Ca.* Thalassarchaeaceae genes and was enriched when genus
372 O1 was transcriptionally active (Fig. 3). Halocyanins are involved in the electron transport chain
373 and have been posited to increase the energy yield of aerobic respiration in *Ca.*
374 Thalassarchaeaceae to stimulate rapid growth [12]. The similar transcription patterns of
375 proteorhodopsin and halocyanin suggests proteorhodopsin activity in coastal *Ca.*
376 Thalassarchaeaceae may function to increase growth rates during respiration, aiding rapid
377 population growth when conditions permit.

378 The *hppA* gene from all *Ca* Poseidoniales MAGs was highly transcribed (Fig. 2; Table 1).
379 This gene (which is found across many domains of life [86], including a diverse array of marine
380 microbes [87]) putatively encodes a membrane-bound pyrophosphatase, which generates a
381 proton gradient via hydrolysis of pyrophosphate, a byproduct of numerous cellular processes
382 [86]. In metatranscriptomes from a phytoplankton bloom, enrichment of *hppA* transcripts
383 suggested high pyrophosphate-based energy conservation in oligotrophic waters [87], and *hppA*
384 transcripts were similarly abundant at night in an oligotrophic lake [88]. Although widespread in
385 MAGs from *Ca.* Poseidoniales [16], *hppA* has not been previously recognized as a potentially

386 important part of their metabolism. Our data suggest *Ca.* Poseidoniales may be capable of using
387 pyrophosphatase (along with proteorhodopsin) to generate a protonmotive force (Table 1).
388
389 POTENTIAL IMPORTANCE OF MARINE DOM IN *CA.* POSEIDONIALES METABOLISM
390 Although the T_0 metatranscriptomes from summer versus fall were dominated by
391 transcripts from different *Ca.* Poseidoniales families, a 24-hour dark incubation consistently
392 favored transcription by *Ca.* Thalassarchaeaceae when samples were collected at HT (Fig. 4).
393 This tidal stage-linked increase in *Ca.* Thalassarchaeaceae transcription could relate to
394 differences in DOM availability between HT and LT, consistent with numerous studies
395 implicating DOM in shaping *Ca.* Poseidoniales populations [9,10,89]. The HT Sapelo Island
396 DOM pool is primarily of marine origin while LT DOM is more riverine- and marsh-derived
397 [70], which may select for growth of different *Ca.* Poseidoniales families in the water masses
398 present at different tidal stages. Furthermore, the depletion of transcripts encoding two pyruvate
399 dehydrogenase subunits (*pdhA* and *pdhC*) during incubations (Fig. 5) suggests *Ca.* Poseidoniales
400 were metabolizing phytoplankton photosynthate *in situ*. Alternatively, these tidal stage-driven
401 transcriptional patterns may relate to differential light adaptation in populations originating in
402 offshore versus nearshore waters. Inshore populations, potentially adapted to life in turbid
403 waters, may increase transcription upon dark enclosure, whereas offshore populations
404 (transported shoreward during flood tide) may be adapted to clearer waters and reduce
405 transcription in dark conditions.
406 Multiple lines of evidence indicate coastal *Ca.* Poseidoniales may have been
407 metabolizing proteins and fatty acids. High transcription of genes encoding proteases or
408 peptidases from all *Ca.* Poseidoniales MAGs (Fig. 2) suggests likely metabolism of proteins or

409 peptides by both families (Table 1). Furthermore, decreased transcription of protease genes
410 mapping to the O1 MAG during dark incubations (Fig. 5) suggests protein metabolism by *Ca.*
411 Thalassarchaeaceae may have been active *in situ* prior to incubation. While some of these genes
412 could be involved in intracellular recycling (particularly lon protease and cytosol
413 aminopeptidase), active protein metabolism is consistent with previous experiments
414 demonstrating protein assimilation [10] and high transcription of *Ca.* Poseidoniales peptidase
415 genes in other marine regions [15,21]. In addition to genes encoding protein catabolism, high
416 transcription of the 3-hydroxyacyl-CoA dehydrogenase gene from all three MAGs (Fig. 2), and
417 its enrichment in O1-active field samples (Fig. 3), suggests a likely importance of fatty acid
418 metabolism for both *Ca.* Poseidoniales families (Table 1), matching widespread ß-oxidation
419 genes in *Ca.* Poseidoniales MAGs [16,17] and transcriptional data from the deep ocean [15].
420

421 DISTINCT PATTERNS OF AMINO ACID AND NUCLEOTIDE UPTAKE AND
422 METABOLISM

423 Transcription of *livK* and *aapJ* appears to differentiate *Ca.* Poseidoniales families in the
424 coastal ocean (Table 1). These genes putatively encode ligand-binding receptors for ABC
425 transporters: *aapJ* for a general L-amino acid transporter and *livK* for a branched-chain amino
426 acid transporter [90]. Both are commonly present in *Ca.* Thalassarchaeaceae (MGIIb) but not *Ca.*
427 Poseidoniaceae (MGIIa) [17] and were among the most highly transcribed *Ca.*
428 Thalassarchaeaceae genes (Fig. 2). Previous studies noted high transcription of euryarchaeal *livK*
429 and *aapJ* genes in the water column of the Red Sea [20,21], at the Mid-Cayman Rise [15], and
430 throughout the Atlantic Ocean [22]. Our data suggest this activity was probably associated with
431 *Ca.* Thalassarchaeaceae.

432 The *aapJ* and *livK* genes were collocated with genes putatively encoding the full
433 transporters in the *Ca.* Thalassarchaeaceae MAGs (Table S3). Unfortunately, it is difficult to
434 guess their substrates from sequence data alone: AAP transporters are typically capable of
435 transporting a range of L-amino acids [90] while LIV transporters can be highly specific for
436 leucine, specific for branched-chain amino acids, or transport diverse amino acids [90-92]. In soil
437 bacteria grown under inorganic nitrogen limitation, elevated transcription of *aapJ* is linked to
438 organic nitrogen use [93], but it is unclear whether this mechanism translates to *Ca.*
439 Thalassarchaeaceae since amino acids could be used for numerous anabolic or catabolic
440 processes. In addition to these binding proteins, the depletion of transcripts from O1 genes
441 putatively involved in the shikimate pathway of aromatic amino acid synthesis (3-
442 phosphoshikimate 1-carboxyvinyltransferase and anthranilate synthase) during incubations (Fig.
443 5) suggests *Ca.* Thalassarchaeaceae in this genus may have been synthesizing aromatic amino
444 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 [94-96], potentially allowing access to DON. However, *pbuG* and
450 xanthine dehydrogenase (*xdh*) are also transcribed when xanthine is catabolized by marine
451 bacteria [97]. Both *Ca.* Thalassarchaeaceae MAGs contain putative xanthine dehydrogenase
452 genes (*xdhC* and *yagS*; Table S3), 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 S3), only *Ca.* Poseidoniaceae *carA* transcription was high. Since carbamoyl
456 phosphate synthetase is a key enzyme for arginine and pyrimidine synthesis from bicarbonate
457 [98], high *carA* transcription suggests these pathways may be important for *Ca.* Poseidoniaceae
458 growth or survival, though it is not clear whether this transcription is related to synthesis of
459 amino acids, nucleotides, or both.
460

461 **CONCLUSIONS**

462 Our metatranscriptomic data and associated experiments provide a novel window into the
463 activity of *Ca.* Poseidoniales families (formerly "MGIIa" and "MGIIb"). They indicate an
464 important role for *Ca.* Thalassarchaeaceae (MGIIb) as coastal photoheterotrophs, particularly in
465 warm waters. High transcription of proteorhodopsin and membrane-bound pyrophosphatase
466 genes suggested common methods for establishing proton gradients. Furthermore, high
467 transcription of genes involved in protein/peptide metabolism and ß-oxidation of fatty acids
468 confirmed peptide and lipid metabolism as a common trait. However, high transcription of *Ca.*
469 Thalassarchaeaceae genes encoding amino acid binding proteins and nucleotide transporters
470 suggests uptake of these substrates may distinguish the two families. These data confirm the
471 importance of DOM metabolism by *Ca.* Poseidoniales and suggest a potential role for organic
472 nitrogen in *Ca.* Thalassarchaeaceae metabolism.
473

474 **ACKNOWLEDGMENTS**

475 We thank Qian Liu, Bradley Tolar, the Georgia Coastal Ecosystems LTER field team, the
476 University of Georgia Marine Institute at Sapelo Island staff, and the captain and crew of the
477 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), NSF OCE grant 1832178 (to the Georgia Coastal Ecosystems LTER), as well as by resources and 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.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

REFERENCES

1. Fuhrman JA, McCallum K, Davis AA. Novel major archaebacterial group from marine plankton. Nature 1992;356:148–149.
2. DeLong EF. Archaea in coastal marine environments. Proceedings of the National Academy of Sciences 1992;89:5685–5689.
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;11:131–158.
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;2:516–529.
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;6:959–969.
7. Frigaard N-U, Martinez A, Mincer TJ, DeLong EF. Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. Nature 2006;439:847–850.
8. Alonso C, Pernthaler J. Concentration-dependent patterns of leucine incorporation by coastal picoplankton. Applied and Environmental Microbiology 2006;72:2141–2147.
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;9:1747–1763.
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;10:2158–2173.
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;335:587–590.
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;9:1619–1634.

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:e01017–7.

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;13:3024–3036.

15. 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.

16. 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;13:663–675.

17. Tully BJ. Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. Nature Communications 2019;10:271.

18. 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;7:1962–1973.

19. 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;345:207–212.

20. 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;80:223–242.

21. 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.

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–E408.

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;20:3012–3030.

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;12:1473–1485.

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. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Scientific Data 2018;5:170103.

28. Tully BJ, Sachdeva R, Graham ED, Heidelberg JF. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. PeerJ 2017;5:e3558.

29. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nature Microbiology 2018;3:804–813.

30. Haroon MF, Thompson LR, Parks DH, Hugenholtz P, Stingl U. A catalogue of 136 microbial draft genomes from Red Sea metagenomes. Scientific Data 2016;3:160050.

31. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Ben J Woodcroft, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nature Microbiology 2017;2:1533–1542.

564 32. Haro-Moreno JM, López Pérez M, la Torre de JR, Picazo A, Camacho A, Rodriguez-Valera F. Fine
565 metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and
566 recruitment. Microbiome 2018;6:128.

567 33. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of
568 90K prokaryotic genomes reveals clear species boundaries. Nature Communications 2018;9:5114.

569 34. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of
570 life. Nature Microbiology 2016;1:16048.

571 35. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis
572 and visualization platform for 'omics data. PeerJ 2015;3:e1319.

573 36. Eddy SR. Accelerated profile HMM searches. PLoS Computational Biology 2011;7:e1002195.

574 37. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic
575 Acids Research 2004;32:1792–1797.

576 38. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large
577 alignments. PLoS ONE 2010;5:e9490.

578 39. Satinsky BM, Gifford SM, Crump BC, Moran MA. Use of Internal Standards for Quantitative
579 Metatranscriptome and Metagenome Analysis. In: DeLong EF, editor. *Methods in Enzymology Microbial*
580 *Metagenomics, Metatranscriptomics, and Metaproteomics*. Elsevier, Inc.; 2013. pages 237–250.

581 40. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods 2012;9:357–359.

582 41. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format
583 and SAMtools. Bioinformatics 2009;25:2078–2079.

584 42. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical
585 Computing, Vienna, Austria. https://www.R-project.org/. 2019.

586 43. de Mendiburu F. agricolae: Statistical procedures for agricultural research. R package version 1.3-1.
587 https://CRAN.R-project.org/package=agricolae. 2019.

588 44. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community ecology
589 package. R package version 2.5-5. https://CRAN.R-project.org/package=vegan. 2019.

590 45. Legendre P, Legendre L. Numerical Ecology. 3rd ed. Elsevier; 2012.

591 46. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–2069.

592 47. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nature Methods
593 2014;12:59–60.

594 48. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional
595 Characterization of Genome and Metagenome Sequences. Journal of Molecular Biology 2016;428:726–
596 731.

597 49. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes
598 database (CAZy) in 2013. Nucleic Acids Research 2014;42:D490–D495.

599 50. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated
600 carbohydrate-active enzyme annotation. Nucleic Acids Research 2018;46:W95–W101.

601 51. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of
602 proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the
603 PANTHER database. Nucleic Acids Research 2017;46:D624–D632.

604 52. Saier MH Jr, Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. The Transporter Classification
605 Database (TCDB): recent advances. Nucleic Acids Research 2016;44:D372–D379.

606 53. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with
607 DESeq2. Genome Biology 2014;15:550.

608 54. Damashek J, Tolar BB, Liu Q, Okotie Oyekan AO, Wallsgrove NJ, Popp BN, et al. Microbial oxidation of
609 nitrogen supplied as selected organic nitrogen compounds in the South Atlantic Bight. Limnology and
610 Oceanography 2019;64:982–995.

55. 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;63:50–56.

56. 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;70:4411–4414.

57. 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;55:2117–2125.

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. New York: Springer; 2016.

60. 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;22:1725–1731.

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;80:929–954.

62. 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;71:2303–2309.

63. 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.

64. 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;48:1893–1902.

65. 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;9:1162–1175.

66. 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.

67. 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;20:734–754.

68. 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;25:1917–1941.

69. 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;124:1545–1559.

70. 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;62:1718–1731.

71. 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;62:242–257.

72. 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;110:6004–6009.

73. 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;70:781–789.

74. Parada AE, Fuhrman JA. Marine archaeal dynamics and interactions with the microbial community over 5 years from surface to seafloor. The ISME Journal 2017;11:2510–2525.

75. 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.

76. 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;289:1902–1906.

77. 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;105:3805–3810.

78. 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;11:1358–1375.

79. 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;111:11085–11090.

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

81. 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;22:3823-3837.

82. Giovannoni SJ, Bibbs L, Cho J-C, Stapels MD, Desiderio R, Vergin KL, et al. Proteorhodopsin in the ubiquitous marine bacterium SAR11. Nature 2005;438:82–85.

83. 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;445:210–213.

84. 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.

85. 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.

86. Baykov AA, Malinen AM, Luoto HH, Lahti R. Pyrophosphate-fueled Na^+ and H^+ transport in prokaryotes. Microbiology and Molecular Biology Reviews 2013;77:267–276.

87. Rinta-Kanto JM, Sun S, Sharma S, Kiene RP, Moran MA. Bacterial community transcription patterns during a marine phytoplankton bloom. Environmental Microbiology 2012;14:228–239.

88. Vila-Costa M, Sharma S, Moran MA, Casamayor EO. Diel gene expression profiles of a phosphorus limited mountain lake using metatranscriptomics. Environmental Microbiology 2013;15:1190–1203.

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

90. Hosie AHF, Poole PS. Bacterial ABC transporters of amino acids. Research in Microbiology 2001;152:259–270.

91. 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;130:384–392.

705 92. Montesinos ML, Herrero A, Flores E. Amino acid transport in taxonomically diverse cyanobacteria and
706 identification of two genes encoding elements of a neutral amino acid permease putatively involved in
707 recapture of leaked hydrophobic amino acids. Journal of Bacteriology 1997;179:853–862.

708 93. Walshaw DL, Reid CJ, Poole PS. The general amino acid permease of *Rhizobium leguminosarum* strain
709 3841 is negatively regulated by the Ntr system. FEMS Microbiology Letters 1997;152:57–64.

710 94. Berg GM, Shrager J, Glöckner G, Arrigo KR, Grossman AR. Understanding nitrogen limitation in
711 *Aureococcus anophagefferens* (Pelagophyceae) through cDNA and qRT-PCR analysis. Journal of
712 Phycology 2008;44:1235–1249.

713 95. Wurch LL, Gobler CJ, Dyhrman ST. Expression of a xanthine permease and phosphate transporter in
714 cultures and field populations of the harmful alga *Aureococcus anophagefferens*: tracking nutritional
715 deficiency during brown tides. Environmental Microbiology 2014;16:2444–2457.

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

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

720 98. Charlier D, Le Minh PN, Roovers M. Regulation of carbamoylphosphate synthesis in *Escherichia coli*: an
721 amazing metabolite at the crossroad of arginine and pyrimidine biosynthesis. Amino Acids 2018;50:1647–
722 1661.

723

724

725 **TABLES AND FIGURES**

726 **Table 1** Transcriptional traits shared or distinct among *Ca.* Poseidoniales 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).

727

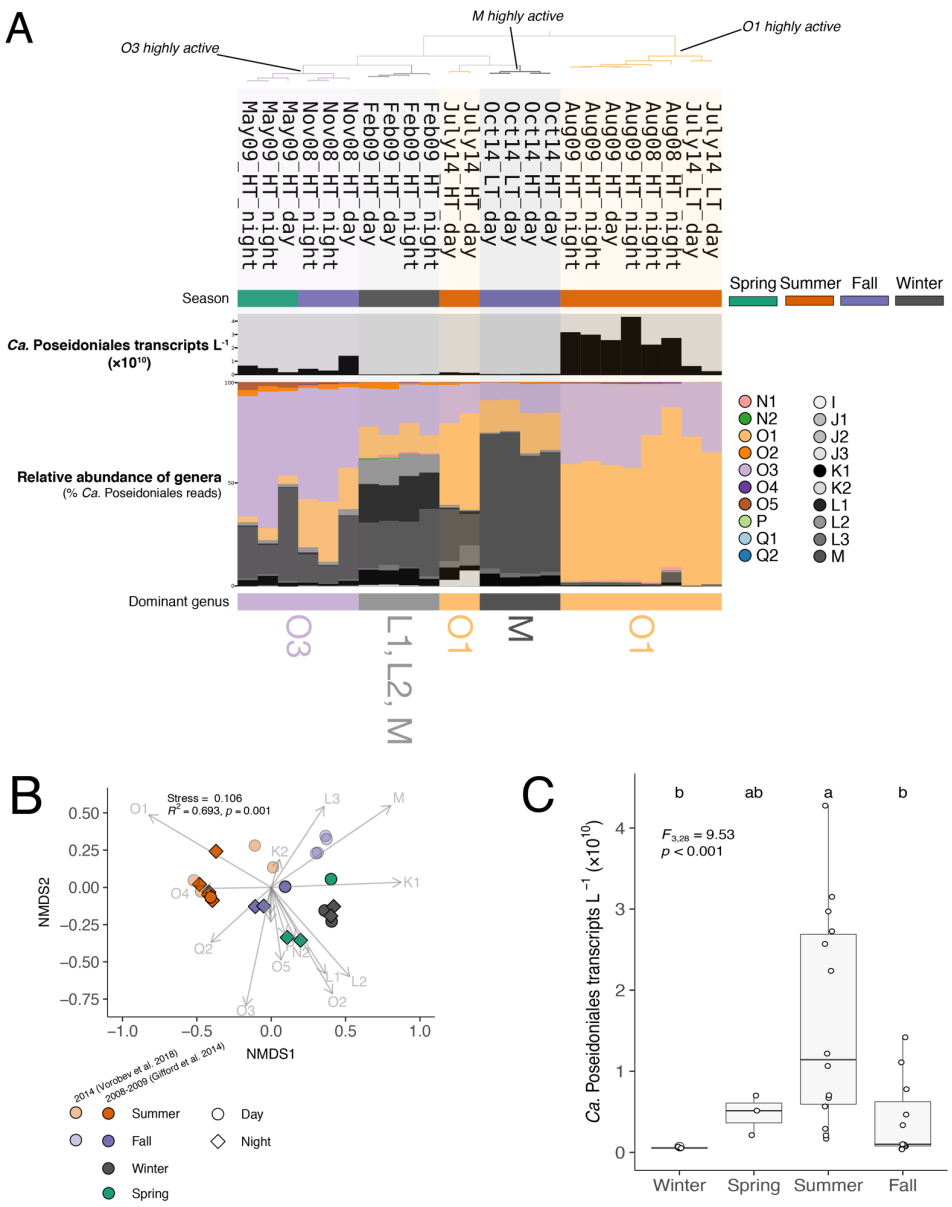

Fig. 1 A) Relative abundance of *Ca.* Poseidoniales genera in Sapelo Island metatranscriptomes (*n*=24). The dendrogram (top) shows grouping by similarity. Season is indicated by color beneath sample names. The bar chart shows the abundance of *Ca.* Poseidoniales transcripts L^{-1} and the stacked bar 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)** 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.

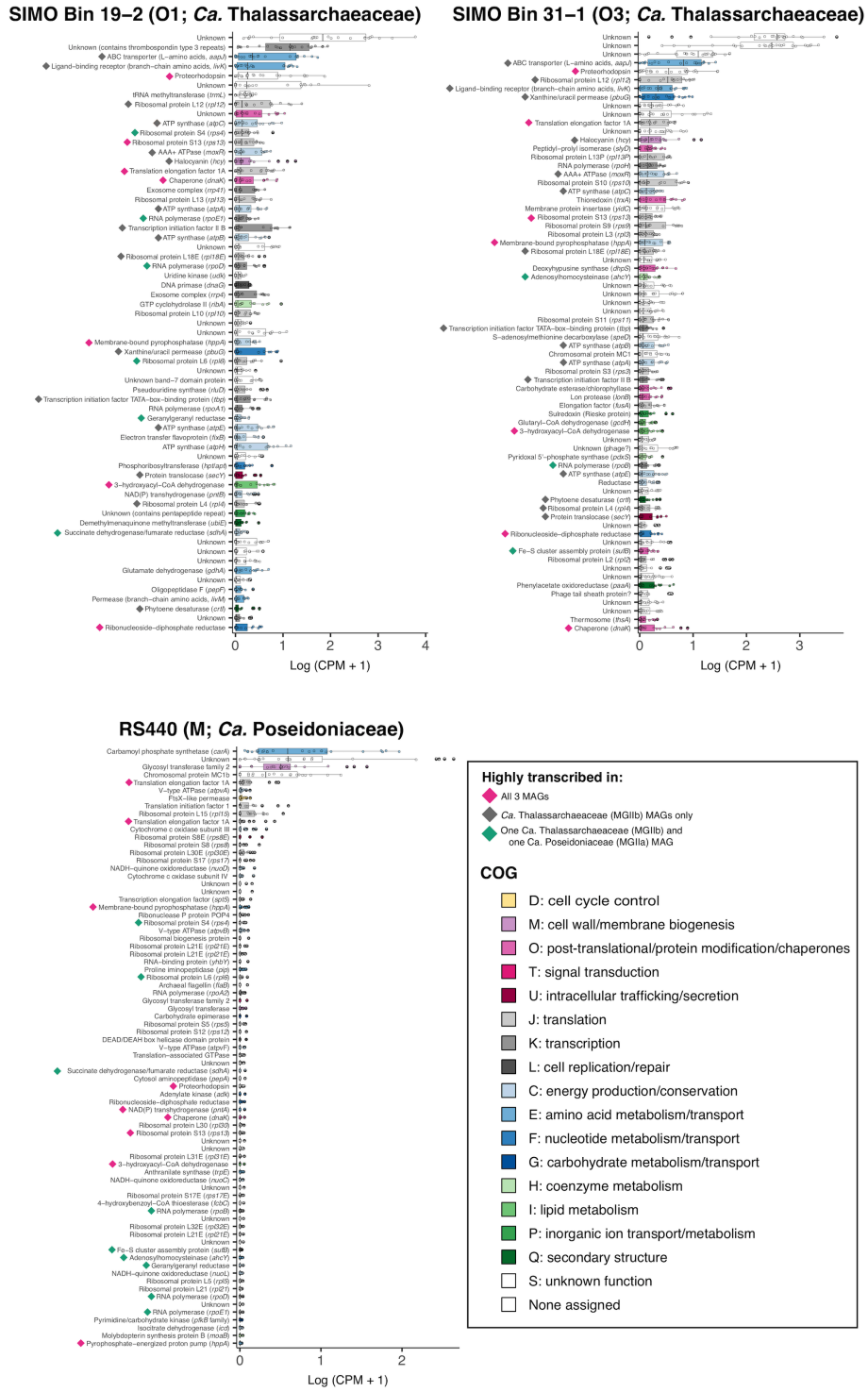

Fig. 2 Boxplots of highly-transcribed MAG genes (top 5%) in Sapelo Island field metatranscriptomes. Overlain points show CPM for individual metatranscriptomes (*n*=24). Shading indicates COG functional category assigned by eggNOG-mapper (genes assigned to group S were similar to proteins of unknown function 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).

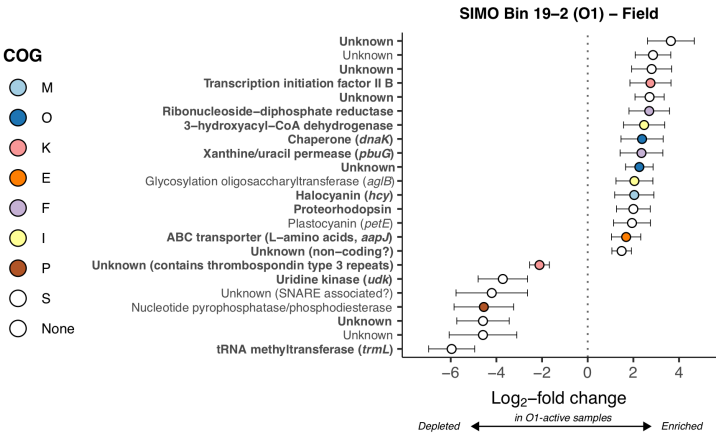

Fig. 3 Log_2-fold change of SIMO Bin 19-2 (genus O1) genes differentially transcribed in field metatranscriptomes where transcriptional activity of *Ca.* Poseidoniales was dominated by genus O1 (see Fig. 1, Table S4), calculated with DESeq2. Error bars show estimated standard error. Only genes with adjusted *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).

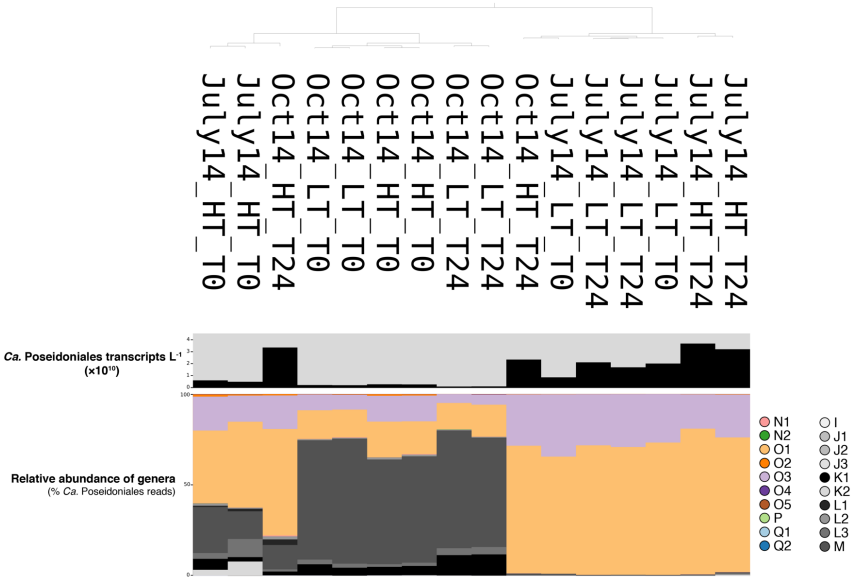

Fig. 4 Comparisons of competitive read mapping to *Ca.* Poseidoniales genera at the beginning and end of 24-hour incubations of Sapelo Island water conducted by Vorobev et al. [23]. The dendrogram (top) shows grouping by similarity. The bar chart below the dendrogram is the abundance of *Ca.* Poseidoniales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

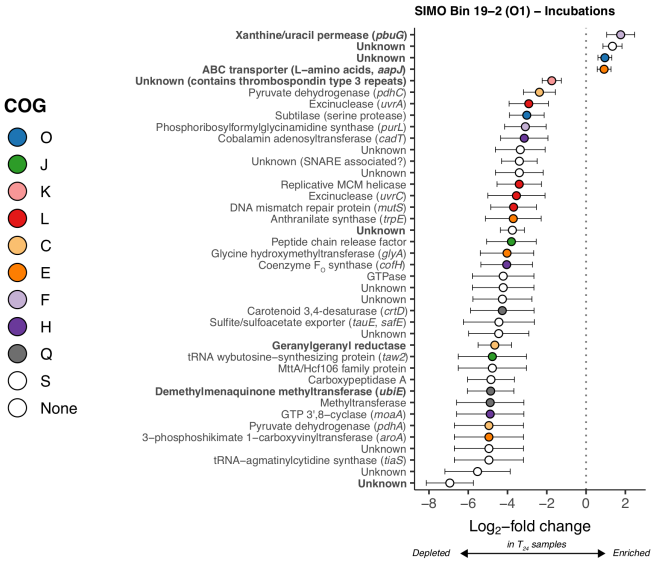

Fig. 5 Log_2-fold change of SIMO Bin 19-2 (genus O1) genes differentially abundant in T_{24} versus T_0 metatranscriptomes from Sapelo Island high tide waters. Error bars show estimated standard error. Only genes with adjusted *p*-values<0.1 are shown. Color indicates COG functional category (see Fig.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

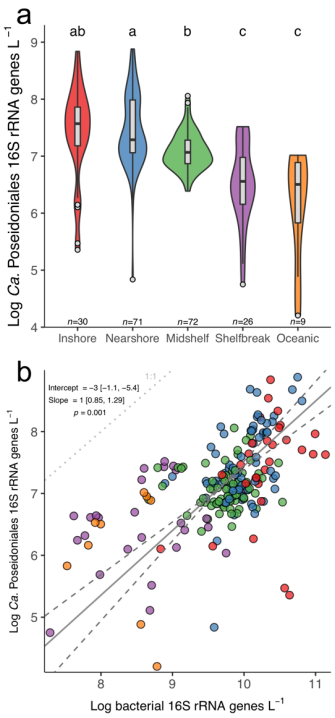

Fig. 6 A) Violin plots of log-transformed *Ca.* Poseidoniales 16S rRNA gene abundances across regions in the SAB, with overlain boxplots. Width of the violin plot corresponds to data probability density. Color denotes sampling region. Letters above boxes denote post-hoc grouping according to Tukey's 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.