Differential gene expression supports a resource-intensive, defensive role for colony production in the bloom-forming haptophyte, *Phaeocystis globosa*

Margaret Mars Brisbin[a] and Satoshi Mitarai[a]

a Marine Biophysics Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

Correspondence
M. Mars Brisbin, Marine Biophysics Unit, Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa, Japan
e-mail: margaret.marsbrisbin@oist.jp

ABSTRACT

Phaeocystis globosa forms dense, monospecific blooms in temperate, northern waters. Blooms are usually dominated by the colonial morphotype—non-flagellated cells embedded in a secreted mucilaginous mass. Colonial *Phaeocystis* blooms significantly affect food-web structure and function and negatively impact fisheries and aquaculture, but factors initiating colony production remain enigmatic. Destructive *P. globosa* blooms have been reported in tropical and subtropical regions more recently and warm-water blooms could become more common with continued climate change and coastal eutrophication. We therefore assessed genetic pathways associated with colony production by investigating differential gene expression between colonial and solitary cells in a warm-water *Phaeocystis globosa* strain. Our results illustrate a transcriptional shift in colonial cells with most of the differentially expressed genes downregulated, supporting a reallocation of resources associated with colony production. Dimethylsulfide and acrylate production and pathogen interaction pathways were upregulated in colonial cells, suggesting a defensive role for colony production. We identify several protein kinase signaling pathways that may influence the transition between morphotypes, providing targets for future research into factors triggering colony production. This study provides novel insights into genetic mechanisms involved in *Phaeocystis* colony formation and provides new evidence supporting a defensive role for *Phaeocystis* colonies.

Keywords
Plankton; phytoplankton; DMS; DMSP; algal bloom; algae bloom; RNA-seq; transcriptome; transcriptomics; colonial morphotype

INTRODUCTION

Phaeocystis is a cosmopolitan bloom-forming haptophyte genus encompassing 6 species (Andersen et al. 2015; Schoemann et al 2005). Most *Phaeocystis* species (*P. globosa, P. antarctica, P. pouchetii*, and *P. jahnii*) exhibit a polymorphic life-cycle, alternating between colonial and free-living morphotypes. *Phaeocystis* blooms are usually dominated by the colonial morphotype and are typically very dense, produce large biomasses, and impact food-web structure and function (Schoemann et al. 2005). *Phaeocystis* is a major contributor to dimethylsulfoniopropionate (DMSP) and dimethylsulfide (DMS) production globally (Liss et al.

1994) and regional peaks in DMS production are closely correlated with colonial *Phaeocystis* blooms (Van Duyl et al. 1998). DMS produced in the surface ocean is aerosolized and its oxidation products promote cloud formation, increase albedo, and affect global climate (Charlson et al. 1987). In algal cells, DMSP and its cleavage products, DMS and acrylate, contribute to osmotic balance, neutralize reactive oxygen species, and deter grazing (Noordkamp et al. 2000; Sunda et al. 2002). Despite the ecological importance of colony formation in *Phaeocystis*, triggers for transition to the colonial morphotype remain enigmatic, and the functional role of colony formation in the *Phaeocystis* life-cycle is not clearly delineated (Peperzak & Gabler-Schwarz 2012).

Myriad factors have been studied in regard to their roles in instigating colony formation in *Phaeocystis* species, including nutrient and light availability (Bender et al. 2018; Cariou et al. 1994; Wang et al. 2011), temperature (Verity & Medlin 2003), mechanical stress (Cariou et al. 1994), grazing cues (Long et al. 2007; Tang 2003; Wang et al. 2015), and viral infection (Brussaard et al. 2005; Brussaard et al. 2007). However, these studies used different *Phaeocystis* species, strains, and morphotypes with a range of experimental conditions, which yielded variable and sometimes contradictory results. Nonetheless, several lines of evidence suggest that colony formation serves a defensive role. First, while viruses can cause 30-100% cell lysis in solitary *Phaeocystis,* viruses rarely infect colonial cells, which lyse primarily due to nutrient limitation (Brussaard et al. 2005; Brussaard et al. 2007). Second, ciliates and other microzooplankton that graze solitary *Phaeocystis* are unable to graze on colonies (Tang et al. 2001) and chemical cues from these grazers induce colony formation and promote increased colony size (Long et al. 2007; Tang 2003). Third, acrylate, which is produced with DMS when DMSP is cleaved, accumulates within colonies and may further deter macro- and micro-grazers and heterotrophic bacteria (Hamm et al. 2000; Noordkamp et al. 2000). However, while cellular growth rate increases in colonial cells relative to solitary cells if colonies are induced in nutrient rich conditions, it decreases when colonies are induced under nutrient limiting conditions (Wang et al. 2015). Thus, colony formation can defend against pathogens and grazers, but it is costly (Wang et al. 2015), suggesting that colony formation is likely a complex response to interacting biotic and abiotic factors (Long et al. 2007).

Colony formation may also play a fundamental role in *Phaeocystis* reproduction. *Phaeocystis* has one of the most complex and polymorphic life cycles among phytoplankton genera, and despite extensive study, it remains largely unresolved in most species. Studies have implicated at least 6 different life stages and up to 15 functional components to the life-cycle (Gaebler-Schwarz et al. 2010). In *P. globosa,* four morphotypes are believed to exist: diploid colonial cells devoid of scales and flagella, diploid scale-free flagellates arising from mechanically disrupted colonies, and two types of small, scaled, haploid flagellates—those that produce vesicles containing star-shaped alpha-chitin filaments and those that do not (Rousseau et al. 2007). Haploid flagellates may fuse (syngamy) to produce diploid colony-forming cells, which in turn undergo meiosis and produce haploid flagellates (Rousseau et al. 2013). Haploid flagellates are often observed swarming inside colonies, suggesting that colonial bloom formation may contribute to successful sexual reproduction in *Phaeocystis* (Peperzak et al. 2000; Rousseau et al. 2013). However, neither syngamy nor meiosis have been directly observed in *Phaeocystis* spp. (Peperzak & Gabler-Schwarz 2012), even though both events have been documented in several other haptophyte genera (Houdan et al. 2003). If colonial *Phaeocystis* blooms are necessary for sexual reproduction, it would further justify the resource costs associated with colony formation.

Historically, colonial *Phaeocystis* blooms have been restricted to cold, high-latitude waters—*P. globosa* blooms in the English Channel and North Sea, *P. pouchetii* in the North Atlantic and Arctic, and *P. antarctica* in the Southern Ocean (reviewed in Schoemann et al. 2005). In the last two decades, however, blooms have increasingly been reported in tropical and subtropical regions, including the subtropical N. Atlantic (Long et al. 2007) and the subtropical and tropical South China Sea (Chen et al. 2002; Doan-Nhu et al. 2010; Liu et al. 2015). Decaying colonial biomass sinks and produces anoxic conditions, making *Phaeocystis* blooms detrimental to benthic fisheries and aquaculture (Desroy & Denis 2008; Peperzak & Poelman 2008; Spilmont et al. 2009). In warmer waters, *Phaeocystis globosa* blooms have been especially catastrophic to local aquaculture (Chen et al. 2002; Doan-Nhu et al. 2010), possibly because the hemolytic activity of *P. globosa* liposaccharides increases with temperature (Peng et al. 2005). Global climate change and increasing nutrient pollution in coastal regions may mean harmful *Phaeocystis* blooms will continue to increase in range and frequency. Given the ecological impact of colonial *Phaeocystis* blooms and their complex and enigmatic initiating triggers, particularly in warm waters, it is imperative to better understand the regulation of colony formation.

Transcriptional approaches have become an exceptionally useful tool to illuminate physiological responses to environmental cues and genes associated with specific life-stages in algae and other protists (Caron et al. 2017). In this study, we investigated genetic regulation of colony formation by analyzing gene expression in colonial and flagellated morphotypes of a warm-water *Phaeocystis globosa* strain. Since *Phaeocystis* is an important marine producer of DMSP and DMS—and because these molecules may be associated with colonial defense, we queried our dataset for algal genes involved in DMSP production (*DSYB*, Curson et al. 2018) and its cleavage to DMS and acrylate (*Alma1*, Alcolombri et al. 2015). *DSYB* and *Alma1* are the only algal genes that encode proteins experimentally proven to catalyze DMSP, DMS, and acrylate production, but neither has been identified in *P. globosa* previously. Overall, our results demonstrate a dramatic transcriptional shift in colonial *P. globosa*, with the vast majority of differentially expressed genes downregulated in colonial cells. Such a strong transcriptional shift supports an allocation of resources toward colony formation and away from other cellular processes such as translation, cell growth, and cell division. Genes associated with DMSP production (*DSYB*-like) were not differentially expressed, but an *Alma* family-like gene was upregulated in colonies, suggesting colonies may produce more DMS and acrylate than do solitary cells. Results further indicate that pathogen interaction affects colony formation, further supporting a defensive role for colonies, and implicating several cell-signaling pathways as important to colony formation. This study provides new insights into the functional role of *Phaeocystis* colonies and physiological processes associated with colony formation. These insights will guide future investigations into triggers that initiate colony formation in harmful *Phaeocystis* blooms.

MATERIALS AND METHODS

Culture strain and maintenance

Starter cultures of *Phaeocystis globosa* CCMP1528, a warm-water, colony forming *P. globosa* strain (Wang et al. 2011), were purchased from the National Center for Marine Algae and Microbiota (NCMA, Maine, U.S.A.) in January 2016. CCMP1528 did not initially form colonies in our culture conditions, but it is known that *P. globosa* sometimes stops forming colonies in

139 culture (Janse et al. 1996). We maintained non-colonial cultures in replicate 150 ml Erlenmeyer
140 flasks with 100 ml L1-Si media in ambient light and temperature on a gently rotating twist mixer
141 (TM-300, AS ONE, Osaka, Japan, speed setting 1). Culture media was prepared by enriching 34
142 ppt autoclaved artificial seawater prepared from milliQ water and sea salts (Marine Art SF-1, AS
143 ONE, Osaka, Japan) with the NCMA L1 media kit (-Si) and filtering through sterile 0.22 µm
144 pore-size filters. Cultures were diluted biweekly with freshly prepared media. When one culture
145 replicate began producing colonies in January 2017, experimental culture conditions were
146 promptly initiated.
147
148 **Experimental culture conditions**
149 We prepared four biological replicates each of colony-forming and non-colonial *P. globosa*
150 CCMP1528 by inoculating 45 ml of sterile L1 media with 1 ml stock culture in 50-ml
151 Erlenmeyer flasks. Replicates were placed on a gently rotating twist mixer in a plant growth
152 chamber with cool white fluorescent lamps (CLE-305, TOMY, Tokyo, Japan) set to 22 °C with
153 light level 4 and a 12:12 day:night ratio. A HOBO temperature and light logger (Onset, MA,
154 U.S.A.) was kept in the growth chamber during the experiment. The daytime temperature was 21
155 °C with about 1900-2000 Lux (~30 µmol m^{-2} s^{-1}) light intensity, and the nighttime temperature
156 was 22 °C. Positions of replicates were rotated daily to prevent position in the chamber from
157 systematically affecting replicates. On days 1, 3, and 4, chlorophyll fluorescence was measured
158 at the middle of the light period by transferring 200 µl aliquots to a black 96-well plate
159 (ThermoFisher, MA, U.S.A.) and recording fluorescence (excitation: 440 nm, emission: 685 nm)
160 with a Tecan Ultra Evolution microplate reader (Tecan, Mannedorf, Switzerland). Exponential
161 growth phase was determined by comparing measured fluorescence to a growth curve for
162 *Phaeocystis globosa* CCMP 1528 grown under identical conditions. On day 4 of the first round,
163 1 ml of each replicate was transferred to 45 ml of sterile L1 media and the experimental setup
164 was repeated, which allowed for adaptation to experimental culture conditions. Algal cells were
165 harvested for RNA extraction on day 4 of the second experimental culture round, when replicates
166 were in middle to late exponential growth phase (Fig. S1).
167 Prior to RNA extraction, each culture replicate was imaged with light microscopy
168 (Olympus CKX53, MA, U.S.A.) to ensure that colony-forming replicates were indeed producing
169 colonies and that non-colonial replicates were not (Fig. S2). Flagellates present in the colonial
170 and non-colonial replicates were actively swimming, suggesting that flagellates in this study
171 were scaled haploid flagellates, rather than scale-free diploid flagellates originating from
172 disrupted colonies, but neither flow-cytometry nor electron microscopy were performed
173 (Rousseau et al. 2007). Colony-forming culture replicates were filtered through
174 polytetrafluoroethylene (PTFE) filters (10-µm pore size) (Millipore, NH, U.S.A.) under gentle
175 vacuum. Colonies were visible by eye on the filter surface and swimming flagellates were
176 observed in the flow-through when viewed with light microscopy. The non-colonial culture
177 replicates were first filtered through 50-µm nylon mesh to remove culture debris, and then
178 filtered through 1.0 µm pore-size PTFE filters. No flagellates were visible when the filtrates
179 were viewed with light microscopy. Filters were immediately flash frozen in liquid nitrogen and
180 stored at -80 °C until RNA extraction.
181
182 **RNA extraction, library preparation and sequencing**
183 Total RNA was extracted from filters by following the manufacturer's protocols for the MoBio
184 PowerWater RNA extraction kit (Qiagen, MD, U.S.A.), including the optional initial heating

185 step. Following extraction, we assessed RNA quality and concentration. RNA extracts were
186 diluted so that 10 ng of RNA were used for each sample with the SMART-seq v4 Ultra Low
187 Input RNA Kit (Clonetech/Takara, CA, U.S.A.) along with 2 µl of a 1:10,000 dilution of
188 External RNA Controls Consortium (ERCC) spike-in mix 1 (Ambion, CA, U.S.A.), an internal
189 quality control. The SMART-seq kit employs poly-A priming to target eukaryotic mRNA and to
190 reduce the amount of ribosomal and bacterial RNA present in sequencing libraries. The quality
191 and concentration of the resulting cDNA was assessed before continuing with the manufacturer's
192 protocols for the Nextera XT DNA Library Prep Kit (Illumina, CA, U.S.A.). Finally, we checked
193 cDNA fragment size before submitting libraries to the Okinawa Institute of Science and
194 Technology DNA Sequencing Section for paired-end 150x150 bp sequencing across 8 lanes of
195 an Illumina Hiseq4000 flow-cell.
196
197 **Bioinformatic processing and quality control**
198 Sequencing reads were processed with Trimmomatic software to remove adapter sequences and
199 to filter low-quality sequences (Bolger et al. 2014). Read quality was checked with FastQC
200 before and after trimming to ensure that adapters were removed (Andrews 2010). Remaining
201 reads were mapped to the ERCC reference sequences (Cronin et al. 2004) and mapped reads
202 were counted with RSEM software (Li & Collin 2011). Counts were further analyzed in the R
203 statistical environment (R Core Team 2013). Reads mapping to the ERCC reference sequences
204 were then removed from each sample with SAMtools (Li et al. 2009) and BEDTools (Quinlan &
205 Hall 2010).
206
207 **Transcriptome assembly, assessment, and functional annotation**
208 The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP, Keeling et al.
209 2014) assembled a transcriptome for *Phaeocystis sp.* CCMP2710, which groups with the
210 *Phaeocystis globosa* species complex in phylogenetic analyses (Fig. S3). Only 25% of our reads,
211 however, mapped to this reference transcriptome. We therefore assembled a de novo
212 transcriptome for *Phaeocystis globosa* CCMP1528 to serve as a reference for read mapping in
213 this study. We used Trinity software for transcriptome assembly (Grabherr et al. 2013) and
214 dereplicated the transcriptome by removing reads with 95% similarity using CD-HIT-EST (Fu et
215 al. 2012). Bacterial contamination was removed by performing a blastn query against the NCBI
216 nucleotide database (downloaded March 2018, ncbi-blast v2.6.0+, Camacho et al. 2009) and
217 parsing results to identify and remove bacterial contigs. The final assembly was assessed for
218 completeness with Benchmarking Universal Single-Copy Orthologs (BUSCO v3, Simao et al.
219 2015) and results were compared with those for the MMETSP *Phaeocystis sp.* CCMP2710
220 transcriptome.
221 We annotated the CCMP1528 transcriptome using two different databases, Pfam (Finn et
222 al. 2010) and KEGG (Kanehisa et al. 2016). Pfam annotation was performed with the dammit
223 software (Scott 2018), which wraps Transdecoder to translate transcriptome contigs to the
224 longest possible amino acid sequence (Haas et al. 2013), and HMMER to assign protein
225 homologs to sequences (Eddy 2011). After discarding annotations with e-values greater than 1E-
226 5, the annotation with the lowest e-value was selected for each contig. Gene Ontology (GO)
227 terms were assigned to Pfam annotations using the Gene Ontology Consortium's Pfam2GO
228 mapping (geneontology.org/external2go/pfam2go, version 07/14/2018, Mitchell et al. 2015).
229 KEGG annotation was performed with the GhostKOALA tool and translated amino acid
230 sequences (kegg.jp/ghostkoala, 05/21/2018, Kanehisa et al. 2016). Annotated K numbers were

231 then used to assign KEGG pathways by accessing the KEGG API
232 (kegg.jp/kegg/rest/keggapi.html, July 2018).
233
234 **Differential gene expression analysis**
235 Quality filtered sequences from each sample were mapped to the assembled *P. globosa*
236 CCMP1528 transcriptome and counted with RSEM software. Counts for each sample were
237 imported into the R statistical environment, where differential gene expression between colonial
238 and solitary culture replicates was tested with the DESeq function in the Bioconductor package
239 DESeq2 (Love et al. 2014). Genes that were differentially expressed were considered statistically
240 significant if the False Discovery Rate (FDR) adjusted p-value (padj) was less than 0.05.
241
242 **Gene set enrichment testing**
243 We identified GO terms enriched among significantly upregulated and downregulated genes by
244 applying a hypergeometric test in the R package GOstats (Falcon & Gentleman 2007). GOstats
245 accommodates user-defined GO annotations, which are necessary when studying non-model
246 organisms like *Phaeocystis*. Likewise, a hypergeometric test for significant enrichment of KEGG
247 pathways was applied using the enricher function from the R package ClusterProfiler (Yu et al.
248 2012). Because of the lower annotation rate, KEGG pathway enrichment was further investigated
249 by additionally applying linear model analysis with the kegga function in the R package edgeR
250 (Robinson et al. 2010). GO terms and KEGG pathways were considered significantly enriched
251 when the statistical test returned a p-value less than 0.05.
252
253 **Genes associated with DMSP and DMS production**
254 Because *Phaeocystis* is a profusive producer of DMSP and DMS, we specifically queried our
255 dataset for recently discovered algal genes involved in DMSP production (DSYB) and its
256 cleavage to DMS and acrylate (Alma family genes). We performed blastp queries with curated
257 DSYB protein sequences (provided by Curson et al. 2018) and *E. huxleyi* and *Symbiodinium*
258 Alma family protein sequences downloaded from UniProt (July 2018) against *Phaeocystis*
259 *globosa* CCMP1528 amino acid sequences. Expression levels of putative *Phaeocystis globosa*
260 *DSYB* and *Alma* family genes were then checked in colonial and solitary culture replicates.
261
262 **RESULTS**
263
264 **Bioinformatic processing and quality control**
265 Sequencing for this project produced over 1.9 billion read pairs with 159-383 million read pairs
266 per sample. The reads for each sample were deposited in the Sequence Read Archive (SRA) with
267 accession numbers SRR7811979-SRR7811986. Following quality filtering with Trimmomatic,
268 1.7 billion read pairs remained, with 140-341 million read pairs per sample (Table S1). After
269 mapping reads from each sample to ERCC reference sequences, we plotted the Log_2 FPKM for
270 each sequence against the Log_2 of its concentration in the standard mix. A simple linear
271 regression was fitted for each sample and R^2 values ranged from 0.93-0.937 for each sample
272 (Fig. S4). The strong correlation between observed FPKM and initial concentration for ERCC
273 sequences indicates that minimal bias was introduced during PCR amplification, library
274 preparation, and sequencing.
275
276 **Transcriptome assembly, assessment, and functional annotation**

The final assembly of the *Phaeocystis globosa* CCMP1528 transcriptome included 69,528 contigs and a total of 43.9 Mbp (available for download from https://doi.org/10.5281/zenodo.1476491). The CCMP1528 transcriptome was about 3 times larger than the MMETSP CCMP2710 transcriptome, but the minimum, maximum, and mean contig lengths were about the same for both (Table S2). When Transrate was used to align the two transcriptomes, only 18% of CCMP1528 contigs aligned to the CCMP2710 transcriptome, but 55% of the CCMP2710 contigs aligned to the CCMP1528 transcriptome. BUSCO software was utilized to assess completeness of the *P. globosa* CCMP1528 transcriptome. It included more complete eukaryote and protist BUSCOs than the MMETSP *Phaeocystis sp.* CCMP2710 transcriptome (Fig.1). Together, these results demonstrate that the transcriptome generated in this study is more complete than the MMETSP transcriptome and is a better reference for this study. The results also indicate that CCMP2710 and CCMP1528 are more genetically distant than was expected based upon ribosomal RNA gene sequences (Fig. S3).

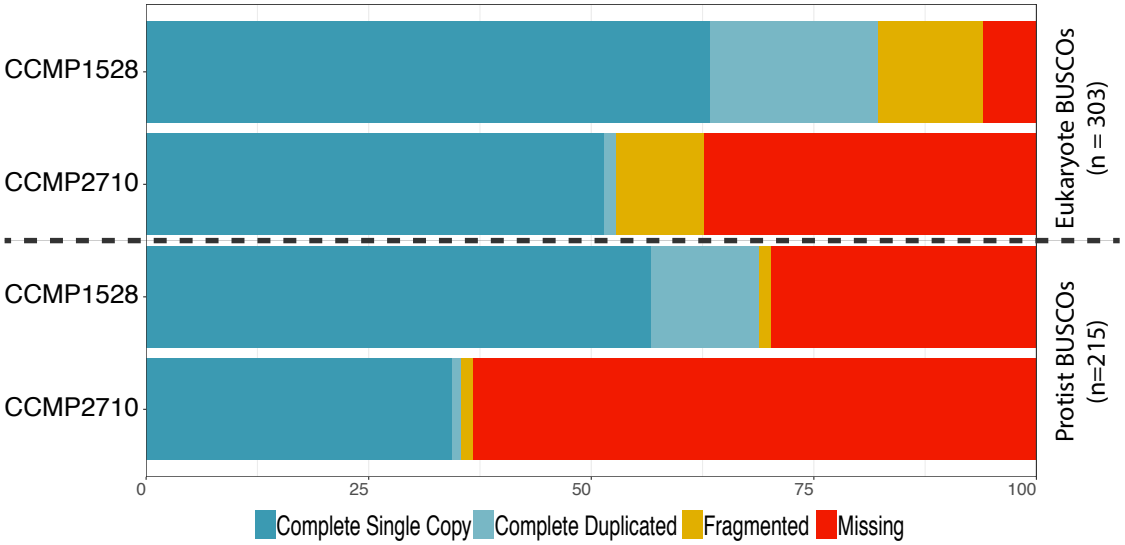


Figure 1. Percent of eukaryote and protist Benchmarking Universal Single Copy Orthologs (BUSCOs) complete, fragmented, or missing in *Phaeocystis globosa* CCMP1528 and *Phaeocystis sp.* CCMP2710 transcriptomes. BUSCO software was used to determine the percent of eukaryotic and protistan BUSCOs represented by complete single copies, complete but duplicated copies, copies that were fragmented or missing in the *Phaeocystis globosa* CCMP1528 transcriptome assembled for this study and the *Phaeocystis sp.* CCMP2710 transcriptome assembled for the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP). More eukaryotic and protistan BUSCOs were represented in the *Phaeocystis globosa* CCMP1528 transcriptome than the *Phaeocystis sp.* CCMP2710 transcriptome. The plot was rendered with the R package ggplot2.

Annotation was possible for relatively few of the contigs in the *P. globosa* CCMP1528 transcriptome assembly, but more genes were annotated with the Pfam and GO annotation pipeline (26%) than with the KEGG pipeline (14%). Additionally, both annotation methods annotated the significantly differentially expressed (DE) genes at a higher rate than the whole transcriptome (Table 1).

Table 1. Pfam, Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation statistics for the *Phaeocystis globosa* CCMP1528 transcriptome assembly and differentially expressed (DE) genes.

	Pfam / GO	KEGG
Total genes annotated	17,826 (26%)	9,967 (14%)
Total w. pathway or GO	8,962 (13%)	5,585 (8%)
DE genes annotated	5,180 (66%)	3,305 (43%)
DE genes w. pathway or GO	2,764 (35%)	1,883 (24%)

Differential gene expression analysis

Gene expression patterns in colonial and solitary replicates were explored with a principal component analysis (PCA) and an expression heatmap. Initial data exploration revealed the colonial replicate 'C2' as an outlier to other colonial replicates and solitary replicates (Fig. S5A, B) and this sample was excluded from further analyses. The remaining 3 colonial replicates clustered separately from the 4 solitary replicates in a PCA plot (Fig. 2A). Differential expression analysis identified 535 genes as significantly upregulated and 7,234 genes as significantly downregulated in colonial replicates. An expression heatmap of the most differentially expressed genes sorted by FDR adjusted p-value (padj) clearly illustrates the overall expression pattern—the majority of significantly differentially expressed genes are downregulated in colonial replicates (Fig. 2B).

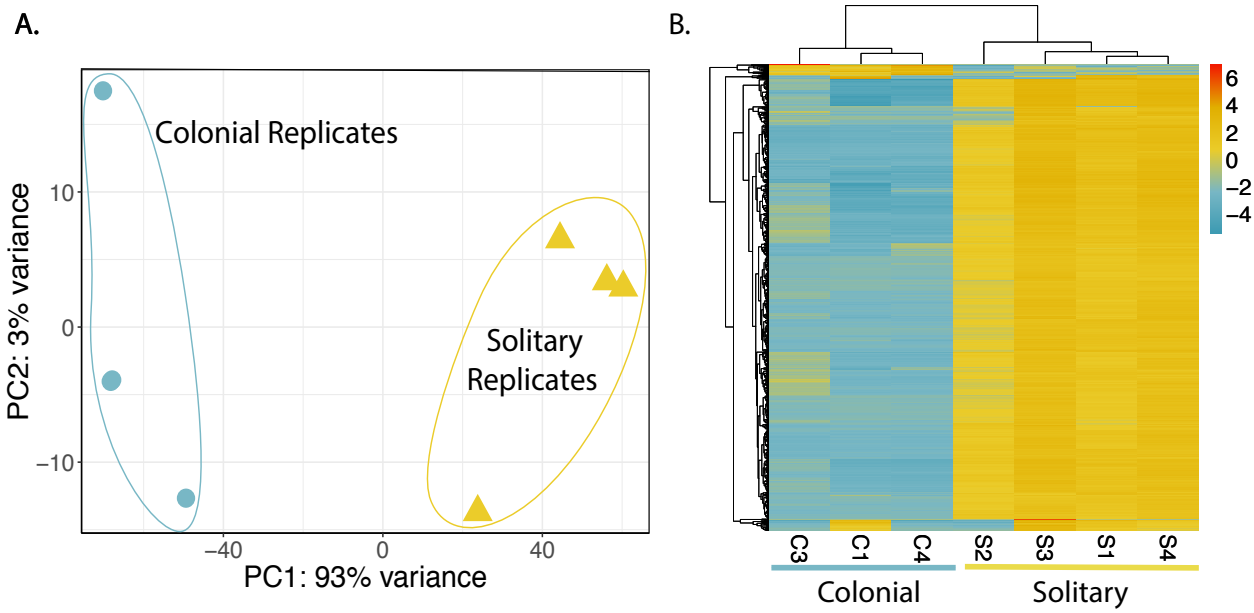


Figure 2. Principal component analysis (PCA) and heatmap demonstrating gene expression patterns in colonial and solitary *Phaeocystis globosa*. A. PCA performed on distances between samples derived from regularized log transformed counts. Colonial and solitary replicates cluster separately, and the majority of variance is between sample type rather than within replicates. Results plotted with R package ggplot2. **B.** Heatmap includes the 1000 significantly differentially expressed genes with the lowest FDR adjusted p-values. Heatmap color represents difference from the mean regularized log transformed count for each contig in each sample. The majority of differentially expressed genes are downregulated in colonial replicates, and replicates cluster by sample type. Results plotted with R package pheatmap.

Gene set enrichment analysis

In order to identify Biological Process (BP) GO terms over-represented in significantly up- and downregulated gene sets, we applied a hypergeometric test with a significance cut-off of $p < 0.05$. Twenty BP GO term were over-represented among significantly upregulated genes and were primarily involved in cell signal transduction in response to external stimuli (Fig. 3; Table S3). Notably, GO terms involving arabinose, a component of the colonial matrix, were also enriched among upregulated genes. In the downregulated gene set, 48 BP GO terms were enriched, including several involved in cation transport, response to oxygen-containing compounds, translation and protein transport, and vacuolar transport and exocytosis (Fig. 3; Table S4). REVIGO software was used to remove redundant GO terms from lists of enriched terms and to visualize results in a Multidimensional Scaling (MDS) plot (Fig. 3) based on GO term semantic similarities (Supek et al. 2011) as determined by shared ancestry (Pesquita et al. 2009).

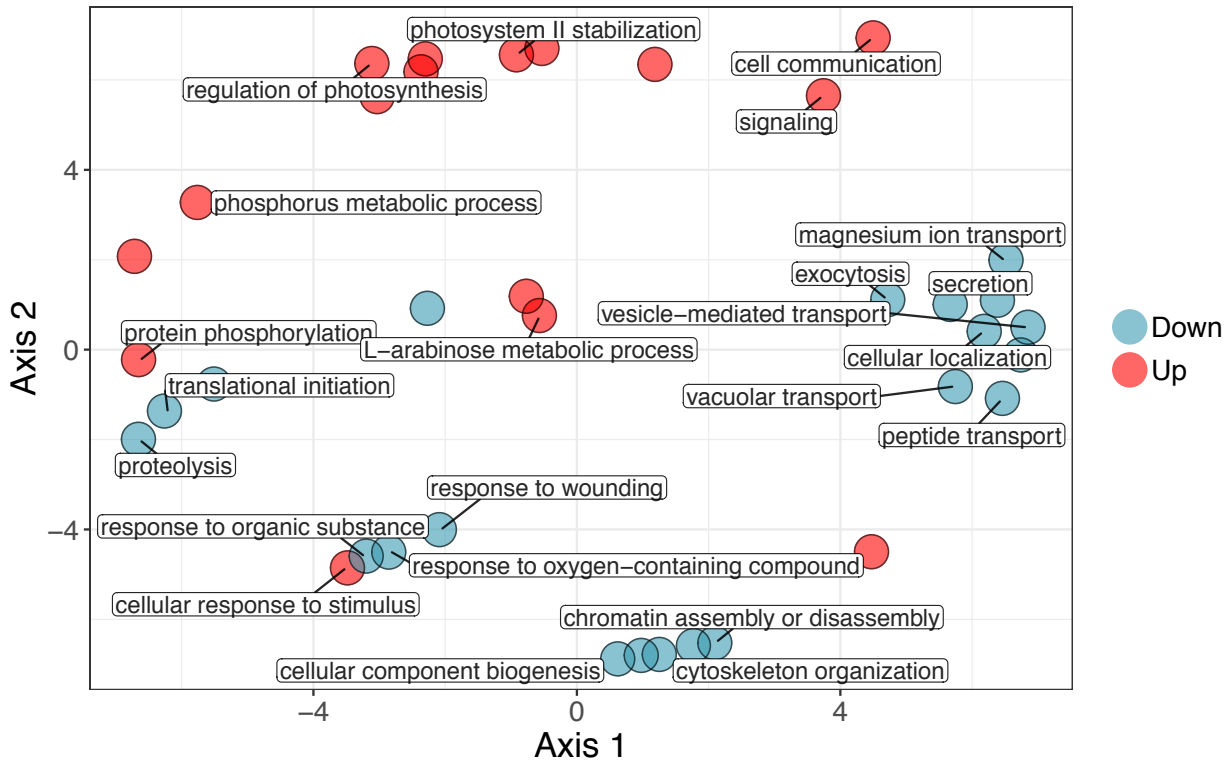


Figure 3. Multidimensional scaling plot of semantic similarities between non-redundant GO terms over-represented in significantly up- and downregulated gene sets. Analysis performed using the REVIGO tool (http://revigo.irb.hr/) with the allowed similarity set to 0.7 (to remove redundant GO terms) and the SimRel metric selected to calculate similarities. REVIGO results were exported to the R statistical environment and plotted with ggplot2. Representative GO terms were manually selected and labeled on the plot. To view all GO term labels, an interactive version of the plot made with R package ggplotly is available at: https://brisbin.shinyapps.io/shinycolsol/.

When we applied hypergeometric testing to KEGG pathways, only the cGMP-PKG signaling pathway (cyclic guanosine monophosphate-protein kinase G pathway) was enriched in the upregulated gene set ($p<0.05$) (Table S5). Five pathways were enriched among downregulated genes: Lysosome, Autophagy, MAPK signaling pathway (mitogen activated protein kinase signaling pathway), AMPK signaling pathway (adenosine monophosphate-activated protein

kinase signaling pathway), and Epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor resistance (Table S6). The lower annotation rate for KEGG pathways compared with GO terms contributed to the difference in enrichment testing results. We therefore also applied a linear model test for KEGG pathway enrichment, which identified several additional pathways as being significantly enriched in the up- (6) and downregulated (7) gene sets. With this additional test, the PI3K-Akt signaling (phosphoinositide 3-kinase-protein kinase B signaling pathway), Glycosphingolipid biosynthesis, Ferroptosis, Plant-pathogen interaction, Circadian rhythm, Viral carcinogenesis pathways were also enriched among upregulated genes (Table S7). The Protein processing in endoplasmic reticulum, Oxidative phosphorylation, Ras signaling pathway, Sphingolipid metabolism, Steroid biosynthesis, Fatty acid degradation, and Taste transduction pathways were additionally identified as enriched in downregulated genes (Table S8).

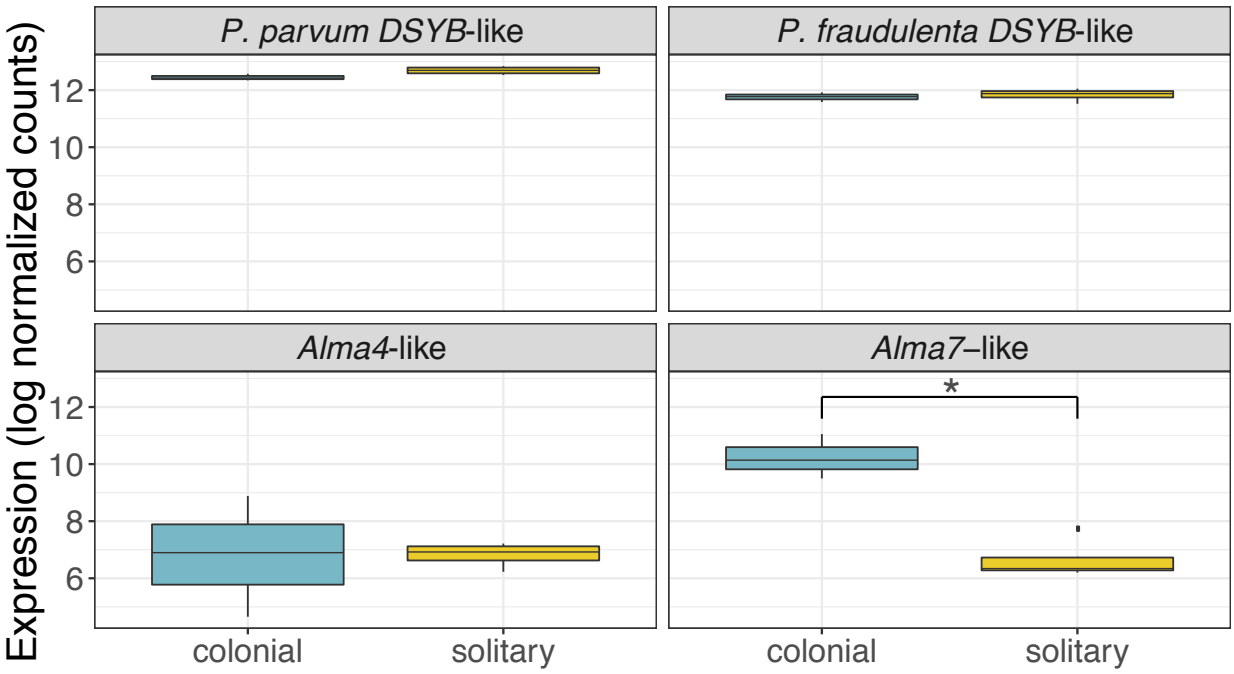
Genes associated with DMSP and DMS production

A blastp query against curated DSYB protein sequences from Curson et al. (2018) and Alma family protein sequences from Alcolombri et al. (2015) identified 4 *Phaeocystis globosa* contigs as putative *DSYB* or *Alma* family genes (Table 2). The *P. globosa* amino acid (AA) sequence translated from Transcript_30752 aligned with the sequence for *Prymnesium parvum* CCAP946/1B DSYB protein, which is experimentally proven to be highly active. It is therefore likely that this gene is actively involved in DMSP biosynthesis in *Phaeocystis globosa*. A second *P. globosa* AA sequence, from Transcript_31221, aligned with the *Pseudonitzchia fraudulenta* DYSB protein sequence, making it also a possible *DSYB* gene. The *Pseudonitzchia* DSYB protein has not been experimentally proven to be active, but its sequence is phylogenetically close to the *Fragillariopsis* DSYB, which has been proven to be active. Neither putative *P. globosa DSYB* genes were differentially expressed between solitary and colonial culture replicates in this study, but both were expressed at relatively high levels in both sample types (Fig. 4).

Two *Alma* family-like genes were identified in the *Phaeocystis globosa* transcriptome. One *P. globosa* AA sequence, from Transcript_36000, aligned with *Emiliania huxleyi* Alma7. All 4 *Alma* homologs identified from the MMETSP *Phaeocystis antarctica* transcriptome are phylogenetically closest to the *E. huxleyi Alma7*, but *E. huxleyi* Alma7 has not been proven to have DMSP-lyase activity. Another *P. globosa* AA sequence, from Transcript_68879, aligned with *E. huxleyi* Alma4, which also has not been experimentally proven active. Both putative *P. globosa Alma* family genes were expressed at lower rates than *DSYB*-like genes. The *P. globosa Alma4*-like gene was not differentially expressed in solitary and colonial culture replicates in this study (Fig. 4). The *P. globosa Alma7*-like gene, however, was significantly upregulated in colonial replicates (padj = 1.44E-12, logFC = 3.55), suggesting that DMSP biosynthesis is occurring in both colonial and solitary cells, but colonial cells may be cleaving DMSP to DMS and acrylate more actively than solitary cells.

410 **Table 2. Top *Phaeocystis globosa* CCMP1528 blastp results against DSYB and Alma family reference**
411 **sequences (e-values <1E-30).**

	Database Sequence	%ID	Alignment Length (AAs)	Gaps	E-value	Bit Score
Transcript_30752	*Prymnesium parvum* CCAP946/1B DSYB	74	314	2	1.15E-170	472
Transcript_31221	*Pseudonitzschia fraudulenta* WWA7 DSYB	45	263	8	1.92E-65	210
Transcript_36000	*Emiliania huxleyi* Alma7	61	272	4	3.70E-119	336
Transcript_68879	*Emiliania huxleyi* Alma4	40	253	4	2.64E-48	166



412
413 **Figure 4. Normalized expression levels of *DSYB*-like and *Alma*-like genes in *Phaeocystis globosa* colonial and**
414 **solitary cell cultures.** Box plots show the range, quartiles and median of the log normalized counts for each gene in
415 colonial and solitary culture replicates. Only the *E. huxleyi Alma7*-like gene was significantly differentially
416 expressed and was upregulated in colonial samples (padj = 1.44E-12, log fold-change = 3.55) and is represented
417 with an asterisk (*) on the plot. Plots made with R package ggplot2.
418

419 **DISCUSSION**

420 Colonial *Phaeocystis* blooms widely impact ecosystem function and can be extremely
421 detrimental in some systems, particularly in subtropical and tropical coastal regions. Although
422 many hypotheses exist, factors initiating colonial blooms and the ecological function of colonial
423 formation remain enigmatic. We investigated gene expression associated with colony formation
424 in a warm-water colony-forming strain of *Phaeocystis globosa* to identify cellular processes
425 associated with colony formation and potentially provide clues for what initiates colony
426 formation and the functional role of colonies in the *Phaeocystis* life-cycle. Overall, we observed

427 a transcriptional shift in colonial cultures compared to solitary cell cultures, with vastly more
428 genes significantly downregulated in colonial cells than upregulated (Fig. 2). This shift suggests
429 that there are trade-offs associated with colony production and resources must be diverted to
430 construct and maintain the colonial matrix. A relatively small number of genes are upregulated to
431 produce colonies, but the low annotation rate of these genes, and the transcriptome overall, make
432 it challenging to fully interpret the results (Table 1). Gene set enrichment analyses inherently
433 relies on how many and which genes are annotated and systematic biases likely in gene
434 annotation will influence results (Haynes et al. 2018). However, these analyses still assist in
435 identifying pathways and functions that may be important to the question at hand and indicate
436 genes and pathways that should be followed up in future studies. The results presented here
437 highlight genes involved in constructing the colonial matrix, changes in cellular morphology,
438 responding to external stimuli, cellular proliferation, and producing DMSP, DMS, and acrylate.
439 Results from this study support a defensive role for colony formation in *Phaeocystis globosa*.
440
441 **Colony matrix carbohydrates and colonial cell morphology**
442 Differential expression of genes associated with clearly observable changes between treatment
443 groups can serve to "ground-truth" results from RNA-seq experiments and therefore increase
444 confidence in expression changes detected for genes for less observable traits. In this study,
445 changes in cellular morphology and colony formation itself are clearly observable differences for
446 which several associated genes are differentially expressed. The observed expression patterns for
447 these genes can additionally provide new insight into the construction of the colonial matrix and
448 pathways associated with morphological changes in colonial *Phaeocystis* cells. Many different
449 polysaccharides are recognized as contributors to the matrix structure of *Phaeocystis globosa*
450 colonies, including arabinose, rhamnose, xylose, mannose, galactose, glucose, gluconuronate,
451 and O-methylated pentose sugars (Janse et al. 1996). *Phaeocystis* isolated from different
452 locations tends to have distinct matrix carbohydrate fingerprints, which may be due to genetic
453 attributes of different strains or which could arise from different environmental conditions, such
454 as light or nutrient availability. For example, arabinose is the most abundant matric carbohydrate
455 in *P. globosa* sampled from the North Sea (Janse et al. 1996). In this study, the GO term for
456 Arabinose metabolic process was enriched among upregulated genes in colonial cells (Fig. 3;
457 Table S3). These results indicate that arabinose is likely the dominant matrix polysaccharide in
458 *P. globosa* CCMP1528 and that arabinose production is specifically associated with colony
459 formation in this strain. The colonial matrix also contains nitrogen (Hamm 2000), which is likely
460 included in amino sugars (Solomon et al. 2003). Our results, however, did not indicate that
461 amino sugar biosynthesis or metabolism was upregulated in colonial cells.
462 Divalent cations, particularly Mg^{2+} or Ca^{2+}, are required for colonial polymers to gel and
463 contribute to the stability of the colonial matrix (van Boekel 1992). GO terms for Divalent
464 inorganic cation transport, Magnesium ion transport, and Divalent metal ion transport, however
465 were enriched in downregulated genes in colonial replicates (Fig. 3; Table S4). Similarly, Bender
466 et al. (2018) found that *Phaeocystis antarctica* produces more calcium-binding proteins when
467 iron limitation decreases colony formation. These results may be due to the importance of
468 divalent cations for flagellate motility. Actively swimming *Phaeocystis globosa* flagellates, as
469 observed in this study, may require continuous transport of divalent cations to the point of
470 masking their shared importance in colony formation. Calcium signaling also induces secretion
471 of vesicles containing gel forming polymers (Chin et al. 2004), which further confounds the
472 observed downregulation of divalent cation transport genes in colonies. However, flagellates also

473 secrete vesicles, but instead of gel polymers they contain star-shaped structures composed of
474 chitinous filaments (Chretiennot-Dinet et al. 1997). The exact function of these structures is
475 unknown, but they may be involved in mating or defense (Dutz & Koski 2006). In addition to
476 divalent cation transport, a number of other GO-terms enriched in downregulated genes may be
477 involved in secreting these structures, such as Exocytosis, Secretion, Vesicle mediated transport,
478 and Vacuolar transport (Fig. 3; Table S3). Alternatively, these GO terms may be involved in
479 scale formation and secretion (Taylor et al. 2007), as scales are only observed on *Phaeocystis*
480 *globosa* flagellates and not colonial cells (Rousseau et al. 2007).
481
482 **A defensive role for colony formation: resource allocation, pathogen interaction, and**
483 **DMS/acrylate production**
484 Out of 7,769 genes that were significantly differentially expressed between colonial and solitary
485 replicates, 7,234 genes were downregulated in colonial cells. This dramatic transcriptional shift
486 in colonial cells supports a high resource cost associated with producing colonies (Wang et al.
487 2015). Specifically, our results indicate that resources are being diverted from protein translation
488 and transport and cell division in order to produce the colonial matrix. Several GO terms
489 involved in the synthesis of larger nitrogenous compounds and their transport, including
490 Translation initiation, Protein metabolic process, Protein N-linked glycosylation, and Protein
491 transport were significantly enriched in downregulated genes in colonial cells (Fig. 3; Table S4).
492 Similarly, the KEGG pathway, Protein processing in endoplasmic reticulum, was also enriched
493 in downregulated genes in colonies (Table S7). Likewise, several mitosis-associated GO terms
494 (Chromatin assembly and disassembly, Cytoskeleton organization, Cellular component
495 biogenesis) were also enriched among downregulated genes in colonial cells (Fig. 3; Table S4).
496 However, the downregulation of mitosis-associated genes in colonial cells conflicts with
497 observations in previous studies. Veldhius et al. (2005) observed that colonial cells divide at a
498 higher rate than solitary cells and proposed that in addition to experiencing less grazing and viral
499 lysis, colonial cells may dominate blooms because they outgrow solitary cells. We believe the
500 difference in our results may be due to the type of solitary cells observed—the solitary cells in
501 previous studies could have been diploid flagellates, especially if they were derived from
502 disrupted colonies, whereas solitary cells in our study are likely haploid flagellates, which have
503 been reported to divide extremely rapidly (Rousseau et al. 2007).
504 There were also several signaling pathways represented in the results suggesting that
505 colonial cells are exposed to fewer general stressors, but may be responding to more strongly to
506 specific pathogens. In plants, the MAPK (mitogen activated protein kinase) pathway primarily
507 transduces signals from extracellular stressors to the nucleus or cytoplasm and initiates an
508 appropriate response (Taj et al. 2010). In our results, the MAPK signaling pathway was
509 significantly enriched in genes downregulated in colonies. The downregulated genes in this
510 pathway encode MAP3Ks, MAP2Ks, and MAPKs, which are activated in response to pathogen
511 attack and infection, phytohormones, cold and salt stress, and reactive oxygen species (Taj et al.
512 2010). Downregulation of genes associated with stress response is also evidenced by related GO
513 terms enriched among downregulated genes, specifically Response to wounding and Response to
514 oxygen-containing compounds. These results support the hypothesis that colony formation
515 serves a defensive purpose. Defense responses regulated by the MAPK pathway are unneeded
516 because the colony skin is protecting cells from these stressors. However, the Plant-pathogen
517 interaction pathway was enriched in upregulated genes in colonial cells, indicating that specific
518 pathogens may penetrate the colonial fortress or that pathogen interaction may play a role in

519 stimulating colony formation. Genes upregulated in this pathway were for calcium-dependent
520 protein kinases and calcium-binding protein CML (calmodulin-like protein), immune response
521 genes that are activated following recognition of specific pathogen-associated molecular patterns
522 (Cheval et al. 2013). Specific bacterial interactions are known to influence transitions between
523 life-cycle stages in several other protists: specific bacteria stimulate growth in marine diatoms
524 (Amin et al. 2015) and specific bacterial signaling molecules are responsible for inducing both
525 colony formation (Woznica et al. 2016) and sexual reproduction (Woznica et al. 2017) in
526 choanoflagellates, another single-celled colony-forming marine plankton. In *Phaeocystis*, axenic
527 cultures exhibit decreased growth rates (Solomon et al. 2003), but the effects of specific bacteria
528 on colony formation have not yet been investigated.
529 *Phaeocystis* is a copious producer of DMSP and its cleavage products, DMS and acrylate.
530 DMS and acrylate have been indicated as grazer-deterrents and antimicrobials (Hamm 2000;
531 Noordkamp et al. 2000; Wolfe & Steinke 1996). While algal genes associated with DMSP
532 biosynthesis (*DSYB*, Curson et al. 2018) and cleavage (*Alma* family genes, Alcolombri et al.
533 2015) have been identified in many algal transcriptomes, including *Phaeocystis antarctica*, this
534 study is the first to identify these genes for *Phaeocystis globosa*. We found that colonial and
535 solitary *P. globosa* expressed *DSYB*-like genes at the similar levels, suggesting that the two cell
536 types produce similar amounts of DMSP. *DSYB* expression in *Prymnesium parvum*, the
537 haptophyte in which DSYB was discovered, is affected only by salinity, potentially indicating
538 that DMSP production functions primarily in osmoregulation rather than as a defensive or stress
539 response (Curson et al. 2018). Similar *DSYB*-like gene expression levels in colonial and solitary
540 *P. globosa* cells support a basic, shared function for DMSP in the two cell types. Contrastingly,
541 an *Alma* family gene was upregulated in colonial cells. Acrylate accumulates in *Phaeocystis*
542 colonies and may serve to deter grazers and pathogens from disrupting the colonial matrix
543 (Noordkamp et al. 2000). While acrylate may accumulate in colonies simply because it cannot
544 escape through the colonial skin, the upregulation of an *Alma*-like gene in colonies suggests that
545 colonial *Phaeocystis* cells may actively produce excess acrylate and DMS than solitary cells,
546 further supporting a defensive role for colony production in *P. globosa*.
547
548 **Role of colonies in *Phaeocystis* reproduction**
549 Colony formation is believed to be involved in sexual reproduction in *Phaeocystis* since
550 swarming flagellates have been observed within senescent colonies (Peperzak et al. 2000;
551 Rousseau et al. 2013). However, we did not find meiosis or sexual reproduction GO terms or
552 KEGG pathways enriched in up- (or down-) regulated genes in this study. These results may
553 arise from RNA being extracted during mid- to late exponential growth phase. Previous
554 observations suggest that colonies produce flagellates during bloom decay, so we might have
555 found meiosis genes upregulated in colonial cells if we had sampled toward late stationary phase
556 instead of exponential phase.
557
558 **Signaling pathways associated with colony formation**
559 Processes and pathways involved in cell-signaling, cell communication, and response to stimuli
560 that are enriched in upregulated-genes are particularly interesting because they shed some light
561 on factors stimulating colony formation in *Phaeocystis globosa*. The cGMP-PKG signaling
562 pathway was the only KEGG pathway significantly enriched among upregulated genes when a
563 hypergeometric test was used. Three genes in this pathway were upregulated: 1) cGMP-
564 dependent protein kinase, which phosphorylates biologically important targets, has been

565 implicated in cell division and nucleic acid synthesis, and reduces cytoplasmic Ca^{2+}
566 concentrations (Lincoln et al. 2001); 2) cAMP-dependent protein kinase regulator; and 3) a
567 cAMP-responsive element-binding protein (CREB), which binds to DNA to increase or decrease
568 transcription and is associated with increased cell survival (Chrivia et al. 1993). The PI3K-Akt
569 signaling pathway was also significantly enriched in upregulated genes when the additional
570 linear model test was used. Within this pathway, two Extracellular Matrix (ECM) focal adhesion
571 genes, for Tenascin (a glycoprotein) and Type IV collagen, were significantly upregulated. Focal
572 adhesion proteins connect cells to extracellular matrices both literally and figuratively, by
573 holding cells in place and by initiating cellular responses to external conditions (Wozniak et al.
574 2004). Bender et al. (2018) also found focal adhesion proteins, specifically glycoproteins,
575 upregulated in colonial *Phaeocystis antarctica*. It is therefore likely that these proteins have an
576 important function in structurally maintaining cell positions in the colonial matrix and signaling
577 between colonial cells. Focal adhesion proteins may be mediating interactions with protein
578 kinases in colonial cells, which go on to promote cell proliferation and differentiation into the
579 colonial morphotype. These signaling pathways represent important candidates for continued
580 study of molecular mechanisms regulating colony formation.
581
582 **Conclusions and future directions**
583 This study investigated gene expression associated with colony formation in *Phaeocystis globosa*
584 for the first time and discovered a large transcriptional shift associated with colony production.
585 Differentially expressed genes were mostly downregulated in colonies, providing evidence for
586 extensive resource allocation toward colony formation. Together, activation of pathogen
587 interaction pathways, reduced expression of stress-response pathways, and increased expression
588 of a DMSP-lyase, which produces DMS and acrylate, supporting a defensive role for colony
589 formation. Future studies may extend this work by investigating *P. globosa* gene expression in
590 colonial and solitary cells in a time course study through the waxing and waning of a bloom and
591 under different nutrient and grazing regimes, potentially by using mesocosms or
592 metatranscriptomic methods in natural communities. While our ability to fully interpret the
593 results was inhibited by an overall lack of annotated genomes and transcriptomes for diverse
594 protist lineages, this study represents a step in the right direction by contributing a new and
595 deeply sequenced transcriptome for *Phaeocystis globosa*. Identification of *DSYB* and *Alma*
596 family-like genes in this transcriptome will additionally allow for further investigation into *P.*
597 *globosa* DMSP and DMS production in the oceans. We were also able to identify several protein
598 kinase signaling pathways that are potentially important for regulating colony formation and
599 should be experimentally investigated in follow-up studies. The results presented here will guide
600 and facilitate continued efforts to unravel the complex factors responsible for triggering harmful
601 colonial *Phaeocystis* blooms, which will likely increase with continued climate change and
602 nutrient pollution in the future.
603
604 **ACKNOWLEDGEMENTS**
605 This work was supported by funding from the Marine Biophysics Unit of the Okinawa Institute
606 of Science and Technology Graduate University. MMB is supported by a JSPS DC1 graduate
607 student fellowship. We thank Angela Ares Pita for valuable discussion and suggestions and for
608 helpful feedback on manuscript drafts. Hiroki Goto and the OIST sequencing section provided
609 important guidance for RNA library preparation and other sequencing considerations. Steven D.
610 Aird edited the manuscript and provided helpful comments.

611
612 **DATA AVAILABILITY**
613 Sequence data is available in the NCBI Sequence Read Archive (SRA) with accession numbers
614 SRR7811979–SRR7811986. The *Phaeocystis globosa* CCMP1528 *de novo* transcriptome
615 produced and used in this study, data files, and data analysis scripts can be accessed at
616 https://doi.org/10.5281/zenodo.1476491.
617
618 **LITERATURE CITED**
619 Alcolombri, U., Ben-dor, S., Feldmesser, E., Levin, Y., Tawfik, D. S., & Vardi, A. 2015.
620 Identification of the algal dimethyl sulfide-releasing enzyme: a missing link in the marine
621 sulfur cycle. *Science,* 348:1-4.
622 Andrews S. 2010. *FastQC: a quality control tool for high throughput sequence data*. Available
623 online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
624 Amin, S. A., Hmelo, L. R., van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K. R., Morales,
625 R. L., Berthiaume, C. T., Parker, M. S., Djunaedi, B., Ingalls, A. E., Parsekm M. R., Moran,
626 M. A., & Ambrust, E. V. 2015. Interaction and signalling between a cosmopolitan
627 phytoplankton and associated bacteria. *Nature,* 522:98-101. doi:10.1038/nature14488.
628 Andersen, R. A., Bailey, J. C., Decelle, J., & Probert, I. 2015. *Phaeocystis rex* sp. nov.
629 (Phaeocystales, Prymnesiophyceae): a new solitary species that produces a multilayered
630 scale cell covering. *Eur. J. Phycol.,* 50:207-222. doi:10.1080/09670262.2015.1024287.
631 Bender, S. J., Moran, D. M., McIlvin, M. R., Zheng, H., McCrow, J. P., Badger, J., DiTullio, G.
632 R., Allen, A. E., & Saito, M. A. 2018. Iron triggers colony formation in *Phaeocystis*
633 *antarctica*: connecting molecular mechanisms with iron biogeochemistry. *Biogeosciences*,
634 15:4923-4942. doi:10.5194/bg-15-4923-2018.
635 Bolger, A. M., Lohse, M., & Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina
636 sequence data. *Bioinformatics,* 30:2114-2120. doi:10.1093/bioinformatics/btu170.
637 Brussaard, C. P. D., Bratbak, G., Baudoux, A. C., & Ruardij, P. 2007. *Phaeocystis* and its
638 interaction with viruses. *Phaeocystis, Major Link Biogeochem. Cycl. Clim. Elem.*, 201–215.
639 doi:10.1007/978-1-4020-6214-8_15.
640 Brussaard, C. P. D., Kuipers, B., & Veldhuis, M. J. W. 2005. A mesocosm study of *Phaeocystis*
641 *globosa* population dynamics: I. regulatory role of viruses in bloom control. *Harmful Algae,*
642 4:859-874. doi:10.1016/j.hal.2004.12.015.
643 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. J.
644 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:1-9.
645 doi:10.1186/1471-2105-10-421.
646 Cariou, V., Casotti, R., Birrien, J. L., & Vaulot, D. 1994. The initiation of *Phaeocystis* colonies.
647 *J. Plankton Res.,* 16:457-470. doi:10.1093/plankt/16.5.457.
648 Caron, D. A., Alexander, H., Allen, A. E., Archibald, J. M., Armbrust, E. V., Bachy, C., Bell, C.
649 J., Bharti, A., Dyhrman, A. T., Guida, S. M., Heidelberg, K. B., Kaye, J. Z., Metzner, J.,
650 Smith, S. R., & Worden, A. Z. 2017. Probing the evolution, ecology and physiology of
651 marine protists using transcriptomics. *Nat. Rev. Microbiol.*, 15:6-20.
652 doi:10.1038/nrmicro.2016.160.
653 Charlson, R. J., Lovelock, J. E., Andreae, M. O., & Warren, S. G. 1987. Oceanic phytoplankton,
654 atmospheric sulphur, cloud albedo and climate. *Nature* 326, 655–661.
655 doi:10.1038/326655a0.
656 Chen, Y. Q., Wang, N., Zhang, P., Zhou, H., & Qu, L. H. 2002. Molecular evidence identifies

657 bloom-forming *Phaeocystis* (Prymnesiophyta) from coastal waters of southeast China as
658 *Phaeocystis globosa*. *Biochem. Syst. Ecol.,* 30:15-22. doi:10.1016/S0305-1978(01)00054-0.
659 Cheval, C., Aldon, D., Galaud, J.-P., & Ranty, B. 2013. Calcium/calmodulin-mediated regulation
660 of plant immunity. *Biochim. Biophys. Acta - Mol. Cell Res.*, 1833:1766-1771. doi:
661 10.1016/j.bbamcr.2013.01.031.
662 Chin, W. C., Orellana, M. V., Quesada, I., & Verdugo, P. 2004. Secretion in unicellular marine
663 phytoplankton: demonstration of regulated exocytosis in P*haeocystis globosa*. *Plant Cell*
664 *Physiol.,* 45:535-542. doi:10.1093/pcp/pch062.
665 Chrétiennot-Dinet, M.-J., Giraud-Guille, M.-M., Vaulot, D., Putaux, J.-L., Saito, Y., & Chanzy,
666 H. 1997. The chitinous nature of filaments ejected by *Phaeocystis* (Prymnesiophyceae). *J.*
667 *Phycol.,* 33:666-672.
668 Chrivia, J. C., Kwok, R. P. S., Lamb, N., Hagiwara, M., Montminy, M. R., & Goodman, R. H.
669 1993. Phosphorylated CREB binds specifically to the nuclear protein CBP. *Nature*,
670 365:855-859.
671 Cronin, M., Ghosh, K., Sistare, F., Quackenbush, J., Vilker, V., & O'Connell, C. 2004. Universal
672 RNA reference materials for gene expression. *Clin. Chem.,* 50:1464-1471.
673 doi:10.1373/clinchem.2004.035675.
674 Curson, A. R. J., Williams, B. T., Pinchbeck, B. J., Sims, L. P., Martínez, A. B., Rivera, P. P. L.,
675 Kumaresan, D., Mercade, E., Spurgin, L. G., Carrion, O., Moxon, S., Cattolico, R. A.,
676 Kuzhiumparambil, U., Guagliardo, P., Clode, P. L., Raina, J.-B., & Todd, J. D. 2018. DSYB
677 catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton.
678 *Nat. Microbiol,.* 3:430-439. doi:10.1038/s41564-018-0119-5.
679 Desroy, N., & Denis, L. 2004. Influence of spring phytodetritus sedimentation on intertidal
680 macrozoobenthos in the eastern English Channel. *Mar. Ecol. Prog. Ser.,* 270:41-53.
681 doi:10.3354/meps270041.
682 Doan-Nhu, H., Nguyen-Ngoc, L., & Dippner, J. W. 2010. Development of *Phaeocystis globosa*
683 blooms in the upwelling waters of the South Central coast of Vietnam. *J. Mar. Syst.,*
684 83:253-261. doi:10.1016/j.jmarsys.2010.04.015.
685 Dutz, J., & Koski, M. 2006. Trophic significance of solitary cells of the prymnesiophyte
686 *Phaeocystis globosa* depends on cell type. *Limnol. Oceanogr.,* 51:1230-1238.
687 doi:10.4319/lo.2006.51.3.1230.
688 Eddy, S. R. 2011. Accelerated Profile HMM Searches. *PLOS Comput. Biol.,* 7, e1002195.
689 doi:10.1371/journal.pcbi.1002195.
690 Falcon, S., & Gentleman, R. 2007. Using GOstats to test gene lists for GO term association.
691 *Bioinformatics,* 23:257-258. doi:10.1093/bioinformatics/btl567.
692 Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. 2012. CD-HIT: Accelerated for clustering the next-
693 generation sequencing data. *Bioinformatics,* 28:3150-3152.
694 doi:10.1093/bioinformatics/bts565.
695 Gaebler-Schwarz, S., Davidson, A., Assmy, P., Chen, J., Henjes, J., Nöthig, E. M., Lunau, M., &
696 Medlin, L. K. 2010. A new cell stage in the haploid-diploid life cycle of the colony-forming
697 haptophyte *Phaeocystis antarctica* and its ecological implications. *J. Phycol,.* 46:1006-
698 1016. doi:10.1111/j.1529-8817.2010.00875.x.
699 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X.,
700 Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A.,
701 Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., &
702 Regev, A. 2013. Trinity: reconstructing a full-length transcriptome without a genome from

RNA-Seq data. *Nat. Biotechnol.*, 29:644-652. doi:10.1038/nbt.1883.Trinity.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Philip, D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., & Regev, A. 2013. De novo transcript sequence reconstruction from RNA-seq: reference generation and analysis with Trinity. *Nat Protoc.,* 8:1-43. doi:10.1038/nprot.2013.084.De.

Hamm, C. E. 2000. Architecture, ecology and biogeochemistry of *Phaeocystis* colonies. *J. Sea Res.,* 43:307-315. doi:10.1016/S1385-1101(00)00014-9.

Haynes, W. A., Tomczak, A., & Khatri, P. 2018. Gene annotation bias impedes biomedical research. *Sci. Rep.,* 8:1-7. doi:10.1038/s41598-018-19333-x.

Houdan, A., Billard, C., Marie, D., Not, F., Sâez, A. G., Young, J. R., & Probert, I. 2004. Holococcolithophore- heterococcolithophore (Haptophyta) life cycles: Flow cytometric analysis of relative ploidy levels. *Syst. Biodivers.,* 1:453-465. doi:10.1017/S1477200003001270.

Janse, I., van Rijssel, M., Gottschal, Jan C., Lancelot, C., & Gieskes, W. W. C. (1996). Carbohydrates in the North Sea during spring bloom of *Phaeocystis*: A specific fingerprint. *Aquat Microb Ecol,* 10:97-103.

Kanehisa, M., Sato, Y., & Morishima, K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.,* 428:726-731. doi:10.1016/j.jmb.2015.11.006.

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., et al. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* 12, e1001889. doi:10.1371/journal.pbio.1001889.

Li, B., & Dewey, C. N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics,* 12:323. doi:10.1186/1471-2105-12-323.

Lincoln, T. M., Dey, N., & Sellak, H. 2001. Signal transduction in smooth muscle invited review: cGMP-dependent protein kinase signaling mechanisms in smooth muscle: from the regulation of tone to gene expression. *J. Appl. Physiol.,* 91:1421-1430. doi:10.1152/japplphysiol.01002.2001.

Liss, P. S., Malin, G., Turner, S. M., & Holligan, P. M. 1994. Dimethyl sulphide and *Phaeocystis*: A review. *J. Mar. Syst.,* 5:41-53. doi:10.1016/0924-7963(94)90015-9.

Liu, H.-X., Huang, H.-H., Xu, S.-N., Dai, M., & Shen, P.-P. 2015. Planktonic community structure during a harmful bloom of *Phaeocystis globosa* in a subtropical bay, with special reference to the ciliate assemblages. *Ecotoxicology,* 24:1419-1429. doi:10.1007/s10646-015-1464-2.

Long, J. D., Smalley, G. W., Barsby, T., Anderson, J. T., & Hay, M. E. 2007. Chemical cues induce consumer-specific defenses in a bloom-forming marine phytoplankton. *Proc. Natl. Acad. Sci.,* 104:10512-10517. doi:10.1073/pnas.0611600104.

Love, M. I., Huber, W., & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.,* 15:1-21. doi:10.1186/s13059-014-0550-8.

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S.-Y., Bateman, A., Punta, M., Attwood, T. K. Sigrist, C. J. A., Redaschi, N.,

Rivoire, C., Xenarios, I., Kahn, D., Guyot, D., Bork, P., Letunic, I., Gough, J., Oates, M., Haft, D., Huang, H., Natale, D.A., Wu, C. H., Orengo, C., Sillitoe, I., Mi, H., Thomas, P. D., & Finn, R. D. 2015. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.*, 43:213-221. doi:10.1093/nar/gku1243.

Noordkamp, D. J. B., Gieskes, W. W. C., Gottschal, J. C., Forney, L. J., & Van Rijssel, M. 2000. Acrylate in *Phaeocystis* colonies does not affect the surrounding bacteria. *J. Sea Res.,* 43:287-296. doi:10.1016/S1385-1101(00)00021-6.

Peng, X. C., Yang, W. D., Liu, J. S., Peng, Z. Y., Lü, S. H., & Ding, W. Z. 2005. Characterization of the hemolytic properties of an extract from *Phaeocystis globosa* Scherffel. *J. Integr. Plant Biol.,* 47:165-171. doi:10.1111/j.1744-7909.2005.00039.x.

Peperzak, L., Colijn, F., Vrieling, E. G., Gieskes, W. W. C., & Peeters, J. C. H. 2000. Observations of flagellates in colonies of *Phaeocystis globosa* (Prymnesiophyceae); a hypothesis for their position in the life cycle. *J. Plankton Res.,* 22:2181-2203. doi:10.1093/plankt/22.12.2181.

Peperzak, L., & Gabler-Schwarz, S. 2012. Current knowledge of the life cycles of *Phaeocystis globosa* and *Phaeocystis antarctica* (prymnesiophyceae). *J. Phycol.,* 48:514-517. doi:10.1111/j.1529-8817.2012.01136.x.

Peperzak, L., & Poelman, M. 2008. Mass mussel mortality in The Netherlands after a bloom of *Phaeocystis globosa* (prymnesiophyceae). *J. Sea Res.,* 60:220-222. doi:10.1016/j.seares.2008.06.001.

Pesquita, C., Faria, D., Falcão, A. O., Lord, P., & Couto, F. M. 2009. Semantic similarity in biomedical ontologies. *PLoS Comput. Biol.,* 5, e1000443. doi:10.1371/journal.pcbi.1000443.

Quinlan, A. R., & Hall, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26:841-842. Available at: http://dx.doi.org/10.1093/bioinformatics/btq033.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26:139-140. doi:10.1093/bioinformatics/btp616.

Rousseau, V., Chrétiennot-Dinet, M.-J., Jacobsen, A., Verity, P., & Whipple, S. 2007. The life cycle of *Phaeocystis*: state of knowledge and presumptive role in ecology. *Biogeochemistry,* 83:29-47.

Rousseau, V., Lantoine, F., Rodriguez, F., LeGall, F., Chretiennot-Dinet, M. J., & Lancelot, C. 2013. Characterization of *Phaeocystis globosa* (Prymnesiophyceae), the blooming species in the Southern North Sea. *J. Sea Res.,* 76:105-113. doi:10.1016/j.seares.2012.07.011.

Schoemann, V., Becquevort, S., Stefels, J., Rousseau, V., & Lancelot, C. 2005. *Phaeocysti*s blooms in the global ocean and their controlling mechanisms: A review. *J. Sea Res.,* 53:43-66. doi:10.1016/j.seares.2004.01.008.

Scott, C. 2018. Dammit: an open and accessible de novo transcriptome annotator. *in prep.* Available at: www.camillescott.org/dammit.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics,* 31:3210-3212. doi:10.1093/bioinformatics/btv351.

Solomon, C. M., Lessard, E. J., Keil, R. G., & Foy, M. S. 2003. Characterization of extracellular polymers of *Phaeocystis globosa* and *P. antarctica*. *Mar. Ecol. Prog. Ser.,* 250:81-89. doi:10.3354/meps250081.

Spilmont, N., Denis, L., Artigas, L. F., Caloin, F., Courcot, L., Créach, A., Desroy, N., Gevaert, F., Hacquebart, P., Hubas, C., Janquin, M.-A., Lemoine, Y., Luczak, C., Migne, A., Rauch, M., & Davoult, D. 2009. Impact of the *Phaeocystis globosa* spring bloom on the intertidal benthic compartment in the eastern English Channel: A synthesis. *Mar. Pollut. Bull.,* 58:55-63. doi:10.1016/j.marpolbul.2008.09.007.

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. 2011. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One,* 6, e0021800. doi:10.1371/journal.pone.0021800.

Taj, G., Agarwal, P., Grant, M., & Kumar, A. 2010. MAPK machinery in plants. *Plant Signal. Behav.*, 5:1370-1378. doi:10.4161/psb.5.11.13020.

Tang, K. W. 2003. Grazing and colony size development in *Phaeocysis globosa* (Prymnesiophyceae): the role of chemical signal. *J. Plankton Res.,* 35:831-842.

Tang, K. W., Jakobsen, H. H., & Visser, A. W. 2001. *Phaeocystis globosa* (Prymnesiophyceae) and the planktonic food web: feeding, growth, and trophic interactions among grazers. *Limnol. Oceanogr.*, 46:1860-1870. doi:10.4319/lo.2001.46.8.1860.

Taylor, A. R., Russell, M. A., Harper, G. M., Collins, T. F. T., & Brownlee, C. 2007. Dynamics of formation and secretion of heterococcoliths by *Coccolithus pelagicus* ssp. braarudii. *Eur. J. Phycol.,* 42:125-136. doi:10.1080/09670260601159346.

Team, R Core. 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.r-project.org/.

Van Boekel, W. H. M. 1992. *Phaeocystis* colony mucus components and the importance of calcium ions for colony stability. *Mar. Ecol. Prog. Ser.,* 87:301-305. doi:10.3354/meps087301.

Van Duyl, F. C., Gieskes, W. W. C., Kop, A. J., & Lewis, W. E. 1998. Biological control of short-term variations in the concentration of DMSP and DMS during a *Phaeocystis* spring bloom. *J. Sea Res.,* 40:221-231. doi:10.1016/S1385-1101(98)00024-0.

Veldhuis, M. J. W., Brussaard, C. P. D., & Noordeloos, A. A. M. 2005. Living in a *Phaeocystis* colony: a way to be a successful algal species. *Harmful Algae,* 4:841-858. doi:10.1016/j.hal.2004.12.013.

Verity, P. G., & Medlin, L. K. 2003. Observations on colony formation by the cosmopolitan phytoplankton genus *Phaeocystis*. *J. Mar. Syst.,* 43:153-164. doi:10.1016/j.jmarsys.2003.09.001.

Wang, X., Wang, Y., Ou, L., He, X., & Chen, D. 2015. Allocation costs associated with induced defense in *Phaeocystis globosa* (Prymnesiophyceae): the effects of nutrient availability. *Sci. Rep.,* 5:10850. doi:10.1038/srep10850.

Wang, X., Wang, Y., & Smith, W. O. 2011. The role of nitrogen on the growth and colony development of *Phaeocystis globosa* (Prymnesiophyceae). *Eur. J. Phycol.* 46, 305–314. doi:10.1080/09670262.2011.602430.

Wolfe, G. V, & Steinke, M. 1996. Grazing-activated production of dimethyl sulfide (DMS) by two clones of *Emiliania huxleyi*. *Limnol. Oceanogr.,* 41:1151-1160.

Wozniak, M. A., Modzelewska, K., Kwong, L., & Keely, P. J. 2004. Focal adhesion regulation of cell behavior. *Biochim. Biophys. Acta - Mol. Cell Res.,* 1692:103-119. doi:10.1016/j.bbamcr.2004.04.007

Woznica, A., Cantley, A. M., Beemelmanns, C., Freinkman, E., Clardy, J., & King, N. 2016. Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc. Natl. Acad. Sci.,* 113:7894-7899. doi:10.1073/pnas.1605015113.

Woznica, A., Gerdt, J. P., Hulett, R. E., Clardy, J., & King, N. 2017. Mating in the closest living

relatives of animals is induced by a bacterial chondroitinase. *Cell,* 170:1175-1183.e11. doi:10.1016/j.cell.2017.08.005.

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omi. a J. Integr. Biol.,* 16:284-287.