1 **Genome-wide analysis of cell cycle-regulating genes in the symbiotic dinoflagellate**

2 ***Symbiodinium minutum*.**

3 Michael L. Cato^{*}, Hallie D. Jester[†], Adam Lavertu[†], Audrey Lyman[†], Lacey M. Tallent^{*}, Geoffrey C.

4 Mitchell^{*†}.

5 ^{*}Department of Biology, Wofford College, Spartanburg, SC, 29303

6 [†]Department of Biology, Colby College, Waterville, ME, 04901

7 **Cell cycle-regulating genes in *Symbiodinium*.**

8 **Keywords:** cell cycle, cyclins, cyclin-dependent kinases, dinoflagellate, *Symbiodinium*

9 **Corresponding Author:**

10 Geoffrey C. Mitchell

11 Department of Biology

12 Wofford College

13 429 N. Church St.

14 Spartanburg, SC, 29303

15 (864) 597-4625

16 mitchellgc@wofford.edu

ABSTRACT

A critical relationship exists between reef-building corals and their photosynthetic endosymbionts. As important as this relationship is for reef health, it is exquisitely delicate—exposure to temperatures only marginally above the average summer maximum can cause corals to bleach, expelling their resident algae. Interestingly, several studies indicate that failure of corals to properly regulate symbiont cell divisions at high temperatures may cause bleaching. This needs to be further investigated, but first, it is necessary to decipher the molecular mechanisms controlling the cell division cycle in these organisms. As a first step toward this goal, we identified key cell cycle-regulating genes in the recently published genome of the symbiotic dinoflagellate *Symbiodinium minutum*. We then correlated expression of these genes with cell cycle phase in diurnally growing *S. minutum* in culture. Of particular interest, this approach allowed us to identify cyclins and cyclin-dependent kinases that are involved in the G1/S transition—a likely point for coral cells to exert control over algal cell divisions.

INTRODUCTION

A critical relationship exists between reef-building corals and the symbiotic algae residing within them. These dinoflagellate algae (zooxanthellae), from the genus *Symbiodinium*, are photosynthetic, harvesting energy from sunlight and sharing that energy with their coral hosts. In return, corals provide them with metabolic intermediates, a stable position in the water column, and protection from grazing. This relationship, as important as it is for reef health, is delicate—exposure to temperatures only marginally above the average summer maximum can cause corals to bleach, expelling their resident algae (reviewed in (Brown 1997)). Massive bleaching due to global warming will drastically and irreversibly alter coral reef

ecosystems around the world, adversely impacting fisheries, coastal ecosystems, and placing financial strain on developing economies that depend on tourism (reviewed in (Moberg and Folke 1999)).

Remarkably, studies suggest that failure of the host to properly regulate zooxanthellae cell divisions at high temperatures may cause bleaching (Bhagooli and Hidaka 2002; Strychar *et al.* 2005). In fact, in many corals, the optimal growth temperature for the symbiont exceeds the bleaching threshold. For example, clade C zooxanthellae isolated from *Montipora verrucosa* proliferate better when reared at 31° than when reared at lower temperatures (Kinzie *et al.* 2001); this temperature, however, is lethal to the host (Jokiel and Coles 1977).

To establish symbiosis, corals produce a chemical signal that forces symbionts into a non-motile, dividing state (Koike *et al.* 2004). Unfortunately, it is unclear how corals coordinate host and algal cell divisions to ensure that the proper density of symbionts is maintained. Several pre-mitotic mechanisms have been proposed in a variety of Cnidarians, including factors produced by coral cells that inhibit the algal cell cycle (Smith and Muscatine 1999) and limited access to nutrients (Falkowski *et al.* 1993), which may be controlled by the host or simply by competition among symbionts (consistent with (Mcauley and Darrah 1990)). Post-mitotic regulation might include digestion of algae *in situ* (Titlyanov *et al.* 1996) or expulsion of excess algae (Jones and Yellowlees 1997). In the anemone *Aiptasia pulchella*, rates of algal expulsion increase with temperature and expelled algae, surprisingly, have a much higher rate of mitosis than algae that are retained within their hosts (Baghdasarian and Muscatine 2000). Regardless of the mechanism(s) responsible for maintaining a healthy symbiont density, it is evident that controlling algal cell divisions in hospite is important.

The eukaryotic cell cycle is divided into four distinct phases—two gap phases (G1 and G2) interrupted by DNA replication (synthesis/S phase) and followed by mitosis (M), which ends with cytokinesis. Moving successively from one stage to another is governed through phosphorylation of substrates by the cyclin-dependent kinases (CDKs). As this name implies, these proteins are active only when paired with a partner cyclin that can target them to the appropriate substrates. The CDKs are additionally activated by the CDK-activating kinases (CAKs; reviewed in (Kaldis 1999)) and Cdc25 phosphatases and inhibited by members of the WEE1/MYT1 kinase family (reviewed in (Perry and Kornbluth 2007)) . Further, the CDK-interacting protein CKS1 is essential for both the G1/S and G2/M transitions (Tang and Reed 1993). Meanwhile, the cyclin-dependent kinase inhibitors (CKIs)—e.g. the mammalian proteins p21^{Cip1/Waf1} and p27^{Kip1}—can initiate cell cycle arrest in response to internal or external stimuli (El-Deiry *et al.* 1994; Polyak *et al.* 1994).

While checkpoints regulate each cell cycle transition, eukaryotic cells commit to division during G1. In yeast, this point of commitment is called START (Hartwell *et al.* 1974), while in metazoans it is referred to as the restriction point (R-point) (Pardee 1974). In general, the cell cycle is only sensitive to external factors prior to this; once a cell has committed to divide, it can do so without any further input. In light of this, if a Cnidarian host is regulating the proliferation of its symbionts, it makes sense that it would do so by blocking the R-point. In fact, symbiotic algae in *Aiptasia pulchella* spend considerably longer in G1 than the same species of algae (*Symbiodinium pulchrorum*) in culture, despite the durations of S, G2, and M remaining relatively consistent (Smith and Muscatine 1999).

As a first step toward our goal of understanding how coral cells regulate the divisions of their symbiotic algae, we identified key cell cycle regulators in the recently published genome of *Symbiodinium minutum* (Shoguchi *et al.* 2013). We then correlated the expression of those genes with cell cycle phase in diurnally growing *S. minutum* in culture. This approach allowed us to find specific cyclins and CDKs that are involved in the G1/S transition—a likely point for coral cells to exert control over algal cell divisions.

MATERIALS AND METHODS

Identification of Putative Cell Cycle-Regulating Genes: Amino acid sequences for all annotated CDKs in *H. sapiens*, *A. thaliana*, and *S. cerevisiae* were downloaded from GenBank. An alignment, constructed with Probcons (Do *et al.* 2005), was used to generate a CDK-specific HMMER model (HMMER 3.1b1, hmmer.org). Again using HMMER, this model was queried against the predicted proteome of *S. minutum* (Shoguchi *et al.* 2013); available at marinegenomics.oist.jp/symb/viewer/info?project_id=21) to identify novel CDKs. Initial results were culled by removing all high scoring pairs (HSPs) with fewer than 200 residues and any HSPs that didn't begin within 55 bases of the start of the CDK alignment. To identify novel cyclins, Cdc25s, MAT1 homologs, and CKS1 homologs in *S. minutum*, hidden Markov models were retrieved from PFAM (pfam.xfam.org) and queried against the predicted proteome of *S. minutum*, again using HMMER. For the cyclins, a PFAM model describing the characteristic N-terminal domain that is common to all cyclins was chosen. All potential CDK, cyclin, Cdc25, MAT1, and CKS1 sequences were then reciprocally queried against the UniProtKB database using the HMMER web server (Finn *et al.* 2011). Those sequences that had CDKs, cyclins, Cdc25s, MAT1, or CKS1 as top hits were used for further analysis unless they carried additional

domains that made them likely members of another protein family (e.g. an N-terminal cyclin domain in a putative CDK).

Phylogenetic Analysis: Probcons was used to create an alignment of putative *S. minutum* CDKs with representative CDKs from *H. sapiens*, *A. thaliana*, and *S. cerevisiae*. Gaps were removed from the alignment before generating PhyML (Guindon *et al.* 2010) and BioNJ (Gascuel 1997) trees with 1000 bootstraps that were rooted with a related kinase—MAPK-overlapping kinase isoform 1 (MOK1) from *A. thaliana*. For the cyclins, an alignment of the N-terminal domains of the novel *S. minutum* sequences and representative cyclins from *H. sapiens*, *A. thaliana*, and *S. cerevisiae* was created using ClustalW ((Larkin *et al.* 2007; Goujon *et al.* 2010); accessed through phylogeny.fr). Gaps were removed from the alignment and trees were constructed as they were for the CDKs.

Cell Culture: Cultures of *S. minutum* were acquired from the National Center for Marine Algae (Boothbay Harbor, ME; accession number CCMP3345). They were grown in flasks containing silica-free f/2 made from filtered sea water collected from the Gulf of Maine supplemented with 50 mg/ml each of kanamycin, ampicillin, and streptomycin. A period of 13h of light (approximately $100 \mu\text{mol} \cdot \text{photons} \cdot \text{m}^2 \cdot \text{s}^{-1}$) followed by 11h of darkness was maintained in an environmental growth chamber set to 25°. Experiments were performed on cultures undergoing log-phase expansion ($1-2 \times 10^6$ cells/ml).

Flow Cytometry: To determine the fraction of cultured cells in each cell cycle phase across a 24h day-night cycle, 1×10^6 cells were collected. They were centrifuged for 5 minutes at 200 x g, fixed in 1 ml of 70% ethanol, and resuspended in a staining buffer (PBS, 0.1% Triton X-100, 20 $\mu\text{g}/\text{ml}$ propidium iodide (Sigma-Aldrich), and 10 $\mu\text{g}/\text{ml}$ RNase A). Cells stained with propidium

iodide were run through a FACSCalibur (BD Biosciences, San Jose, CA) set to detect 585 nm fluorescence upon excitation at 488 nm. Data were collected for 10,000 cells and analyzed using FlowJo v.10.1 (Ashland, OR).

Quantitative RT-PCR: To measure the expression of putative cell cycle regulators in cultured cells across a 24h day-night cycle, 2×10^7 cells were collected by centrifuging at $5,500 \times g$ for 15 min., media was removed, and pellets were flash frozen in liquid nitrogen. RNA extraction combined lysis and homogenization in TRIzol (Life Technologies, Carlsbad, CA) with a kit-based cleanup following the protocol established by Rosic and Hoegh-Guldberg (Rosic and Hoegh-Guldberg 2010). Briefly, pellets were resuspended in 1 mL of TRIzol, mixed thoroughly by pipetting, and transferred to screw-cap tubes with 0.3 g of glass beads (Sigma-Aldrich, St. Louis, MO). Cells were then disrupted twice for 90 sec. in a MagNA Lyser (Roche Life Science, Basel, Switzerland) at 4,500 rpm. Debris was removed by centrifuging at $12,000 \times g$ for 1 min., and then supernatant was run through the RNeasy Plus Mini kit following manufacturer's instructions (Qiagen, Hilden, Germany). RNA was eluted in 30 μ l of nuclease-free water and quantified on a NanoDrop Lite (Thermo Fisher Scientific, Waltham, MA). 1 μ g of RNA was reverse transcribed using a QuantiTect Reverse Transcription Kit (Qiagen) following manufacturer's instructions; this kit includes a genomic DNA wipeout step. cDNA was diluted 1:5 in nuclease-free water before being used for quantitative PCR (qPCR). qPCR reactions were setup as follows: 10 μ l QuantiNova SYBR Green PCR Master Mix (Qiagen), 1.5 μ l mixed forward and reverse primer (10 μ M each, Table S3), 4.5 μ l of nuclease-free water, and 4 μ l of template. Reactions were run on a CFX96 (Bio-Rad, Hercules, CA) at 95°C for 2 min. for initial activation following by 40 cycles of 95°C for 5 sec. then 60°C for 10 sec. Data were normalized to the

expression of S-adenosyl methionine synthase (SAM) (Rosic *et al.* 2011) and analyzed using the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen 2001). Average expression levels relative to the 08:00 time-point were clustered using MeV 4.9.0.

Data Availability: File S1 contains flow cytometry data (FCS files) for Figure 1. Individual FCS files are ordered by time-point (three files for each time-point from 0:00 to 20:00). File S2 contains all real-time RT-PCR analysis with a separate tab for each gene.

RESULTS AND DISCUSSION

Amino acid sequences for annotated cell cycle regulators (belonging to the following families: CDKs, cyclins, CAKs, CKS1, WEE1/MYT1, Cdc25, and CKIs) in *H. sapiens*, *A. thaliana*, and *S. cerevisiae* were used to identify homologous sequences in the genome of *S. minutum*. The expression of these novel cell cycle genes was monitored throughout the cell cycle to help validate their function as predicted by homology.

Division of free-living *Symbiodinium* is regulated in response to light, so cells grown on a diurnal cycle (13h of light to 11h of darkness) were sampled every 4h over one 24h period to determine the percentage of cells in each cell cycle phase at a given time. Cells were stained with the nuclear stain propidium iodide and nuclear content was assessed by flow cytometry. The highest percentage of G1 cells—about 70%—was observed at the onset of the light period (time 00:00) and persisted until the onset of DNA replication at 04:00 (Figure 1, Figure S1). By 08:00, the fraction of G2 cells began to rise later as the fraction of G1 cells fell. S phase peaked around the initiation of the dark period (13:00), with the largest percentage of G2 cells and lowest percentage of G1 cells observed a few hours later. Cytokinesis occurred in the hours immediately prior to the start of the light period. These results are consistent with published

data, which show that the vast majority of free-living *Symbiodinium* are in G1 early in the day and that the proportion of G2 cells peaks during the dark period (Smith and Muscatine 1999; Wang *et al.* 2013). It is worth noting that our results indicate a background presence of S phase cells (about 10%) at all time points. Most likely, this background is an artifact of the modeling that was done to fit a histogram of DNA content to specific cell cycle phases in FlowJo. These cells probably belong to either the G1 or G2 population.

When attempting to assign molecular events, such as changes in gene expression, to a particular cell cycle phase, it is commonplace to use a synchronized cell population; this is often achieved by treating with a mitosis inhibitor such as nocodazole. While this is generally desirable, we chose not to use a synchronized population. First of all, nocodazole is cytotoxic and is known to affect gene expression (Cho *et al.* 2006). Second, while darkness leads to the accumulation of cultured *Symbiodinium* cells in G1, we were unable to increase the percentage of G1 cells with prolonged darkness (unpublished data). Still, with a large fraction of cells entering S phase at 04:00, we would expect to see increased expression of genes involved in the G1/S transition at that time. Likewise, the proportion of G2 cells drops precipitously within the last 4 hours of darkness, so we would expect to see increased expression of genes involved in mitosis during this period.

CDKs: Nine CDKs were identified in *S. minutum*; to group them within accepted subclasses (e.g. CDKA, CDKB), phylogenetic trees were constructed using both maximum likelihood (PhyML) and neighbor joining (BioNJ) methods. Both methods yielded similar topologies, so the BioNJ tree was chosen for publication. Of the nine CDKs, two clustered with the CDKA/CDK1-3 family (Figure 2). CDKs from this family all contain the highly conserved PSTAIRE motif that seems to

be important for their association with the cyclins that are most responsible for cell cycle progression (Figure 3A). Further supporting their role in cell cycle regulation, the mRNA expression for both of these *S. minutum* CDKs is highest from 04:00 to 16:00 (Figure 4). It is precisely during this time that we observe an increase in the percentage of S phase cells (starting at 04:00) and subsequently G2 cells (starting at 08:00) (Figure 1). Based on these data, and sticking with the nomenclature established for cell cycle genes in other protists (e.g. (Huysman *et al.* 2010)), we are confident in classifying these genes as *S. minutum* CDKA1 and CDKA2. Interestingly, mRNA expression for CDKA1 begins to increase earlier than the mRNA expression for CDKA2, suggesting that CDKA1 may be important for S phase entry and CDKA2 for progression into G2.

In addition to the two CDKAs, our phylogenetic analysis revealed two CDKCs, a CDKF that clustered with human CDK20, and a cluster of four *S. minutum*-specific CDKs (Figure 2). The CDKC family, including CDK9 in animals and Ctk1p in *S. cerevisiae*, is involved in transcription initiation (Wallenfang and Seydoux 2002). Of the two CDKCs identified computationally, only one of them, which we call CDKC1, had detectable mRNA expression at any point over a 24h time-course (Figure 4), with expression increasing immediately after the onset of the light period and remaining high through at least 08:00. This expression probably reflects a general increase in gene expression as cells prepare to enter S phase and mitosis. As in other eukaryotes, expression of CDKC1 is not likely to depend on cell cycle phase. As for the CDKF homolog, its expression peaks at or near the end of the light period before dropping continuously through the dark hours (Figure 4). Interestingly, in the green algae, *Chlamydomonas reinhardtii*, a CDKF homolog is critical for assembly of flagella (Tam *et al.*

2007). In *Symbiodinium*, it has been established that a non-motile dividing cell, lacking flagella, will give rise to two motile daughter cells (Fitt and Trench 1983). The highest fraction of motile cells is observed within the first few hours of the light period (Fitt and Trench 1983), which can be explained by the fact that cytokinesis occurs within the last few hours of darkness (Figure 1). Increased expression of CDKF as cells enter G2 may indicate that the daughter flagella in *S. minutum* are assembled prior to mitosis; this would be an interesting area for further investigation. Finally, the *S. minutum*-specific CDKs (CDKSm1-4) are each expressed differently throughout the day (Figure 4). CDKSm1 and CDKSm2, for instance, reach peak expression during the dark period, suggesting they may have a role in preparing the cell for mitosis. Interestingly, CDKSm4 has minimal expression except for the 4h immediately prior to the onset of the light period—the time during which mitosis occurs. Finally, CDKSm3 has a similar expression pattern to CDKF; its expression peaks at or near the end of the light period then decreases through the dark hours. It is possible that CDKSm3 partners with CDKF in flagellar assembly or plays some other role critical during G2.

Cyclins: Fifteen cyclins were identified in *S. minutum*. From the N-terminal domain that is present in all cyclins, phylogenetic trees were constructed using both maximum likelihood (PhyML) and neighbor joining (BioNJ) methods. Both trees had similar topologies, so the BioNJ tree was chosen for publication. Based on this phylogeny, the *S. minutum* cyclins were grouped within accepted subclasses (e.g. cyclin A, cyclin B). This grouping did not reveal any sequences with a high degree of similarity to known cyclin A family members (Figure 5). Analysis of gene expression, however, revealed that a particular cyclin transcript (identified as cyclin A) reaches its peak expression at 12:00 (Figure 6). During this time the percentage of S phase cells also

peaks (Figure 1). It is known that A-type cyclins are active in S phase and are, in fact, necessary to initiate DNA replication (Girard *et al.* 1991). Further supporting our identification of a cyclin A homolog in *S. minutum*, hierarchical clustering of gene expression indicated that its expression most closely matches the expression of CDKA1—the CDK that appears to be involved in S phase entry (Figure S2). Additionally, three cyclin B family members were identified (Figure 5). B-type cyclins are the “mitotic cyclin” that is necessary for proper spindle assembly (Michel *et al.* 2004) and whose rapid degradation precipitates anaphase (Holloway *et al.* 1993). As such, we would expect to see high levels of cyclin B mRNA from 16:00 to 20:00 (i.e. when the highest percentage of cells is in G2 or mitosis) (Figure 1). This, however, is not the case; mRNA expression of both cyclin B1 and B3 peaks around 12:00, while cyclin B2 transcript remains steady throughout the day before dropping off during the dark hours (Figure 6). It is possible that the bulk of transcription in *Symbiodinium* occurs during the daylight hours, though this is certainly not the case for all transcripts we investigated. Another possibility is that while expression of the B-type cyclins drops off before the majority of cells make it into G2, the protein has accumulated to sufficiently high levels and will remain stable until the cells are ready to complete mitosis. Undoubtedly, post-translational control of cyclins is critical for proper coordination of the cell cycle. For example, the ubiquitin proteasome pathway is critical for causing the rapid degradation of cyclin B that triggers anaphase (Holloway *et al.* 1993). Oddly, no cyclin D homolog was uncovered. This family of cyclins is involved in bringing cells out of quiescence and driving them through G1 (Resnitzky *et al.* 1994). Cyclin D appears to have arisen early in the eukaryotes, so the lack of a homolog in *S. minutum* likely reflects a deletion somewhere in its lineage (Cao *et al.* 2014). Similarly, no cyclin E homolog was detected in *S.*

minutum. Initial work suggested that cyclin E is absent in protists (Ma *et al.* 2013), however, analysis of several unicellular choanoflagellates uncovered cyclin E homologs (Cao *et al.* 2014). Cyclin E seems to have arisen in the lineage leading to choanoflagellates and animals after splitting from the lineage that led to dinoflagellates.

In addition to these canonical cell cycle-regulating cyclins, we identified six P-type cyclins, three L-type cyclins, and two putative cyclins that do not fit phylogenetically into the well-established families (Figure 5). A reciprocal query of the two unidentified sequences against the UniProtKB database using the HMMER web server yielded no clues about their identity; nevertheless, a variety of cyclin sequences were returned as top hits, leading us to classify these as *Symbiodinium*-specific cyclins. Both—cyclin Sm1 and Sm2—are expressed in a seemingly cell cycle-specific way. Cyclin Sm1 is expressed similarly to cyclin A, peaking at 12:00 (Figure 6). Similarly, cyclin Sm2 is at its highest level of expression from 08:00 to 12:00 (Figure 6). During this time, cells go through S phase (Figure 1), suggesting roles for cyclins Sm1 and Sm2 in DNA replication. The six P-type cyclins are particularly interesting because cyclins belonging to this family are thought to be involved in phosphate signaling (Torres Acosta *et al.* 2004). Such a link connecting phosphate signaling to cell cycle regulation would be unsurprising in *Symbiodinium* since concurrent nitrogen and phosphate limitation can induce G1 arrest in these algae (Smith and Muscatine 1999). Under the conditions we tested, the cyclins P1, P2, and P3 have relatively steady expression across 24h (Figure 7). Cyclin P4 is present at a level that is too low for analysis, while cyclin P5 is undetectable. Since inorganic phosphate is abundant in f/2 media, steady expression of these P-type cyclins is unsurprising. Cyclin P6, however, has an interesting expression pattern, peaking from 04:00 to 08:00 before decreasing

substantially. It is during this time that cells are transitioning from G1 to S phase. Perhaps cyclin P6 is functioning as a nutrient sensor that allows cells to enter S phase if there is abundant phosphate. Finally, the L-type cyclins are important for RNA splicing (Dickinson *et al.* 2002). Transcripts for the three L-type cyclins that were identified computationally, however, were undetectable.

CDK Activators and Inhibitors: In addition to the presence or absence of a cyclin partner, CDKs are regulated by numerous other modulators. The CAKs and Cdc25 phosphatases, for instance, work in concert to add activating phosphorylations and remove inhibitory phosphorylations from CDKs (reviewed in (Kaldis 1999; Perry and Kornbluth 2007)). In contrast, members of the WEE1/MYT1 kinase family phosphorylate CDKs in a way that renders them inactive (reviewed in (Perry and Kornbluth 2007)). Further regulating CDK activity, the CDK-interacting protein CKS1 functions as a docking protein that allows CDKs to interact with their substrates (Tang and Reed 1993). Finally, the CKIs can block CDK activity and initiate cell cycle arrest in response to internal or external stimuli (El-Deiry *et al.* 1994; Polyak *et al.* 1994).

From this list of CDK activators and inhibitors, only a partial CAK subunit and three CKS1 homologs were identified in *S. minutum*. The CAK subunit was identified as MAT1—one part of a trimeric CAK complex (reviewed in (Kaldis 1999)). Surprisingly, however, only the N-terminus of this MAT1 homolog aligned with MAT1 sequences from other species. For example, MAT1 in *H. sapiens* is 309 amino acids long but only the first 140 residues aligned with the *S. minutum* protein, which itself is 245 amino acids long. The transcript for this partial MAT1 is expressed at relatively steady levels throughout the day, but drops just before the onset of the dark period (Figure 8). Canonically, MAT1 functions as an assembly factor, forming a CAK with cyclin H and a

D-type CDK. Since there are no cyclin H or CDKD homologs in *S. minutum*, it is likely that this partial MAT1 plays another role, perhaps independent of the cell cycle. The evolutionary history of this sequence might hold clues to its function. As for the three CKS1 homologs, only two—CKS1A and CKS1B—had detectable transcripts (Figure 8). The expression of CKS1B is highest just before darkness and has a similar expression pattern to cyclin B1 (Figure S2), perhaps indicating that these two bind together to a target CDK. Expression of CKS1A is less predictable across the cell cycle.

No members of the WEE1/MYT1 family of CDK-inhibiting kinases were identified; thus, it is not surprising that their antagonists—the Cdc25 phosphatases—were also absent. Finally, no CKIs—small CDK inhibitors such as the mammalian proteins p21^{Cip1/Waf1} and p27^{Kip1}—were uncovered in the *S. minutum* genome. It is possible that one or more CKIs are present but are simply undetectable using our methods, which rely on a large degree of sequence similarity that is uncharacteristic of the CKIs.

Conclusion: Unsurprisingly, our analysis of key cell cycle regulators in the genome of *S. minutum* has revealed that control of mitotic cell divisions in these dinoflagellates shares many features with other eukaryotes. For instance, the two CDKAs that were identified both contain a highly conserved PSTAIRE motif, suggesting that they regulate cell cycle progression directly. Further, CDKA1 has a similar expression profile to cyclin A, suggesting that these proteins form a cyclin/CDK pair that may be involved in driving cells through S phase, much like cyclin A/CDK2 in mammals. Similarly, three cyclin B family members were identified that may pair with one or both CDKAs and play a role in G2 and mitosis—exactly what happens in the mammalian cell cycle. Knowing the identity of these cell cycle regulators will allow us to design experiments to

determine the molecular events that are responsible for maintaining Cnidarian control over symbiont life cycles. The cyclin A/CDKA1 combination is particularly interesting because it seems to be the pair that is necessary for initiating S phase. Eukaryotic cells commit to division at a point in G1 or at the G1/S transition. As a general rule, the cell cycle is only sensitive to external factors prior to this commitment point (Pardee 1974; Hartwell *et al.* 1974). Thus, if a Cnidarian host is regulating cell divisions in its endosymbiont, it is likely that it does so by controlling the G1/S transition that seems to be regulated by the cyclin A/CDKA1 pair.

ACKNOWLEDGEMENTS

Special thanks to Natalie Spivey for critical review of this manuscript.

LITERATURE CITED

- Baghdasarian, G., and L. Muscatine, 2000 Preferential expulsion of dividing algal cells as a mechanism for regulating algal-cnidarian symbiosis. *Biol. Bull.* 199: 278–286.
- Bhagooli, R., and M. Hidaka, 2002 Physiological fascicularis responses of the coral *Galaxea* and its algal symbiont to elevated temperatures. *J. Japanese Coral Reef Soc.* 4: 33–42.
- Brown, B. E., 1997 Coral bleaching: causes and consequences. *Coral Reefs* 16: S129–S138.
- Cao, L., F. Chen, X. Yang, W. Xu, J. Xie *et al.*, 2014 Phylogenetic analysis of CDK and cyclin proteins in premetazoan lineages. *BMC Evol. Biol.* 14: 10.
- Cho, S. G., C.-R. Sihn, S. J. Yoo, K. K. Cho, H. Lee *et al.*, 2006 Analysis of gene expression induced by microtubule-disrupting agents in HeLa cells using microarray. *Cancer Lett.* 241: 110–117.
- Dickinson, L. A., A. J. Edgar, J. Ehley, and J. M. Gottesfeld, 2002 Cyclin L is an RS domain protein involved in pre-mRNA splicing. *J. Biol. Chem.* 277: 25465–73.
- Do, C. B., M. S. P. Mahabhashyam, M. Brudno, and S. Batzoglou, 2005 ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res.* 15: 330–340.
- El-Deiry, W. S., J. W. Harper, P. M. O ’connor, W. S. Ei-Deiry, V. E. Velculescu *et al.*, 1994 Arrest and Apoptosis 1 -mediated G p53 Is Induced in WAF1/CIP1 WAF1/CIP1 Is Induced in /?53-mediated G! Arrest and Apoptosis1. *Cancer Res [CANCER Res.* 54: 1169–1174.
- Falkowski, P. G., Z. Dubinsky, L. Muscatine, and L. Mccloskey, 1993 Control Population Symbiotic Corals in ammonium ions and organic materials maintain the density of zooxanthellae. *Bioscience* 43: 606–611.
- Finn, R. D., J. Clements, and S. R. Eddy, 2011 HMMER web server: interactive sequence

similarity searching. *Nucleic Acids Res.* 39: W29–W37.

Fitt, W. K., and R. K. Trench, 1983 The Relation of Diel Patterns of Cell Division to Diel Patterns of Motility in the Symbiotic Dinoflagellate *Symbiodinium microadriaticum* Freudenthal in Culture. *New Phytol.* 94: 421–432.

Gascuel, O., 1997 BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14: 685–695.

Girard, F., U. Strausfeld, A. Fernandez, and N. J. C. Lamb, 1991 Cyclin a is required for the onset of DNA replication in mammalian fibroblasts. *Cell* 67: 1169–1179.

Goujon, M., H. McWilliam, W. Li, F. Valentin, S. Squizzato *et al.*, 2010 A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* 38: W695–W699.

Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk *et al.*, 2010 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59: 307–21.

Hartwell, L. H., J. Culotti, J. R. Pringle, and B. J. Reid, 1974 Genetic Control of the Cell Division Cycle in Yeast. *Science* (80-.). 183: 46 LP-51.

Holloway, S. L., M. Glotzer, R. W. King, and A. W. Murray, 1993 Anaphase is initiated by proteolysis rather than by the inactivation of maturation-promoting factor. *Cell* 73: 1393–1402.

Huysman, M. J. J., C. Martens, K. Vandepoele, J. Gillard, E. Rayko *et al.*, 2010 Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. *Genome Biol.* 11: R17.

Jokiel, P. L., and S. L. Coles, 1977 Effects of temperature on the mortality and growth of

377 Hawaiian reef corals. Mar. Biol. 43: 201–208.

378 Jones, R. J., and D. Yellowlees, 1997 Regulation and control of intracellular algae (=

379 zooxanthellae) in hard corals. Philos. Trans. R. Soc. B Biol. Sci. 352: 457–468.

380 Kaldis, P., 1999 The cdk-activating kinase (CAK): From yeast to mammals. Cell. Mol. Life Sci. 55:

381 284–296.

382 Kinzie, R. A., M. Takayama, S. R. Santos, M. Alice, R. A. K. Iii *et al.*, 2001 The Adaptive Bleaching

383 Hypothesis : Experimental Tests of Critical Assumptions. 200: 51–58.

384 Koike, K., M. Jimbo, R. Sakai, M. Kaeriyama, K. Muramoto *et al.*, 2004 Octocoral chemical

385 signaling selects and controls dinoflagellate symbionts. Biol. Bull. 207: 80–86.

386 Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan *et al.*, 2007 Clustal W

387 and Clustal X version 2.0. Bioinformatics 23: 2947–8.

388 Livak, K. J., and T. D. Schmittgen, 2001 Analysis of Relative Gene Expression Data Using Real-

389 Time Quantitative PCR and the 2– $\Delta\Delta$ CT Method. Methods 25: 402–408.

390 Ma, Z., Y. Wu, J. Jin, J. Yan, S. Kuang *et al.*, 2013 Phylogenetic analysis reveals the evolution and

391 diversification of cyclins in eukaryotes. Mol. Phylogenet. Evol. 66: 1002–1010.

392 Mcauley, P. J., and P. R. Darrah, 1990 Regulation of Numbers of Symbiotic Chlorella by Density-

393 Dependent Division. Philos. Trans. R. Soc. B Biol. Sci. 329: 55–63.

394 Michel, L., E. Diaz-Rodriguez, G. Narayan, E. Hernando, V. V. V. S. Murty *et al.*, 2004 Complete

395 loss of the tumor suppressor MAD2 causes premature cyclin B degradation and mitotic

396 failure in human somatic cells. Proc. Natl. Acad. Sci. U. S. A. 101: 4459 LP-4464.

397 Moberg, F., and C. Folke, 1999 Ecological goods and services of coral reef ecosystems. Ecol.

398 Econ. 29: 215–233.

399 Pardee, a B., 1974 A restriction point for control of normal animal cell proliferation. Proc. Natl.
400 Acad. Sci. U. S. A. 71: 1286–90.

401 Perry, J. A., and S. Kornbluth, 2007 Cdc25 and Wee1: analogous opposites? Cell Div. 2: 12.

402 Polyak, K., J. Kato, M. J. Solomon, C. J. Sherr, J. Massague *et al.*, 1994 Transforming Growth
403 Factor-13 and Contact Inhibition To Cell Cycle Arrest. 9–22.

404 Resnitzky, D., M. Gossen, H. Bujard, and S. I. Reed, 1994 Acceleration of the G1/S phase
405 transition by expression of cyclins D1 and E with an inducible system. Mol. Cell. Biol. 14:
406 1669–1679.

407 Rosic, N. N., and O. Hoegh-Guldberg, 2010 A method for extracting a high-quality RNA from
408 Symbiodinium sp. J. Appl. Phycol. 22: 139–146.

409 Rosic, N. N., M. Pernice, M. Rodriguez-Lanetty, and O. Hoegh-Guldberg, 2011 Validation of
410 Housekeeping Genes for Gene Expression Studies in Symbiodinium Exposed to Thermal
411 and Light Stress. Mar. Biotechnol. 13: 355–365.

412 Shoguchi, E., C. Shinzato, T. Kawashima, F. Gyoja, S. Mungpakdee *et al.*, 2013 Draft Assembly of
413 the Symbiodinium minutum Nuclear Genome Reveals Dinoflagellate Gene Structure. Curr.
414 Biol. 23: 1399–1408.

415 Smith, G. J., and L. Muscatine, 1999 Cell cycle of symbiotic dinoflagellates: variation in G1
416 phase-duration with anemone nutritional status and macronutrient supply in the Aiptasia
417 pulchella--Symbiodinium pulchrorum symbiosis. Mar. Biol. 134: 405–418.

418 Strychar, K. B., M. Coates, P. W. Sammarco, T. J. Piva, and P. T. Scott, 2005 Loss of
419 Symbiodinium from bleached soft corals Sarcophyton ehrenbergi, Sinularia sp. and Xenia
420 sp. J. Exp. Mar. Bio. Ecol. 320: 159–177.

421 Tam, L.-W., N. F. Wilson, and P. A. Lefebvre, 2007 A CDK-related kinase regulates the length and
422 assembly of flagella in *Chlamydomonas*. *J. Cell Biol.* 176: 819–829.

423 Tang, Y., and S. I. Reed, 1993 The Cdk-associated protein Cks1 functions both in G 1 and G2 in
424 *Saccharomyces cerevisiae*. *Genes Dev.* 8: 822–832.

425 Titlyanov, E. A., T. V. Titlyanova, V. A. Leletkin, J. Tsukahara, R. Van Woesik *et al.*, 1996
426 Degradation of zooxanthellae and regulation of their density in hermatypic corals. *Mar.*
427 *Ecol. Prog. Ser.* 139: 167–178.

428 Torres Acosta, J. A., J. de Almeida Engler, J. Raes, Z. Magyar, R. De Groodt *et al.*, 2004 Molecular
429 characterization of *Arabidopsis* PHO80-like proteins, a novel class of CDKA;1-interacting
430 cyclins. *Cell. Mol. Life Sci.* 61: 1485–1497.

431 Wallenfang, M. R., and G. Seydoux, 2002 cdk-7 is required for mRNA transcription and cell cycle
432 progression in *Caenorhabditis elegans* embryos. *Proc. Natl. Acad. Sci.* 99: 5527–5532.

433 Wang, L.-H., H.-H. Lee, L.-S. Fang, A. B. Mayfield, and C.-S. Chen, 2013 Fatty Acid and
434 Phospholipid Syntheses Are Prerequisites for the Cell Cycle of Symbiodinium and Their
435 Endosymbiosis within Sea Anemones. *PLoS One* 8: e72486.

436

Figure 1: Cell cycle distribution of free-living *S. minutum* over 24h. *S. minutum* maintained on a diurnal cycle were collected from a culture every 4h. Cells were stained with propidium iodide and nuclear content was assessed by flow cytometry. Data were analyzed using the FlowJo Cell Cycle platform and plotted as the average percentage of gated cells in each cell cycle phase +/- standard deviation (n = 3). Data are representative of multiple biological replicates. Grey background represents periods of darkness.

444

445 **Figure 2: Phylogenetic analysis of the CDKs of *S. minutum*.** Protein sequences for 9 CDKs
 446 identified in the genome of *S. minutum* (labelled Sm) were aligned with representative CDK
 447 sequences from *H. sapiens* (Hs), *A. thaliana* (At), and *S. cerevisiae* (Sc) using Probcons.
 448 Phylogenetic trees created using either neighbor-joining (BioNJ) or maximum likelihood
 449 (PhyML) methods yielded similar topologies. The tree shown here was constructed using BioNJ
 450 with 1,000 replicates. NCBI accession numbers or specific genome identifiers (for *S. minutum*)
 451 are indicated next to the sequence ID.

A)	AtCdkA1	...EGVPSTAI	REISLLKE...
	HsCdk1-3	...EGVPSTAI	REISLLKE...
	ScCdc28	...EGVPSTAI	REISLLKE...
	ScPho85	...EGTPSTAI	REISLMKE...
	SmCdkA1	...EGIPSTAI	REISLLKE...
	SmCdkA2	...EGIPSTAI	REISLLKE...
		**	***** **

B)	SmCdkSm1	...--LSSF	SLREISALTA...
	SmCdkSm2	...---TW	EREICALKC...
	SmCdkSm3	...DGVPA	VAMREISLLKG...
	SmCdkSm4	...FGAST	QALREISLFRM...
	SmCdkC1	...GGFPL	SLLREITILRS...
	SmCdkF	...GLDFD	PWFKSAERERD...

452

453 **Figure 3: The cyclin binding motif in the CDK family.** Protein sequences for the CDKA/1-3
 454 family members from *H. sapiens* (Hs), *A. thaliana* (At), and *S. cerevisiae* (Sc) were aligned to the
 455 CDK protein sequences identified in *S. minutum* (Sm) using ClustalW. Sixteen residues flanking
 456 the cyclin binding motif are shown. Conserved amino acids are marked by an asterix in the
 457 bottom row. **A)** CDKA/1-3 family members from each species contain a highly conserved cyclin
 458 binding motif (PSTAIRE). **B)** Comparison of the cyclin binding motif in all *Symbiodinium* CDKs.

Figure 4: Hierarchical clustering of the gene expression profiles of CDKs in *S. minutum*.

Messenger RNAs were isolated from free-living *S. minutum* maintained on a diurnal cycle every 4h. Following reverse transcription, qPCR was run with primers for specific CDKs. Data were normalized to the expression of SAM and analyzed using the $2^{-\Delta\Delta Ct}$ method. Average expression levels relative to the 0:00 time-point were clustered using MeV 4.9.0 (n = 3). Data are representative of multiple biological replicates. Grey boxes below indicate the predominant cell cycle phase at a given time-point.

Figure 5: Phylogenetic analysis of the cyclins of *S. minutum*. Protein sequences of the N-terminal domains for 15 cyclins identified in the genome of *S. minutum* (labelled Sm) were aligned with representative cyclin N-terminal domain sequences from *H. sapiens* (Hs), *A. thaliana* (At), and *S. cerevisiae* (Sc) using ClustalW. Phylogenetic trees created using either neighbor-joining (BioNJ) or maximum likelihood (PhyML) methods yielded similar topologies. The tree shown here was constructed using BioNJ with 1,000 replicates. NCBI accession numbers or specific genome identifiers (for *S. minutum*) are indicated next to the sequence ID.

Figure 6: Hierarchical clustering of the gene expression profiles of cyclins in *S. minutum*.

Messenger RNAs were isolated from free-living *S. minutum* maintained on a diurnal cycle every 4 hr. Following reverse transcription, qPCR was run with primers for specific cyclins. Data were normalized to the expression of SAM and analyzed using the $2^{-\Delta\Delta Ct}$ method. Average expression levels relative to the 0:00 time-point were clustered using MeV 4.9.0 (n = 3). Data are representative of multiple biological replicates. Grey boxes below indicate the predominant cell cycle phase at a given time-point.

Figure 7: Hierarchical clustering of the gene expression profiles of P-type cyclins in *S. minutum*. Messenger RNAs were isolated from free-living *S. minutum* maintained on a diurnal cycle every 4 hr. Following reverse transcription, qPCR was run for specific cyclins P family members. Data were normalized to the expression of SAM and analyzed using the $2^{-\Delta\Delta C_t}$ method. Average expression levels relative to the 0:00 time-point were clustered using MeV 4.9.0 (n = 3). Data are representative of multiple biological replicates. Grey boxes below indicate the predominant cell cycle phase at a given time-point.

Figure 8: Hierarchical clustering of the gene expression profiles of cyclin/CDK interacting proteins in *S. minutum*. Messenger RNAs were isolated from free-living *S. minutum* maintained on a diurnal cycle every 4 hr. Following reverse transcription, qPCR was run with primers for specific cyclins. Data were normalized to the expression of SAM and analyzed using the $2^{-\Delta\Delta Ct}$ method. Average expression levels relative to the 0:00 time-point were clustered using MeV 4.9.0 (n = 3). Data are representative of multiple biological replicates. Grey boxes below indicate the predominant cell cycle phase at a given time-point.

Figure S1: Cell cycles distribution of free-living *S. minutum*. These data were collected over 24h as described for Figure 1. Here, however, the data are duplicated for a second 24h period to show cyclical cell divisions over 48h. Grey background represents periods of darkness.

Figure S2: Hierarchical clustering of the gene expression profiles of putative cell cycle regulators in *S. minutum*. Messenger RNAs were isolated from free-living *S. minutum* maintained on a diurnal cycle every 4h. Following reverse transcription, qPCR was run with primers for specific cyclins. Data were normalized to the expression of SAM and analyzed using the $2^{-\Delta\Delta C_t}$ method. Average expression levels relative to the 0:00 time-point were clustered using MeV 4.9.0 (n = 3). Data are representative of multiple biological replicates. Grey boxes below indicate the predominant cell cycle phase at a given time-point.

Gene	Transcript ID	Efficiency	Forward Sequence	Reverse Sequence
Cdk A1	symbB.v1.2.017277.t3	91%	GGTGAATGGCAGTCCTTTAT	CTGCTGAAGGTGTTCTTAAG
Cdk A2	symbB.v1.2.002415.t1	103%	GGCCCACGATCACTGAATTA	GCTTTCAAGGTGGGTGAATA
Cdk C1	symbB.v1.2.006023.t1	101%	CATCTCCTTCAGCCCAGAAATA	CCACGATCTCCTTCGTTTCA
Cdk C2	symbB.v1.2.024114.t2	No Amplification	CGTGGAGCATCCTTACTTTC	GTCCGCCATGTGCTTAAA
Cdk F	symbB.v1.2.028644.t1	97%	AGCACTGTCGCACATGAA	GCCAAATCCCATGCCAAATAG
Cdk Sm1	symbB.v1.2.029202.t1	95%	CTGCCGTGAGAGGTTGATT	GGAGGATCAGAACCGAATGAAG
Cdk Sm2	symbB.v1.2.000750.t1	100%	TTGTGAGCTCCTGGCTATTG	GGGCGTCTGAGAGATGTTAAG
Cdk Sm3	symbB.v1.2.008214.t1	99%	CAGGTGCAGGCTTTGATAGA	CCAGAGTTCAGCAAGGGTATAA
Cdk Sm4	symbB.v1.2.012451.t1	104%	GCTGAACTGGAAGCATCAATC	GGTGAGTAAGCAGCACAGAA
Cyclin A	symbB.v1.2.024265.t1	100%	TGGTTACTGGAAGCGTGTAAG	GCGATTTCCAAGGTGTTCAAATA
Cyclin B1	symbB.v1.2.033807.t1	99%	CAAAGATGAGGGCGATTCTAGT	CAGGCTCACCGCCATAAATA
Cyclin B2	symbB.v1.2.002641.t1	100%	TGAAGAAGCAATGGAGGAAGTC	GTCTCCTGGGCAAAGACATAAA
Cyclin B3	symbB.v1.2.020228.t1	97%	CAAGGAGCAGGCGAAGTATT	ACGCTCAACTCGCAATCTAC
Cyclin Sm1	symbB.v1.2.010386.t1	91%	CACTGCAGAGTCCCAAAGAA	GCGTCACCACCTGTCTTAAT
Cyclin Sm2	symbB.v1.2.008772.t1	90%	CAGGTGCAGGCTTTGATAGA	CCAGAGTTCAGCAAGGGTATAA
Cyclin P1	symbB.v1.2.021350.t1	96%	CCTGGAGCGCATCGATAAATA	GATAGCTTGCTTGCTGACTAAAC
Cyclin P2	symbB.v1.2.034711.t1	99%	GTCACAAGGCTTTCGTCTACT	CAAGGATGTAGCCCATCTTCTC
Cyclin P3	symbB.v1.2.024186.t1	99%	AATACCCACGAGTCCATCAATC	GACGGATTGGATTGGCTTTG
Cyclin P4	symbB.v1.2.031064.t1	98%	GTGGAGTTCGAACAAAG	GGGAGAAAGATCCAAATG
Cyclin P5	symbB.v1.2.005785.t1	No Amplification	CTTGGTAGTCGGATTGA	GGTCGTGAGAGGATAAG
Cyclin P6	symbB.v1.2.004739.t1	130%	CCGTCCAAAGAGACTAT	CTCACCATCCACAAATAC
MAT1	symbB.v1.2.023842.t1	93%	TGACAAGGAGATGAAGG	CTCCCTCAACATCAAATAC
CKS1A	symbB.v1.2.010262.t4	102%	CATGGAAGAATGGGAATG	AGGACTCGTCGAAATAG
CKS1B	symbB.v1.2.018958.t1	109%	AGGTATGGGCTTGATTT	GCTTCAACGACTTCTTC
Putative CKS1	symbB.v1.2.026341.t1	No Amplification	CAGAGTCAGAGGTTGTT	CCAGATTTCCTTAGAT

Figure S3: Real-time RT-PCR primers used to amplify potential cell cycle genes in *S. minutum*.
Primer sets with efficiency values between 90% and 105% were considered to be useful assays.