Department of Biology
Department of Plant Biotechnology and Genetics
VIB Department of Plant Systems Biology

Identification and functional characterization of cell cycle genes in the pennate diatom *Phaeodactylum tricornutum*

Marie Huysman

Promoter: Prof. Dr. Wim Vyverman
Co-promoter: Prof. Dr. Lieven De Veylder

Dissertation submitted in fulfillment of the requirements for the degree of Doctor (PhD) in Sciences, Biotechnology.
February 4th 2011

Academic year: 2010-2011

Onderzoek gefinancierd met een onderzoeksbeurs van het agentschap voor Innovatie door Wetenschap en Technologie (IWT).

This work was supported by a PhD grant from the agency for Innovation by Science and Technology (IWT).

Members of the examination committee

Prof. Dr. Ann Depicker (chairwoman)

Department of Plant Biotechnology and Genetics, Ghent University

Prof. Dr. Wim Vyverman (promotor, secretary)

Department of Biology, Ghent University

Prof. Dr. Lieven De Veylder (co-promotor)

Department of Plant Biotechnology and Genetics, Ghent University

Dr. Ir. Gino Baart

Department of Plant Biotechnology and Genetics / Department of Biology, Ghent University

Prof. Dr. Ir. Gerrit Beemster

Laboratory for Molecular Plant Physiology and Biotechnology, University of Antwerpen

Prof. Dr. Chris Bowler

Department of Biology, Ecole Normale Supérieure, Paris (France)

Prof. Dr. Dirk Inzé

Department of Plant Biotechnology and Genetics, Ghent University

Dr. Steven Maere

Department of Plant Biotechnology and Genetics, Ghent University

Prof. Dr. Koen Sabbe

Department of Biology, Ghent University

Dr. Heroen Verbruggen

Department of Biology, Ghent University

Acknowledgements

The work presented in this thesis could not be done without the support and help from several people. Therefore, a few words of thanks for all people that helped me to succeed this PhD.

First of all I would like to express my gratitude to **Wim Vyverman** and **Dirk Inzé**, for giving me the opportunity to join the Laboratory of Protistology and Aquatic Ecology and the Plant Systems Biology department five years ago. Wim, I really appreciate the confidence and trust you have in me, and the opportunities you gave me throughout the years. Your advice and support have been really valuable and certainly helped me to succeed my PhD. Dirk, thank you for your enthusiasm and your support during my PhD years.

Because of the joint project between both labs, I have been working at different places and with different people during those five years. At first, I was mainly working at "de Sterre" where **Jeroen Gillard** thought me everything about diatoms, those small algae that I vaguely remembered from one of the plant courses during my bachelor training. Jeroen reminded me that diatoms contain those beautifully shaped silica (glass) cell walls and recommended me to watch a spectacular DVD (Diatoms: life in glass houses by Jeremy Pickett-Heaps) that showed in detail the process that I would be studying, i.e. cell division. However, how to express my disappointment when I first saw under the microscope the diatom species that I would be working with; *Phaeodactylum tricornutum* turned out to be an incredibly small, naked - so no beautiful glass cell wall - and, let's be honest, a really ugly diatom - no wonder it avoids sex. Nevertheless, it was used as a model species in the "DIATOMICS" project that the lab participated in, meaning that I was stuck with it. I quickly realized that it would get difficult to explain to other scientists – especially to the IWT evaluation members - why this diatom was a good model species, so I started to focus on the good properties. While most of my colleagues at "de Sterre" were mocking about the fact that I was working with - what they called - "that weed, that probably isn't even a diatom", I met some true *Phaeodactylum* believers during my first meeting in Paris. **Chris Bowler**, project leader on the DIATOMICS project introduced me into the molecular secrets that were waiting to be discovered as the

genome of *Phaeodactylum* was about to be fully sequenced. Thanks Chris, for the wonderful collaboration and for sharing your enthusiasm and passion for diatom research. The collaboration with Chris' lab on studying diatom cell division helped me to set-up an optimal cultivation space and get acquainted with the cell biology of and molecular techniques for *Phaeodactylum*. In this respect, I would especially like to thank **Marc Heijde** and **Alessandra De Martino** for providing very useful tips and tricks.

Back in Gent, I could always count on **Cindy Martens** and **Klaas Vandepoele** for all my issues related to the bioinformatics part of the project. I am very grateful to them for the time they took to do several analyses over and over again and explain all those things that first seemed very abstract to me. I would also like to thank **Yves Van de Peer** for his support over the years and for suggesting the right journal for my first paper.

By the time I was getting familiar with taking care of my diatoms, I started the molecular work at the Plant Systems Biology Department at campus Technologiepark in Gent. At first, I joined the group of **Marnik Vuylsteke**. But Marnik quickly suggested that, as I was specifically working on cell cycle regulation, there was another group in the department more suited for me. Nevertheless, I enjoyed the years that I was a member of Marnik's group and I would like to thank him and all the **colleagues of the Quantitative Genomics** group for their hospitality and all the nice moments we had together in and outside the lab.

Thanks to Marnik's recommendation, I became a member of the Cell Cycle group, headed by **Lieven De Veylder**. Though I was still physically located in the lab of Marnik, I could go and disturb Lieven now at any time for questions. Which I did. Especially when preparing for my IWT defense. But Lieven was really patient and probably thanks to his tremendous coaching skills, I obtained the funding in the end - I probably was the only one who came out smiling from the IWT selection session that day. Lieven, I would really like to thank you for adopting me in your lab (and ultimately becoming my co-promoter), for your interest in my project (although not related to plants), the nice discussions and your support and guidance throughout the years. I would also like to thank **all the members of the Cell Cycle group** for the pleasant atmosphere throughout the years and especially our lunch-team (Sara, Véronique, Toon and Lieven V.).

Furthermore, to **all colleagues from PSB and PAE** and especially to all **support staff** (Leen, Wilson, Nino, Raf, Jackie, Kristof, David, Martine, Bernard, Hilde, Nathalie and many more), my **colleagues from the office at the PAE** (Aaike, Pieter, Jeroen VW, Renaat and

Els), my student **Joris** and my ***Phaeodactylum* colleagues** Gino, Michele and Michiel, thanks for all your support!

Natuurlijk zou dit alles niet gelukt zijn zonder de onvoorwaardelijke steun van het thuisfront. Daarom wil ik ook graag mijn familie, schoonfamilie en vrienden bedanken. In het bijzonder **mams en paps**, voor de kansen die jullie me geboden heb tijdens mijn opleiding en voor de uitstekende opvoeding die ik van jullie heb gekregen. Jullie zeggen het misschien niet met zoveel woorden, maar ik weet dat jullie in me geloven en enorm trots zijn. Hetzelfde geldt voor zus **Nathalie** (en **de Dejonghe bende**) en broer **Elkana**, enorm bedankt voor jullie mentale en fysieke steun (om me wakker te houden tijdens nachtelijke staalnames, wanneer Stephen een keertje niet kon, en om mijn pc te redden op het ultieme cruciale moment). Mijn liefje **Stephen** wil ik speciaal bedanken, om me telkens weer de rust en liefde te brengen die ik nodig had en me te steunen op welke manier dan ook. Ik ben er zeker van dat je enorm hebt afgezien met mij de laatste maanden, maar dat maak ik wel goed! En oké, ook al klinkt het misschien wat raar, ik wil ook mijn newfy **Bucky** bedanken. Een beter anti-stress middel bestaat er niet dan zo een levensgrote lieve knuffelbeer.

Tenslotte ook een woordje voor mijn beste vrienden. **Kim en Jens, Sarah en Christophe, Anne-Marieke en Laszlo, Séverine en Hans, en Chloë en Tom**, bedankt voor de vriendschap, steun en de nodige verstrooiing tussendoor. En ik hoop dat ik jullie nu eindelijk terug wat meer ga kunnen zien.

Marie

Abbreviations

3AT	3-amino-1,2,4-triazole
AD	Activator domain
BF	Bright field
bp	Base pairs
BY-2	Bright-Yellow 2
bZIP	Basic Leucine Zipper
CAK	Cyclin-dependent kinase activating kinase
Cdc	Cell division cycle
CDK	Cyclin-dependent kinase
cDNA	Complementary DNA
CHX	Cycloheximide
CKI	Cyclin-dependent kinase inhibitor
CKS	Cyclin-dependent kinase subunit
CPF1	Cryptochrome/Photolyase Family member 1
CYC	Cyclin
DAPI	4',6-diamidino-2-phenylindole
DBD	DNA binding domain
DCMU	3-(3,4-dichlorophenyl)-1,1-dimethylurea
DD	Dimerization domain
DEL	DP-E2F like
DNA	Deoxyribonucleic acid
DP	Dimerization partner
dsCYC	Diatom-specific cyclin
DTT	1,4-dithiothreitol
E2F	Adenovirus E2 promotor binding factor
EF1	Elongation Factor 1
EST	Expressed sequence tag
eYFP	Enhanced yellow fluorescent protein

fcpB	Fucoxanthin chlorophyll binding protein B
fluc	Firefly luciferase
G1 phase	Gap 1 phase
G2 phase	Gap 2 phase
GFP	Green fluorescent protein
GO	Gene Ontology
GUS	β-glucuronidase
H4	Histone H4
HA	Hemagglutinin
hCDK	Hypothetical cyclin-dependent kinase
His, H	Histidine
HMM	Hidden Markov Model
HSF	Heat shock factor
LacZ	β-galactosidase
Leu, L	Leucine
LOV	Light, oxygen or voltage
LRE	Light-responsive element
M phase	Mitosis
MC	Microtubule center
MCM	Minichromosome maintenance
mRNA	Messenger RNA
MT	Microtubule
MTOC	Microtubule organizing center
MYB	Myeloblastosis
NLS	Nuclear localization signal
PC	Polar complex
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PRC1	Protein regulator of cytokinesis 1
prom	Promoter
qPCR	Quantitative polymerase chain reaction
Rb	Retinoblastoma
RBR1	Retinoblastoma-related 1
RNA	Ribonucleic acid

S phase	DNA synthesis phase
SD	Synthetic dextrose
SDV	Silica deposition vesicle
TA	Transcription activator
TBP	TATA-box binding protein
TF	Transcription factor
Trp, T	Tryptophane
TubA	Tubulin A
UBI-4	Polyubiquitin gene 4
WT	Wild-type
X-gal	Bromo-chloro-indolyl-galactopyranosid
Y2H	Yeast-two-hybrid

Contents

Aims

Accounting for almost one fifth of the primary production on Earth, diatoms play a key ecological and biogeochemical role in our contemporary oceans (Van den Hoek, 1995; Smetacek, 1999; Falkowski et al., 2004). Furthermore, as producers of various lipids and pigments, and characterized by their finely ornamented silica cell wall, diatoms gained an emerging interest of different industrial fields, including biofuel production, nanotechnology and pharmaceutics (Lebeau and Robert, 2003; Kröger, 2007). However, despite their major ecological importance and their high commercial value, little is known about the mechanisms that control their life cycle. Their ability to live and dominate in highly unstable and sometimes harsh environments, suggests that diatoms have evolved specific strategies to adapt to and survive in such conditions. Unraveling the regulatory mechanisms that underlie their unique life cycle strategies will therefore be of crucial importance to understand diatom ecology and evolution and to further exploit their industrial potential.

The major aim of this thesis is to gain insights in the genetic mechanisms and environmental factors that control the diatom cell cycle. Because at the onset of this work, little was known about cell cycle regulation in diatoms, except for the elegant observatory microscopic studies in some large diatom species and some physiological work describing the effects of various environmental conditions on diatom cell cycle progression (Reviewed in **Chapter 1**), we started this study examining the newly sequenced genome of *Phaeodactylum tricornutum* (Bowler et al., 2008) for conserved and novel cell cycle regulators.

Chapter 2 describes the genome-wide identification and annotation of the diatom cell cycle genes in *P. tricornutum*, including the most important eukaryotic cell cycle regulators, the cyclin-dependent kinases (CDKs) and the cyclins (Morgan, 1997; Inzé and De Veylder, 2006; Doonan and Kitsios, 2009). To gain insight in their biological function we developed a cell cycle synchronization method that allowed us to monitor their transcript expression

during synchronized growth. Building on these results, a set of genes was defined that can be used as markers for specific cell cycle phases in diatoms.

From the first comparative genome analyses, it became clear that the cyclin family in diatoms represents an expanded gene family (Bowler et al., 2008). In **Chapter 2**, we examined if this expansion is specific to diatoms by comparing the cyclin family size of diatoms with those of closely related groups. Furthermore, the putative role of a novel type of cyclins, the diatom-specific cyclins, as integrators of environmental signals was assessed by nutrient starvation-repletion experiments.

When nutrients are not limiting, one of the limiting factors of diatom growth is light. **Chapter 3** describes the strict light-dependent regulation and functional characterization of *diatom-specific cyclin 2 (dsCYC2)*. Its putative role as a crucial regulator of the cell cycle upon changing light conditions was examined by studying the effects of *dsCYC2* overexpression and silencing on cell cycle progression.

In addition to the identification of the more general cell cycle regulatory mechanisms described in Chapter 2, **Chapter 4** elaborates specifically on the G1/S transition. Genome-wide analysis identified all members of the Retinoblastoma (Rb)-pathway, a conserved pathway controlling G1/S transition in plants and animals (Weinberg, 1995; de Jager and Murray, 1999). The functionality of a diatom-specific E2F motif, detected through *de novo* motif detection, was investigated by transactivation assays.

Chapter 5 deals with the ambiguous nature of *CDKA2*. This CDK was originally assigned to the A-type CDKs (Chapter 2), although its mitotic transcription is characteristic of the plant-specific B-type CDKs. To uncover its real identity and function, we determined the subcellular localization of CDKA2, identified its interaction partners and analyzed the effect of its overexpression on cell cycle progression.

The main conclusions and future perspectives of this thesis are summarized and discussed in **Chapter 6** and a summary is provided in **Chapter 7**.

REFERENCES

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kroger N, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin-Jezequel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Oudot-Le Secq MP, Napoli C, Obornik M, Parker MS, Petit JL, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature **456:** 239-244

de Jager SM, Murray JA (1999) Retinoblastoma proteins in plants. Plant Mol Biol **41:** 295-299

Doonan JH, Kitsios G (2009) Functional Evolution of Cyclin-Dependent Kinases. Molecular Biotechnology **42:** 14-29

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJ (2004) The evolution of modern eukaryotic phytoplankton. Science **305:** 354-360

Inzé D, De Veylder L (2006) Cell cycle regulation in plant development. Annu Rev Genet **40:** 77-105

Kröger N (2007) Prescribing diatom morphology: toward genetic engineering of biological nanomaterials. Curr Opin Chem Biol **11:** 662-669

Lebeau T, Robert JM (2003a) Diatom cultivation and biotechnologically relevant products. Part II: current and putative products. Appl Microbiol Biotechnol **60:** 624-632

Morgan DO (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu Rev Cell Dev Biol **13:** 261-291

Smetacek V (1999) Diatoms and the ocean carbon cycle. Protist **150:** 25-32

Van den Hoek C, Mann, D.G., and Jahns, H.M. (1995) Algae: An Introduction to Phycology.

Weinberg RA (1995) The retinoblastoma protein and cell cycle control. Cell **81:** 323-330

1

Introduction

Diatoms: ecology and economy

Life on Earth relies mainly on primary production, the production of organic compounds and oxygen from carbon dioxide and water, through the process of photosynthesis driven by the radiant energy of the sun. Although marine phytoplankton represents only 1% of the total photosynthetic biomass, they account for about 45% of the global primary production (Field et al., 1998; Falkowski et al., 2004). Diatoms are one of the most species-rich phytoplankton classes, responsible for approximately 20% of the total carbon fixation on Earth (Nelson et al., 1995; Van den Hoek, 1995). These single-celled eukaryotes belong to the heterokont algae (Chromalveolates), a group originated through secondary endosymbiosis when a photosynthetic red algae was engulfed by a heterotrophic host that presumably already contained a green genomic "footprint" from an earlier secondary symbiosis event (Li et al., 2006; Moustafa et al., 2009). Based on their morphology, diatoms can be divided into three major groups: the radial centrics (Coscinodiscophyceae) which represent the earliest fossil diatom records, followed by the bi- and multipolar centrics (Mediophyceae), and the youngest group of pennates (Bacillariophyceae) which diverged about 90 million years ago (Mya) (Medlin and Kaczmarska, 2004; Sims et al., 2006; Kooistra et al., 2007; Sorhannus, 2007). Genome analysis of a representative of two groups, *Thalassiosira pseudonana* (a multipolar centric) and *Phaedactylum tricornutum* (a pennate) (Figure 1.1), revealed the presence of a substantial set of genes of bacterial origin acquired through horizontal gene transfer that provide diatoms with novel metabolic capacities and signaling mechanisms (Armbrust et al., 2004; Bowler et al., 2008). Their complex evolutionary history, with multiple endosymbiotic events and an unprecedented level of horizontal gene transfer, is believed to have supplied the diatoms with the genetic potential to become one of the most ecologically successful and dominant primary producers in our contemporary oceans on a relative short period of time (Bowler et al., 2008; Moustafa et al., 2009). Although the oldest diatom fossil records come from the Early Jurassic, about 190 million years ago (Mya), molecular dating estimates indicated that diatoms probably

arose about 250 Mya during the Triassic period (Sims et al., 2006; Sorhannus, 2007). Diatoms began to colonize and dominate the open ocean after the Cretaceous-Tertiary mass extinction (65 Mya) and diatom species diversity peaked during the Eocene-Oligocene boundery, followed by a decline in diversity during the Oligocene (34 Mya) which has only partially been recovered since (Falkowski et al., 2004; Rabosky and Sorhannus, 2009).

Diatoms play a major role in the biogeochemical cycling of carbon and silica. They thrive in the photic zone of nutrient-rich upwelling regions, but once dissolved silicate is exhausted they sink out of the photic zone. In this way, they export carbon dioxide from the atmosphere (i.e. export production) and sequester it effectively to the ocean interior where it remains for centuries to millennia (Dugdale and Wilkerson, 1998; Falkowski et al., 1998; Allen et al., 2005). Generally, annual export production in the oceans is in the order of a third (16 Gt/yr) of the of the total ocean production (45-50 Gt per year) and diatoms would account for about 50% of the organic carbon export to the ocean interior (Dugdale and Wilkerson, 1998; Falkowski et al., 1998; Smetacek, 1999). Over the past 200 years, atmospheric CO_2 concentrations have increased about 100 ppm (from the relatively constant concentration of 280 ppm during the last 10000 year to the present concentration of 380 ppm) due to human activities such as burning of fossil fuels, deforestation and cement production (Lorius et al., 1990; Falkowski et al., 2000; Feely et al., 2008). About one-third of the anthropogenic CO_2 emissions has been absorbed by the oceans since the beginning of the industrial revolution (Sabine et al., 2004; Feely et al., 2008). Although this CO_2 uptake process reduces atmospheric greenhouse gas concentrations, thereby mitigating the climate effects of global warming, it has a serious impact on the ocean chemistry and biology due to acidification of the oceans. The expected continuous rise of CO2 levels at an increasing rate has been predicted to result in a decrease of surface ocean pH of about 0.4 units by the end of this century, a change that is likely to affect marine biota (Caldeira and Wickett, 2003; Feely et al., 2008). Besides ocean acidification, rising atmospheric CO_2 levels are expected to stimulate global warming resulting in melting ice caps, global sea level rise and increased vertical stratification of the ocean (Sarmiento et al., 2004). The latter could lead to a decrease in export production due to the isolation of the phytoplankton in the photic upper zone from the nutrients needed for photosynthesis and growth available in the deeper layers (Bopp et al., 2001). Because of the importance of diatoms in marine ecosystems and biogeochemical cycles, it will be of primary interest to understand the effects of lower pH, higher temperature and increased stratification on the physiology and productivity of diatoms. Therefore, understanding the mechanisms that

regulate the diatom cell and life cycle, characterized by obligate cell size reduction-restitution cycles (see below), will be of major importance to understand diatom population dynamics in changing conditions and to predict the biogeochemical impact of diatoms on the global carbon flow. Furthermore, by the generation of genetic diversity, diatom life histories also influence diatom species diversity, and in turn have an impact on their ecological role. For example, the genus *Pseudo-Nitzschia* comprises closely related species which either do or do not produce domoic acid, a neurotoxin causing Amnesic Shellfish Poisoning (Bates, 2000; Erdner et al., 2008).

Besides their enormous ecological importance, diatoms are also interesting from a biotechnological perspective (Reviewed in Lebeau and Robert, 2003a; Bozarth et al., 2009). Diatoms are used in aquaculture as food for mollusk larvae and shrimps, but also as bioremediators to treat effluent water from fish farms enriched in phosphate and nitrogen or water contaminated by heavy metals. As producers of a variety of metabolites, including oils, polyunsaturated fatty acids (e.g. the omega-3 fatty acid eicosapentaenoic acid), pigments, amino acids and antibiotics, diatoms can have applications in biodiesel production and in pharmaceutical and cosmetic industries as well (Lebeau and Robert, 2003a; Ramachandra et al., 2009). Moreover, their highly structured mesoporous cell wall (called frustule), made of amorphous silica, has drawn the attention of nanotechnology and material scientists (Kröger, 2007; Gordon et al., 2009). Diatoms are capable of creating a diversity of unique patterns and structures with micro- to nanoscale dimensions, which allows researchers to either select from the available species or attempt to modify their morphogenesis to apply in the nanofabrication of new materials based on diatom silica. Because of their structural, mechanical and optical properties, diatom frustules and silica-based structures have found widespread applications including biomineralization, biophotonics, photoluminescence, microfluidics, filtration, detection of trace gases, controlled drug delivery and computer design. (Gordon et al., 2009). Furthermore, the recent identification of proteins involved in the silica biomineralization process, including silaffins and long chain polyamines, lead to the development of synthetic biopolymers for the in vitro formation of silica and other inorganic materials from precursor solutions (Sun et al., 2004; Kröger, 2007; Kröger and Sandhage, 2010).

A major bottleneck for exploiting the industrial potential of diatoms lies in the cost-effectiveness of cultivation strategies. In general, large-scale production of microalgae for industrial applications occurs outdoors, either in open raceway ponds or closed tubular

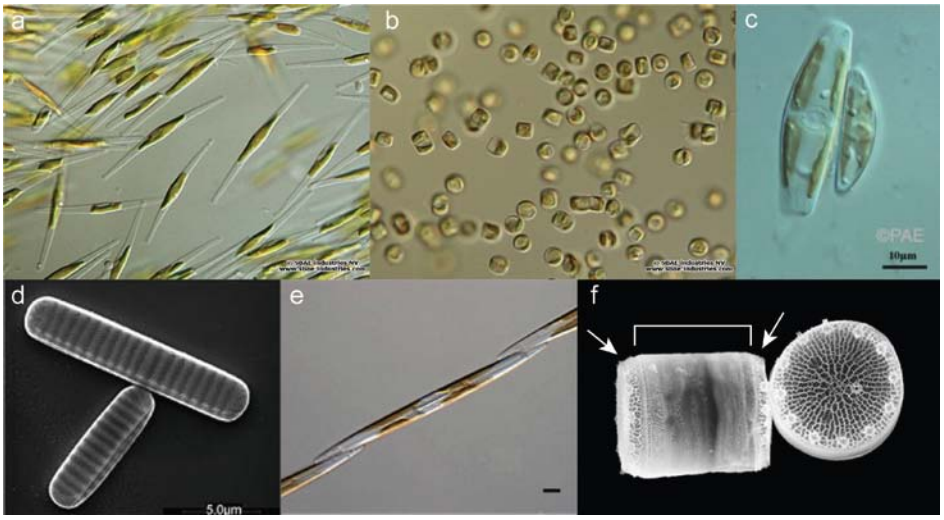


Figure 1.1: Microscopic images of the five diatom species for which the whole genome sequence is already available or is expected to be available soon: **a.** The pennate *Phaeodactylum tricornutum*, **b.** The centric *Thalassiosira pseudonana*, **c.** The pennate *Seminavis robusta*, **d.** The pennate *Fragilariopsis cylindrus*, and **e.** The pennate *Pseudo-nitzschia multiseries.* **f.** Electron micrograph of *Thalassiosira pseudonana* showing how the two valves (indicated by the arrows) are held together by siliceous girdle bands (brackets). (Picture courtesy of SBAE (sustainable bio-engineering for the aquatic environment) industries, Belgium (**a, b**); PAE (Protistology and Aquatic Ecology) laboratory, Ghent University, Belgium (**c**); H. Lange and G. Dieckmann, Alfred-Wegener Institute for Polar and Marine Research, Germany (**d**); K. Holtermann (**e**); and N. Kroger, Georgia Institute of Technology, Atlanta (**f**)).

photobioreactors, or a combination of both (Pulz, 2001; Chisti, 2007; Williams and Laurens, 2010). For both systems, the costs associated with energy supply, environmental control of light, nutrients, temperature, pH and gas, mixing, system set-up and harvesting are still high as compared to the profit that can be made from the final product (Lebeau and Robert, 2003b). However, the biomass productivity per volume is higher in photobioreactors compared to raceway systems resulting in considerably lower recovery costs (Chisti, 2007). While many efforts are ongoing to further optimize bioreactor design, improving the economics of diatom cultivation may greatly benefit from the generation and use of genetically or metabolically engineered diatom strains that have e.g. enhanced growth rate, or less growth-limiting properties associated with light (Lebeau and Robert, 2003b; Chisti, 2007). An understanding of the molecular mechanisms that control the diatom life cycle and the signaling pathways that link the environment with the growth process will therefore be necessary to optimize diatom cultivation. Because of their life in highly fluctuating environments (e.g. light and nutrient supply), diatoms most probably have evolved specific mechanisms to control their life cycle strategies.

The recent availability of the fully sequenced genomes of *T. pseudonana* (Armbrust et al., 2004) and *P. tricornutum* (Bowler et al., 2008) and the ongoing sequencing of three other pennate diatom species, the cold-loving *Fragilariopsis cylindrus*, the toxigenic *Pseudo-nitzschia multiseries* (http://www.jgi.doe.gov/genome-projects/), and the life cycle-model *Seminavis robusta* (personal communication A. Bones) (Figure 1.1), will provide the basis for exploring the conservation and the identification of unique molecular components controlling the diatom cell division process and will help us to decipher the mechanisms underlying the ecological success of diatoms.

The diatom life cycle

The diatom life cycle typically comprises two principal phases: a prolonged vegetative stage, lasting months to years, during which diploid cells divide mitotically, and a relatively short phase of sexual reproduction involving haploid gametes (Chepurnov et al., 2004). Vegetative reproduction in diatoms is inevitably connected to their main characteristic feature, the diatom cell wall. This silicified frustule is composed of two unequal halves: a larger epivalve and a smaller hypovalve, which fit into each other like the two halves of a Petri dish, and are held together by siliceous structures, called girdle bands (Round et al., 1990) (Figure 1.1f). Because diatom cells always divide in the valvar plane and mitotic division is accompanied by the formation of a new hypovalve within the confines of the parental cell wall, one of the two daughter cells will attain the same size of the mother cell, while the other will always be slightly smaller. As a result, the mean cell size in a proliferating diatom population decreases while the standard deviation increases, a phenomenon known as the McDonald-Pfeitzer rule (Chepurnov et al., 2004). This cell size reduction mechanism implies that cells will keep on dividing mitotically until they lose their viability due to extreme size diminuation. However, in general, when cells become small enough they can enter the sexual phase of the life cycle and, as such, escape their fatal fate. The maximum size of cells that can be sexually induced is referred to as the sexual size threshold and this is a species-specific property. In addition, the secondary cues to switch from vegetative to meiotic division and the specific modes of sexual reproduction itself are fundamentally different between centrics and pennates. Auxosporulation in centric diatoms is influenced by external species-specific factors such as

light irradiance, day length and temperature. Centrics form within one clonal culture large egg cells and motile, flagellated sperm cells (oogamy). Although auxosporulation in pennate diatoms is also dependent on environmental factors, the primary determinant of gametogenesis onset in pennates seems to be cell-cell interaction between vegetative cells from different sexually compatible clones (mating types). The gametes produced by most pennates are, in contrast to those produced by centric diatoms, non-flagellated and morphologically identical (isogamy) (Chepurnov et al., 2004). Some diatoms have evolved other strategies besides auxosporulation to restore their original cell size, e.g. through vegetative enlargement (Van den Hoek, 1995; Chepurnov et al., 2004). In this thesis we will focus on the regulation of vegetative cell division in diatoms. Essentially, vegetative reproduction in diatoms occurs through mitotic divisions, regulated by a process known as the cell cycle.

Molecular control of the eukaryotic cell cycle

The eukaryotic cell cycle comprises the coordinated succession of a phase of DNA replication (the DNA synthesis phase or S-phase), and a phase of physical separation of both copies of the genomes (mitosis or M-phase) and cell division itself (cytokinesis). Both S- and M-phase are separated from each other in time by two gap phases, one preceding S-phase (G1-phase) and the other preceding M-phase (G2-phase) (Figure 1.2). A tight coordination of the cell cycle process is essential to the reproduction and development of every living organism. Different cell cycle checkpoints (e.g. at the G1-S and G2-M transition) ensure that the genetic information is inherited correctly by inhibiting the replication and distribution of incomplete or damaged chromosomes to the daughter cells. The major cell cycle control points represent the onset of DNA replication (the G1-S transition) and mitosis itself (the G2-M transition) (Buchanan et al., 2000). In addition, most organisms show during the mid-to-late G1 phase a commitment point (known as START in yeast, restriction point in animals or commitment point in *Chlamydomonas*), before which a number of decisions, depending on intra- and extracellular information, must be integrated (Oakenfull et al., 2002).

Generally, in eukaryotes, cell cycle regulation is controlled at multiple points by an evolutionarily conserved set of proteins, the cyclin-dependent kinases (CDKs) and cyclins, that can form functional kinase complexes (Reviewed in Morgan, 1997; Inzé and De Veylder, 2006). In these complexes, the CDKs and cyclins act as catalytic and regulatory subunits,

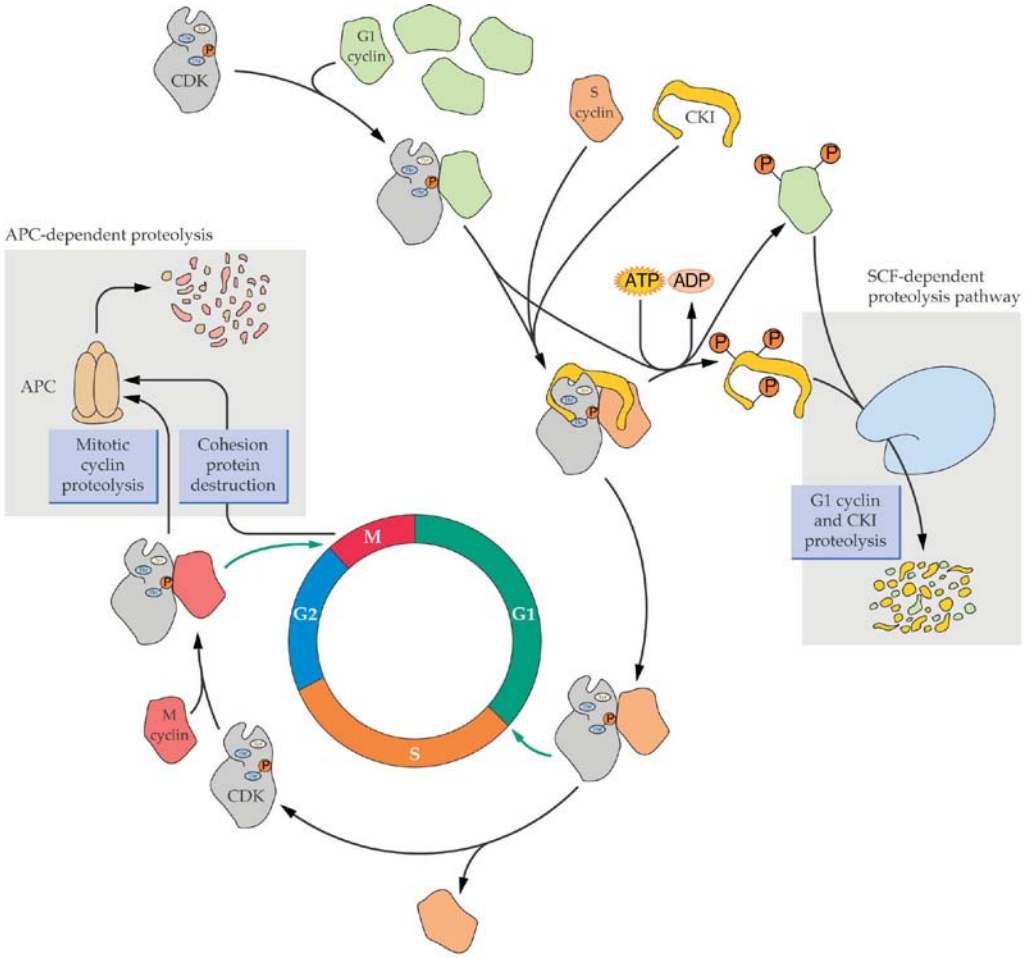


Figure 1.2: Overview of the eukaryotic cell cycle regulation (redrafted from Buchanan et al., 2000). The cell cycle consist of four consecutive phases: G1 phase, S phase (DNA synthesis), G2 phase and M (mitosis) phase. The main checkpoints are situated at the G1-S and G2-M transition (green arrows). The activity of the central regulators, the cyclin dependent kinase (CDK)/cyclin complexes, is controlled through phosphorylation and dephosporylation of the CDK subunit and interaction with CDK inhibitors (CKI). At the G1-S transition G1 cyclins and CKIs are targeted for degradation by the SCF (Skp1/Cdc53/F-box) protein complex. Exit from mitosis is induced by the degradation of mitotic cyclins and chesion proteins by the APC (Anaphase Promoting Complex) complex.

respectively. CDKs were first discovered in *Xenopus* eggs as the active component of maturation-promoting factor (MPF), a complex that expresses kinase activity and enables entry in mitosis (Gautier et al., 1988). Since then, multiple CDKs have been identified in different eukaryotic organisms and based on the cyclin-binding motif present the CDK family can be subdivided in different groups (Doonan and Kitsios, 2009). The most conserved type of CDK holds a PSTAIRE cyclin-binding motif (cdc2/cdc28 in yeast, CDK1/2 in animals and CDKA in plants and algae), while the plant-specific B-type CDKs displays a PPTALRE/PPTTLRE motif. Other CDK subfamilies (CDK9/CDKC, CDK7/CDKD, CDK8/CDKE in animals/plants

respectively) are related to the PSTAIRE CDK, but some have been shown to be involved in transcriptional control or splicing rather then cell cycle regulation (Coqueret, 2002; Kitsios et al., 2008). Specific functions of the different CDK groups during the cell cycle are discussed more in detail in Chapter 2.

The regulatory partners of the CDKs, the cyclins, were initially identified in fertilized sea urchin eggs as proteins with particular oscillatory patterns of gene expression and protein destruction during the cell cycle (Evans et al., 1983). Because cyclins are involved in determining the substrate specificity of the CDK complex, the fluctuation of cyclin abundance creates different kinase specificities and activities of the CDK/cyclin complex at different stages of the cell cycle. In general, eukaryotes express two main groups of cyclins: the G1 cyclins that regulate G1-S transition (called Cln in yeast and D-type cyclins in animals and plants) and the mitotic cyclins that control G2-M transition (called Clb or cdc13 in yeast and A- or B-type cyclins in animals and plants) (Oakenfull et al., 2002). CDK/cyclin complex activity is not only controlled by cyclin association, but also by phosphoregulation of the CDK subunits by regulatory proteins (see further) and interaction with CDK inhibitors (CKIs) or scaffolding proteins like CKS1/Suc1 (CDK subunit) (Pines, 1996; Harper, 2001; De Clercq and Inzé, 2006) (Figure 1.2).

All eukaryotes except yeasts share the retinoblastoma (Rb)-mediated pathway for G1-S regulation, involving the Rb protein and E2F/DP transcription factors (Figure 1.3) (Weinberg, 1995; Claudio et al., 2002). Upon mitogenic stimulation (by external and/or internal factors) G1 cyclins are produced that associate with a CDK partner (Figure 1.2) (Oakenfull et al., 2002).

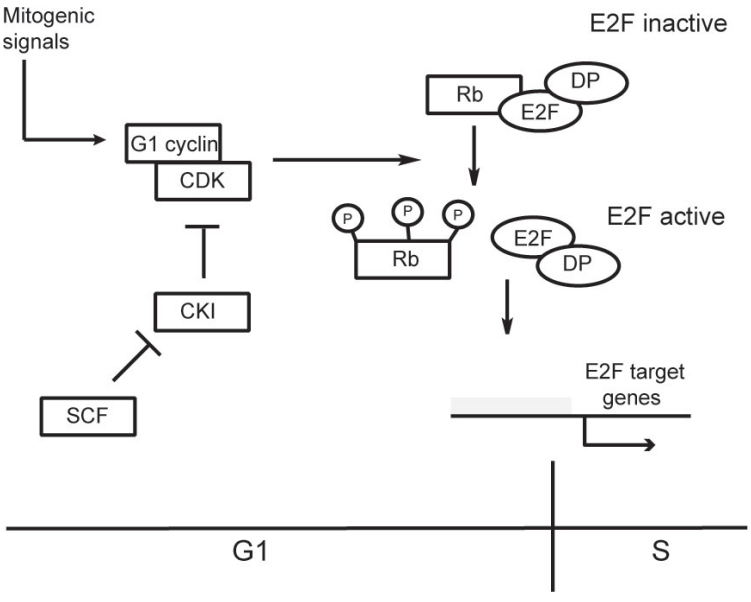


Figure 1.3: Regulation of the G1-S phase transition by the Rb-mediated pathway. Mitogenic signals trigger the transcription of G1 cyclins that form functional complexes with CDKs and promote cell cycle progression by inactivating the Rb protein through phosphorylation, resulting in the activation of the E2F/DP transcription factor complex and S-phase gene expression. Arrows indicate activation and T-shaped lines indicate inhibition.

In turn, the CDK/cyclin complex is activated upon dissociation and degradation of CDK inhibitors (CKIs) mediated by the SCF (Skp1/Cdc53/F-box protein) ubiquitination complex (Figure 1.2 and 1.3) (Deshaies and Ferrell, 2001). The activated CDK/cyclin complex then phosphorylates the Rb protein. In hypophosphorylated form the Rb protein binds to the E2F and DP transcription factors, rendering them inactive. Phosphorylation and hence dissociation of the Rb protein results in the activation of the E2F/DP complex and the transcriptional activation of genes involved in DNA replication and S-phase onset (Figure 1.3) (Weinberg, 1995; de Jager and Murray, 1999; Claudio et al., 2002).

Activation of the CDK/cyclin complex responsible for G2-M transition depends on phosphorylation and dephosphorylation of specific amino acid residues of the CDK subunit (Figure 1.4) (Doonan and Kitsios, 2009). A dual-specific Thr/Tyr phosphatase (CDC25 in yeast and animals) dephosphorylates the inhibitory phosphorylation sites Tyr^{15} en Thr^{14}, while a CDK-activating kinase (CAK) phosphorylates the conserved Thr^{160} residue on the CDK subunit. In contrast, inactivation of the CDK/cyclin complex is accomplished by phosphorylation of the Tyr^{15} en Thr^{14} residues by a family of inhibitory kinases (WEE1/MIK1/MYT1 family) (Figure 1.3) (Perry and Kornbluth, 2007). The activation of the CDK/cyclin complex eventually results in the induction of events that precede cytokinesis, such as spindle synthesis, chromatin condensation and nuclear envelope breakdown. Exit from mitosis is enhanced by ubiquitin-mediated degradation of anaphase inhibitors and mitotic cyclins by the anaphase-promoting complex (APC) (Figure 1.2) (Stals and Inzé, 2001; Capron et al., 2003).

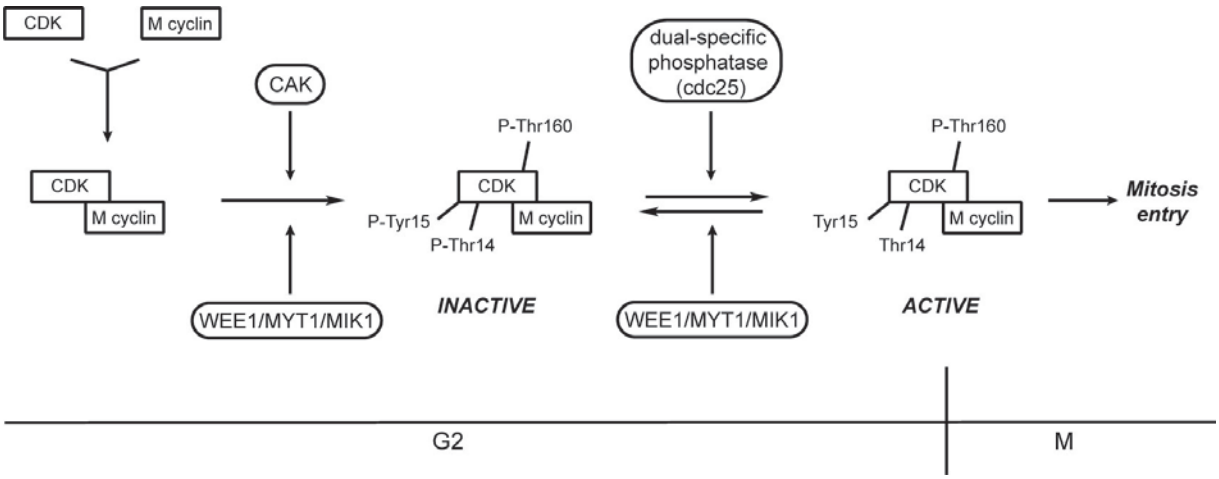


Figure 1.4: General control mechanism of the G2-M transition in eukaryotes. Before mitosis onset CDKs are kept at an inactive state by phosphorylation. Inhibitory phosphorylation at Thr^{14} and Tyr^{15} is removed by a dual-specific phosphatase, resulting in activation of mitotic CDKs and mitosis entry. Arrows indicate activation.

The regulatory mechanisms of G1-S and G2-M transition described above have been shown to apply for all higher eukaryotes for which cell cycle regulation has been studied so far (Reviewed in Morgan, 1997; Inzé and De Veylder, 2006) and genome-wide analyses of different alga species (*Ostreococcus tauri*, *Chlamydomonas reinhardtii* and *Ectocarpus siliculosus*) have revealed the general conservation of the main cell cycle regulatory components and mechanisms in these organisms as well (Bisova et al., 2005; Robbens et al., 2005; Bothwell et al., 2010). In addition, passage through G1-S in yeast is regulated by a mechanism similar to the Rb-mediated pathway in animals and plants but it involves distinct regulatory components, including WHI5, an inhibitor of the G1-S transcription factor SBF/MBF (Mendenhall and Hodge, 1998).

The diatom cell cycle

Organellar and structural organization during diatom mitosis

Although the first light microscopy studies of diatom mitosis at the end of the 19th century by Robert Lauterborn already revealed that cell division in diatoms displays major differences compared to the classical paradigms of animal and plant cells (Lauterborn, 1896; translated and summarized in Pickett-Heaps et al., 1984), many mysteries have remained. Because of the rigid nature of the frustule, only a few studies of the internal structures of diatom cells have been made by transmission electron microscopy, further impeding studies of their basic biology. However, these studies have shown that diatoms contain unique ultrastructural characteristics pointing to the likely existence of diatom-specific features important for cell division (reviewed in De Martino et al., 2009).

Chloroplast division and development are tightly linked with cell cycle progression in a number of diatom species (Chepurnov et al., 2002) and this likely contributes to an equal distribution of the daughter chloroplasts on each side of the future division plane. In fact, by cytological observations and transcriptional analysis of synchronized cultures, chloroplast division has been shown to precede karyokinesis (i.e. nuclear division) and cytokinesis (i.e. cellular division) in the pennate diatoms *Seminavis robusta* (Gillard et al., 2008) and *P. tricornutum* (De Martino et al., Manuscript in preparation) (Figure 1.5). However, the molecular mechanisms of chloroplast movement and division in diatoms remain unclear. In diatoms containing more than two chloroplasts, in the red alga *Cyanidioschyzon merolae* and in plants

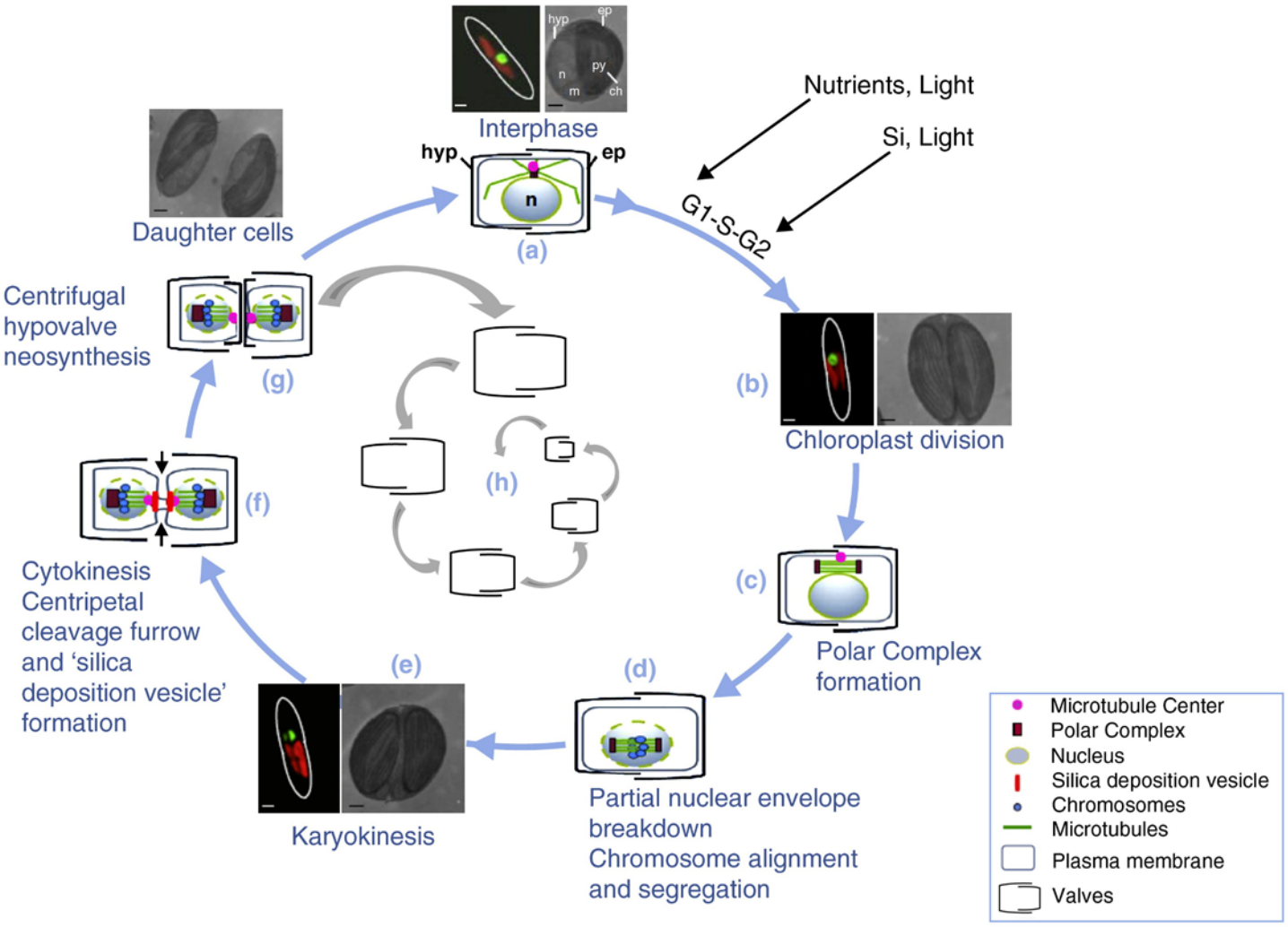


Figure 1.5: Overview of the diatom cell cycle (redrafted from Bowler et al. 2010). Diatoms divide principally asexually, through mitosis **(a–g)**. Diatom cells are confined within a rigid cell wall consisting of two silicified valves organized with the smaller (hypovalve (hyp)) fitting into the larger (epivalve (ep)). Because mitotic division is accompanied by the formation of a new hypovalve within the confines of the parental cell wall, one of the two daughter cells will attain the same size of the mother cell, while the other will always be slightly smaller **(g–a)**. When a critical species-specific size is reached, the sexual cycle can be induced to restore the maximal initial cell size **(h)**. Chloroplast division **(b)** precedes karyokinesis **(e)** and cytokinesis **(f)**. One chloroplast segregates in two, positioned on each side of the future plane of division. Mitosis is open, with partial nuclear envelope breakdown, and involves a unique MTOC consisting of the MC in interphase cells and the PCs in pre-mitotic and mitotic cells **(a, c, d, f, g)**. At cytokinesis, cells divide by centripetal invagination of the plasma membrane, which also involves the MC **(f)**. A new hypovalve is created within the SDV **(f)** which extends centrifugally before being exocytosed. Flow cytometric analyses performed on different diatom species have indicated the presence of two checkpoints, in G1 and G2, dependent on light and nutrient availability (see text). Progression through G2 depends on silicate availability for those species requiring it. The electron micrographs represent transverse sections of *P. tricornutum* cells at different stages of cell division. **(a)** Interphase cell with one nucleus (n), one chloroplast (ch) with one pyrenoid (py), mitochondria (m), epivalve (ep), and hypovalve (hyp); **(b)** Cell with two daughter chloroplasts; **(e)** Cells after karyokinesis; **(g)** Daughter cells. Fluorescent images represent confocal images of *P. tricornutum* cells in

interphase **(a)**, after chloroplast division **(b)**, and with two daughter nuclei **(e)**. Red: chlorophyll autofluorescence; Green: histone H4-GFP fluorescence in nuclei. Scale bars are 1 μm in electron micrographs and 2 μm in the fluorescent images. (Bowler et al., 2010).

chloroplast movements have been shown to be actin-dependent (de Francisco and Roth, 1977; Nishida et al., 2005; Krzeszowiec et al., 2007). However, the transcriptional induction of beta-tubulin in *S. robusta* and the tight microtubule (MT) network surrounding the constriction sites on the mother chloroplast in *P. tricornutum* suggest the involvement of MTs during chloroplast segregation and rearrangement, but does not exclude the possibility that actin may also be involved (Gillard et al., 2008; De Martino et al., Manuscript in preparation).

Diatom cells typically possess one microtubule center (MC), a dense spherical body positioned next to the interphase nucleus, from which MTs radiate (Figure 1.5a). The diatom MC resembles the microtubule organizing center (MTOC) of animal cells, the centrosome (Lloyd and Chan, 2006), but without centrioles (reviewed in Pickett-Heaps and Tippit, 1978 and De Martino et al., 2009). Besides its structural differences, the diatom MC also behaves differently during the cell cycle. Whereas animal centrosomes, present as single entities in the cell, duplicate before mitosis and give rise to the spindle poles (Azimzadeh and Bornens, 2007), the diatom MC is involved in initial spindle creation, after which it disintegrates (Figure 1.6a-b). During mitosis, new functional MTOCs appear that cap the poles of the diatom spindle, designated the polar complexes (PCs) (Figure 1.5c and 1.6a-b). At cytokinesis, the PCs dissappear while a new MC arises (Pickett-Heaps and Tippit, 1978) (Pickett-Heaps, 1991) (Figure 1.5d-g). These observations differ from what is known in animals and plants and therefore diatoms appear to undergo cell division in a rather unique way (De Martino et al., 2009). More particularly, diatom mitosis involves the action of an unusual, highly organized "central spindle" that is initially formed outside the nucleus (Figure 1.2c and 1.6a) and consists of parallel MTs arranged as two interdigitated half-spindles creating a central overlap region of MTs (Figure 1.6a, c and d) (Tippit and Pickett-Heaps, 1977; Pickett-Heaps and Tippit, 1978; Wordeman et al., 1986). The diatom spindle is encircled by a dense matrix, the "collar" (Figure 1.6c), which is supposed to help attach the spindle to an atypical type of "presumptive kinetochores", the sites of chromosome attachment (Figure 1.6d) (Tippit and Pickett-Heaps, 1977; Tippit et al., 1980). Reconstruction of the kinetochore fiber in diatoms has shown that unlike in convential kinetochores, most of the MTs do not nucleate at this site, but extend past the kinetochore, suggesting that the diatom kinetochores function by associating to MTs

originating from the poles, rather than by nucleating MTs (Tippit et al., 1980).

Cytokinesis of diatoms also differs from that of animals and plants: whereas plant cells build their new cell membrane and cell wall centrifugally (inside-out), diatom cells, as animal cells, first divide centripetally (outside-in) by invagination of their plasma membrane and then create the new hypovalve centrifugally inside an intracellular acidic tubular compartment, the silica deposition vesicle (SDV), a cellular structure unique to diatoms (Pickett-Heaps and Tippit, 1978; Pickett-Heaps, 1991; De Martino et al., 2009) (Figure 1.5f-g). During this process, the new MC is positioned adjacent to the SDV, suggesting a role during valve synthesis (Figure 1.5f) (Boyle et al., 1984; Edgar and Pickett-Heaps, 1984). When the new hypovalve is fully synthesized in the SDV it is secreted from the cell by exocytosis, after which the daughter cells can separate (Figure 1.2g).

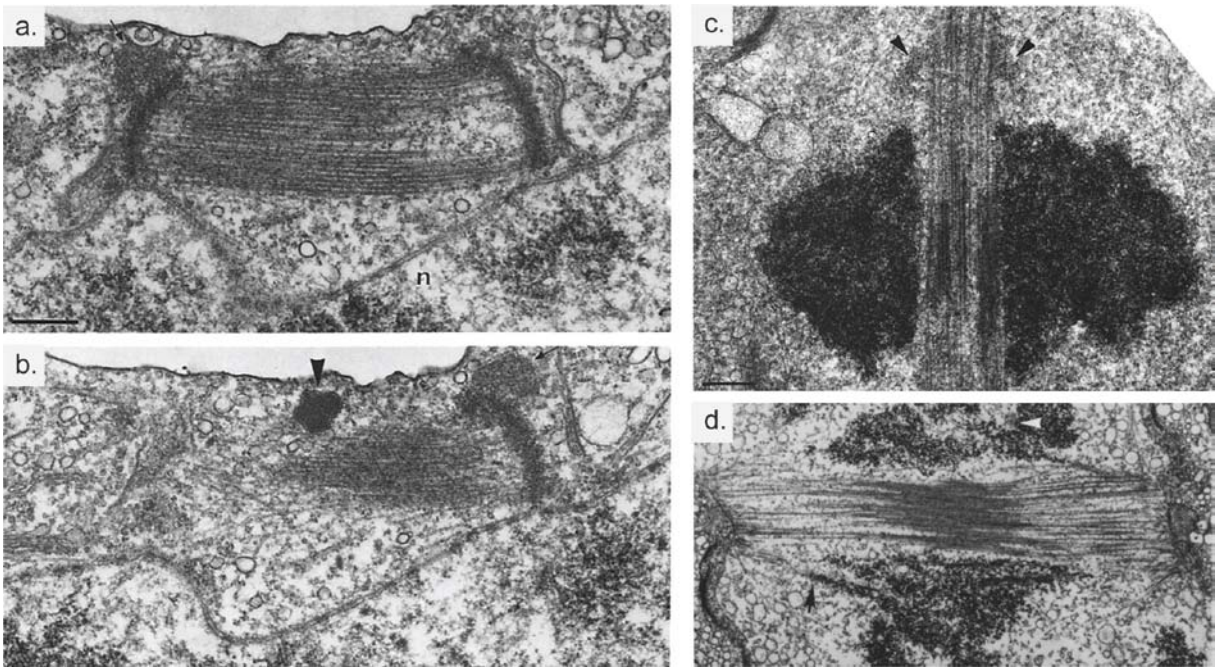


Figure 1.6: Transmission electron micrographs illustrating the unique mitotic structures in diatoms. **a.** and **b.** Two sections through the same late prophase spindle in the diatom *Surirella ovalis*. Each polar complex (PC) shows a thickening at one end (small arrows). The MTs of the central spindle appear to run from pole to pole, while other MTs invaginate the nuclear envelope. The MC (large arrow) starts to disintegrate at this stage, and is not observed at metaphase and anaphase. Bar, 0.5 µm. Image taken from Tippit and Pickett-Heaps, 1977. **c.** High voltage micrograph of a thick section of late metaphase *S. ovalis* cell showing "the collar" (arrows) surrounding the central spindle between the chromatin and the poles. Bar, 0.5 µm. Image taken from Tippit and Pickett-Heaps, 1977. **d.** Attachment of chromosomes to the central spindle via ”presumptive kinetochores” (black arrow) during mid-prometaphase in *Hantzschia*. Some chromosomes still have no MTs associated with their kinetochores (white arrow). x 11000. Image taken from Tippit et al., 1980.

Diatom cell cycle regulation: clues from the genome

The molecular mechanisms underlying the complex and tightly coordinated production and orientation of the unique mitotic and cytokinetic structures in diatoms, as well as the decision of the start of the division cycle still need to be uncovered. Homology searches in the genomes of both sequenced diatoms for mitotic checkpoint and spindle-associated regulatory molecules resulted in the identification of genes important for chromosome segregation (including a kinetochore protein ZW10, several chromokinesins and extra-spindle pole-like proteins) and genes involved in spindle assembly and elongation (diatom spindle kinesin 1 and several dyneins) (De Martino et al., 2009). Furthermore a centromere-specific histone H3 variant was detected in *P. tricornutum*, but not in *T. pseudonana*, suggesting a low level of similarity of centromeric proteins among diatoms (De Martino et al., 2009). Interestingly, comparative genomics revealed that the cyclin gene family represents one of the expanded gene families in diatoms, next to histidine kinases and heat-shock transcription factors (Bowler et al., 2008). In addition to members of each of the canonical families of cyclins, 10 and 42 diatom-specific cyclin genes were found in *P. tricornutum* and *T. pseudonana*, respectively (Bowler et al., 2008; see also Chapter 2). The dramatic expansion of this gene family may reflect the unusual characteristics of diatom life cycles due to the rigid nature of their cell walls, such as the control of cell size reduction, the activation of sexual reproduction at a critical size threshold, and life in rapidly changing and unpredictable environments (Kooistra et al., 2007). In addition, it may be significant that genes encoding RCC1 proteins (RCC, regulator of chromosome condensation), which are also involved in cell cycle control, have been expanded in both diatom genomes (Ohtsubo et al., 1989; Bowler et al., 2008).

Physiological and molecular responses to life in unstable environments

The dominance of diatoms in phytoplankton assemblages of marine and freshwater ecosystems when conditions are favorable suggests that they possess efficient sensing and signaling mechanisms that allow them to respond or adapt adequately to environmental fluctuations such as light and nutrient supply (Falciatore et al., 2000). The use of the recently advanced molecular techniques that can be applied to diatoms, including genetic transformation and genome-wide transcript analyses (Falciatore et al., 1999; Siaut et al., 2007; Gillard et al., 2008; De Riso et al., 2009), can now complement physiological studies in order

to study the link between the environment and control of diatom life strategies.

Heat shock factors

Regulation of gene expression is an important mechanism in every organism to control growth, development, cell cycle progression and adaptation to variable environmental conditions. Transcription factors (TFs) play a key role in this regulation, by binding specific DNA sequences, *cis*-acting elements, in the promoter region of their target gene and thereby promote (activator) or block (repressor) the recruitment of the transcriptional machinery to bind and transcribe the gene. Alternatively, TFs may also interact directly with the transcriptional machinery or they may introduce chromatin modifications to regulate transcriptional activity. Comparative genome analyses of the genomes of *P. tricornutum* and *T. pseudonana* revealed that the heat shock factor (HSF) family is expanded and represents the most abundant class of TFs (Bowler et al., 2008) with 69 copies in *P. tricornutum* and 89 copies in *T. pseudonana* (Montsant et al., 2007), respectively. Together, these HSFs represent almost half of all TFs identified in both sequenced diatoms. The significance of this expansion is unclear, but expressed sequence tag (EST) data indicates that the majority is expressed and that some members are induced specifically in response to certain growth conditions (Bowler et al., 2008; Maheswari et al., 2009).

Light

For any photosynthetic organism, including diatoms, light is an extremely important factor that influences growth. Because diatoms can grow over a wide range of light intensities and wavelengths, they are believed to have developed specific photo-acclimation and photo-adaptation mechanisms (Huisman et al., 2004; Lavaud et al., 2004; Lavaud et al., 2007). As in most other phytoplankton species, timing of diatom cell division can be entrained by alternating periods of light and dark, implying that the cell cycle consists of light-dependent and light-independent segments (Vaulot et al., 1986). Accordingly, both by light limitation and deprivation experiments, light-controlled restriction points have been identified in several diatom species, either only during the G1 phase, or both during the G1 and G2-M phases of the cell division cycle (Olson et al., 1986; Vaulot et al., 1986; Brzezinski et al., 1990; Gillard et al., 2008) (Figure 1.5).

Detailed investigation of the main TF families in diatoms led to the identification of putative orthologs of TFs involved in light signaling (Rayko et al., 2010). These include a

group of Myb proteins containing a single Myb domain (Myb1R) that belong to the SHAQKYF-like family described in plants and green algae that include the clock genes *CCA1* (*circadian clock associated 1*) and *LHY* (*late elongated hypocothyl*) (Wang et al., 1997). Furthermore, aureochrome-like sequences and bZIP-PAS proteins were found, which might both represent classes of photoreceptors that contain putative light-sensitive and DNA-binding domains. Of particular interest, aureochromes contain a bZIP (basic Leucine Zipper) domain and a LOV (light, oxygen or voltage) domain, also found in the phototropin family of blue-light photoreceptors (Takahashi et al., 2007; Ishikawa et al., 2009).

Blue light (350-500nm) is the most prevailing band below the surface of oceanic waters (MacIntyre et al., 2000), so efficient blue light sensing and signaling mechanisms are expected to play a crucial role in control of diatom growth. Related to this, a blue-light sensor (CPF1) has recently been identified and characterized in *P. tricornutum*. Overexpression of this cryptochrome/photolyase family member affected blue-light induced expression of genes involved in cell cycle regulation and DNA repair, suggesting a role for *CPF1* in perception and signaling of environmental light conditions and linking these to cell cycle progression (Coesel et al., 2009). In addition to the small-scale analyses of light-regulated gene expression in diatoms (Leblanc et al., 1999; Coesel et al., 2008; Zhu and Green, 2010), Nymark and colleagues demonstrated by global transcriptional profiling that diatoms are efficiently capable of coping with the damaging effects of excess light by induction of photoprotective mechanisms. Upon high-light exposure of *P. tricornutum*, they found an initial response of rapid induction of genes involved in photosynthesis, pigment metabolism and reactive oxygen species (ROS) scavenging systems (Nymark et al., 2009).

Nutrients

In addition to light, nutrient availability can also strongly affect diatom population dynamics. Nitrogen, together with the micronutrient iron, is generally considered to be a major limiting factor of primary production in the oceans (Falkowski et al., 1998). The effects of nitrogen limitation and starvation on diatom cell cycle progression have been demonstrated by flow cytometry to induce an expansion of the G1 phase or an arrest at multiple G1 checkpoints, respectively (Olson et al., 1986; Vaulot et al., 1987). Moreover, in *T. weissflogii* administration of periodic nitrogen pulses can override the photoperiod-determined cell division phasing, suggesting that nutrient control of the cell cycle precedes the light control in this diatom (Olson and Chisholm, 1983).

Recent studies illustrate how diatoms have developed different strategies to survive upon and acclimate rapidly to iron limitation through transcriptional and biochemical reconfiguration of iron requirement and acquisition pathways and by the use of ferritin to maintain internal iron storage (Allen et al., 2008; Marchetti et al., 2009). Diatoms can also respond efficiently to phosphate limitation, which is considered to be less common in the marine environment as compared to nitrogen and iron limitation. Nevertheless, phosphate limitation has been reported in certain oceanic areas (Wu et al., 2000) and has been hypothesized to have been more wide-spread during the glacial periods (Pichevin et al., 2009). Van Mooy and colleagues showed that diatoms reduce their phosphorus demand upon phosphorus limitation, and maintain their growth by substituting phospholipids with non-phosphorus membrane lipids (Van Mooy et al., 2009). Furthermore, differential transcription of some diatom-specific cyclin genes during nutrient starvation-repletion experiments in *P. tricornutum* showed the importance of phosphate as a cell cycle rate-limiting nutrient (see Chapter 2).

Silicon

As mentioned before, dissolved silicon can be a major limiting factor for diatom reproduction. Silicon limitation in diatom cultures induces cell cycle arrest at the G1-S boundary and during the G2-M phases (Vaulot et al., 1987; Brzezinski et al., 1990) and these arrest points have been linked to the requirement for silica during DNA replication and frustules formation, respectively (Coombs et al., 1967; Darley and Volcani, 1969; Okita and Volcani, 1980; Vaulot et al., 1987). Furthermore, in some species that deposit siliceous setae, like *Chaetoceros* species, silicon limitation experiments revealed an additional checkpoint early at the G1-phase, related to setae formation (Brzezinski et al., 1990). Whole genome expression profiling of the diatom *P. tricornutum* in response to different silica conditions led to the identification of novel compounds putatively involved in silicic acid storage and transport (Sapriel et al., 2009).

Phaeodactylum tricornutum: a suitable model organism for studying the diatom cell cycle

When studying a particular biological process, the choice of model organism needs to be considered well. In this study, we have chosen to work with the model organism *P.*

tricornutum for various reasons. First, the laboratory convenience of this species, including a short generation time of about one division per day, ease of cultivation due to the absence of a sexual phase during its life cycle, the availability of its full genome sequence and the presence of a wide range of molecular tools for genetic transformation, including overexpression and knock-down (Brzezinski et al., 1990; Falciatore et al., 1999; Siaut et al., 2007; Bowler et al., 2008; De Riso et al., 2009). In addition, for decades this species has been the subject of several cytological, physiological and biochemical experiments. A considerable advantage arises from the absence of a sexual phase in *P. tricornutum*. Although this complicates the generation of knock-out lines in a diploid organism like *P. tricornutum* by classical crossing techniques, it constrains reproduction to the vegetative phase, without the typical cell size reduction. This allows us to study the diatom cell division process independent from any other processes that in other diatoms are inevitably linked to life cycle regulation, e.g. cell size sensing and induction of sexual reproduction.

Although not considered to be of major ecological significance, *P. tricornutum* has been found all over the world, especially in dynamic coastal and estuarine waters (De Martino et al., 2007; Lavaud et al., 2007). *P. tricornutum* is a pleiomorphic species that can appear in three main morphotypes (oval, fusiform and triradiate) (Figure 1.7) (Lewin et al., 1958). The plasticity of its shape relates to the rather atypical nature of its cell wall, which is mainly organic and only poorly silicified (Borowitzka and Volcani, 1978). As a consequence, *P. tricornutum* does not have an obligate requirement for silica during growth (Brzezinski et al., 1990). Morphological changes of *P. tricornutum* within clonal cultures can occur and these can be influenced by specific culture conditions, such as nutrient depletion, transfer to solid medium and low temperature, while morphological stability might be dependent on the genotype (Borowitzka and Volcani, 1978; Tesson et al., 2009). These interchangeable

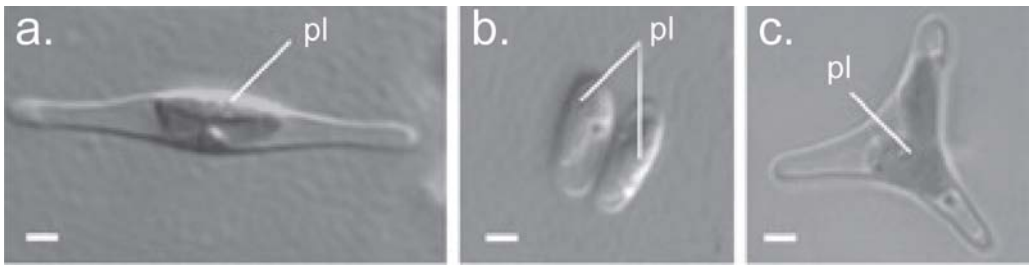


Figure 1.7: The three different morphotypes of *P. tricornutum*: **a.** A fusiform cell consisting of a central body with two thinner arms at both ends. **b.** Two oval cells. **c.** A triradiate cell consisting of a central body and three arms. pl: chloroplast. Scale bars: 3 µm. Image adapted from De Martino et al., 2007.

morphotypes and its facultative silicon biogenesis of silicified frustules provide opportunities to explore the molecular mechanisms of cell shape control and silica-based pattern and structure formation in diatoms.

Different *P. tricornutum* accessions isolated from several locations worldwide have been characterized and described, allowing the examination of natural variation of cellular responses or gene expression (De Martino et al., 2007). The strain used during this project is a fusiform strain called Pt1 clone 8.6 and represents the accession that was used for genome sequencing (De Martino et al., 2007). The *P. tricornutum* genome is approximately 27,4 megabases (Mb) in size, which is slightly smaller than *T. pseudonana* (32.4 Mb), and the number of predicted *P. tricornutum* genes is about 10000 (as opposed to about 12000 genes in *T. pseudonana* (Armbrust et al., 2004; Bowler et al., 2008). In addition, gene identification and functional analysis in *P. tricornutum* is being facilitated by the availability of more than 130000 expressed sequence tags (ESTs) generated from cells grown under 16 different conditions (Bowler et al., 2008; Maheswari et al., 2009; Maheswari et al., 2010).

REFERENCES

Allen AE, Laroche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C (2008) Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. Proc Natl Acad Sci U S A **105:** 10438-10443

Allen JT, Brown L, Sanders R, Moore CM, Mustard A, Fielding S, Lucas M, Rixen M, Savidge G, Henson S, Mayor D (2005) Diatom carbon export enhanced by silicate upwelling in the northeast Atlantic. Nature **437:** 728-732

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WW, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science **306:** 79-86

Azimzadeh J, Bornens M (2007) Structure and duplication of the centrosome. Journal of Cell Science **120:** 2139-2142

Bates SS (2000) Domoic-acid-producing diatoms: another genus added! Journal of Phycology **36:** 978-983

Bisova K, Krylov DM, Umen JG (2005) Genome-wide annotation and expression profiling of cell cycle regulatory genes in Chlamydomonas reinhardtii. Plant Physiol **137:** 475-491

Bopp L, Monfray P, Aumont O, Dufresne JL, Le Treut H, Madec G, Terray L, Orr JC (2001) Potential impact of climate change on marine export production. Global Biogeochemical Cycles **15:** 81-99

Borowitzka MA, Volcani BE (1978) The Polymorphic Diatom Phaeodactylum tricornutum: Ultrastructure of Its Morphotypes. Journal of Phycology **14:** 10-21

Bothwell JH, Marie D, Peters AF, Cock JM, Coelho SM (2010) Role of endoreduplication and apomeiosis during parthenogenetic reproduction in the model brown alga Ectocarpus. The New phytologist **188:** 111-121

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kroger N, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin-Jezequel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Oudot-Le Secq MP, Napoli C, Obornik M, Parker MS, Petit JL, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature **456:** 239-244

Bowler C, De Martino A, Falciatore A (2010) Diatom cell division in an environmental context. Current Opinion in Plant Biology **13:** 623-630

Boyle JA, Pickett-Heaps JD, Czarnecki DB (1984) Valve Morphogenesis in the Pennate Diatom Achnanthes-Coarctata. Journal of Phycology **20:** 563-573

Bozarth A, Maier UG, Zauner S (2009) Diatoms in biotechnology: modern tools and applications. Appl Microbiol Biotechnol **82:** 195-201

Brzezinski MA, Olson RJ, Chisholm SW (1990) Silicon availability and cell-cycle progression in marine diatoms. Marine Ecology-Progress Series **67:** 83-96

Buchanan BB, Gruissem W, Jones RL, eds (2000) Biochemistry and Molecular Biology of Plants. American Society of Plant Physiologists, Rockville, Maryland

Caldeira K, Wickett ME (2003) Oceanography: anthropogenic carbon and ocean pH. Nature **425:** 365

Capron A, Okresz L, Genschik P (2003) First glance at the plant APC/C, a highly conserved ubiquitin-protein ligase. Trends in Plant Science **8:** 83-89

Chepurnov VA, Mann DG, Sabbe K, Vyverman W (2004) Experimental studies on sexual reproduction in diatoms. International review of cytology **237:** 91-154

Chepurnov VA, Mann DG, Vyverman W, Sabbe K, Danielidis DB (2002) Sexual reproduction, mating system, and protoplast dynamics of Seminavis (Bacillariophyceae). Journal of Phycology **38:** 1004-1019

Chisti Y (2007) Biodiesel from microalgae. Biotechnology advances **25:** 294-306

Claudio PP, Tonini T, Giordano A (2002) The retinoblastoma family: twins or distant cousins? Genome Biol **3:** reviews3012

Coesel S, Mangogna M, Ishikawa T, Heijde M, Rogato A, Finazzi G, Todo T, Bowler C, Falciatore A (2009) Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. EMBO Rep **10:** 655-661

Coesel S, Obornik M, Varela J, Falciatore A, Bowler C (2008) Evolutionary origins and functions of the carotenoid biosynthetic pathway in marine diatoms. PLoS One **3:** e2896

Coombs J, Darley WM, Holmhans O, Volcani BE (1967) Studies on Biochemistry and Fine Structure of Silica Shell Formation in Diatoms. Chemical Composition of Navicula Pelliculosa during Silicon-Starvation Synchrony. Plant Physiology **42:** 1601-1606

Coqueret O (2002) Linking cyclins to transcriptional control. Gene **299:** 35-55

Darley WM, Volcani BE (1969) Role of Silicon in Diatom Metabolism . A Silicon Requirement for Deoxyribonucleic Acid Synthesis in Diatom Cylindrotheca-Fusiformis Reimann and Lewin. Experimental Cell Research **58:** 334-342

De Clercq A, Inzé D (2006) Cyclin-dependent kinase inhibitors in yeast, animals, and plants: a functional comparison. Crit Rev Biochem Mol Biol **41:** 293-313

de Francisco AD, Roth LE (1977) Marine Diatom, Striatella-Unipunctata .1. Cytoplasmic Fine-Structure with Emphasis on Golgi Apparatus. Cytobiologie **14:** 191-206

de Jager SM, Murray JA (1999) Retinoblastoma proteins in plants. Plant Mol Biol **41:** 295-299

De Martino A, Amato A, Bowler C (2009) Mitosis in diatoms: rediscovering an old model for cell division. Bioessays **31:** 874-884

De Martino A, Amato A, Meichenin A, Montsant A, Mathieu B, Huysman MJJ, De Veylder L, Vyverman W, Bowler C (Manuscript in preparation) Dissection of mitotic features in a pennate diatom using molecular tools: *Phaeodactylum tricornutum* as a new model for cell division.

De Martino A, Meichenin A, Shi J, Pan KH, Bowler C (2007) Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions. Journal of Phycology **43:** 992-1009

De Riso V, Raniello R, Maumus F, Rogato A, Bowler C, Falciatore A (2009) Gene silencing in the marine diatom Phaeodactylum tricornutum. Nucleic Acids Research **37:** e96

Deshaies RJ, Ferrell JE, Jr. (2001) Multisite phosphorylation and the countdown to S phase. Cell **107:** 819-822

Doonan JH, Kitsios G (2009) Functional Evolution of Cyclin-Dependent Kinases. Molecular Biotechnology **42:** 14-29

Dugdale RC, Wilkerson FP (1998) Silicate regulation of new production in the equatorial Pacific upwelling. Nature **391:** 270-273

Edgar LA, Pickett-Heaps JD (1984) Valve Morphogenesis in the Pennate Diatom Navicula-Cuspidata. Journal of Phycology **20:** 47-61

Erdner DL, Dyble J, Parsons ML, Stevens RC, Hubbard KA, Wrabel ML, Moore SK, Lefebvre KA, Anderson DM, Bienfang P, Bidigare RR, Parker MS, Moeller P, Brand LE, Trainer VL (2008) Centers for Oceans and Human Health: a unified approach to the challenge of harmful algal blooms. Environmental health : a global access science source **7 Suppl 2:** S2

Evans T, Rosenthal ET, Youngblom J, Distel D, Hunt T (1983) Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. Cell **33:** 389-396

Falciatore A, Casotti R, Leblanc C, Abrescia C, Bowler C (1999) Transformation of Nonselectable Reporter Genes in Marine Diatoms. Mar Biotechnol (NY) **1:** 239-251

Falciatore A, d'Alcala MR, Croot P, Bowler C (2000) Perception of environmental signals by a marine diatom. Science **288:** 2363-2366

Falkowski P, Scholes RJ, Boyle E, Canadell J, Canfield D, Elser J, Gruber N, Hibbard K, Hogberg P, Linder S, Mackenzie FT, Moore B, 3rd, Pedersen T, Rosenthal Y, Seitzinger S, Smetacek V, Steffen W (2000) The global carbon cycle: a test of our knowledge of earth as a system. Science **290:** 291-296

Falkowski PG, Barber RT, Smetacek VV (1998) Biogeochemical controls and feedbacks on ocean primary production. Science **281:** 200-207

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJ (2004) The evolution of modern eukaryotic phytoplankton. Science **305:** 354-360

Feely RA, Sabine CL, Hernandez-Ayon JM, Ianson D, Hales B (2008) Evidence for upwelling of corrosive "acidified" water onto the continental shelf. Science **320:** 1490-1492

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. Science **281:** 237-240

Gautier J, Norbury C, Lohka M, Nurse P, Maller J (1988) Purified maturation-promoting factor contains the product of a Xenopus homolog of the fission yeast cell cycle control gene cdc2+. Cell **54:** 433-439

Gillard J, Devos V, Huysman MJJ, De Veylder L, D'Hondt S, Martens C, Vanormelingen P, Vannerum K, Sabbe K, Chepurnov VA, Inzé D, Vuylsteke M, Vyverman W (2008) Physiological and transcriptomic evidence for a close coupling between chloroplast ontogeny and cell cycle progression in the pennate diatom Seminavis robusta. Plant Physiol **148:** 1394-1411

Gordon R, Losic D, Tiffany MA, Nagy SS, Sterrenburg FAS (2009) The Glass Menagerie: diatoms for novel applications in nanotechnology. Trends in Biotechnology **27:** 116-127

Harper JW (2001) Protein destruction: adapting roles for Cks proteins. Curr Biol **11:** R431-435

Huisman J, Sharples J, Stroom JM, Visser PM, Kardinaal WEA, Verspagen JMH, Sommeijer B (2004) Changes in turbulent mixing shift competition for light between phytoplankton species. Ecology **85:** 2960-2970

Inzé D, De Veylder L (2006) Cell cycle regulation in plant development. Annu Rev Genet **40:** 77-105

Ishikawa M, Takahashi F, Nozaki H, Nagasato C, Motomura T, Kataoka H (2009) Distribution and phylogeny of the blue light receptors aureochromes in eukaryotes. Planta **230:** 543-552

Kitsios G, Alexiou KG, Bush M, Shaw P, Doonan JH (2008) A cyclin-dependent protein kinase, CDKC2, colocalizes with and modulates the distribution of spliceosomal components in Arabidopsis. Plant J **54:** 220-235

Kooistra WH, Gersonde R, Medlin LK, Mann DG (2007) The Origin and Evolution of the Diatoms: Their Adaptation to a Planktonic Existence. *In* PG Falkowski, AH Knoll, eds, Evolution of Primary Producers in the Sea. Academic Press, pp 207-249

Kröger N (2007) Prescribing diatom morphology: toward genetic engineering of biological nanomaterials. Curr Opin Chem Biol **11:** 662-669

Kröger N, Sandhage KH (2010) From Diatom Biomolecules to Bioinspired Syntheses of Silica- and Titania-Based Materials. Mrs Bulletin **35:** 122-126

Krzeszowiec W, Rajwa B, Dobrucki J, Gabrys H (2007) Actin cytoskeleton in Arabidopsis thaliana under blue and red light. Biology of the Cell **99:** 251-260

Lauterborn R (1896) Untersuchugen über Bau, Kernteilung und Bewegung der Diatomeen. Verlag von Willelm Engelmann, Leipzig

Lavaud J, Rousseau B, Etienne AL (2004) General features of photoprotection by energy dissipation in planktonic diatoms (Bacillariophyceae). Journal of Phycology **40:** 130-137

Lavaud J, Strzepek RF, Kroth PG (2007) Photoprotection capacity differs among diatoms: Possible consequences on the spatial distribution of diatoms related to fluctuations in the underwater light climate. Limnology and Oceanography **52:** 1188-1194

Lebeau T, Robert JM (2003a) Diatom cultivation and biotechnologically relevant products. Part II: current and putative products. Appl Microbiol Biotechnol **60:** 624-632

Lebeau T, Robert JM (2003b) Diatom cultivation and biotechnologically relevant products. Part I: cultivation at various scales. Appl Microbiol Biotechnol **60:** 612-623

Leblanc C, Falciatore A, Watanabe M, Bowler C (1999) Semi-quantitative RT-PCR analysis of photoregulated gene expression in marine diatoms. Plant Mol Biol **40:** 1031-1044

Lewin JC, Lewin RA, Philpott DE (1958) Observations on Phaeodactylum tricornutum. Journal of General Microbiology **18:** 418-426

Li S, Nosenko T, Hackett JD, Bhattacharya D (2006) Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. Mol Biol Evol **23:** 663-674

Lloyd C, Chan J (2006) Not so divided: the common basis of plant and animal cell division. Nature reviews. Molecular cell biology **7:** 147-152

Lorius C, Jouzel J, Raynaud D, Hansen J, Letreut H (1990) The Ice-Core Record - Climate Sensitivity and Future Greenhouse Warming. Nature **347:** 139-145

MacIntyre HL, Kana TM, Geider RJ (2000) The effect of water motion on short-term rates of photosynthesis by marine phytoplankton. Trends in Plant Science **5:** 12-17

Maheswari U, Jabbari K, Petit JL, Porcel BM, Allen AE, Cadoret JP, De Martino A, Heijde M, Kaas R, La Roche J, Lopez PJ, Martin-Jezequel V, Meichenin A, Mock T, Schnitzler Parker M, Vardi A, Armbrust EV, Weissenbach J, Katinka M, Bowler C (2010) Digital expression profiling of novel diatom transcripts provides insight into their biological functions. Genome biology **11:** R85

Maheswari U, Mock T, Armbrust EV, Bowler C (2009) Update of the Diatom EST Database: a new tool for digital transcriptomics. Nucleic Acids Res **37:** D1001-1005

Marchetti A, Parker MS, Moccia LP, Lin EO, Arrieta AL, Ribalet F, Murphy ME, Maldonado MT, Armbrust EV (2009) Ferritin is used for iron storage in bloom-forming marine pennate diatoms. Nature **457:** 467-470

Medlin LK, Kaczmarska I (2004) Evolution of the diatoms: V. Morphological and cytological support for the major clades and a taxonomic revision. Phycologia **43:** 245-270

Mendenhall MD, Hodge AE (1998) Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast Saccharomyces cerevisiae. Microbiol Mol Biol Rev **62:** 1191-1243

Montsant A, Allen AE, Coesel S, De Martino A, Falciatore A, Mangogna M, Siaut M, Heijde M, Jabbari K, Maheswari U, Rayko E, Vardi A, Apt KE, Berges JA, Chiovitti A, Davis AK, Thamatrakoln K, Hadi MZ, Lane TW, Lippmeier JC, Martinez D, Parker MS, Pazour GJ, Saito MA, Rokhsar DS, Armbrust EV, Bowler C (2007) Identification and comparative genomic analysis of signaling and regulatory components in the diatom Thalassiosira pseudonana. J Phycol **43:** 585-604

Morgan DO (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu Rev Cell Dev Biol **13:** 261-291

Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D (2009) Genomic footprints of a cryptic plastid endosymbiosis in diatoms. Science **324:** 1724-1726

Nelson DM, Treguer P, Brzezinski MA, Leynaert A, Queguiner B (1995) Production and Dissolution of Biogenic Silica in the Ocean - Revised Global Estimates, Comparison with Regional Data and Relationship to Biogenic Sedimentation. Global Biogeochemical Cycles **9:** 359-372

Nishida K, Yagisawa F, Kuroiwa H, Nagata T, Kuroiwa T (2005) Cell cycle-regulated, microtubule-independent organelle division in Cyanidioschyzon merolae. Molecular Biology of the Cell **16:** 2493-2502

Nymark M, Valle KC, Brembu T, Hancke K, Winge P, Andresen K, Johnsen G, Bones AM (2009) An integrated analysis of molecular acclimation to high light in the marine diatom Phaeodactylum tricornutum. PLoS One **4:** e7743

Oakenfull EA, Riou-Khamlichi C, Murray JA (2002) Plant D-type cyclins and the control of G1 progression. Philosophical transactions of the Royal Society of London. Series B, Biological sciences **357:** 749-760

Ohtsubo M, Okazaki H, Nishimoto T (1989) The Rcc1 Protein, a Regulator for the Onset of Chromosome - Condensation Locates in the Nucleus and Binds to DNA. Journal of Cell Biology **109:** 1389-1397

Okita TW, Volcani BE (1980) Role of Silicon in Diatom Metabolism .10. Polypeptide Labeling Patterns during the Cell-Cycle, Silicate Starvation and Recovery in Cylindrotheca-Fusiformis. Experimental Cell Research **125:** 471-481

Olson RJ, Chisholm SW (1983) Effects of Photocycles and Periodic Ammonium Supply on 3 Marine-Phytoplankton Species .1. Cell-Division Patterns. Journal of Phycology **19:** 522-528

Olson RJ, Vaulot D, Chisholm SW (1986) Effects of environmental stresses on the cell cycle of two marine phytoplankton species. Plant Physiol **80:** 918-925

Perry JA, Kornbluth S (2007) Cdc25 and Wee1: analogous opposites? Cell Div **2:** 12

Pichevin LE, Reynolds BC, Ganeshram RS, Cacho I, Pena L, Keefe K, Ellam RM (2009) Enhanced carbon pump inferred from relaxation of nutrient limitation in the glacial ocean. Nature **459:** 1114-1117

Pickett-Heaps J (1991) Cell-Division in Diatoms. International Review of Cytology-a Survey of Cell Biology **128:** 63-108

Pickett-Heaps JD, Schmid AMM, Tippit DH (1984) Cell-Division in Diatoms. Protoplasma **120:** 132-154

Pickett-Heaps JD, Tippit DH (1978) The diatom spindle in perspective. Cell **14:** 455-467

Pines J (1996) Cell cycle: reaching for a role for the Cks proteins. Curr Biol **6:** 1399-1402

Pulz O (2001) Photobioreactors: production systems for phototrophic microorganisms. Applied Microbiology and Biotechnology **57:** 287-293

Rabosky DL, Sorhannus U (2009) Diversity dynamics of marine planktonic diatoms across the Cenozoic. Nature **457:** 183-186

Ramachandra TV, Mahapatra DM, Karthick B, Gordon R (2009) Milking Diatoms for Sustainable Energy: Biochemical Engineering versus Gasoline-Secreting Diatom Solar Panels. Industrial & Engineering Chemistry Research **48:** 8769-8788

Rayko E, Maumus F, Maheswari U, Jabbari K, Bowler C (2010) Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. The New phytologist **188:** 52-66

Robbens S, Khadaroo B, Camasses A, Derelle E, Ferraz C, Inzé D, Van de Peer Y, Moreau H (2005) Genome-wide analysis of core cell cycle genes in the unicellular green alga Ostreococcus tauri. Mol Biol Evol **22:** 589-597

Round FE, Crawford RM, Mann DG (1990) The Diatoms: Biology and Morphology of the Genera. Cambridge University Press, Cambridge

Sabine CL, Feely RA, Gruber N, Key RM, Lee K, Bullister JL, Wanninkhof R, Wong CS, Wallace DW, Tilbrook B, Millero FJ, Peng TH, Kozyr A, Ono T, Rios AF (2004) The oceanic sink for anthropogenic CO2. Science **305:** 367-371

Sapriel G, Quinet M, Heijde M, Jourdren L, Tanty V, Luo G, Le Crom S, Lopez PJ (2009) Genome-wide transcriptome analyses of silicon metabolism in Phaeodactylum tricornutum reveal the multilevel regulation of silicic acid transporters. PLoS One **4:** e7458

Sarmiento JL, Slater R, Barber R, Bopp L, Doney SC, Hirst AC, Kleypas J, Matear R, Mikolajewicz U, Monfray P, Soldatov V, Spall SA, Stouffer R (2004) Response of ocean ecosystems to climate warming. Global Biogeochemical Cycles **18:** GB3003

Siaut M, Heijde M, Mangogna M, Montsant A, Coesel S, Allen A, Manfredonia A, Falciatore A, Bowler C (2007) Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum. Gene **406:** 23-35

Sims PA, Mann DG, Medlin LK (2006) Evolution of the diatoms: insights from fossil, biological and molecular data. Phycologia **45:** 361-402

Smetacek V (1999) Diatoms and the ocean carbon cycle. Protist **150:** 25-32

Sorhannus U (2007) A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. Marine Micropaleontology **65:** 1-12

Stals H, Inzé D (2001) When plant cells decide to divide. Trends in Plant Science **6:** 359-364

Sun QY, Vrieling EG, van Santen RA, Sommerdijk NAJM (2004) Bioinspired synthesis of mesoporous silicas. Current Opinion in Solid State & Materials Science **8:** 111-120

Takahashi F, Yamagata D, Ishikawa M, Fukamatsu Y, Ogura Y, Kasahara M, Kiyosue T, Kikuyama M, Wada M, Kataoka H (2007) AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles. Proceedings of the National Academy of Sciences of the United States of America **104:** 19625-19630

Tesson B, Gaillard C, Martin-Jézéquel V (2009) Insights into the polymorphism of the diatom Phaeodactylum tricornutum Bohlin. Botanica Marina **52:** 104-116

Tippit DH, Pickett-Heaps JD (1977) Mitosis in the pennate diatom Surirella ovalis. The Journal of cell biology **73:** 705-727

Tippit DH, Pickett-Heaps JD, Leslie R (1980) Cell division in two large pennate diatoms Hantzschia and Nitzschia III. A new proposal for kinetochore function during prometaphase. The Journal of cell biology **86:** 402-416

Van den Hoek C, Mann, D.G., and Jahns, H.M. (1995) Algae: An Introduction to Phycology.

Van Mooy BA, Fredricks HF, Pedler BE, Dyhrman ST, Karl DM, Koblizek M, Lomas MW, Mincer TJ, Moore LR, Moutin T, Rappe MS, Webb EA (2009) Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. Nature **458:** 69-72

Vaulot D, Olson RJ, Chisholm SW (1986) Light and Dark Control of the Cell-Cycle in 2 Marine-Phytoplankton Species. Experimental Cell Research **167:** 38-52

Vaulot D, Olson RJ, Merkel S, Chisholm SW (1987) Cell-cycle response to nutrient starvation in 2 phytoplankton species, Thalassiosira-Weissflogii and Hymenomonas-Carterae. Marine Biology **95:** 625-630

Weinberg RA (1995) The retinoblastoma protein and cell cycle control. Cell **81:** 323-330

Williams PJL, Laurens LML (2010) Microalgae as biodiesel & biomass feedstocks: Review & analysis of the biochemistry, energetics & economics. Energy & Environmental Science **3:** 554-590

Wordeman L, McDonald KL, Cande WZ (1986) The distribution of cytoplasmic microtubules throughout the cell cycle of the centric diatom Stephanopyxis turris: their role in nuclear migration and positioning the mitotic spindle during cytokinesis. The Journal of cell biology **102:** 1688-1698

Wu J, Sunda W, Boyle EA, Karl DM (2000) Phosphate depletion in the western North Atlantic Ocean. Science **289:** 759-762

Zhu SH, Green BR (2010) Photoprotection in the diatom Thalassiosira pseudonana: role of LI818-like proteins in response to high light stress. Biochimica Et Biophysica Acta **1797:** 1449-1457

2

Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling

Marie J.J. Huysman[1,2,3], Cindy Martens[2,3], Klaas Vandepoele[2,3], Jeroen Gillard[1], Edda Rayko[4], Marc Heijde[4], Chris Bowler[4], Dirk Inzé[2,3], Yves Van de Peer[2,3], Lieven De Veylder[2,3] and Wim Vyverman[1]

[1]Protistology and Aquatic Ecology, Department of Biology, Ghent University, 9000 Gent, Belgium
[2]Department of Plant Systems Biology, Flanders Institute for Biotechnology (VIB), 9052 Gent, Belgium
[3]Department of Plant Biotechnology and Genetics, Ghent University, 9052 Gent, Belgium
[4]Département de Biologie, Ecole Normale Supérieure, CNRS, Unité Mixte de Recherche 8186, 75230 Paris Cedex 05, France

Published article: Genome Biology 2010, 11 (2)

Authors' contributions

MJJH performed the synchronization and expression experiments, analyzed the data and wrote the manuscript. CM was involved in the genome-wide annotation of cell cycle genes in *P. tricornutum* and *T. pseudonana* and helped write the manuscript. KV and ER were involved in the genome-wide annotation of diatom cell cycle genes. MJJH, CM, KV, JG, ER, MH, CB, DI, YVDP, LDV and WV helped to conceive and design the study, and read and approved the manuscript.

ABSTRACT

Background - Despite the enormous importance of diatoms in aquatic ecosystems and their broad industrial potential, little is known about their life cycle control. Diatoms typically inhabit rapidly changing and unstable environments, suggesting that cell cycle regulation in diatoms must have evolved to adequately integrate various environmental signals. The recent genome sequencing of *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* allows us to explore the molecular conservation of cell cycle regulation in diatoms.

Results - By profile-based annotation of cell cycle genes, counterparts of conserved as well as new regulators were identified in *T. pseudonana* and *P. tricornutum*. In particular, the cyclin gene family was found to be expanded extensively compared to that of other eukaryotes and a novel type of cyclins was discovered, the diatom-specific cyclins. We established a synchronization method for *P. tricornutum* that enabled assignment of the different annotated genes to specific cell cycle phase transitions. The diatom-specific cyclins are predominantly expressed at the G1-to-S transition and some respond to phosphate availability, hinting at a role in connecting cell division to environmental stimuli.

Conclusion - The discovery of highly conserved and new cell cycle regulators suggests the evolution of unique control mechanisms for diatom cell division, probably contributing to their ability to adapt and survive under highly fluctuating environmental conditions.

INTRODUCTION

Diatoms (Bacillariophyceae) are unicellular photosynthetic eukaryotes responsible for approximately 20% of the global carbon fixation (Van den Hoek et al., 1995; Field et al., 1998). They belong to the Stramenopile algae (chromists) that most probably arose from a secondary endosymbiotic process in which a red eukaryotic alga was engulfed by a heterotrophic eukaryotic host approximately 1.3 billion years ago (Kooistra et al., 2003; Li et al., 2006). This event led to an unusual combination of conserved features with novel metabolism and regulatory elements, as recently confirmed by whole-genome analysis of *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (Armbrust et al., 2004; Montsant et al., 2005; Bowler et al., 2008), which are representatives of the two major architectural diatom types, the centrics and the pennates, respectively.

Besides their huge ecological importance, diatoms are interesting from a biotechnological perspective as producers of relatively high amounts of a variety of metabolites (including oils, poly-unsaturated fatty acids (such as eicosapentaenoic acid), and pigments (such as xanthophylls and marennine) (Lebeau and Robert, 2003b; Bozarth et al., 2009), and because of their highly structured mesoporous cell wall, made of amorphous silica (Kröger, 2007). Thus, understanding the basic mechanisms controlling the diatom life cycle will be important to comprehend their ecological success in aquatic ecosystems and to control and optimize diatom growth for commercial applications.

As predominant organisms in marine and freshwater ecosystems, diatoms often encounter rapid and intense environmental fluctuations (for example, light and nutrient supply) (Round et al., 1990) that might have dramatic effects on cell physiology and viability. Therefore, cell cycle regulation in diatoms most probably involves efficient signalling of different environmental cues (Falciatore et al., 2000). Recent studies illustrate how diatoms can acclimate rapidly to iron limitation (Allen et al., 2008; Marchetti et al., 2009) and phosphorus scarcity (Van Mooy et al., 2009) through biochemical reconfiguration or maintenance of internal reservoirs and how their cell fate can be determined by perception of diatom-derived reactive aldehydes (Vardi et al., 2006; Vardi et al., 2008). Furthermore, in *P. tricornutum*, a new blue light sensor (cryptochrome/photolyase family member 1) has been discovered with dual activity as a 6-4 photolyase and a blue-light-dependent transcription regulator (Coesel et al., 2009). Thus, diatoms are expected to possess complex fine-tuned signalling networks that integrate diverse stimuli with the cell cycle. The recent availability of

genome data of *T. pseudonana* (Armbrust et al., 2004) and *P. tricornutum* (Bowler et al., 2008) now provides the basis to explore how the cell cycle machinery has evolved in diatoms.

Efficient molecular regulation of the cell cycle is crucial to ensure that structural rearrangements during cell division are coordinated and that the genetic material is replicated and distributed correctly. In eukaryotes, the mitotic cell cycle comprises successive rounds of DNA synthesis (S phase) and cell division (mitosis or M phase) separated from each other by two gap (G1 and G2) phases (Inzé and De Veylder, 2006). Passage through the different cell cycle phases is controlled at multiple checkpoints by an evolutionarily conserved set of proteins, the cyclin-dependent kinases (CDKs) and cyclins (Reviewed in Morgan, 1997; Inzé and De Veylder, 2006). Together, these proteins can form functional complexes, in which the CDKs and cyclins act as catalytic and regulatory subunits, respectively. Various types of CDKs and cyclins exist and they generally regulate the cell cycle, but some can be involved in other processes, such as transcriptional control or splicing (Coqueret, 2002; Kitsios et al., 2008).

In eukaryotes, activity of CDK-cyclin complexes is mainly controlled by (de)phosphorylation of the CDK subunits and interaction with inhibitors or scaffolding proteins (De Clercq and Inzé, 2006). Regulators include CDK-activating kinases (CAKs) (Kaldis, 1999; Umeda et al., 2005), members of the WEE1/MYT1/MIK1 kinase family and CDC25 phosphatases that carry out inhibitory phosphorylation and dephosphorylation (Perry and Kornbluth, 2007), as well as CDK inhibitors (CKIs) (De Clercq and Inzé, 2006) and the scaffolding protein CKS1/Suc1 (Pines, 1996; Harper, 2001).

The aim of this work was to reveal the molecular network of cell cycle regulators in *P. tricornutum*, a species used for decades as a model diatom for physiological studies. *P. tricornutum* is a coastal diatom, typically found in highly unstable environments, and its cells can easily acclimate to environmental changes (De Martino et al., 2007; Allen et al., 2008). Key cell cycle regulators (CDKs, CDK interactors, and cyclins) were annotated and their transcript expression profiled during synchronized growth in *P. tricornutum*. The results indicate that diatom cell division is controlled by a combination of conserved molecules found in yeast, animals and/or plants, and novel components, including diatom-specific cyclins that probably transduce the environmental status of the cells to the cell cycle machinery.

RESULTS AND DISCUSSION

Annotation of the cell cycle genes in diatoms

The following cell cycle gene families were selected for comprehensive analysis: CDKs, cyclins, CKS1/suc1, WEE1/MYT1/MIK1, CDC25, and CKIs. These gene families were annotated functionally on the basis of their homology with known cell cycle genes in other organisms (see Materials and methods). The results of this family-wise annotation are discussed below and summarized in Table 2.1 and Additional Table 2.1. The nomenclature of all identified proteins is according to that used in other protists for which cell cycle gene annotation was available (Bisova et al., 2005; Robbens et al., 2005).

Table 2.1: Overview and evolutionary conservation of the different core cell cycle gene families.

Cell cycle gene	Number of copies				
	Phatr[a]	**Arath**[a,b]	**Ostta**[a, c]	**Sacce**[a, c]	**Homsa**[a, c]
CDKA	2[d]	1	1	1	3
CDKB	-	4	1	-	-
CDKC	2	2	1	1	1
CDKD	1	3	1	-	1
CDKF	-	1	-	1	1
CYCA	1?	10	1	NA	NA
CYCB	2?	9	1	NA	NA
CYCD	1?	10	1	NA	NA
CYCH	1?	1	1	NA	NA
CDC25	-	-	1[?]	1	3
WEE1/MYT1/MIK1	1	1	2	2	2
CKS	1	2	1	1	2
CKI	-	7	1	1	8

[a]Abbreviations: Phatr, *Phaeodactylum tricornutum*; Arath, *Arabidopsis thaliana*; Ostta, *Ostreococcus tauri* Sacce, *Saccharomyces cerevisiae*; Homsa, *Homo sapiens.*

[b]Data taken from (Vandepoele et al., 2002)

[c]Data taken from (Robbens et al., 2005)

[d]One of these genes shows some *CDKB* characteristics.

[?]Classification uncertain because of weak phylogeny.

NA, not available due to other classification nomenclature.

Cell cycle synchronization and expression analysis

To validate the predicted functions of the annotated genes, we examined their transcript expression during the cell cycle. To synchronize cell division in *P. tricornutum*, we subjected exponentially growing cells to a prolonged dark period, which arrests the cells in the G1 phase (Brzezinski et al., 1990) (Figure 2.1; Additional figure 2.1), and released the cells synchronously from this arrest point by illumination. A comparable method had been applied successfully to synchronize growth in a closely related diatom, *Seminavis robusta* (Gillard et al., 2008). Microscopic observations of the dark-arrested *P. tricornutum* cultures showed that all cells contained a single undivided chloroplast (Figure 2.1a, upper panel). Accordingly, in flow cytometric histograms, the dark-arrested cells showed only a 2C peak (Figure 2.1b and Additional figure 2.1, t = 0), confirming the G1 phase identity of cells containing a single chloroplast. When cells were released from the dark arrest, the population of bi-chloroplastidic cells steadily increased and cells entered the S phase, as observed by flow cytometry (Additional figure 2.1, upper panel). However, the level of synchrony decreased at later time points (from 10 h after the dark release onward), probably because cells entered the next cell division cycle at the moment other cells still had to pass through M phase (Additional figure 2.1). To circumvent this problem and to obtain an enrichment of cells in M phase during the later time points (Additional figure 2.1), the metaphase blocker nocodazole was added at the time of re-illumination (Ng et al., 1998), but without major effect on cell cycle progression (Additional figure 2.1).

To monitor gene expression during the different cell cycle phases, exponentially growing cells were synchronized in the presence of nocodazole (Figure 2.1b-c). Automated analysis of the flow histograms indicated that G1-phase cells were dominant during the first 4 h of re-illumination; from 4 to 7 h, cells went through S phase, as seen by the broadening and lowering of the 2C peak, while cells went mainly through the G2 and M phases at 8 to 12 h (Figure 2.1b-c). In *S. robusta*, chloroplast division had been found to take place only after S-phase onset (Gillard et al., 2008). Chloroplast division in *P. tricornutum* was observed starting from 5 h after illumination, confirming the S-phase timing determined by flow cytometry (Figure 2.1a lower panel, and 2.1c). The duration of the cell cycle after the synchronization procedure was comparable with that of cultures grown under standard conditions (approximately one division per day; Additional figure 2.2). For downstream analysis, at hourly intervals after illumination, samples were taken for expression analysis by real-time quantitative polymerase chain reaction

(qPCR).

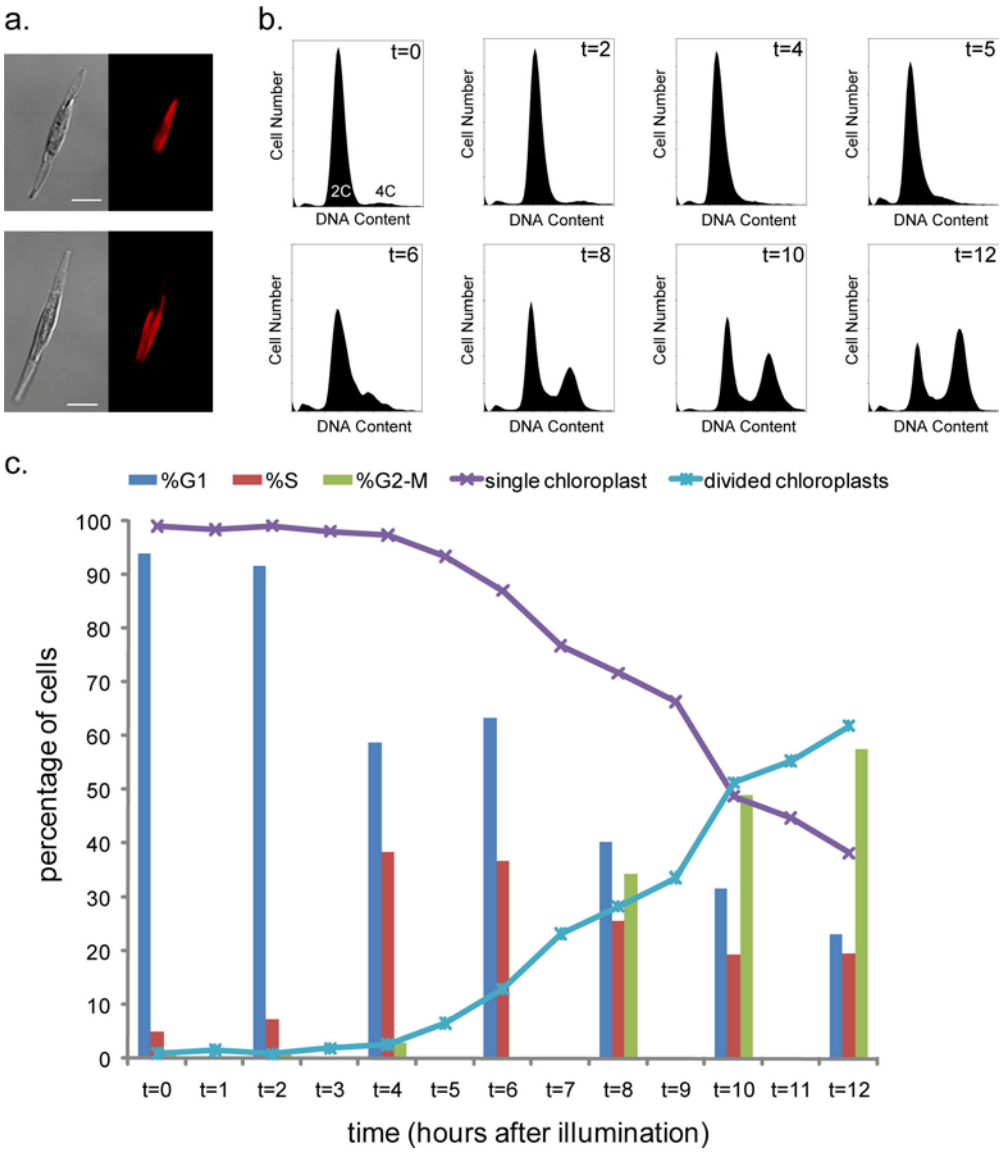


Figure 2.1: Synchronization of the cell cycle in *P. tricornutum*. **a.** Confocal images of a dark-arrested cell (upper panel) showing a single parietal chloroplast and a cell after 12 h illumination (lower panel) showing divided and translocated daughter chloroplasts. Red, autofluorescence of the chloroplast. Scale bar: 5 µm. **b.** Validation of synchronization of the cell cycle of *P. tricornutum* by flow cytometry. DNA content (abscissa) is plotted against cell number (ordinate). After a 20-h dark period, most of the cells are blocked in G1 phase (t = 0 to 4 h), indicated by the single 2C peak. After reillumination, cells proceed synchronously with their cell cycle, going through S-phase (between t = 4 and 7 h), visible as the broadening and lowering of the 2C peak, and G2-M phase (t = 8 to 12 h), indicated by the accumulation of 4C cells. **c.** Histogram indicating the proportion of cells in a certain cell cycle phase and chloroplast conformation during the cell cycle. Divided chloroplasts were observed starting from 5 h after illumination, after S-phase onset.

CDKs and CDK interactors

CDKs

CDKs are serine/threonine kinases that play a central role in cell cycle regulation and other processes, such as transcriptional control. Yeast uses only one single PSTAIRE-containing CDK for cell cycle progression (Mendenhall and Hodge, 1998; Moser and Russell, 2000), while higher organisms encode different CDKs implicated in cell division. The most conserved CDKs contain a PSTAIRE motif in their cyclin-binding domain (Morgan, 1997; Inzé and De Veylder, 2006). In plants, the PSTAIRE-containing CDK had been designated CDKA and is active during both G1-to-S and G2-to-M transitions (Inzé and De Veylder, 2006). The plant-specific B-type CDKs contain a P[P/S]T[A/T]LRE motif and are active during the G2 and M phases (Inzé and De Veylder, 2006). In animals, three PSTAIRE (Cdk1, Cdk2, and Cdk3) and two P(I/L)ST(V/I)RE (Cdk4 and Cdk6) CDKs are involved in cell cycle control, although evidence has been found recently that only Cdk1 is really required to drive cell division (Morgan, 1997; Santamaria et al., 2007).

Five CDKs could be identified in *P. tricornutum* (Table 2.1), of which two clustered together with the CDKA (plant)/CDK1-2 (animal) family in the phylogenetic tree (Figure 2.2). CDKA1 contains the typical PSTAIRE cyclin-binding motif (Figure 2.3) and its mRNA levels were high during late G1 and S phase (Figure 2.4a), suggesting a role at the G1-to-S transition. CDKA2 shows a PSTALRE motif (Figure 2.3), which is a midway motif between the CDKA hallmark PSTAIRE and the plant-specific CDKB hallmark P[P/S]T[A/T]LRE. The mRNA levels of CDKA2 were elevated in G2/M cells (Figure 2.4a). No homologs of the metazoan CDK4/6 family were found in *P. tricornutum*.

CDKC, CDKD and CDKE (designated Cdk9, Cdk7 and Cdk8 in animals, respectively) are kinases related to CDKA (Joubes et al., 2000). C-type CDKs (CDKC and Cdk9) and Cdk8 have been shown to associate with transcription initiation complexes and, thus, to play a role in transcriptional control (Oelgeschlager, 2002; Barroco et al., 2003). Additionally, CDKC2 is active in spliceosomal dynamics in plants (Kitsios et al., 2008) and CDKE controls floral cell differentiation (Wang and Chen, 2004). We identified two C-type CDKs (Table 2.1), CDKC1 and CDKC2 (Figure 2.2a) with PITALRE and PLQFIRE cyclin-binding motifs, respectively (Figure 2.3). No CDKE homolog was found in *P. tricornutum*. Both *CDKC* genes had relatively low mRNA levels throughout the cell cycle without any discernible cell cycle phase pattern (data not shown). Thus, like in other eukaryotes, *CDKC* expression probably does not depend

on the cell cycle phase in *P. tricornutum*, but it might be involved in other processes, such as transcription or splicing. One CDKD was identified (*CDKD1*) in *P. tricornutum* (Table 2.1 and Figure 2.2a). D-type CDKs are known to interact with H-type cyclins to form a CAK complex (Umeda et al., 2005). We found that *CDKD1* mRNA levels were high at the G1-to-S phase transition (Figure 2.4a). Another CDK variant, CDKF, has only been found in plants, where it functions as a CAK-activating kinase (CAKAK) (Umeda et al., 2005). No members of the CDKF family were identified in *P. tricornutum*, confirming that the CAKAK pathway is specific to plants and should have evolved within the green lineage (Table 2.1).

In addition, we identified seven hypothetical CDKs (hCDKs; Additional Table 2.1) with divergent cyclin-binding domains (Figure 2.3) that could not be integrated into the phylogenetic tree due to high sequence divergence. The expression levels of several of these hCDKs were modulated during the cell cycle (Figure 2.4a). The *hCDK1* mRNA levels were the highest during G2-M, whereas those of *hCDK6* were up-regulated during G1 phase and *hCDK2*, *hCDK3*, *hCDK4*, and *hCDK5* were predominantly expressed at G1 and/or S phase. For *hCDK7*, no reproducible expression pattern was found (data not shown).

CDK subunit

CDK subunit (CKS) proteins act as docking factors that mediate the interaction of CDKs with putative substrates and regulatory proteins (Pines, 1996). In *P. tricornutum*, one *CKS* gene was found (*CKS1*; Table 2.1) of which the mRNA levels were mainly high in G2/M cells (Figure 2.4b).

WEE1/MYT1/MIK1 kinases

WEE1/MYT1/MIK1 kinases inhibit cell cycle progression through phosphorylation of CDKs (Perry and Kornbluth, 2007). In yeast and animals, MYT1 is a membrane-associated kinase that phosphorylates Thr14 of Cdc2 proteins, as well as Tyr15, which is also a target of WEE1, a nucleus-localized kinase (Mueller et al., 1995; Liu et al., 1997). A single CDK inhibitory kinase could be identified in *P. tricornutum*, belonging to the MYT1 family (Table 2.1; Additional figure 2.3) (Mueller et al., 1995). In *Arabidopsis thaliana*, the inhibitory kinase corresponds to WEE1 (De Schutter et al., 2007), while the green alga *Ostreococcus tauri* expresses both *WEE1* (Robbens et al., 2005) and *MYT1* (unpublished data), like animals do (Mueller et al., 1995) (Table 2.1). Expression of the *P. tricornutum MYT1* kinase was not associated with a specific cell cycle phase (data not shown). Because *MYT1* is probably

implicated in stress responses during the cell cycle (Zhou and Elledge, 2000), it is possible that the imposed dark arrest or addition of nocodazole influenced the mRNA levels of *MYT1*, with too much variability in its expression profile as a consequence.

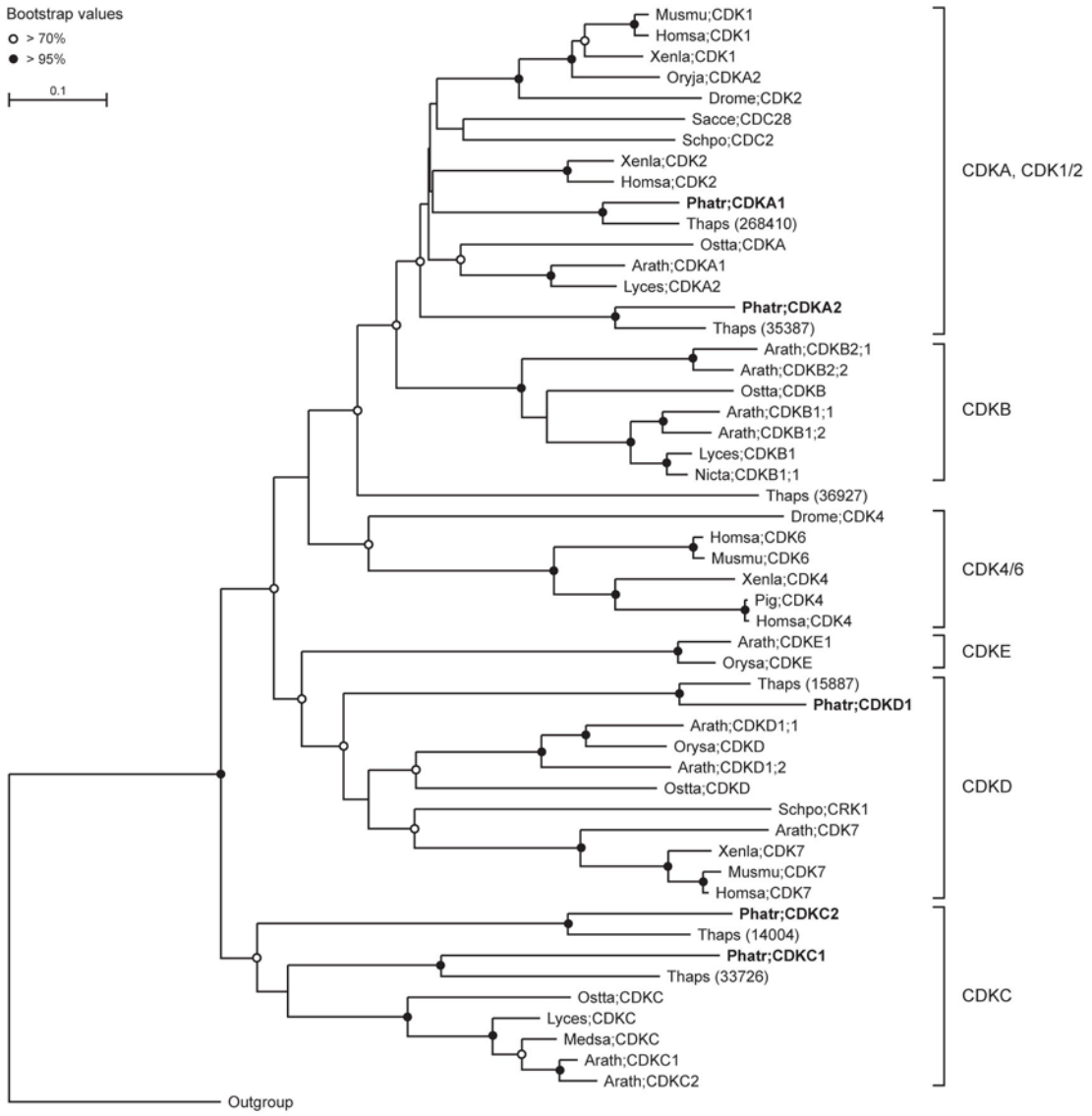


Figure 2.2: Phylogenetic analysis of the cyclin-dependent kinases of *P. tricornutum*. Neighbor-joining tree (TREECON, Poisson correction, 1000 replicates) of the CDK family. The *P. tricornutum* sequences are shown in bold. Abbreviations: Arath, *Arabidopsis thaliana*; Drome, *Drosophila melanogaster*; Homsa, *Homo sapiens*; Lyces, *Lycopersicon esculentum*; Medsa, *Medicago sativa*; Musmu, *Mus musculus*; Nicta, *Nicotiana tabacum*; Oryja, *Oryza japonica*; Orysa, *Oryza sativa*; Ostta, *Ostreococcus tauri*; Phatr, *Phaeodactylum tricornutum*; Sacce, *Saccharomyces cerevisiae*; Schpo, *Schizosaccharomyces pombe*; Thaps, *Thalassiosira pseudonana*; and Xenla, *Xenopus laevis*. The outgroup is represented by human CDK10.

```
CDKC1      WGMPLQFIREIKI
hCDK1      -IDAKRILREIKL
hCDK7      -VDAVRLYREIHI
hCDK4      ---KVVPMRELQS
CDKC2      -GFPITALREVKI
CDKD       -GVNFTAVREIKL
CDKA2      -GIPSTALREISL
CDKA1      -GIPSTAIREISL
hCDK3      -GVPCNVIREISL
hCDK5      -GFPVTALREINV
hCDK2      -GFPVTTLREIQS
hCDK6      -KVLQNLEIEISI
                      *
```

Figure 2.3: Cyclin-dependent kinase cyclin-binding motifs. Alignment of the cyclin-binding motifs of all annotated CDKs in *P. tricornutum*. The motifs are indicated in the green box. Conserved residues are marked by an asterix in the bottom line.

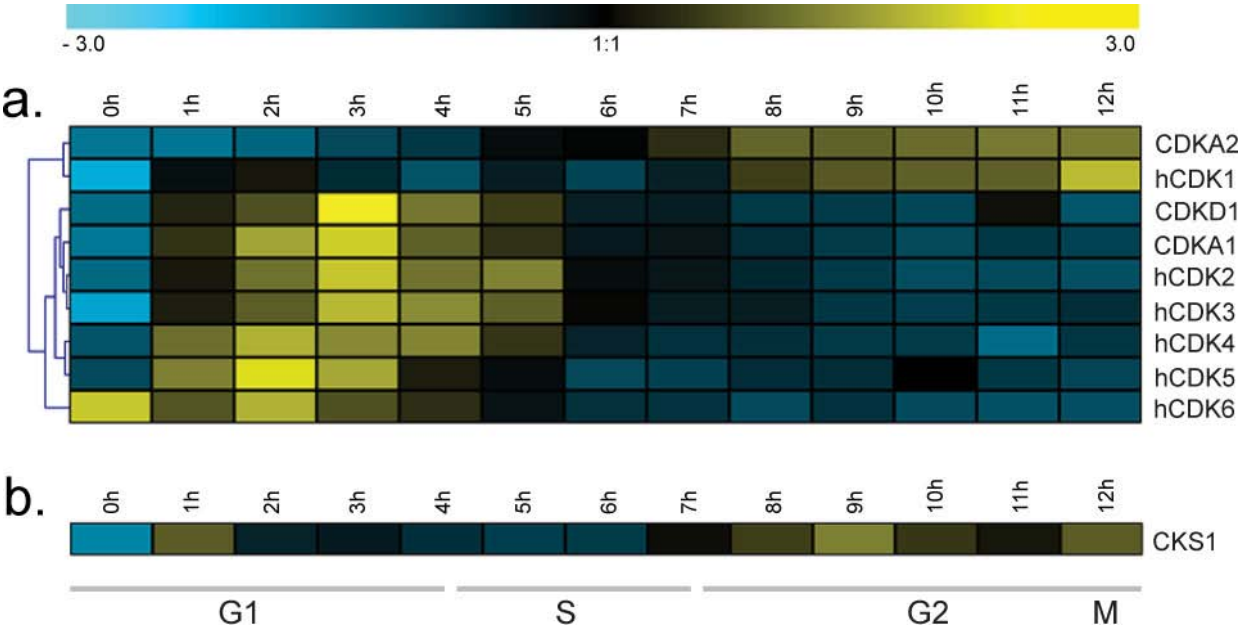


Figure 2.4: Hierarchical average linkage clustering of the expression profiles of the differentially expressed cyclin-dependent kinases and the scaffolding protein CKS in *P. tricornutum*. **a.** Members of the CDK family. **b.** CKS1. h, hypothetical. Yellow: Relative upregulation. Blue: Relative downregulation. Approximate timing of the different cell cycle phases is indicated at the bottom. Values are the mean of three independent experiments normalized against the values of the consitutively expressed *histone H4* gene. No reproducible expression patterns could be found for CDKC1, CDKC2, hCDK7 and MYT1 (data not shown).

CDC25 phosphatase

As antagonists of the WEE1/MYT1/MIK1 kinases, CDC25 phosphatases activate CDKs (Perry and Kornbluth, 2007). In contrast to the presence of a counteracting kinase, no CDC25 phosphatase could be identified in *P. tricornutum* (Table 2.1) or in *T. pseudonana*. Both *Arabidopsis* and *Oryza sativa* also lack a functional CDC25 (Bleeker et al., 2006; Duan et al., 2007) and, in plants, CDC25-mediated regulatory mechanisms have been proposed to be replaced by a mechanism governed by the plant-specific B-type CDKs (Boudolf et al., 2006). In *P. tricornutum*, no true B-type CDK homolog could be found, but CDKA2, classified by weak homology as A-type CDK class, possessed a PSTALRE cyclin-binding motif (Figure 2.3), which is halfway between the CDKA and CDKB hallmarks. This motif also occurred in the *Dictyostelium discoideum* CDC2 homolog (Michaelis and Weeks, 1992) and in the *O. tauri* CDKB protein (Robbens et al., 2005). The PSTALRE motif is present as well in the CDKA2 homolog of *T. pseudonana* (Thaps3_35387; Figure 2.2a), confirming that this subtype could generally be found in diatoms. Moreover, *CDKA2* was expressed during G2-M (Figure 2.4a), the expected time of action of a B-type CDK. Although further in-depth biochemical research will be required to determine its true physiological function, the presence of this A/B-type CDK might explain the absence of a CDC25 phosphatase in diatoms. Alternatively, if the sequence of the CDC25 phosphatase had diverged to such an extent in diatoms, it might be not detectable by sequence homology, as already suggested for higher plants as well (Khadaroo et al., 2004).

CDK inhibitors

CDK-cyclin complexes can be inactivated by CKIs, including the members of the INK4 family and the Cip/Kip family in animals (Sherr and Roberts, 1999), or Kip-related proteins and SIAMESE proteins in plants (Verkest et al., 2005; Churchman et al., 2006). CKIs are mainly low-molecular-weight proteins that inhibit CDK activity by tight association in response to developmental or environmental stimuli (Sherr and Roberts, 1995, 1999; De Clercq and Inzé, 2006). Despite extensive sequence similarity searches for CKIs, no homologs could be identified in *P. tricornutum*, which is not so surprising given the high sequence diversity of this cell cycle family (Verkest et al., 2005). These inhibitory proteins are most probably present in *P. tricornutum*, but their identification will require more advanced molecular techniques.

Cyclins

The cyclin gene family is expanded in diatoms

We found a large number of highly diverged cyclin genes in diatoms, of which 24 proteins in *P. tricornutum* (Additional Table 2.1). Due to their high divergence, indicated by the low bootstrap values in the phylogenetic tree, the classification into different subclasses was not clear (Figure 2.5), as it was for the 52 putative cyclins identified in *T. pseudonana* (Montsant et al., 2007). Moreover, many represent a novel class of cyclins, which we designated diatom-specific cyclins (dsCYCs).

To investigate whether the expansion of the cyclin gene family is specific to diatoms, we compared cyclin abundance among a representative set of Chromalveolates (Stramenopiles, Apicomplexa, and Ciliates; Table 2.2) for which genome data are available (Carlton et al., 2002; Gardner et al., 2002; Abrahamsen et al., 2004; Xu et al., 2004; Gardner et al., 2005; Pain et al., 2005; Aury et al., 2006; Eisen et al., 2006; Tyler et al., 2006) and have been pre-processed in a previous study (Martens et al., 2008). Because of the lack of cell cycle gene annotation in all investigated species, we first screened for cyclin genes, which allowed us to create a reference dataset for analyzing cyclin evolution. We searched the different genomes for proteins that showed similarity to our cyclin HMMER profile and determined the number of proteins that contained an InterPro cyclin domain (Table 2.2). Generally, both detection methods yielded comparable results within all species (Table 2.2). An indication of the putative subclasses and function of the detected proteins is given by specific cyclin InterPro domains (Table 2.2). The proportion of the detected cyclin proteins relative to the predicted total gene number of each species revealed that, in the diatom genomes, cyclins are overrepresented compared to all investigated species, except for both *Cryptosporidium* species (Abrahamsen et al., 2004; Xu et al., 2004) and *Paramecium tetraurelia* (Aury et al., 2006) (Table 2.2). However, the total number of cyclins found in *Cryptosporidium* (12) is low compared to that in diatoms (28 in *P. tricornutum* and 57 in *T. pseudonana*). *Cryptosporidium* species are protozoan pathogens that depend on their hosts for nutrients. Moreover, Gene Ontology distribution for *Cryptosporidium* and *Plasmodium* is similar, indicating that no functional specialization of conserved gene families has occurred (Xu et al., 2004). In *Paramecium tetrauleria*, the cyclin family is expanded as well. However, this species has a complex genome structure, possessing silent diploid micronuclei and polyploid macronuclei. Furthermore, *P. tetraurelia* underwent at least

three whole-genome duplications, resulting in an apparent expansion of almost every gene family (Aury et al., 2006).

In conclusion, the large number of cyclin genes in both diatoms does not seem to be shared with its closest related species, indicating that diatom cyclins could have evolved separately to acquire new specific functions. Although the cyclin family has been found to be expanded in both diatoms, the size of the cyclin gene family in *T. pseudonana* is larger than that in *P. tricornutum*, which seems to result mainly from the presence of a larger number of diatom-specific cyclins in *T. pseudonana* (Figure 2.5). The biological cause of the changes in the cyclin family size remains unknown, although natural selection due to differential habitats might have played a role, or alternatively, random gene loss or gain might have occurred over long time stretches, as both species diverged at least 90 million years ago (Bowler et al., 2008). Genome sequence data of other diatom species are currently being generated (for example, for *Fragilariopsis cyclindrus* and *Pseudo-nitzschia multiseries*) and will help to shed light on cyclin gene family evolution in diatoms.

Conserved cyclins

Cyclins can be functionally classified into two major groups: the cell cycle regulators and the transcription regulators. Generally, during the cell cycle, specific cyclins are associated with G1 phase (cyclin D), S phase (cyclins A and E), and mitosis (cyclins A and B) (Sanchez and Dynlacht, 2005). In *P. tricornutum*, we identified a single A/B-type cyclin gene (*CYCA/B;1*; Figure 2.5), which gradually accumulated its mRNA transcript during the G2 and M phases (Figure 2.6a). Both B-type cyclin genes (encoded by *CYCB1* and *CYCB2*) (Figure 2.5) were predominately expressed in G2/M cells, but mRNA levels of *CYCB2* accumulated earlier than those of *CYCB1* (Figure 2.6a). The single D-type cyclin (encoded by *CYCD1*; Figure 2.2b) was mainly expressed during S and G2/M phase progression (Figure 2.6a). As in plants, CYCE seems to be absent in diatoms (Vandepoele et al., 2002).

Cyclins with a regulatory role during transcription include those belonging to the classes C, H, K, L, and T (Oelgeschlager, 2002). However, some cyclins involved in transcriptional control might also have a function in cell cycle regulation. For example, besides being a transcriptional regulator, the human C-type cyclin is also involved in the control of cell cycle transitions (Liu et al., 1998) and H-type cyclins can regulate the cell cycle through interaction with D-type CDKs, thereby forming a CAK complex (Fisher and Morgan, 1994; Yamaguchi et al., 2000; Umeda et al., 2005). The latter is probably also true

for the *P. tricornutum CYCH1* (Figure 2.5) because it was coexpressed with *CDKD1* during the cell cycle (Figure 2.6a). The single L-type cyclin (encoded by *CYCL1*; Figure 2.5) showed elevated mRNA levels at G1 and during S phase (Figure 2.6a). In animals, cyclin L (also called Ania-6) has previously been demonstrated to be an immediate early gene that could be involved in cell cycle re-entry (Iyer et al., 1999; Berke et al., 2001).

Six cyclins in *P. tricornutum* clustered together with P-type cyclins (PHO80-like proteins, also called U-type cyclins; Additional Table 2.1 and Figure 2.5) that are believed to play a role in phosphate signalling (Kaffman et al., 1994; Torres Acosta et al., 2004). The mRNA levels of all P-type cyclin genes (*CYCP1*, *CYCP2*, *CYCP3*, *CYCP4*, *CYCP5*, and *CYCP6*) were high early during the time series (Figure 2.6a). One cyclin gene did not cluster with any of the represented classes and was annotated as CYC-like (Figure 2.5). The mRNA levels of this gene peaked during the G1 and S phases (Figure 2.6a).

Most diatom-specific cyclins are expressed early during the cell cycle

Eleven cyclin genes were identified that clustered only with cyclins of *T. pseudonana* (Figure 2.5). Therefore, we assigned these as dsCYC genes. *dsCYC3* and *dsCYC4* showed both high expression at the G2/M phases (Figure 2.6b). The mRNA levels of *dsCYC10* were slightly up-regulated at the G1-to-S transition and reached a peak late during the cell cycle (Figure 2.6b). As the other *dsCYC* genes displayed increased mRNA levels during the G1 and/or S phases (*dsCYC1*, *dsCYC2*, *dsCYC5*, *dsCYC6*, *dsCYC7*, *dsCYC8*, *dsCYC9*, and *dsCYC11*; Figure 2.6b), some might function as immediate early genes controlled by light or mitogens.

Organisms living in aquatic environments, particularly in coastal regions, often have to cope with rapid and broad fluctuations in light intensity, temperature, nutrient availability, oxygen level, and salinity, all of which can have profound consequences on cell cycle progression. Comparative genome analyses of marine phytoplankton have revealed that coastal organisms contain genetic imprints indicative of adaptation to life under variable conditions (Palenik et al., 2006; Peers and Niyogi, 2008), including distinct proteins coding for photosynthesis and light harvesting, additional two-component regulatory systems, novel carbon-concentrating mechanisms, transcription of transporters and assimilation proteins for the uptake of alternative nitrogen sources, and numerous metal transporter families and metal enzymes (Palenik et al., 2006; Peers and Niyogi, 2008). Similar adaptation imprints were also found in the diatom genomes (Armbrust et al., 2004; Bowler et al., 2008). Nevertheless, because diatoms generally dominate the microplankton in temperate waters and coastal

upwelling regions under favorable conditions (Irigoien et al., 2002), we expect diatoms to possess additional sophisticated fine-tuning systems enabling them to adjust the pace of the cell division rate in tune with the prevailing conditions.

Although in plants numerous copies of D-type cyclins integrate both external and internal signals into the cell cycle (Inzé and De Veylder, 2006), in *P. tricornutum* only one D-type cyclin (*CYCD1*) was identified that was highly expressed late during the cell cycle (Figure 2.6a). Therefore, in diatoms CYCD probably does not play its classical role of G1-phase signal integrator, but might have acquired an alternative function in the G2-to-M transition as previously proposed for some D-type cyclins in plants (Schnittger et al., 2002). On the other hand, the wide variety of *dsCYC* genes in diatoms expressed early during the cell cycle renders them plausible candidates to fulfil the task of signal integrators. Moreover, diatom-specific genes have been found to evolve faster than other genes in diatom genomes (Bowler et al., 2008), indicating that these cyclin genes might have acquired novel and/or species-specific functions. Interestingly, other gene families expanded in diatoms include histidine kinases and heat shock factors, which are supposed to be involved in environmental sensing and expressed under certain growth conditions (Bowler et al., 2008). Thus, gene family expansion in diatoms could possibly be linked to the development of specific signal responses and adaptations to the environment.

Figure 2.5 (Next page): Phylogenetic analysis of the cyclins of *P. tricornutum*. Neighbor-joining tree (TREECON, Poisson correction, 500 replicates) of the cyclin family. The *P. tricornutum* sequences are shown in bold. Abbreviations: Arath, *Arabidopsis thaliana*; Homsa, *Homo sapiens*; Ostta, *Ostreococcus tauri*; Phatr, *Phaeodactylum tricornutum*; and Thaps, *Thalassiosira pseudonana*.

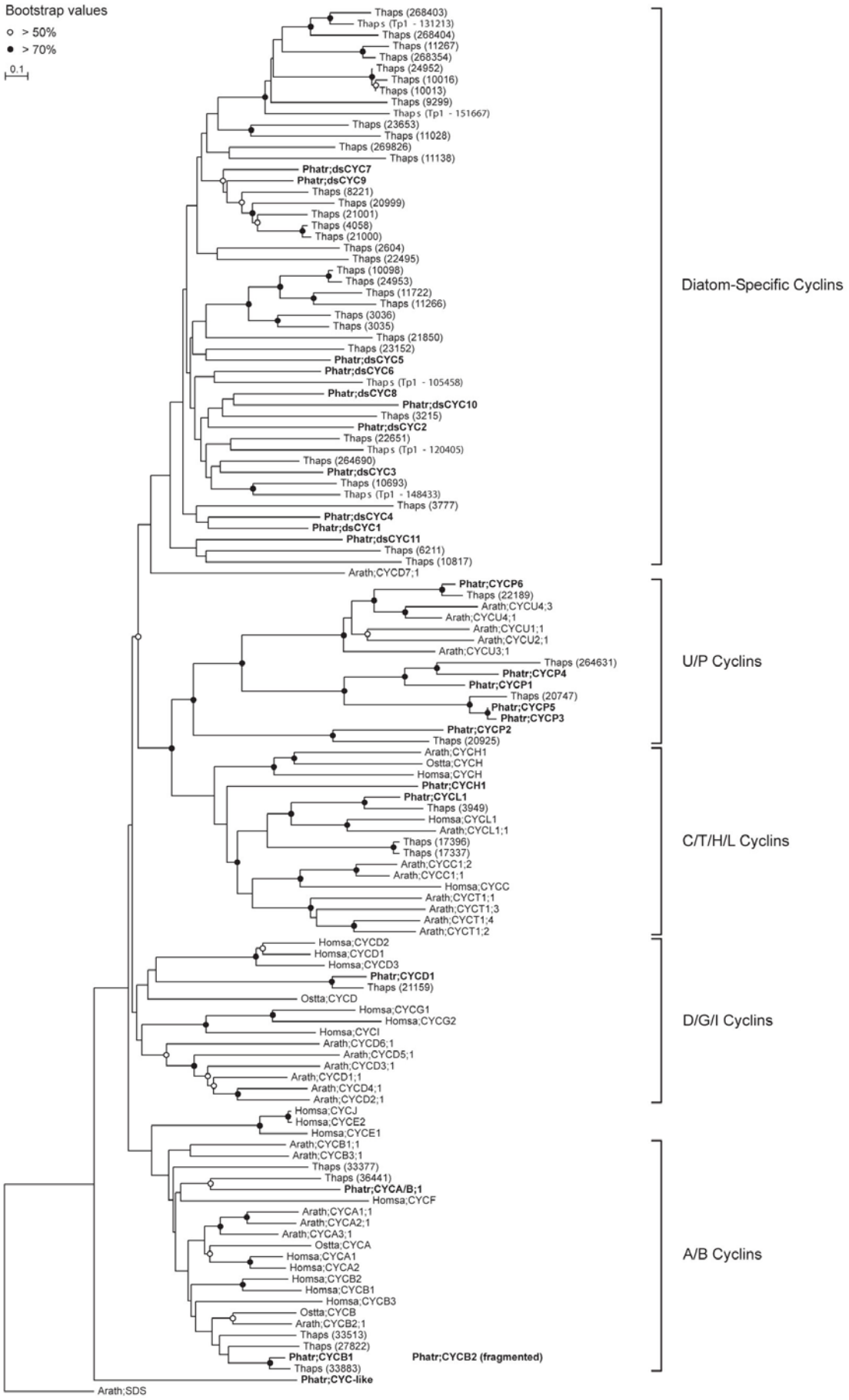

Bootstrap values
> 50%
> 70%
0.1
Thaps (268403)
Thaps (Tp1 - 131213)
Thaps (268404)
Thaps (11267)
Thaps (268354)
Thaps (24952)
Thaps (10016)
Thaps (10013)
Thaps (9299)
Thaps (Tp1 - 151667)
Thaps (23653)
Thaps (11028)
Thaps (269826)
Thaps (11138)
Phatr;dsCYC7
Phatr;dsCYC9
Thaps (8221)
Thaps (20999)
Thaps (21001)
Thaps (4058)
Thaps (21000)
Thaps (2604)
Thaps (22495)
Thaps (10098)
Thaps (24953)
Thaps (11722)
Thaps (11266)
Thaps (3036)
Thaps (3035)
Thaps (21850)
Thaps (23152)
Phatr;dsCYC5
Phatr;dsCYC6
Thaps (Tp1 - 105458)
Phatr;dsCYC8
Phatr;dsCYC10
Thaps (3215)
Phatr;dsCYC2
Thaps (22651)
Thaps (Tp1 - 120405)
Thaps (264690)
Phatr;dsCYC3
Thaps (10693)
Thaps (Tp1 - 148433)
Thaps (3777)
Phatr;dsCYC4
Phatr;dsCYC1
Phatr;dsCYC11
Thaps (6211)
Thaps (10817)
Arath;CYCD7;1
Diatom-Specific Cyclins
Phatr;CYCP6
Thaps (22189)
Arath;CYCU4;3
Arath;CYCU4;1
Arath;CYCU1;1
Arath;CYCU2;1
Arath;CYCU3;1
Thaps (264631)
Phatr;CYCP4
Phatr;CYCP1
Thaps (20747)
Phatr;CYCP5
Phatr;CYCP3
Phatr;CYCP2
Thaps (20925)
U/P Cyclins
Arath;CYCH1
Ostta;CYCH
Homsa;CYCH
Phatr;CYCH1
Phatr;CYCL1
Thaps (3949)
Homsa;CYCL1
Arath;CYCL1;1
Thaps (17396)
Thaps (17337)
Arath;CYCC1;2
Arath;CYCC1;1
Homsa;CYCC
Arath;CYCT1;1
Arath;CYCT1;3
Arath;CYCT1;4
Arath;CYCT1;2
C/T/H/L Cyclins
Homsa;CYCD2
Homsa;CYCD1
Homsa;CYCD3
Phatr;CYCD1
Thaps (21159)
Ostta;CYCD
Homsa;CYCG1
Homsa;CYCG2
Homsa;CYCI
Arath;CYCD6;1
Arath;CYCD5;1
Arath;CYCD3;1
Arath;CYCD1;1
Arath;CYCD4;1
Arath;CYCD2;1
D/G/I Cyclins
Homsa;CYCJ
Homsa;CYCE2
Homsa;CYCE1
Arath;CYCB1;1
Arath;CYCB3;1
Thaps (33377)
Thaps (36441)
Phatr;CYCA/B;1
Homsa;CYCF
Arath;CYCA1;1
Arath;CYCA2;1
Arath;CYCA3;1
Ostta;CYCA
Homsa;CYCA1
Homsa;CYCA2
Homsa;CYCB2
Homsa;CYCB1
Homsa;CYCB3
Ostta;CYCB
Arath;CYCB2;1
Thaps (33513)
Thaps (27822)
Phatr;CYCB1
Phatr;CYCB2 (fragmented)
Thaps (33883)
Phatr;CYC-like
A/B Cyclins
Arath;SDS

Table 2.2: Expansions of the cyclin gene family in different representatives of the Chromalveolata.

	Stramenopiles				Apicomplexa						Ciliates	
	Diatoms		Oomycetes									
	Phatr	Thaps	Phyra	Physo	Cryho	Crypa	Plafa	Playo	Thean	Thepa	Parte	Tetth
General												
# proteins matching the cyclin HMM-profile	28	57	19	19	12	12	5	5	8	8	144	29
# proteins with an InterPro cyclin domain	27	55	18	18	12	12	5	5	4	6	140	27
Specific InterPro domains												
IPR004367 Cyclin, C-terminal	7	9	4	5	-	-	-	-	-	-	19	6
IPR006670 Cyclin	6	18	7	7	2	2	2	1	2	1	94	20
IPR006671 Cyclin, N-terminal	18	45	9	10	1	1	-	-	-	-	96	20
IPR011028 Cyclin-like	27	55	18	18	11	11	5	5	4	6	140	27
IPR013763 Cyclin-related	21	47	13	13	6	6	3	2	2	2	72	22
IPR013922 Cyclin-related 2	1	1	1	1	3	3	1	1	1	1	21	1
IPR014400 Cyclin, A/B/D/E	2	4	2	3	1	1	-	-	-	-	42	3
IPR015429 Transcription regulator cyclin	4	3	6	5	2	2	2	2	2	2	6	3
IPR015432 Cyclin H	1	-	1	1	1	1	1	1	1	1	-	-
IPR015451 Cyclin D	-	4	-	-	1	-	-	-	-	-	-	-
IPR015452 G2/mitotic-specific cyclin B3	-	-	-	-	-	-	-	-	-	-	3	1
IPR015453 G2/mitotic-specific cyclin A	1	1	3	3	-	-	-	-	-	-	2	-
IPR015454 G2/mitotic-specific cyclin B	1	-	-	-	-	-	-	-	-	-	9	1
IPR017060 Cyclin L	-	-	1	1	-	-	1	1	-	-	2	-
Total number of genes	10402	11776	15743	19027	3994	3952	5268	5268	3792	4035	39642	27000
Genome size (Mbp)	27.4	32.4	65.00	95.00	9.16	9.11	22.85	23.1	8.35	8.3		104
Cyclins/genes total (%) [1]	0.27	0.48	0.12	0.10	0.30	0.30	0.09	0.09	0.21	0.20	0.36	0.11
Cyclins/genes total (%) [2]	0.26	0.47	0.11	0.09	0.30	0.30	0.09	0.09	0.11	0.15	0.35	0.10

Abbreviations: Phatr, *Phaeodactylum tricornutum*; Thaps, *Thalassiosira pseudonana*; Phyra, *Phytophthora ramorum*; Physo, *Phytophthora sojae*; Cryho, *Cryptosporidium hominis*; Crypa, *Cryptosporidium parvum*; Plafa, *Plasmodium falciparum*; Playo, *Plasmodium yoelii yoelii*; Thean, *Theileria annulata*; Thepa, *Theileria parva*; Parte, *Paramecium tetraurelia*, Tetth, *Tetrahymena thermophila*.

[1]Number of cyclins versus total number of genes calculated with the number of proteins that match our cyclin HMMer profile.

[2]Number of cyclins versus total number of genes calculated with the number of proteins with a InterPro cyclin domain.

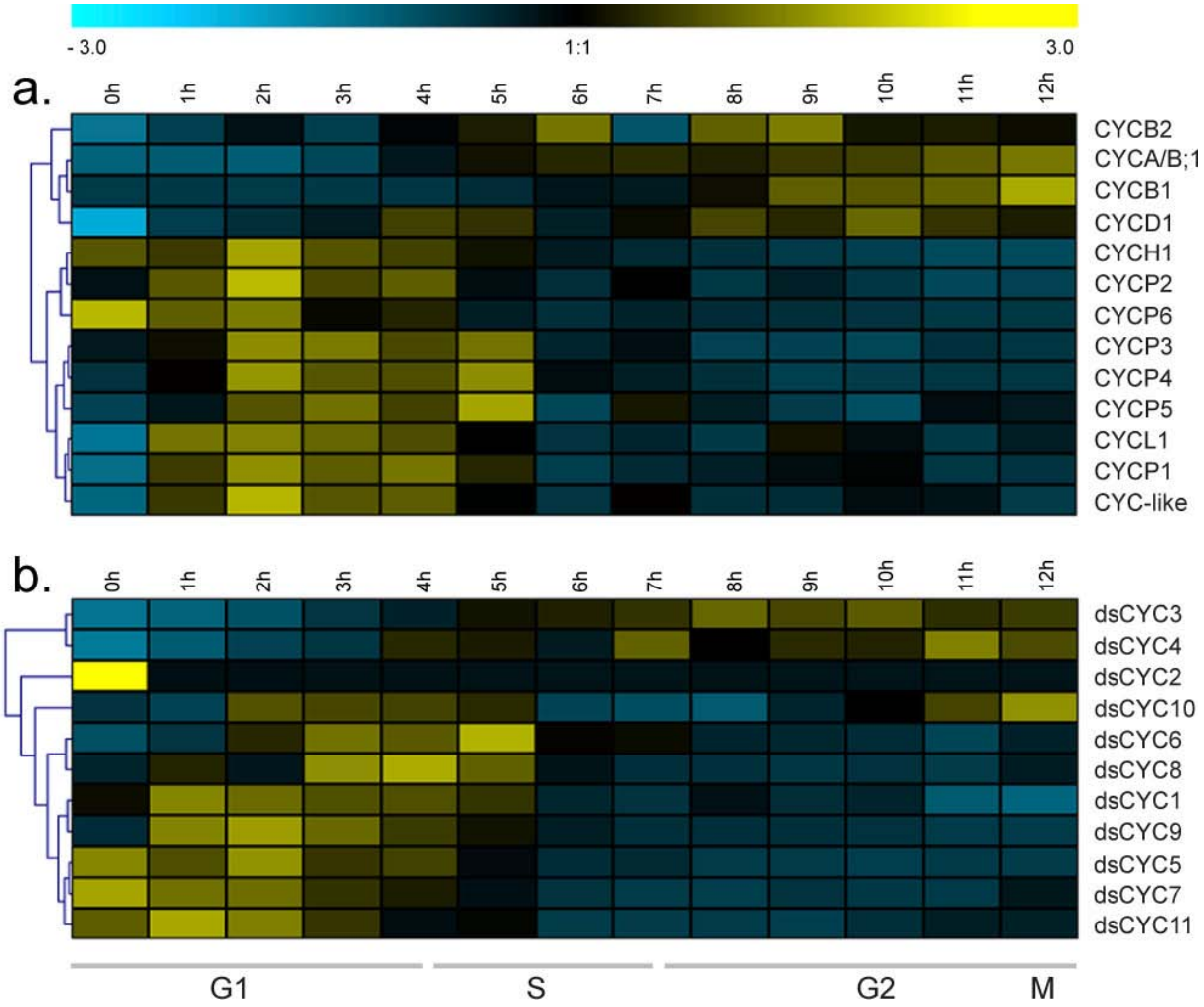


Figure 2.6: Hierarchical average linkage clustering of the expression profiles of cyclin genes in *P. tricornutum*. **a.** Conserved cyclins. **b.** Diatom-specific cyclins (dsCYCs). ds, diatom specific. Yellow: Relative upregulation. Blue: Relative downregulation. Approximate timing of the different cell cycle phases is indicated at the bottom. Values are the mean of three independent experiments normalized against the values of the consitutively expressed *histone H4* gene.

dsCYCs respond to nutrient availability

To investigate the role of the *dsCYC* genes during the cell cycle, we analyzed them in more detail. More specifically, we examined whether their transcription is affected by nutrient deprivation. Analysis of recently published expressed sequence tag data (Maheswari et al., 2009) illustrates the differential expression of *dsCYC3*, *dsCYC7*, and *dsCYC10* across a range of environmental conditions (for example, nitrate-starved, nitrate-repleted, and iron-limited

cultures). Moreover, a microarray analysis revealed that *dsCYC9* transcript levels were higher in cultures grown in the presence of silica than those without silica (Sapriel et al., 2009).

To examine whether *dsCYC* expression could be responsive to nutrient status during the cell cycle, we monitored mRNA levels in parallel with cell growth during nutrient starvation-repletion experiments. Exponentially growing cultures were nutrient-starved for 24 h and re-supplied with only nitrate, phosphate, iron, trace metals, the combination of all nutrients (positive control), or no nutrients (negative control). Three hours after nutrient supply, samples were collected for expression analysis. After nitrate repletion, cells reinitiated cell division at almost comparable levels to the positive control cultures, whereas repletion with phosphate, iron, or trace elements did not differ from the negative control (Figure 2.7a), indicating that nitrate is the most cell cycle rate-limiting nutrient in *P. tricornutum*. Nitrogen starvation in diatoms generally leads to a G1-phase arrest (Olson et al., 1986; Vaulot et al., 1987). Thus, increased mRNA levels of early cell cycle-regulated genes are to be expected at the time of cell cycle reinitiation after nitrate repletion. Accordingly, cell cycle genes that showed early (G1 phase) transcription in the synchronized series (*CYCP6*, *CYCH1*, and *hCDK5*) (Figure 2.4 and 2.6) were induced in the nitrate replete and positive control cultures (Figure 2.7b). To exclude cell cycle effects during sampling, the starvation experiment was repeated for nitrate repletion, but after imposing a 24-h dark arrest after starvation and re-supply of nitrate in complete darkness. In these cultures, the expression of the early cell cycle genes did not differ from that of the negative control after nitrate supply (Figure 2.7c), confirming that expression of *CYCP6*, *CYCH1*, and *hCDK5* is linked to cell cycle re-entry rather than to the nitrate status of the cells.

In contrast to nitrate, cells resupplied with phosphate remained arrested (Figure 2.7a-b). Upon addition of phosphate, mRNA levels of *dsCYC5*, *dsCYC7* and *dsCYC10* were significantly higher than those of the negative control (Figure 2.7d), strongly suggesting that these genes might function as direct cell cycle signal integrators upon increase of phosphate levels. Upon replenishment with nitrate (in the dark), iron or trace elements, no effects on *dsCYC* gene expression were observed (Figure 2.7d and data not shown).

Nitrogen, together with the micronutrient iron, is generally considered to be a major limiting factor of primary production in the oceans (Falkowski et al., 1998). Phosphate limitation, on the other hand, is considered to be less common, although it has been reported in certain oceanic areas (Wu et al., 2000) and has been hypothized recently to have been more wide-spread during the glacial periods (Pichevin et al., 2009). As an important constituent of adenosine triphosphate, nucleic acids, and phospholipids, phosphorus is an important molecule

not only for growth, but for almost all metabolic activities. Recently, diatoms have been shown to reduce their phosphorus demand upon phosphorus limitation, and to maintain growth by substituting phospholipids with non-phosphorus membrane lipids, only when nitrogen is not limiting (Van Mooy et al., 2009).

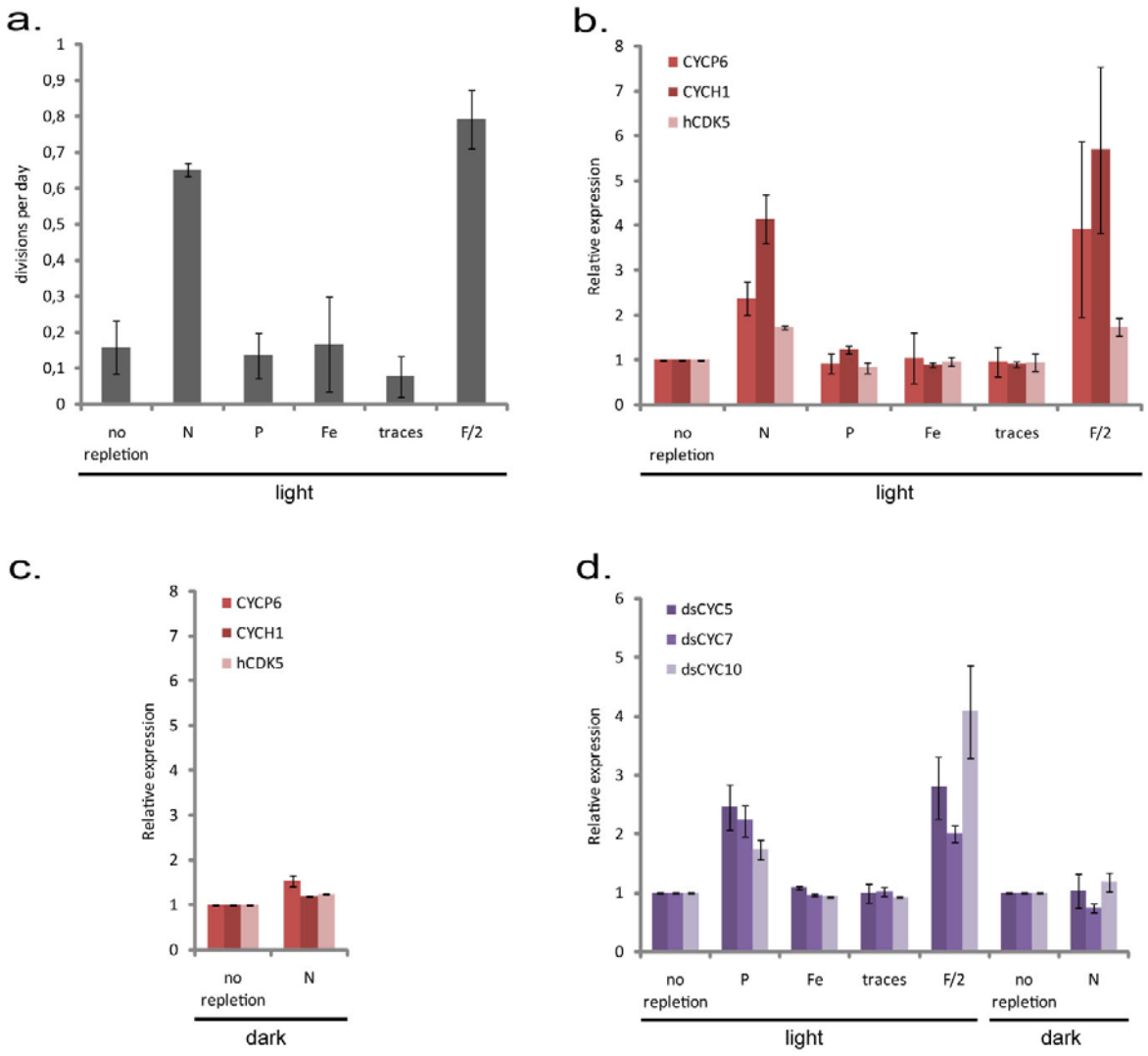


Figure 2.7: Nutrient response of diatom-specific cyclins. **a.** Growth rate of different subcultures after repletion based on average cell density measurements at the time of and 3 days after repletion. These data indicate the ability of the cells to recover from starvation. **b.** Expression profiles of early cell cycle genes *CYCP6* (peak expression at t = 0 in the synchronization series (Figure 2.6)), *CYCH1* and *hCDK5* (both have their peak expression at t = 2 in the synchronization series (Figure 2.4 and 2.6)) at the time of sampling during the light experiment. **c.** Expression profiles of early cell cycle genes at the time of sampling during the dark experiment. **d.** dsCYCs responding to phosphate addition. Error bars represent standard errors of the mean of two biological replicates.

In summary, these results reveal that some dsCYCs might be involved in environmental cell cycle control, functioning as nutrient signal integrators. All phosphate-responding *dsCYC* genes were expressed early during the synchronized time series (Figure 2.6b), fitting with a function in linking nutritional status and cell division start.

Cell cycle biomarkers

The identification of the complete set of major cell cycle regulators in *P. tricornutum*, along with the determination of their temporal expression patterns, generates a basis for studying different cell cycle-related processes in diatoms. Diatom cell cycle biomarkers could be used to observe cell cycle effects in laboratory experiments, but they could also be highly valuable to monitor diatom life cycle events in the natural habitat, like bloom or rest periods.

To validate whether the expression data obtained through the synchronization experiment was applicable in cell cycle-associated studies, we selected diatom cell cycle genes with a defined expression pattern to test their value as cell cycle biomarkers. As a control experiment, we checked the expression of four early (*CYCH1*, *hCDK5*, *CDKA1*, and *CDKD1*) and two late (*CDKA2* and *CYCB1*) cell cycle genes during a 12-h light/12-h dark photoperiod (LD 12:12). Flow cytometry data during this 24-h time course of the grown cultures indicate that the cells show a low degree of 'natural' synchronization of cell division: in the morning, most cells are in the G1 phase, while in the evening, division takes place (Figure 2.8a). Thus, it was to be expected that genes determined as early and as late cell cycle genes would be induced in the morning and in the evening, respectively. Indeed, expression according to the different cell cycle distributions was found for all selected genes (Figure 2.8b-c), indicating that they would perform as good cell cycle markers in cell cycle-related studies and that the expression data obtained from the synchronization studies (Figures 2.4 and 2.6) could serve as a reliable basis to select appropriate marker genes.

In a real case study, we used these cell cycle biomarkers to investigate whether the cell cycle in *P. tricornutum* would be regulated by an endogenous clock or a so-called circadian oscillator. Circadian regulation of cell division is well known to occur in eukaryotes and is particularly well-described for unicellular algae (Goto and Johnson, 1995; Moulager et al., 2007). Although circadian regulation of light-harvesting protein-encoding genes and pigment synthesis has been reported in diatoms (Oeltjen et al., 2004; Ragni and D'Alcala, 2007), we

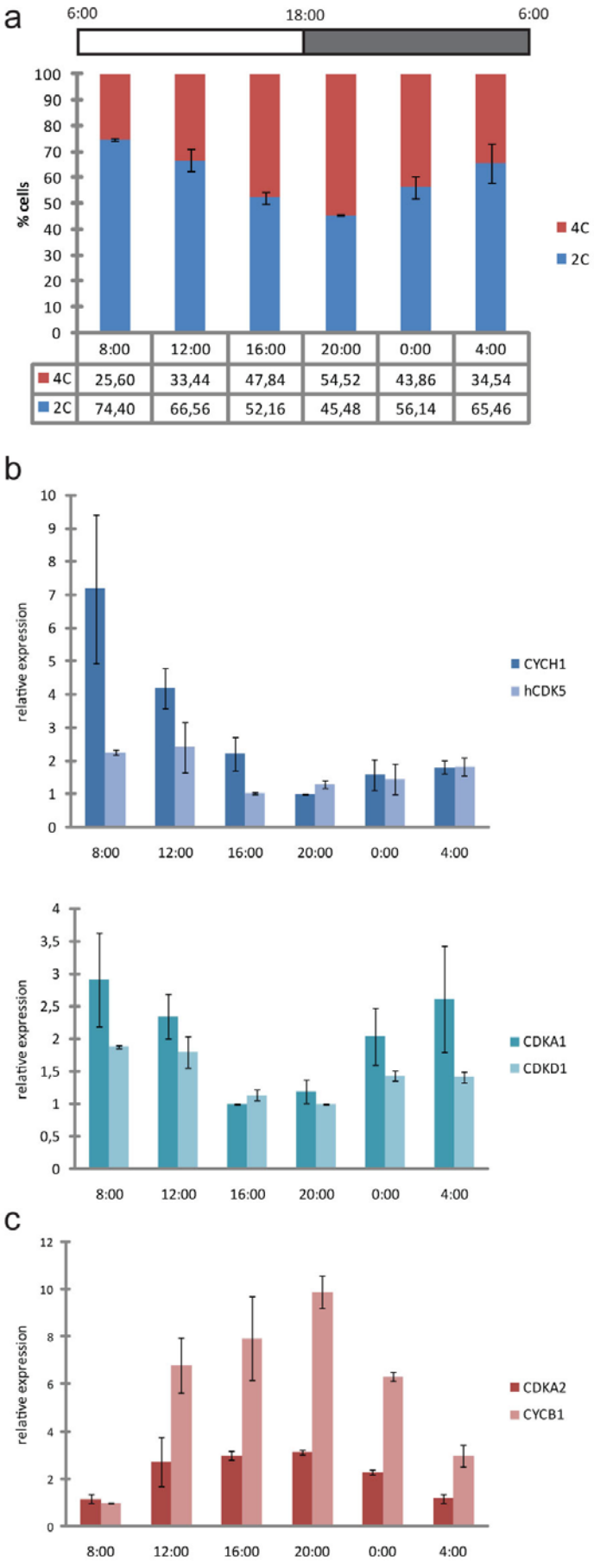


Figure 2.8: Validation of cell cycle marker genes. **a.** DNA distributions (2C versus 4C) of exponentially growing cells entrained by a LD 12:12 photoperiod during the time series **b.** Expression profiles of early cell cycle genes (*CYCH1* and *hCDK5*; peak expression at t = 2 in the synchronization series (Figure 2.4 and 2.6)); and *CDKA1* and *CDKD1* (peak expression at t = 3 in the synchronization series (Figure 2.4)). **c.** Expression profiles of late cell cycle genes *CDKA2* and *CYCB1* (peak expression at t = 12 in the synchronization series (Figure 2.4 and 2.6)). Error bars represent standard errors of the mean of two biological replicates.

did not find any direct evidence that circadian regulation of the cell cycle exists in *P. tricornutum*. Comparison of cell cycle progression and cell cycle biomarker expression in cells under normal 12-h light/12-h dark (LD 12:12) or free-running 12-h light/12-h light (LL 12:12, constant light) light conditions indicate that neither the cell cycle itself nor mRNA accumulation of the main core cell cycle genes depends on a circadian oscillator (Additional figures 2.4 and 2.5). These findings stress even more the importance of the development and use of efficient signaling networks that link environmental cues to cell growth in diatoms.

CONCLUSIONS

From the annotation and expression analyses, we conclude that the diatom cell cycle machinery shares common features with cell cycle regulatory systems present in other eukaryotes, including a PSTAIRE-containing CDK, conserved cyclin classes of types A, B, and D, and a MYT1 kinase. In addition, members of the retinoblastoma pathway for G1-S regulation involving the retinoblastoma protein and E2F/DP transcription factors (Weinberg, 1995; de Jager and Murray, 1999; Claudio et al., 2002) were also found in *P. tricornutum* (unpublished data, see Chapter 4). Components that were expected to be found in diatoms but could not be identified include a CDC25 phosphatase and CKIs. Possibly the function of the CDC25 phosphatase might be taken over by *CDKA2*, given its expression time and sequence similarity with B-type CDKs (Boudolf et al., 2006), whereas the lack of CKI identification by sequence similarity searches might be due to high sequence divergence (Verkest et al., 2005).

Most interestingly, we found a major expansion of the cyclin gene family in diatoms and discovered a new cyclin class, the diatom-specific cyclins. The latter are most probably involved in signal integration to the cell cycle because transcript levels of *dsCYC5*, *dsCYC7*, and *dsCYC10* depended on phosphate (this study), and *dsCYC9* was reported to be induced upon silica availability (Sapriel et al., 2009). Besides their role in nutrient sensing, we hypothesize that transcription of some *dsCYC* genes might also be light-modulated, as illustrated by the high *dsCYC2* mRNA levels in dark-acclimated cells that drastically dropped after 1 h of light exposure (Figure 2.6b). In addition, this gene was recently found to be modulated upon blue light treatment (Coesel et al., 2009). The responsiveness of other *dsCYC* genes to different light conditions is currently under investigation.

The complete set of major diatom key cell cycle regulators identified in this study could serve as a set of marker genes for monitoring diatom growth both in the laboratory and in the field. As cell cycle-regulated transcription cannot be assumed to depict a cell cycle-regulatory role for a gene, the predicted functions of the individual diatom cell cycle genes await further experimental confirmation by molecular and biochemical studies, although they already provide first insights into the manner in which diatoms control their cell division. Therefore, this dataset will form a starting point for future experiments aimed at exploring and manipulating the diatom cell cycle.

MATERIALS AND METHODS

Culture conditions

P. tricornutum (Pt1 8.6; accession numbers CCAP 1055/1 and CCMP2561) (De Martino et al., 2007) was grown in F/2 medium without silica (F/2-Si) (Guillard, 1975), made with filtered and autoclaved sea water collected from the North Sea (Belgium). Cultures were cultivated at 18 to 20°C in a 12-h light/12-h dark regime (50 to 100 μmol·photons·m^{-2}·s^{-1}) and shaken at 100 rpm. Under these conditions, the average generation time of *P. tricornutum* was calculated to be 0.93 ± 0.07 days (Additional figure 2.2).

Family-wise annotation of the diatom cell cycle genes

In a first step, known plant and animal cell cycle genes were selected to construct a reference cell cycle dataset. The members of every cell cycle family were used to build family-specific HMMER profiles (Eddy, 1998). With these profiles, the predicted *P. tricornutum* and *T. pseudonana* proteomes were screened for the presence of core cell cycle families. Missing gene families were also screened against the raw genome sequence (using tBLASTN) to account for annotation errors (that is, missing genes). For each family, the putative *P. tricornutum* homologs found were validated by comparing them with the reference family members in a multiple alignment.

Phylogenetic analysis

Multiple alignments generated with MUSCLE (Edgar, 2004) were manually improved with BioEdit (Hall, 1999). To define subclasses within the gene families, phylogenetic trees were built that included the reference cell cycle genes from plants and animals. Both TREECON (Van de Peer and De Wachter, 1994) and PHYLIP (http://evolution.genetics.washington.edu/phylip/) were used to construct the neighbor-joining trees based on Poisson-corrected distances. To test the significance of the nodes, bootstrap analysis was applied using 1,000 replicates for all trees, except for the cyclin tree (500 replicates).

Synchronization of the cell cycle in *P. tricornutum*

P. tricornutum cells were arrested in the G1 phase by prolonged darkness (20 h). After release of the cells from this G1 checkpoint by reillumination, samples for cell cycle analysis and real-time qPCR were collected during 12 h at hourly intervals, starting at reillumination (t = 0). To prevent cells from entering a second cell cycle, nocodazole (2.5 mg/l; Sigma-Aldrich, St. Louis, Missouri, USA) was added to the cultures at t = 0. Synchronization was validated by flow cytometric analysis on a Partec CyFlow ML platform (with data acquisition software Flomax; Partec GmbH, Münster, Germany) on cells fixed with 70% ethanol, washed three times with 1× phosphate buffered saline and stained with 4',6-diamidino-2-phenylindole (final concentration of 1 ng/ml). For each sample, 10,000 cells were processed. Flow cytograms were analyzed with Multicycle AV for Windows (Phoenix Flow Systems, San Diego, California, USA) software to determine relative representations of the different cell cycle stages in the samples.

Nutrient starvation/repletion experiment

Exponentially growing cells (under constant light, 50 µmol photons m^{-2} s^{-1}) were collected by centrifugation 3 days after medium replenishment, and washed twice with natural seawater (North Sea, Belgium) to starve the cells. After 24 h starvation, the culture was subdivided into six subcultures and supplied with only nitrate (8.82×10^{-4} M $NaNO_3$; N), phosphate (3.62×10^{-5} M NaH_2PO_4 H_2O; P), iron (1.17×10^{-5} M $FeCl_3$ $6H_2O$; Fe), trace metals (3.93×10^{-8} M $CuSO_4$ $5H_2O$, 2.60×10^{-8} M Na_2MoO_4 $2H_2O$, 7.65×10^{-8} M $ZnSO_4$ $7H_2O$, 4.20×10^{-8} $CoCl_2$ $6H_2O$ and 9.10×10^{-7} M $MnCl_2$ $4H_2O$; trace), the combination of all nutrients (concentrations as mentioned above; F/2), or no nutrients (no repletion). Samples were taken for

real-time qPCR after 3 hours of incubation. Cell density and growth rate were monitored during 3 days after repletion using a Bürker counting chamber to assess the degree of starvation in the different subcultures. For each sample, the average cell density was counted from nine large squares (0.1 mm^3) and growth rate was calculated from semi-log linear regression of the cell numbers plotted against time.

To exclude cell cycle effects upon nitrate repletion, the experiment was repeated with cells grown in a LD 12:12 photoperiod. Three days after medium replenishment, the cells were washed twice with natural seawater (North Sea, Belgium) to starve the cells and illuminated for 12 h. The cells were then incubated in the dark for 24 h and no nutrients and nitrate were supplied in the dark as mentioned above. Samples were taken for real-time qPCR after 3 hours of incubation in the dark.

Real-time qPCR

For RNA extraction, 5×10^7 cells were collected at each time point, fast frozen in liquid nitrogen, and stored at -70°C. To lyse the cells and extract RNA, TriReagent (Molecular Research Center, Inc., Cincinnati, Ohio, USA) was used initially. After addition of chloroform, RNA was purified from the aqueous phase by RNeasy purification, according to the manufacturer's instructions (RNeasy MinElute Cleanup kit; Qiagen, Hilden, Germany). Contaminating genomic DNA was removed by DNaseI (GE Healthcare, Little Chalfont, United Kingdom) treatment. RNA concentration and purity were assessed by spectrophotometry (NanoDrop ND-1000, Wilmington, Delaware, USA). Total RNA was reverse transcribed with Superscript II reverse transcriptase (Invitrogen, Carlsbad, California, USA) in a total volume of 40 µl with oligo(dT) primers. Finally, 1.25 ng (synchronization experiment and control experiment) or 10 ng (nutrient starvation/repletion experiment and circadian experiment) of cDNA was used as template for each qPCR reaction.

Samples in triplicate were amplified on the Lightcycler 480 platform with the Lightcycler 480 SYBR Green I Master mix (Roche Diagnostics, Brussels, Belgium) in the presence of 0.5 µM gene-specific primers (Additional Table 2.1). The cycling conditions were 10 minutes polymerase activation at 95°C and 45 cycles at 95°C for 10 s, 58°C for 15 s, and 72°C for 15 s. Amplicon dissociation curves were recorded after cycle 45 by heating from 65°C to 95°C. In qBase (Hellemans et al., 2007), data were analyzed using the ΔC_t relative quantification method with the stably expressed histone H4 as a normalization gene (Additional figure 2.6) (Siaut et al., 2007). Expression profiles of the synchronized cell cycle series were

mean relative expression from three independent sample series. After normalization, the mean profiles were clustered using hierarchical average linkage clustering (analysis software TIGR MultiExperiment Viewer 3D (TMEV3D)).

Image acquisition

Confocal images were obtained with a scanning confocal microscope 100 M (Zeiss, Jena, Germany) equipped with the software package LSM510 version 3.2 (Zeiss, Jena, Germany) and a C-Apochromat 63× (1.2 NA) water-corrected objective. Chlorophyll autofluorescence was excited with HeNe illumination (543 nm).

Accession numbers

Sequence data from this article can be accessed through the Joint Genome Institute (JGI) portal (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html) . Accession numbers of the cell cycle genes are listed in Additional Table 2.1.

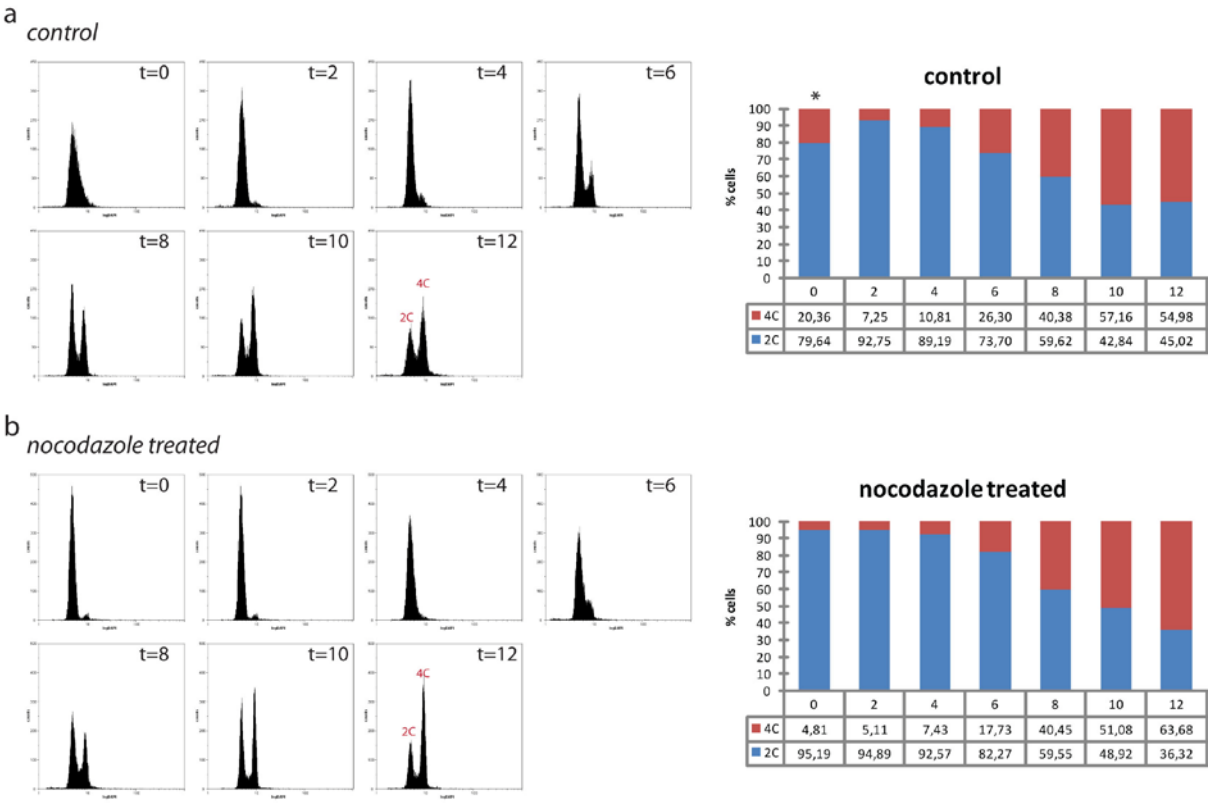
Acknowledgements

The authors thank Magali Siaut who performed some of the initial annotation studies of cell cycle genes in *P. tricornutum*, the colleagues of the cell cycle group (Ghent) for helpful comments and discussions, and Dr Martine De Cock for assistance in preparing the manuscript. This work was partly supported by grants from the Research Fund of Ghent University (‘Geconcerteerde onderzoeksacties’ no. 12050398) and the European Union Framework Program 6 (EU-FP6) Diatomics project (LSHG-CT-2004-512035). MJJH, CM, and JG are indebted to the Agency for Innovation by Science and Technology in Flanders (IWT) for a predoctoral fellowship. KV and LDV are Postdoctoral Fellows of the Research Foundation-Flanders. CB acknowledges funding from the EU-FP6 Marine Genomics Network of Excellence (GOCE-CT-2004-505403), the Centre National de la Recherche Scientifique (CNRS), and the Agence Nationale de la Recherche (France).

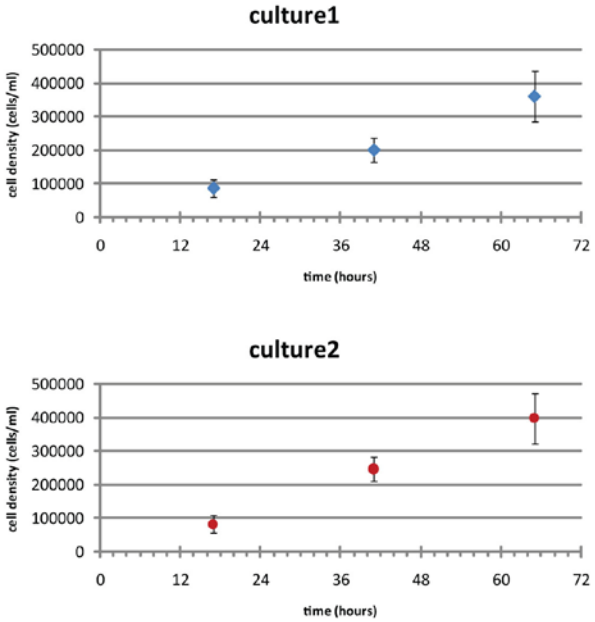
ADDITIONAL DATA

Gene Name	Protein Name	Protein ID (JGI - release 1)	Protein ID (JGI - release 2)	Forward gene specific primer	Reverse gene specific primer	Product length (bp)
Cyclin-dependent kinases						
Phatr;*CDKA1*	CDKA1	Phatr1_28779	Phatr2_20262	AGCGGTATCAAAGGATGGAAAAG	CTTCATCTTCGGCTTCAAGGC	123
Phatr;*CDKA2*	CDKA2	Phatr1_24534	Phatr2_51279	GTTGCCCTCAAGCGTATC	AACACAGTCCTTCAAGTCG	123
Phatr;*CDKC1*	CDKC1	Phatr1_12190	Phatr2_10704	CTGAAGAAACTCACGCATCCG	CGAGAACGAGAAAGAGATTCCC	118
Phatr;*CDKC2*	CDKC2	Phatr1_16981	Phatr2_26925	AACACGAAACAGGAGGAGAACG	ACGCACCTTTGGATGTCACG	121
Phatr;*CDKD1*	CDKD	Phatr1_9317	Phatr2_10160	ATTACTTCTGCGGAGACCATTC	GCGGTAAAGATTTCGTCAAAGG	183
Phatr;*hCDK1*	hCDK1	Phatr1_25351	Phatr2_21410	GTATCTTTGCTGAGTTAATTCTTC	TACTGGAGTCGTGCTTGG	173
Phatr;*hCDK2*	hCDK2	Phatr1_33856	Phatr2_42539	AACTCGCCGATTTTGGTTTGTC	ATTGCGTATATTGCCTGCTTCC	129
Phatr;*hCDK3*	hCDK3	Phatr1_35608	Phatr2_34165	TCTCAAGCCTCATAACCTC	TGCTCCAAACATCAATCG	188
Phatr;*hCDK4*	hCDK4	Phatr1_55230	Phatr2_49025	TCCACTTGAAGAGATTGACG	TCTACACCACGAGCCATC	165
Phatr;*hCDK5*	hCDK5	Phatr1_30237	Phatr2_23198	GGACTGGCACGACGCTTTC	GGCTGGACCGTAACAACTCTC	111
Phatr;*hCDK6*	hCDK6	Phatr1_55668	Phatr2_5526	CCCACAAATGTCGCTGAAACG	CGGTAGGTTCGCAGGATGG	116
Phatr;*hCDK7*	hCDK7	Phatr1_55997	Phatr2_14453	TTTGGGCTGGCTAGAATAG	CGAAGTATATGGCTGGATAAG	192
Cyclins						
Phatr;*CYCA/B;1*	CYCA/B;1	Phatr1_45836	Phatr2_17135	TACACCGCCACTCCAAGAC	TCGGAGGACGGGATGGG	199
Phatr;*CYCB1*	CYCB1	Phatr1_55247	Phatr2_46095	TCCTGGTCCGCTACTTGAAAG	GCTGGCTGGGAAGATAACGC	117
Phatr;*CYCB2*	CYCB2	Phatr1_61099	Phatr2_8598	ACACAAAGTTCCGGCTCAT	AACCAGTTGTAGTCGCTTGC	104
Phatr;*CYCD1*	CYCD	Phatr1_14393	Phatr2_29283	CAGGAGCGTGGTCTTGTAG	AATATCCGACTCCGAATAGCC	213
Phatr;*CYCH1*	CYCH	Phatr1_36311	Phatr2_36892	ATACACGACGAGCGGATAC	AACAAGACGCCGTAAATCG	97
Phatr;*CYCL1*	CYCL	Phatr1_24121	Phatr2_5718	GCTGCCAAAACGGAGGAGTC	CTTCGCCTTTCGGATCAATGC	154
Phatr;*CYCP1*	CYCP1	Phatr1_38529	Phatr2_48210	GCTGCTGCTGCGACTGAC	ATGGCGGAGGCGGCTTC	115
Phatr;*CYCP2*	CYCP2	Phatr1_40691	Phatr2_49817	GGTGAAGGACGGCGATGC	ACGGAAGAACAAGACCAAGAGC	128
Phatr;*CYCP3*	CYCP3	Phatr1_45030	Phatr2_43226	GGCTACTCCACCTTCAACCC	GCTTCCACCGACCGATTCC	136
Phatr;*CYCP4*	CYCP4	Phatr1_49429	Phatr2_48202	AGTGGTTTGCTGGAAGACAG	GAATCCAATTCCAGTCAACG	148
Phatr;*CYCP5*	CYCP5	Phatr1_52534	Phatr2_43226	GGCCATGATTTTACAACAGC	ACCATGGTGCACTTGTAGGT	159
Phatr;*CYCP6*	CYCP6	Phatr1_59602	Phatr2_6231	AGGTGCTTGCTGCTGTTC	ACGAGGCATACTTGTGAATCC	150
Phatr;*dsCYC1*	dsCYC1	Phatr1_33465	Phatr2_31942	TCGCAGTAGAGAAAGTTCAGTC	TTGGCTTCGTGTCGTTGTTG	121
Phatr;*dsCYC;2*	dsCYC2	Phatr1_41834	Phatr2_34956	AAACAGCAACATTCCCAGCAAG	CGCACGCTTCAACCACATAC	157
Phatr;*dsCYC3*	dsCYC3	Phatr1_45204	Phatr2_43046	CTGCGTGCCAATGGAAATGAG	AATAAGGAGGGCTGCGATGG	127
Phatr;*dsCYC4*	dsCYC4	Phatr1_45216	Phatr2_43034	CATCCAATTCCGCTCACAACTC	TGGTGGGCAAACGCAGTC	196
Phatr;*dsCYC5*	dsCYC5	Phatr1_45284	Phatr2_32087	GACGGCGGAATTATCTCTCTGG	GTCGGAGTGAGCAGTGGTC	117
Phatr;*dsCYC6*	dsCYC6	Phatr1_50692	Phatr2_49373	CCGTCCCGTATCGTGAGTTC	CAAAGTGTCCGTGGCTGAG	144
Phatr;*dsCYC7*	dsCYC7	Phatr1_50931	Phatr2_49894	ATGAGTTACTTGGATCGCTACC	TCTTGACGGCTGTGTATAGGC	109
Phatr;*dsCYC8*	dsCYC8	Phatr1_50961	Phatr2_40753	ACACGCTTTTACTGCCTTTGC	TGTCTGTAACGGTTTCGGATTG	121
Phatr;*dsCYC9*	dsCYC9	Phatr1_51633	Phatr2_42423	CGGCTTCGTGGCACATTTTG	AGCGGGAGCAAAGCAACG	198
Phatr;*dsCYC10*	dsCYC10	Phatr1_51766	Phatr2_50251	TGGAGGTGGATCAACTGGTTC	ACGATGTGGTCTTCTTGTTCC	127
Phatr;*dsCYC11*	dsCYC11	Phatr1_51790	Phatr2_41189	CCTTTCCTCCTGGTCCTGTTC	AAGTGGGTGTCGTTGGTGTG	94
Phatr;*CYC-like*	CYC-like	Phatr1_44498	Phatr2_43757	CTCTGGGTCCGTCTTGGC	AAAGATGATGTGCTTACTGTTGTC	105
CDK Interactors and Regulatory proteins						
Phatr;*MYT1*	MYT1	Phatr1_46213	Phatr2_44716	ACCGCCGATAGTGACGAAAC	TCTAAAGTGGACGGATGGTAAGG	101
Phatr;*CKS1*	CKS1	Phatr1_10930	/	GCGGATGGCACCACTATG	GTCCACCTCACCCGTTTG	95

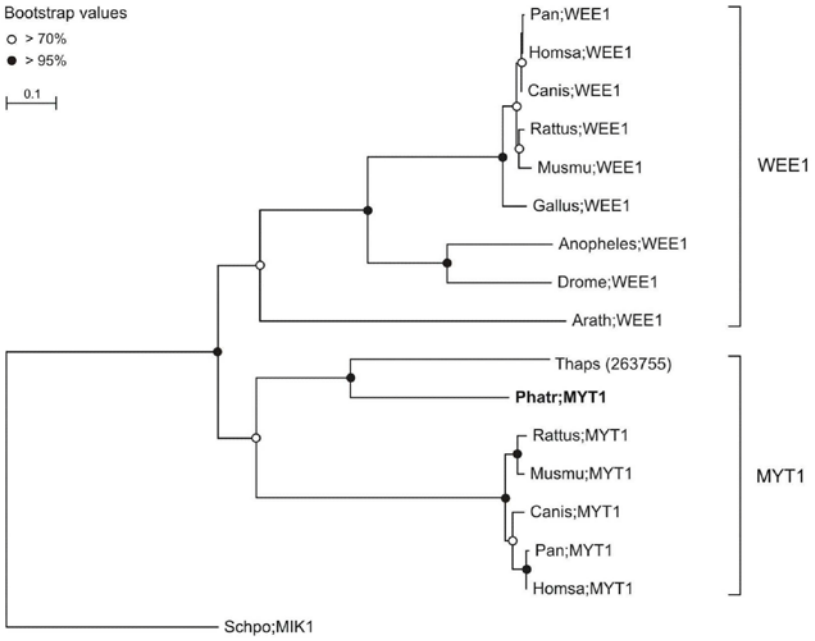
Additional Table 2.1. Overview of the annotated cell cycle genes in *P. tricornutum*.



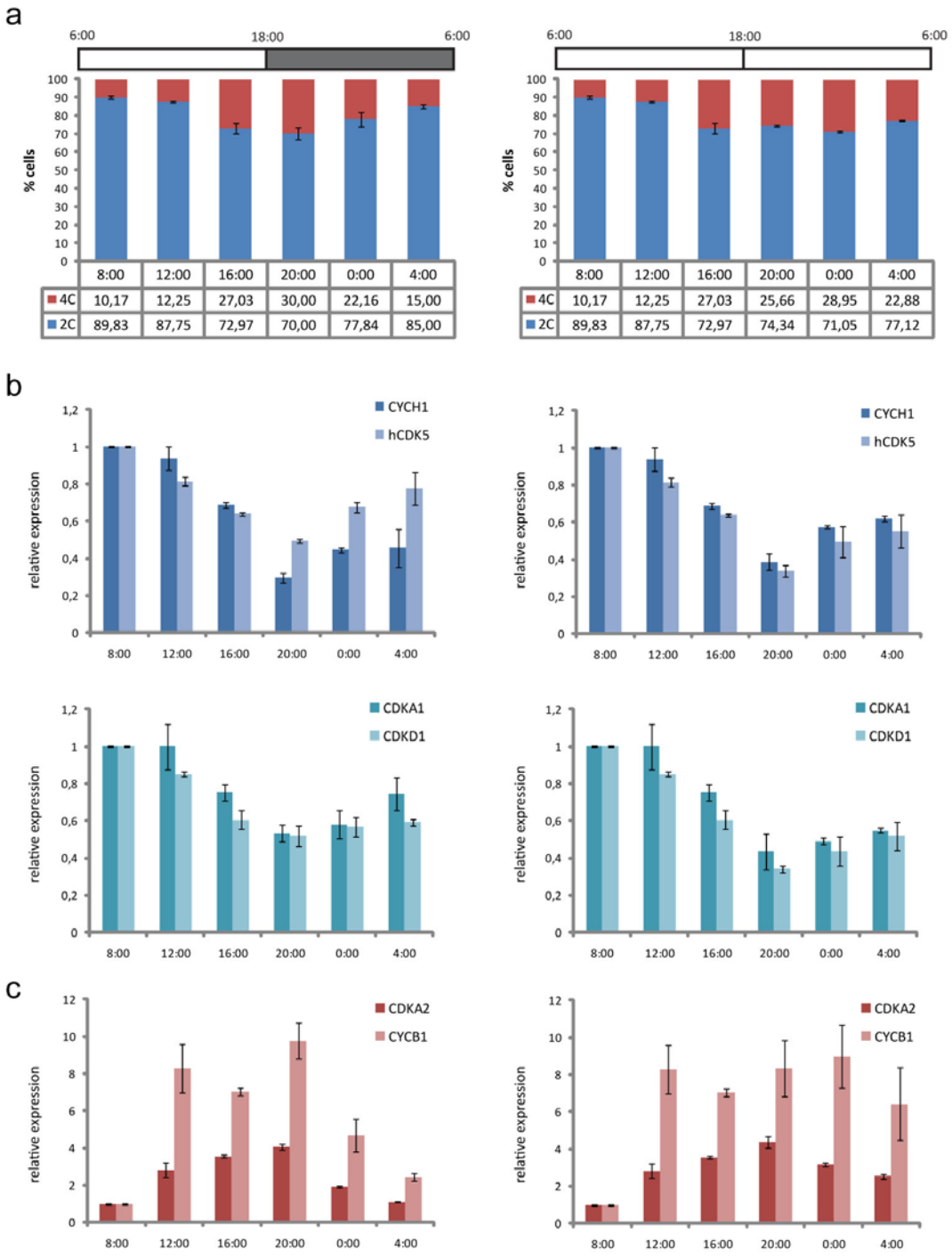
Additional figure 2.1: Cell cycle progression in nocodazole-treated versus untreated cells. **a.** Flow histograms plotting DNA content against cell number (left) and histograms indicating the ploidy distribution (2C versus 4C) (right) during a 12-hour time course of synchronized cells in the absence of nocodazole. At the later time points (t=10-12) the level of synchrony decreased, indicated by the ploidy level equilibrium reached at these time points, probably resulting from cells entering the next cell cycle round, while other cells still have to pass through M phase. **b.** Flow histograms plotting DNA content against cell number (left) and histograms indicating the ploidy distribution (2C versus 4C) (right) during a 12-hour time course of synchronized cells in the presence of nocodazole. At the later time points, an increasing enrichment of 4C cells can be observed, because of a blockage of the cells at metaphase. Asterisk marks the apparently lower proportion of 2C cells after a 20-hour dark treatment in the control series than in the nocodazole series, resulting from an acquisition artefact during flow cytometry, indicated by the increased peak broadness in the respective flow histogram.



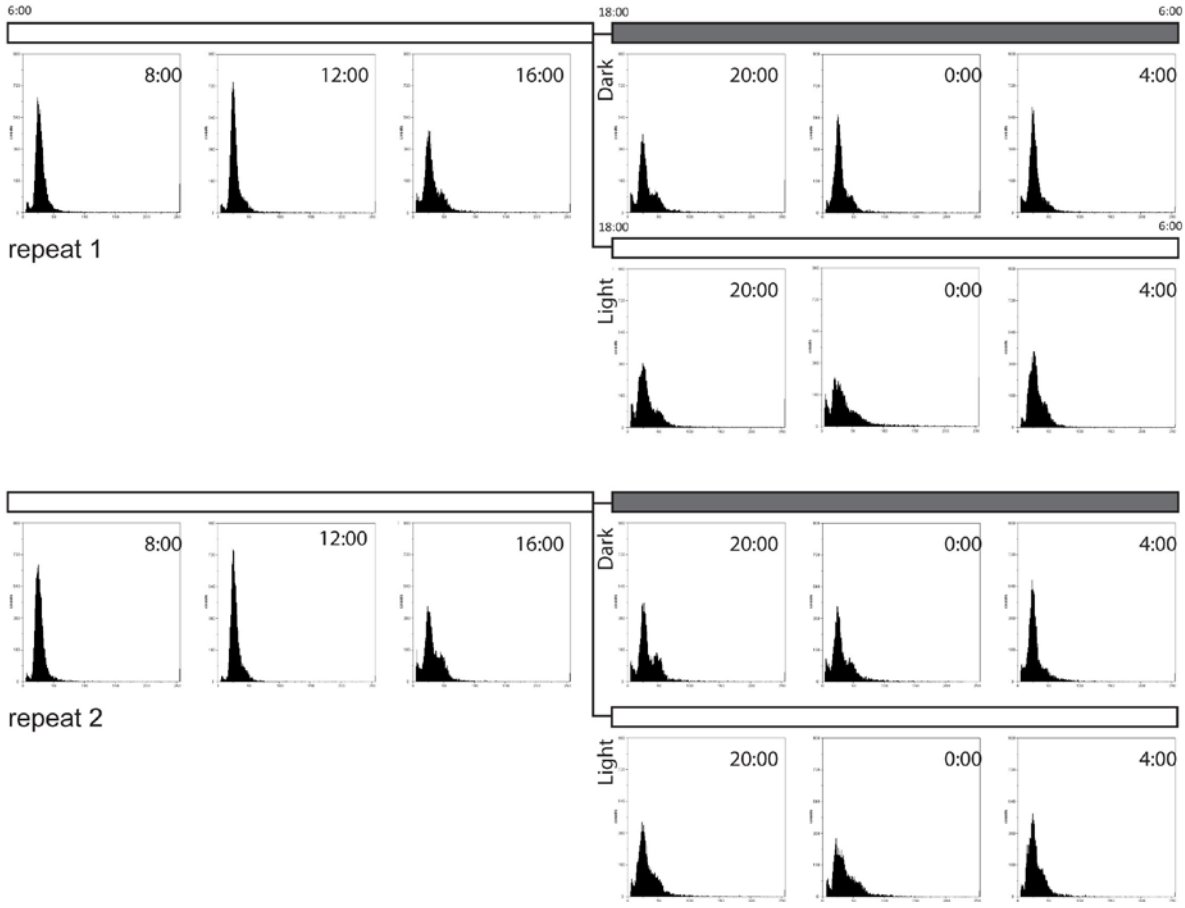
Additional figure 2.2: Growth curves of *P. tricornutum*. Growth curves of *P. tricornutum* cells under standard conditions (18°C, LD 12:12; 50-100 µmol photons m^{-2} s^{-1}). Error bars represent standard deviations.



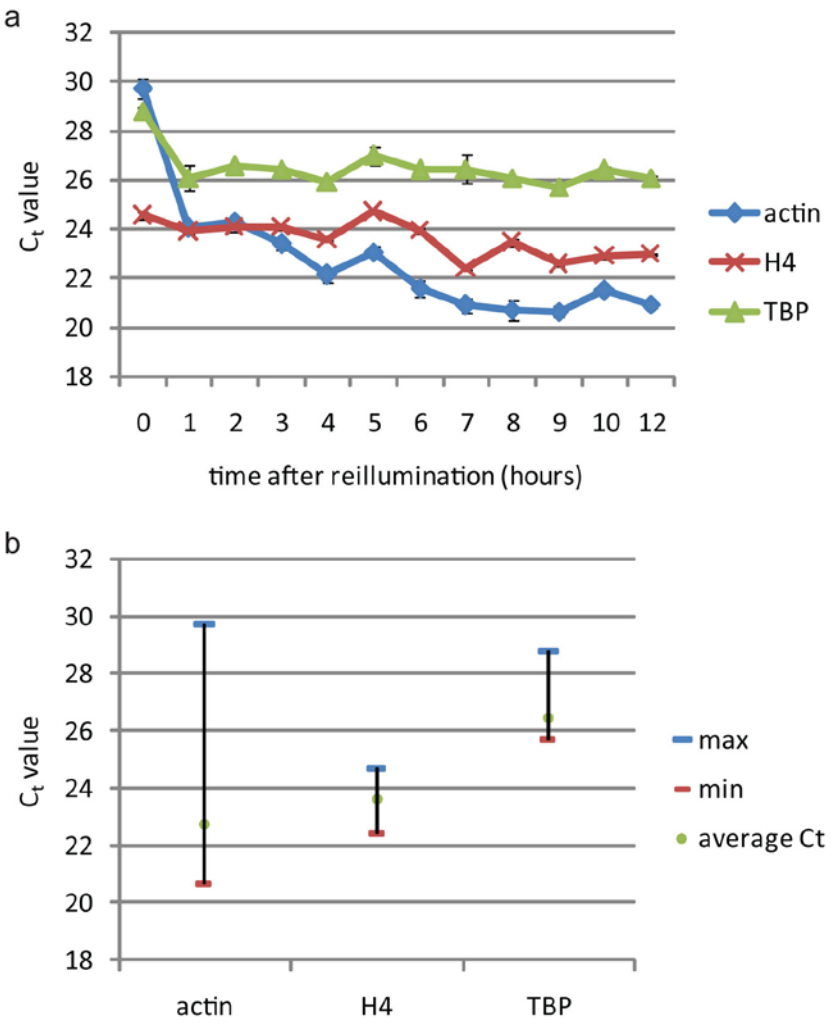
Additional figure 2.3: Phylogenetic tree of WEE1/MYT1/MIK1 family. Neighbor-joining tree (PHYLIP, 1000 replicates) of WEE1/MYT1/MIK1 family. The *P. tricornutum* sequence is shown in bold. Abbreviations: Arath**,** *Arabidopsis thaliana*; Drome, *Drosophila melanogaster*; Homsa, *Homo sapiens*; Musmu, *Mus musculus*; Orysa, *Oryza sativa*; Phatr, *Phaeodactylum tricornutum*; Schpo, *Schizosaccharomyces pombe*; Thaps, *Thalassiosira pseudonana*.



Additional figure 2.4: Cell cycle versus circadian control. Exponentially growing cultures entrained by a LD 12:12 photoperiod were subdivided in two cultures at the end of the light period 3 days after medium replenishment. Left and right, cells experiencing a normal (darkness; grey bar) and subjective (light; white bar) night, respectively. **a.** Histograms plotting DNA distributions (2C vs. 4C) of the cells during the 24-h time series. **b.** Expression profiles of early cell cycle genes. **c.** Expression profiles of late cell cycle genes. Error bars represent standard errors of the mean of two biological replicates.



Additional figure 2.5: Cell cycle versus circadian control: Flow histograms. Flow histograms (DNA content plotted against cell number) of the different sampling points depicted in Additional file 4.



Additional figure 2.6: Normalization gene evaluation. **a.** Real-time qPCR cycle threshold (Ct) values of candidate housekeeping genes during a 12-hour (sampling every hour) synchronization time series. **b.** Variation of Ct values of the candidate housekeeping genes during a 12-hour (sampling every hour) synchronization time series. Error bars represent standard deviations.

REFERENCES

Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S, Buck GA, Xu P, Bankier AT, Dear PH, Konfortov BA, Spriggs HF, Iyer L, Anantharaman V, Aravind L, Kapur V (2004) Complete genome sequence of the apicomplexan, Cryptosporidium parvum. Science **304:** 441-445

Allen AE, Laroche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C (2008) Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. Proc Natl Acad Sci U S A **105:** 10438-10443

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou SG, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The genome of the diatom Thalassiosira pseudonana: Ecology, evolution, and metabolism. Science **306:** 79-86

Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Segurens B, Daubin V, Anthouard V, Aiach N, Arnaiz O, Billaut A, Beisson J, Blanc I, Bouhouche K, Camara F, Duharcourt S, Guigo R, Gogendeau D, Katinka M, Keller AM, Kissmehl R, Klotz C, Koll F, Le Mouel A, Lepere G, Malinsky S, Nowacki M, Nowak JK, Plattner H, Poulain J, Ruiz F, Serrano V, Zagulski M, Dessen P, Betermier M, Weissenbach J, Scarpelli C, Schachter V, Sperling L, Meyer E, Cohen J, Wincker P (2006) Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature **444:** 171-178

Barroco RM, De Veylder L, Magyar Z, Engler G, Inzé D, Mironov V (2003) Novel complexes of cyclin-dependent kinases and a cyclin-like protein from Arabidopsis thaliana with a function unrelated to cell division. Cell Mol Life Sci **60:** 401-412

Berke JD, Sgambato V, Zhu PP, Lavoie B, Vincent M, Krause M, Hyman SE (2001) Dopamine and glutamate induce distinct striatal splice forms of Ania-6, an RNA polymerase II-associated cyclin. Neuron **32:** 277-287

Bisova K, Krylov DM, Umen JG (2005) Genome-wide annotation and expression profiling of cell cycle regulatory genes in Chlamydomonas reinhardtii. Plant Physiol **137:** 475-491

Bleeker PM, Hakvoort HW, Bliek M, Souer E, Schat H (2006) Enhanced arsenate reduction by a CDC25-like tyrosine phosphatase explains increased phytochelatin accumulation in arsenate-tolerant Holcus lanatus. Plant J **45:** 917-929

Boudolf V, Inzé D, De Veylder L (2006) What if higher plants lack a CDC25 phosphatase? Trends Plant Sci **11:** 474-479

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kroger N, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin-Jezequel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Oudot-Le Secq MP, Napoli C, Obornik M, Parker MS, Petit JL, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS,

Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature **456:** 239-244

Bozarth A, Maier UG, Zauner S (2009) Diatoms in biotechnology: modern tools and applications. Appl Microbiol Biotechnol **82:** 195-201

Brzezinski MA, Olson RJ, Chisholm SW (1990) Silicon availability and cell-cycle progression in marine diatoms. Marine Ecology-Progress Series **67:** 83-96

Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL, Peterson JD, Pop M, Kosack DS, Shumway MF, Bidwell SL, Shallom SJ, van Aken SE, Riedmuller SB, Feldblyum TV, Cho JK, Quackenbush J, Sedegah M, Shoaibi A, Cummings LM, Florens L, Yates JR, Raine JD, Sinden RE, Harris MA, Cunningham DA, Preiser PR, Bergman LW, Vaidya AB, van Lin LH, Janse CJ, Waters AP, Smith HO, White OR, Salzberg SL, Venter JC, Fraser CM, Hoffman SL, Gardner MJ, Carucci DJ (2002) Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. Nature **419:** 512-519

Churchman ML, Brown ML, Kato N, Kirik V, Hulskamp M, Inzé D, De Veylder L, Walker JD, Zheng Z, Oppenheimer DG, Gwin T, Churchman J, Larkin JC (2006) SIAMESE, a plant-specific cell cycle regulator, controls endoreplication onset in Arabidopsis thaliana. Plant Cell **18:** 3145-3157

Claudio PP, Tonini T, Giordano A (2002) The retinoblastoma family: twins or distant cousins? Genome Biol **3:** reviews3012

Coesel S, Mangogna M, Ishikawa T, Heijde M, Rogato A, Finazzi G, Todo T, Bowler C, Falciatore A (2009) Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. EMBO Rep **10:** 655-661

Coqueret O (2002) Linking cyclins to transcriptional control. Gene **299:** 35-55

De Clercq A, Inzé D (2006) Cyclin-dependent kinase inhibitors in yeast, animals, and plants: a functional comparison. Crit Rev Biochem Mol Biol **41:** 293-313

de Jager SM, Murray JA (1999) Retinoblastoma proteins in plants. Plant Mol Biol **41:** 295-299

De Martino A, Meichenin A, Shi J, Pan KH, Bowler C (2007) Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions. Journal of Phycology **43:** 992-1009

De Schutter K, Joubes J, Cools T, Verkest A, Corellou F, Babiychuk E, Van Der Schueren E, Beeckman T, Kushnir S, Inzé D, De Veylder L (2007) Arabidopsis WEE1 kinase controls cell cycle arrest in response to activation of the DNA integrity checkpoint. Plant Cell **19:** 211-225

Duan GL, Zhou Y, Tong YP, Mukhopadhyay R, Rosen BP, Zhu YG (2007) A CDC25 homologue from rice functions as an arsenate reductase. New Phytologist **174:** 311-321

Eddy SR (1998) Profile hidden Markov models. Bioinformatics **14:** 755-763

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:** 1792-1797

Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM, Tallon LJ, Delcher AL, Salzberg SL, Silva JC, Haas BJ, Majoros WH, Farzad M, Carlton JM, Smith RK, Jr., Garg J, Pearlman RE, Karrer KM, Sun L, Manning G, Elde NC, Turkewitz AP, Asai DJ, Wilkes DE, Wang Y, Cai H, Collins K, Stewart BA, Lee SR, Wilamowska K, Weinberg Z, Ruzzo WL, Wloga D, Gaertig J, Frankel J, Tsao CC, Gorovsky MA, Keeling PJ, Waller RF, Patron NJ, Cherry JM,

Stover NA, Krieger CJ, del Toro C, Ryder HF, Williamson SC, Barbeau RA, Hamilton EP, Orias E (2006) Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. PLoS Biol **4:** e286

Falciatore A, d'Alcala MR, Croot P, Bowler C (2000) Perception of environmental signals by a marine diatom. Science **288:** 2363-2366

Falkowski PG, Barber RT, Smetacek VV (1998) Biogeochemical controls and feedbacks on ocean primary production. Science **281:** 200-207

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. Science **281:** 237-240

Fisher RP, Morgan DO (1994) A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. Cell **78:** 713-724

Gardner MJ, Bishop R, Shah T, de Villiers EP, Carlton JM, Hall N, Ren Q, Paulsen IT, Pain A, Berriman M, Wilson RJ, Sato S, Ralph SA, Mann DJ, Xiong Z, Shallom SJ, Weidman J, Jiang L, Lynn J, Weaver B, Shoaibi A, Domingo AR, Wasawo D, Crabtree J, Wortman JR, Haas B, Angiuoli SV, Creasy TH, Lu C, Suh B, Silva JC, Utterback TR, Feldblyum TV, Pertea M, Allen J, Nierman WC, Taracha EL, Salzberg SL, White OR, Fitzhugh HA, Morzaria S, Venter JC, Fraser CM, Nene V (2005) Genome sequence of Theileria parva, a bovine pathogen that transforms lymphocytes. Science **309:** 134-137

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. Nature **419:** 498-511

Gillard J, Devos V, Huysman MJJ, De Veylder L, D'Hondt S, Martens C, Vanormelingen P, Vannerum K, Sabbe K, Chepurnov VA, Inzé D, Vuylsteke M, Vyverman W (2008) Physiological and transcriptomic evidence for a close coupling between chloroplast ontogeny and cell cycle progression in the pennate diatom Seminavis robusta. Plant Physiol **148:** 1394-1411

Goto K, Johnson CH (1995) Is the cell division cycle gated by a circadian clock? The case of Chlamydomonas reinhardtii. J Cell Biol **129:** 1061-1069

Guillard RRL (1975) Culture of phytoplankton for feeding marine invertebrates. *In* WL Smith, MH Canley, eds, Culture of Marine Invertebrate animals. Plenum Press, New York, pp 29-60

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser **41:** 95-98

Harper JW (2001) Protein destruction: adapting roles for Cks proteins. Curr Biol **11:** R431-435

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol **8:** R19

Inzé D, De Veylder L (2006) Cell cycle regulation in plant development. Annu Rev Genet **40:** 77-105

Irigoien X, Harris RP, Verheye HM, Joly P, Runge J, Starr M, Pond D, Campbell R, Shreeve R, Ward P, Smith AN, Dam HG, Peterson W, Tirelli V, Koski M, Smith T, Harbour D, Davidson R (2002) Copepod hatching success in marine ecosystems with high diatom concentrations. Nature **419:** 387-389

Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JC, Trent JM, Staudt LM, Hudson J, Jr., Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO (1999) The transcriptional program in the response of human fibroblasts to serum. Science **283:** 83-87

Joubes J, Chevalier C, Dudits D, Heberle-Bors E, Inzé D, Umeda M, Renaudin JP (2000) CDK-related protein kinases in plants. Plant Mol Biol **43:** 607-620

Kaffman A, Herskowitz I, Tjian R, O'Shea EK (1994) Phosphorylation of the transcription factor PHO4 by a cyclin-CDK complex, PHO80-PHO85. Science **263:** 1153-1156

Kaldis P (1999) The cdk-activating kinase (CAK): from yeast to mammals. Cell Mol Life Sci **55:** 284-296

Khadaroo B, Robbens S, Ferraz C, Derelle E, Eychenie S, Cooke R, Peaucellier G, Delseny M, Demaille J, Van de Peer Y, Picard A, Moreau H (2004) The first green lineage cdc25 dual-specificity phosphatase. Cell Cycle **3:** 513-518

Kitsios G, Alexiou KG, Bush M, Shaw P, Doonan JH (2008) A cyclin-dependent protein kinase, CDKC2, colocalizes with and modulates the distribution of spliceosomal components in Arabidopsis. Plant J **54:** 220-235

Kooistra WH, De Stefano M, Mann DG, Medlin LK (2003) The phylogeny of the diatoms. Prog Mol Subcell Biol **33:** 59-97

Kröger N (2007) Prescribing diatom morphology: toward genetic engineering of biological nanomaterials. Curr Opin Chem Biol **11:** 662-669

Lebeau T, Robert JM (2003b) Diatom cultivation and biotechnologically relevant products. Part I: cultivation at various scales. Appl Microbiol Biotechnol **60:** 612-623

Li S, Nosenko T, Hackett JD, Bhattacharya D (2006) Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. Mol Biol Evol **23:** 663-674

Liu F, Stanton JJ, Wu Z, Piwnica-Worms H (1997) The human Myt1 kinase preferentially phosphorylates Cdc2 on threonine 14 and localizes to the endoplasmic reticulum and Golgi complex. Mol Cell Biol **17:** 571-583

Liu ZJ, Ueda T, Miyazaki T, Tanaka N, Mine S, Tanaka Y, Taniguchi T, Yamamura H, Minami Y (1998) A critical role for cyclin C in promotion of the hematopoietic cell cycle by cooperation with c-Myc. Mol Cell Biol **18:** 3445-3454

Maheswari U, Mock T, Armbrust EV, Bowler C (2009) Update of the Diatom EST Database: a new tool for digital transcriptomics. Nucleic Acids Res **37:** D1001-1005

Marchetti A, Parker MS, Moccia LP, Lin EO, Arrieta AL, Ribalet F, Murphy ME, Maldonado MT, Armbrust EV (2009) Ferritin is used for iron storage in bloom-forming marine pennate diatoms. Nature **457:** 467-470

Martens C, Vandepoele K, Van de Peer Y (2008) Whole-genome analysis reveals molecular innovations and evolutionary transitions in chromalveolate species. Proc Natl Acad Sci U S A **105:** 3427-3432

Mendenhall MD, Hodge AE (1998) Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast Saccharomyces cerevisiae. Microbiol Mol Biol Rev **62:** 1191-1243

Michaelis C, Weeks G (1992) Isolation and characterization of a cdc 2 cDNA from Dictyostelium discoideum. Biochim Biophys Acta **1132:** 35-42

Montsant A, Allen AE, Coesel S, De Martino A, Falciatore A, Mangogna M, Siaut M, Heijde M, Jabbari K, Maheswari U, Rayko E, Vardi A, Apt KE, Berges JA, Chiovitti A, Davis AK, Thamatrakoln K, Hadi MZ, Lane TW, Lippmeier JC, Martinez D, Parker MS, Pazour

GJ, Saito MA, Rokhsar DS, Armbrust EV, Bowler C (2007) Identification and comparative genomic analysis of signaling and regulatory components in the diatom Thalassiosira pseudonana. J Phycol **43:** 585-604

Montsant A, Jabbari K, Maheswari U, Bowler C (2005) Comparative genomics of the pennate diatom Phaeodactylum tricornutum. Plant Physiol **137:** 500-513

Morgan DO (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu Rev Cell Dev Biol **13:** 261-291

Moser BA, Russell P (2000) Cell cycle regulation in Schizosaccharomyces pombe. Curr Opin Microbiol **3:** 631-636

Moulager M, Monnier A, Jesson B, Bouvet R, Mosser J, Schwartz C, Garnier L, Corellou F, Bouget FY (2007) Light-dependent regulation of cell division in Ostreococcus: evidence for a major transcriptional input. Plant Physiol **144:** 1360-1369

Mueller PR, Coleman TR, Kumagai A, Dunphy WG (1995) Myt1: a membrane-associated inhibitory kinase that phosphorylates Cdc2 on both threonine-14 and tyrosine-15. Science **270:** 86-90

Ng CKF, Lam CMC, Yeung PKK, Wong JTY (1998) Flow cytometric analysis of nocodazole-induced cell-cycle arrest in the pennate diatom Phaeodactylum tricornutum Bohlin. J Appl Phycol **10:** 569-572

Oelgeschlager T (2002) Regulation of RNA polymerase II activity by CTD phosphorylation and cell cycle control. J Cell Physiol **190:** 160-169

Oeltjen A, Marquardt J, Rhiel E (2004) Differential circadian expression of genes fcp2 and fcp6 in Cyclotella cryptica. Int Microbiol **7:** 127-131

Olson RJ, Vaulot D, Chisholm SW (1986) Effects of environmental stresses on the cell cycle of two marine phytoplankton species. Plant Physiol **80:** 918-925

Pain A, Renauld H, Berriman M, Murphy L, Yeats CA, Weir W, Kerhornou A, Aslett M, Bishop R, Bouchier C, Cochet M, Coulson RM, Cronin A, de Villiers EP, Fraser A, Fosker N, Gardner M, Goble A, Griffiths-Jones S, Harris DE, Katzer F, Larke N, Lord A, Maser P, McKellar S, Mooney P, Morton F, Nene V, O'Neil S, Price C, Quail MA, Rabbinowitsch E, Rawlings ND, Rutter S, Saunders D, Seeger K, Shah T, Squares R, Squares S, Tivey A, Walker AR, Woodward J, Dobbelaere DA, Langsley G, Rajandream MA, McKeever D, Shiels B, Tait A, Barrell B, Hall N (2005) Genome of the host-cell transforming parasite Theileria annulata compared with T. parva. Science **309:** 131-133

Palenik B, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH, Madupu R, Nelson WC, Brinkac LM, Dodson RJ, Durkin AS, Daugherty SC, Sullivan SA, Khouri H, Mohamoud Y, Halpin R, Paulsen IT (2006) Genome sequence of Synechococcus CC9311: Insights into adaptation to a coastal environment. Proc Natl Acad Sci U S A **103:** 13555-13559

Peers G, Niyogi KK (2008) Pond scum genomics: the genomes of Chlamydomonas and Ostreococcus. Plant Cell **20:** 502-507

Perry JA, Kornbluth S (2007) Cdc25 and Wee1: analogous opposites? Cell Div **2:** 12

Pichevin LE, Reynolds BC, Ganeshram RS, Cacho I, Pena L, Keefe K, Ellam RM (2009) Enhanced carbon pump inferred from relaxation of nutrient limitation in the glacial ocean. Nature **459:** 1114-1117

Pines J (1996) Cell cycle: reaching for a role for the Cks proteins. Curr Biol **6:** 1399-1402

Ragni M, D'Alcala MR (2007) Circadian variability in the photobiology of Phaeodactylum tricornutum: pigment content. Journal of Plankton Research **29:** 141-156

Robbens S, Khadaroo B, Camasses A, Derelle E, Ferraz C, Inzé D, Van de Peer Y, Moreau H (2005) Genome-wide analysis of core cell cycle genes in the unicellular green alga Ostreococcus tauri. Mol Biol Evol **22:** 589-597

Round FE, Crawford RM, Mann DG (1990) The Diatoms: Biology and Morphology of the Genera. Cambridge University Press, Cambridge

Sanchez I, Dynlacht BD (2005) New insights into cyclins, CDKs, and cell cycle control. Semin Cell Dev Biol **16:** 311-321

Santamaria D, Barriere C, Cerqueira A, Hunt S, Tardy C, Newton K, Caceres JF, Dubus P, Malumbres M, Barbacid M (2007) Cdk1 is sufficient to drive the mammalian cell cycle. Nature **448:** 811-815

Sapriel G, Quinet M, Heijde M, Jourdren L, Tanty V, Luo G, Le Crom S, Lopez PJ (2009) Genome-wide transcriptome analyses of silicon metabolism in Phaeodactylum tricornutum reveal the multilevel regulation of silicic acid transporters. PLoS One **4:** e7458

Schnittger A, Schobinger U, Bouyer D, Weinl C, Stierhof YD, Hulskamp M (2002) Ectopic D-type cyclin expression induces not only DNA replication but also cell division in Arabidopsis trichomes. Proc Natl Acad Sci U S A **99:** 6410-6415

Sherr CJ, Roberts JM (1995) Inhibitors of mammalian G1 cyclin-dependent kinases. Genes Dev **9:** 1149-1163

Sherr CJ, Roberts JM (1999) CDK inhibitors: positive and negative regulators of G1-phase progression. Genes Dev **13:** 1501-1512

Siaut M, Heijde M, Mangogna M, Montsant A, Coesel S, Allen A, Manfredonia A, Falciatore A, Bowler C (2007) Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum. Gene **406:** 23-35

Torres Acosta JA, de Almeida Engler J, Raes J, Magyar Z, De Groodt R, Inzé D, De Veylder L (2004) Molecular characterization of Arabidopsis PHO80-like proteins, a novel class of CDKA;1-interacting cyclins. Cell Mol Life Sci **61:** 1485-1497

Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RH, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, Chapman J, Damasceno CM, Dorrance AE, Dou D, Dickerman AW, Dubchak IL, Garbelotto M, Gijzen M, Gordon SG, Govers F, Grunwald NJ, Huang W, Ivors KL, Jones RW, Kamoun S, Krampis K, Lamour KH, Lee MK, McDonald WH, Medina M, Meijer HJ, Nordberg EK, Maclean DJ, Ospina-Giraldo MD, Morris PF, Phuntumart V, Putnam NH, Rash S, Rose JK, Sakihama Y, Salamov AA, Savidor A, Scheuring CF, Smith BM, Sobral BW, Terry A, Torto-Alalibo TA, Win J, Xu Z, Zhang H, Grigoriev IV, Rokhsar DS, Boore JL (2006) Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science **313:** 1261-1266

Umeda M, Shimotohno A, Yamaguchi M (2005) Control of cell division and transcription by cyclin-dependent kinase-activating kinases in plants. Plant Cell Physiol **46:** 1437-1442

Van de Peer Y, De Wachter R (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Appl Biosci **10:** 569-570

Van den Hoek C, Mann DG, Jahns HM, eds (1995) Algae: An Introduction to Phycology. Cambridge University Press, Cambridge

Van Mooy BA, Fredricks HF, Pedler BE, Dyhrman ST, Karl DM, Koblizek M, Lomas MW, Mincer TJ, Moore LR, Moutin T, Rappe MS, Webb EA (2009) Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. Nature **458:** 69-72

Vandepoele K, Raes J, De Veylder L, Rouze P, Rombauts S, Inzé D (2002) Genome-wide analysis of core cell cycle genes in Arabidopsis. Plant Cell **14:** 903-916

Vardi A, Bidle KD, Kwityn C, Hirsh DJ, Thompson SM, Callow JA, Falkowski P, Bowler C (2008) A diatom gene regulating nitric-oxide signaling and susceptibility to diatom-derived aldehydes. Curr Biol **18:** 895-899

Vardi A, Formiggini F, Casotti R, De Martino A, Ribalet F, Miralto A, Bowler C (2006) A stress surveillance system based on calcium and nitric oxide in marine diatoms. PLoS Biol **4:** e60

Vaulot D, Olson RJ, Merkel S, Chisholm SW (1987) Cell-cycle response to nutrient starvation in 2 phytoplankton species, Thalassiosira-Weissflogii and Hymenomonas-Carterae. Marine Biology **95:** 625-630

Verkest A, Weinl C, Inzé D, De Veylder L, Schnittger A (2005) Switching the cell cycle. Kip-related proteins in plant cell cycle control. Plant Physiol **139:** 1099-1106

Wang W, Chen X (2004) HUA ENHANCER3 reveals a role for a cyclin-dependent protein kinase in the specification of floral organ identity in Arabidopsis. Development **131:** 3147-3156

Weinberg RA (1995) The retinoblastoma protein and cell cycle control. Cell **81:** 323-330

Wu J, Sunda W, Boyle EA, Karl DM (2000) Phosphate depletion in the western North Atlantic Ocean. Science **289:** 759-762

Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, Puiu D, Manque P, Akiyoshi D, Mackey AJ, Pearson WR, Dear PH, Bankier AT, Peterson DL, Abrahamsen MS, Kapur V, Tzipori S, Buck GA (2004) The genome of Cryptosporidium hominis. Nature **431:** 1107-1112

Yamaguchi M, Fabian T, Sauter M, Bhalerao RP, Schrader J, Sandberg G, Umeda M, Uchimiya H (2000) Activation of CDK-activating kinase is dependent on interaction with H-type cyclins in plants. Plant J **24:** 11-20

Zhou BB, Elledge SJ (2000) The DNA damage response: putting checkpoints in perspective. Nature **408:** 433-439

3

The light-dependent diatom-specific cyclin 2 controls cell cycle progression after dark arrest in *Phaeodactylum tricornutum*

Marie J.J. Huysman[1,2,3], Chris Bowler[4], Angela Falciatore[5,6], Dirk Inzé[2,3], Wim Vyverman[1] and Lieven De Veylder[2,3]

[1]Protistology and Aquatic Ecology, Department of Biology, Ghent University, 9000 Gent, Belgium
[2]Department of Plant Systems Biology, Flanders Institute for Biotechnology (VIB), 9052 Gent, Belgium
[3]Department of Plant Biotechnology and Genetics, Ghent University, 9052 Gent, Belgium
[4]Département de Biologie, Ecole Normale Supérieure, CNRS, Unité Mixte de Recherche 8186, 75230 Paris Cedex 05, France
[5]Laboratory of Ecology and Evolution of Plankton, Stazione Zoologica Anton Dohrn, 80121 Napoli, Italy
[6]Laboratoire de Génomique des Microorganismes, Université Pierre et Marie Curie, CNRS FRE3214, 75006 Paris, France

Manuscript in preparation

Authors' contributions

MJJH performed the experiments, analyzed the data and wrote the chapter. MJJH, DI, WV and LDV conceived and designed the study. WV and LDV revised the manuscript. CB and AF provided overexpression and silencing tools and discussed the data with the other authors.

ABSTRACT

Light is a key environmental cue for growth in diatoms. However, little is known about the cellular signaling cascade connecting the light perception mechanisms and activation of the cell cycle process. Here, we show that within 15 minutes after light exposure, diatom-specific cyclin 2 (*dsCYC2*) displays a transcriptional peak, which is followed instantly by its protein synthesis. Moreover, we show that after illumination a transcriptional repressor is produced through *de novo* protein synthesis that is responsible for rapid decrease in *dsCYC2* transcript levels via inhibition of promoter activity. Introduction of a hairpin construct targeting *dsCYC2* induces a decrease in cell division rate, indicating that *dsCYC2* is important for cell cycle progression after illumination. We show that this cell cycle delay is due to a prolongation of the G1-to-S-phase transition. To our knowledge, this is the first report of a light-regulated cyclin gene that is involved in cell cycle regulation.

INTRODUCTION

Diatoms form an important group of unicellular algae in the contemporary oceans, being responsible for about twenty percent of the primary production on Earth (Van den Hoek, 1995; Field et al., 1998). Although they are widespread throughout the oceans and show extreme flexibility towards continuously varying environmental conditions, little is known about how they perceive and respond or adapt to these external variations, and in particular to changing light conditions. Light is an essential factor for every photosynthetic organism as a source of energy, but also as a trigger that induces several physiological and biochemical responses, and controls different cellular processes, including cell division (Chen et al., 2004). When nutrients are not limiting, diatom growth and distribution is strongly dependent on the prevailing light conditions (Smetacek, 1999). Because diatoms can grow and photosynthesize over a wide range of different light intensities and wavelengths, they are supposed to possess specific light sensing and acclimation strategies. This hypothesis has been supported by the recent discovery of specific photoprotective and acclimation responses of diatoms to high light conditions (Nymark et al., 2009; Zhu and Green, 2010). Absence of light, on the other hand, causes diatom cells to cease their reproduction. Interestingly, the two major diatom groups, the centrics and the pennates (Kooistra et al., 2003; Sims et al., 2006), appear to have developed different light-sensitive phases during their mitotic cell cycle, probably related to the different environments that they inhabit. Centric diatoms display a planktonic lifestyle and are well adapted to growth in deeply mixed water, while pennates commonly inhabit heterogeneous benthic environments. Flow cytometric analysis of dark-adapted cells showed that centric species harbor two light-sensitive stages during their cell cycle, namely during the G1 and G2-M phases (Vaulot et al., 1986; Brzezinski et al., 1990). Pennate diatoms have been reported to show only a G1-arrest, as reported for *Phaeodactylum tricornutum* (Brzezinski et al., 1990; Huysman et al., 2010) and *Seminavis robusta* (Gillard et al., 2008), or both a G1-and a G2-M arrest, like reported for *Cylindrotheca fusiformis* (Brzezinski et al., 1990). It has been suggested that the light-dependent segment in G1 would result from the energy requirement to complete the cell cycle, whereas the light-dependent segment at the G2-M phases would relate to the energetic requirements for the synthesis of the silicon uptake mechanism, silica transport and deposition during valve formation at that stage (Vaulot et al., 1986). However, the latter cannot explain the absence of a dark-sensitive

G2-M segment in some pennate diatoms that require silicon, like *S. robusta* (Gillard et al., 2008). Recently, Gillard et al. suggested that the G2-M phase fraction detected in previous flow cytometric analyses might actually represent post-cytokinetic doublet cells, and therefore only the process of cell separation, and not cytokinesis itself or its related silicon requirement, would be hampered by the absence of light (Gillard et al., 2008). Interestingly, diatom cells that are released from a dark arrest at G1 are able to resume cycling immediately at a normal rate, while G2-M arrested cells (or possibly post-cytokinetic cells) show a retardation of cell cycle progression (Vaulot et al., 1986). For those species with only a light-dependent segment at the G1-phase, the immediate release of dark-arrested cells has been proven a useful characteristic to synchronize and study the cell division process (Gillard et al., 2008; Huysman et al., 2010).

Despite the key role of light for growth in diatoms and other photoautotrophic organisms, little is known about the cellular signaling mechanisms that connect the perception of the light signal through photoreceptors at the surface of the cell membrane with the activation of the main eukaryotic cell cycle machinery in the nucleus, including regulators such as cyclin-dependent kinases (CDKs) and their interaction partners, the cyclins (Morgan, 1997; Inzé and De Veylder, 2006). In a recent study, we revealed a major expansion of the cyclin gene family in the pennate diatom *P. tricornutum* and discovered a new class of diatom-specific cyclins (dsCYCs) involved in environmental signaling (Huysman et al., 2010). One of the strongest and earliest expressed genes during the switch from dark to light in synchronized cultures was *dsCYC2*, hinting at a role for this cyclin in cell cycle activation after dark arrest. To address the latter hypothesis we attempted here to uncover the light-dependent regulation of *dsCYC2* transcription and translation and the possible role of *dsCYC2* at the light-dependent checkpoint of the G1-phase.

The first goal of this study was to document the kinetics of *dsCYC2* transcript and protein abundance upon illumination of dark-arrested *P. tricornutum* cells. We approached this goal by performing gene and protein expression profiling during both a long-term and a finely-resolved short-term time series. Our second goal was to understand the mechanisms that cause the observed changes in transcript levels upon illumination. To address this, we followed the activity of the *dsCYC2* promoter using a reporter line and analyzed the effects of inhibitors interfering with protein translation and redox signaling on *dsCYC2* transcription. Finally, we tried to unravel the precise function of *dsCYC2* during the cell cycle by the identification of interaction partners and by studying the effect of both overexpression and

silencing of *dsCYC2* on cell cycle progression. The results provide a detailed insight into the light-dependent regulation of *dsCYC2* transcription and translation and provide evidence that *dsCYC2* functions as a positive regulator of the light-dependent checkpoint at the G1-phase.

RESULTS

Light-dependent transcriptional and translational control of *dsCYC2*

To be able to monitor simultaneously transcript and protein levels of *dsCYC2*, we generated a transgenic marker line that expressed the full length *dsCYC2* sequence C-terminally fused to a hemagglutinin (HA) tag under control of the *dsCYC2* promoter (Figure 3.1a), which we will refer to as the HA marker line. The HA tag is a short epitope tag that does not appear to interfere with the activity and distribution of the protein fused to it and it facilitates the detection, isolation and purification of the protein.

Initially we sampled during a total of 12 hours to follow the kinetics of *dsCYC2* transcript and protein levels upon light exposure. Therefore, cells were grown exponentially in a 12h light/12h dark (12L/12D) light regime and then transferred to a prolonged dark period of 24 hours which, due to a light-dependent segment at the G1 phase (Brzezinski et al., 1990; Huysman et al., 2010), enriches the cultures for G1-phase cells. When returned to light, cells progress synchronously through the cell cycle (Huysman et al., 2010). After illumination, samples were taken at 0, 0.5, 1, 3, 6 and 12 hours for real-time quantitative PCR (qPCR) and western blot analysis. Starting at relatively high levels, *dsCYC2* transcript increased during the first 30 minutes, to decrease to low basal levels during the later timepoints in both wild-type (WT) and HA marker cells (Figure 3.1b and 3.1c). The latter indicates that the *dsCYC2-HA* fusion reports properly the transcriptional kinetics of the endogenous *dsCYC2* gene (compare Figure 3.1b and 3.1c). Protein levels of HA-tagged dsCYC2 were undetectable at the time of illumination, but reached high levels at 30 to 60 minutes, where after they decreased gradually to become undetectable at 12 hours after light exposure (Figure 3.1e).

To determine how fast *dsCYC2* transcript levels are elevated after light addition, we performed a real-time qPCR analysis on samples during a finely resolved short time course (0, 5, 10, 15, 30, 45 and 60 minutes after light exposure of 24-hour dark-adapted cells).

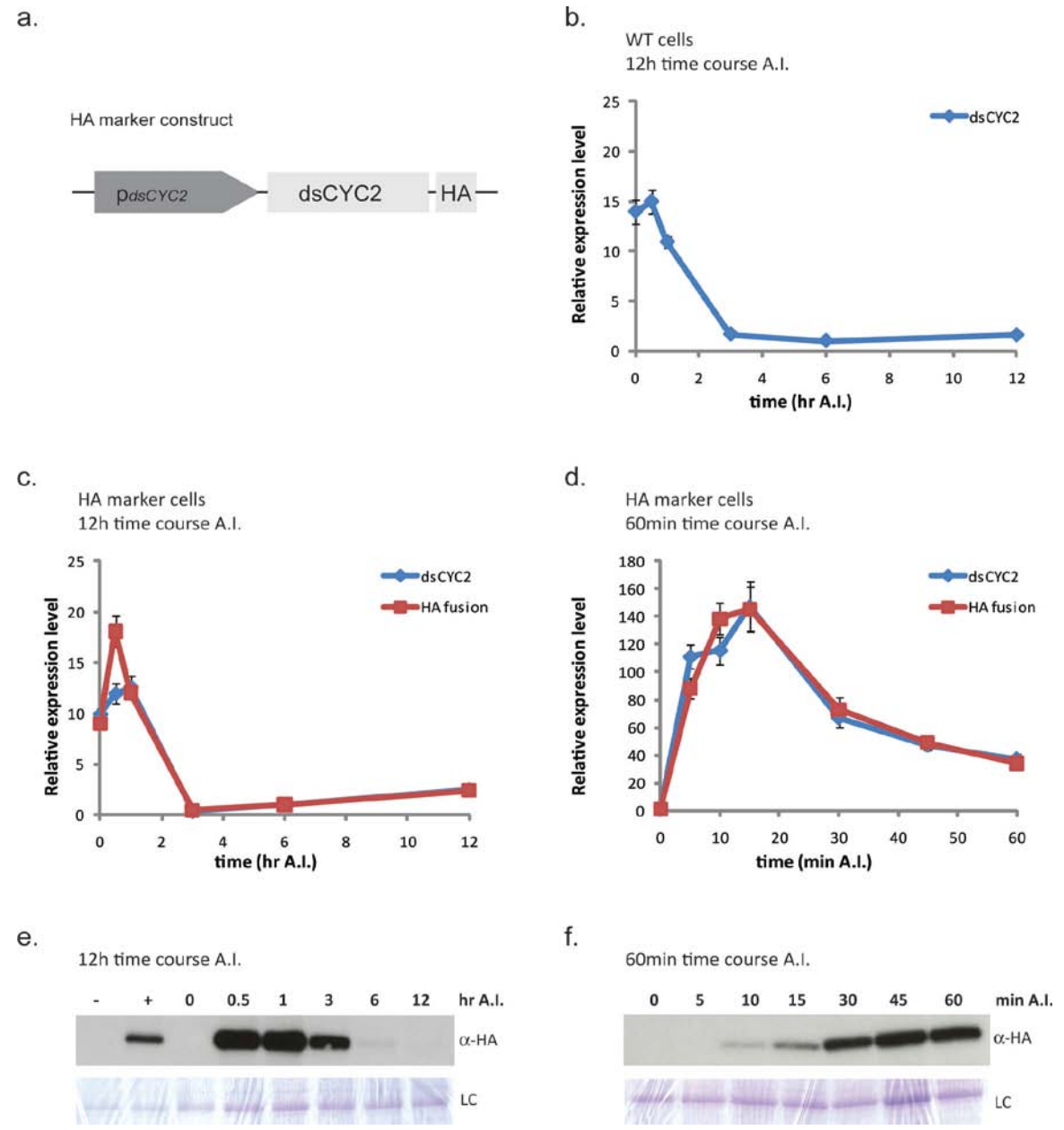


Figure 3.1: Light-dependent transcription and translation of *dsCYC2*. **a.** Schematic representation of the HA marker construct. **b.** Transcript levels during a 12-hour time course after illumination of 24-hour dark-adapted WT cells. **c.** Transcript levels during a 12-hour time course after illumination of 24-hour dark-adapted HA marker cells. In b and c, values were normalized against the gene expression levels at 6hr after illumination (A.I.). Error bars represent standard errors. **d.** Transcript levels over a short time course (1hour) after illumination of 24-hour dark-adapted HA marker cells. Values were normalized against the gene expression levels at 0 min A.I. Error bars represent standard errors. Note that transcript levels at 0 hr A.I. in (c) seem relatively higher compared to those at 0 hr A.I. in (d), but the former is in fact an overestimate resulting from the slower sampling method used during the long-term experiment (see Materials and Methods section) **e.** dsCYC2-HA protein levels during the 12-hour time course. **f.** dsCYC2-HA protein levels during the short time course. -, Negative control (WT 4h light); +, Positive control (HA 4h light). LC: Loading control by Coomassie blue staining.

Interestingly, an initial increase in transcript levels could be observed after only five minutes of illumination, reaching a peak at 15 minutes, after which *dsCYC2* mRNA levels decreased rapidly (Figure 3.1d). Protein levels could be detected starting from 10 minutes after illumination and kept increasing until 60 minutes (Figure 3.1f), although transcript levels were markedly lowered at that timepoint. To summarize, these data show that upon illumination, *dsCYC2* reaches a transcriptional peak within 15 minutes, followed by a translational peak around 30 to 60 minutes.

Promoter responsiveness to light

The fact that the *dsCYC2* transcript levels are markedly affected by light might be due to either a direct effect of light on *dsCYC2* transcript stability or a yet undefined light-dependent signaling pathway targeting the *dsCYC2* promoter sequence. In order to distinguish between both possibilities, we analyzed the long- and short-time response of a reporter gene placed under control of the *dsCYC2* promoter during the light period after dark incubation. To construct this reporter fusion, we combined the 1018 bp region upstream of the translational start of *dsCYC2* and the coding region of enhanced yellow fluorescent protein (*eYFP*) (Figure 3.2a). In general, the *eYFP* transcript was detectable both in the dark and light samples, although shortly after light exposure, higher transcript levels were observed, that dropped again to basal levels after three hours of light exposure (Figure 3.2b). The kinetics of the *eYFP* transcript downregulation showed a slight delay compared to the kinetics of the endogenous *dsCYC2* transcript, most apparent during the short-term experiment (Figure 3.2c). This is probably because of higher stability of *eYFP* versus *dsCYC2* transcripts. Nevertheless, the overall parallel response of the *dsCYC2* transcript and *eYFP* reporter over time suggest that changes in *dsCYC2* mRNA are primarily the consequence of changes in promoter activity rather than mRNA stability. Protein levels were constitutively high during both the light and dark period (Figure 3.2d and 3.2e) and eYFP signals could be detected in cells that were incubated in darkness for at least one week (data not shown). The latter is probably due to high eYFP protein stability in the cells.

Transcript regulation of dsCYC2 is dependent of protein synthesis upon light exposure

To determine if *dsCYC2* transcript regulation during the light period is dependent on protein translation, we treated WT cells just before light exposure with cycloheximide (CHX), an inhibitor of eukaryotic translation. In contrast to the non-treated control cultures that

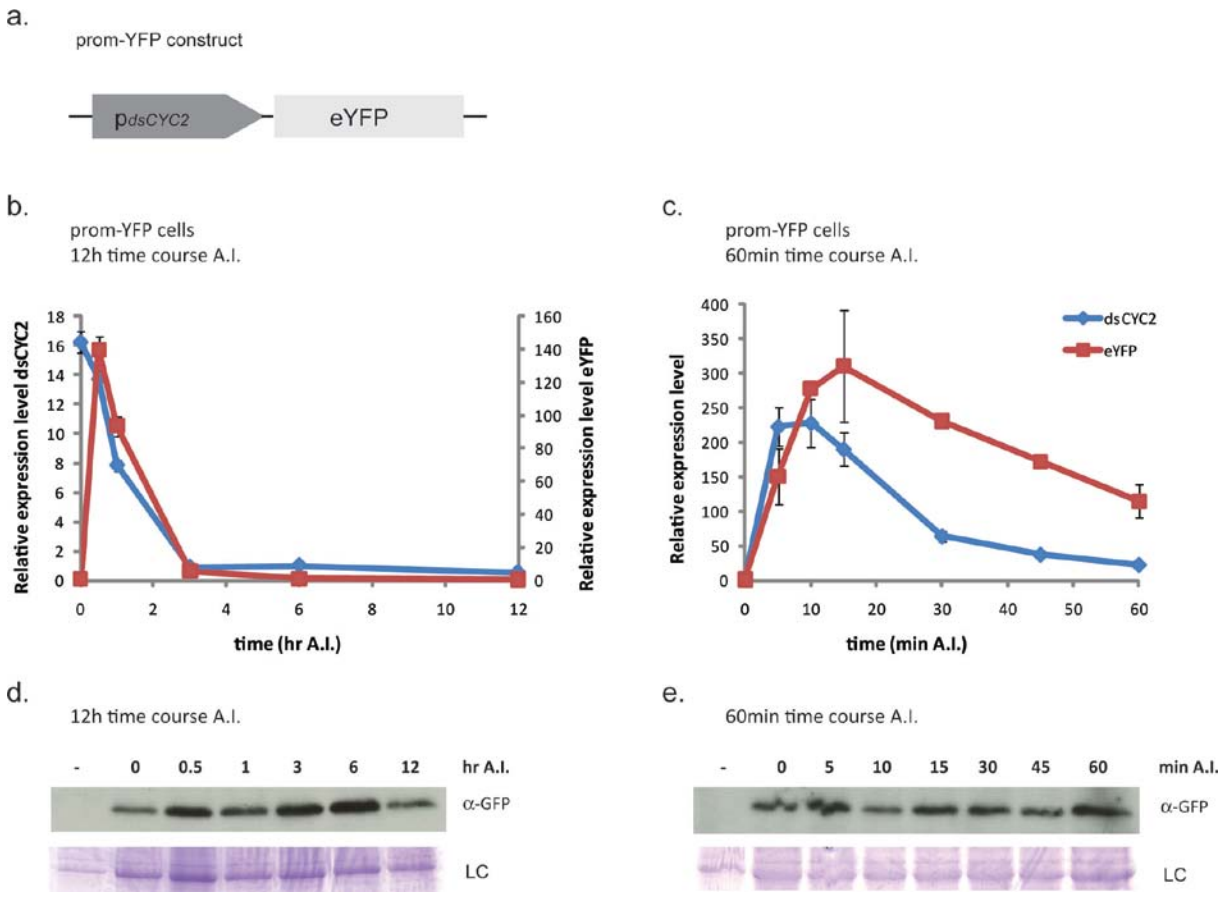


Figure 3.2: Promoter responsiveness to light. **a.** Schematic representation of the prom-YFP reporter construct. **b.** Transcript levels during a 12-hour time course after illumination of 24-hour dark-adapted cells. Values were normalized against the gene expression levels at 6hr after illumination (A.I.). Error bars represent standard errors. **c.** Transcript levels over a short time course (1hour) after illumination of 24-hour dark-adapted cells. Values were normalized against the gene expression levels at 0 min A.I. Error bars represent standard errors of the mean of two independent experiments. Note that transcript levels at 0 hr A.I. in (b) seem relatively higher compared to those at 0 hr A.I. in (c), but the former is in fact an overestimate resulting from the slower sampling method used during the long-term experiment (see Materials and Methods section) **d.** Protein levels during the 12-hour time course. **e.** Protein levels during the short time course. -, Negative control (WT 1h light); LC, Loading control by Coomassie blue staining.

showed, after an initial transcriptional peak, a decrease of transcript levels during the first hours after illumination, *dsCYC2* transcript accumulated to high levels in the CHX-treated cultures upon light exposure (Figure 3.3a). These results suggest that upon illumination a repressor is produced to ensure that *dsCYC2* is expressed during a narrow time window. Such a repressor could act through recognition and binding of a motif in the promoter region of *dsCYC2* or, alternatively, by acting on the stability of the transcript itself. To distinguish

between both possibilities we performed the CHX experiment using the prom-YFP reporter line and found that both the *dsCYC2* and *eYFP* transcript levels were higher in the CHX-treated cells compared to the control cells (Figure 3.3b and 3.3c, respectively), as in the previous experiments with WT cells (Figure 3.3a). These data indicate that the promoter activity of *dsCYC2* is affected when protein synthesis is blocked upon light exposure. In other words, upon illumination a repressor is produced that specifically targets the promoter activity of *dsCYC2*.

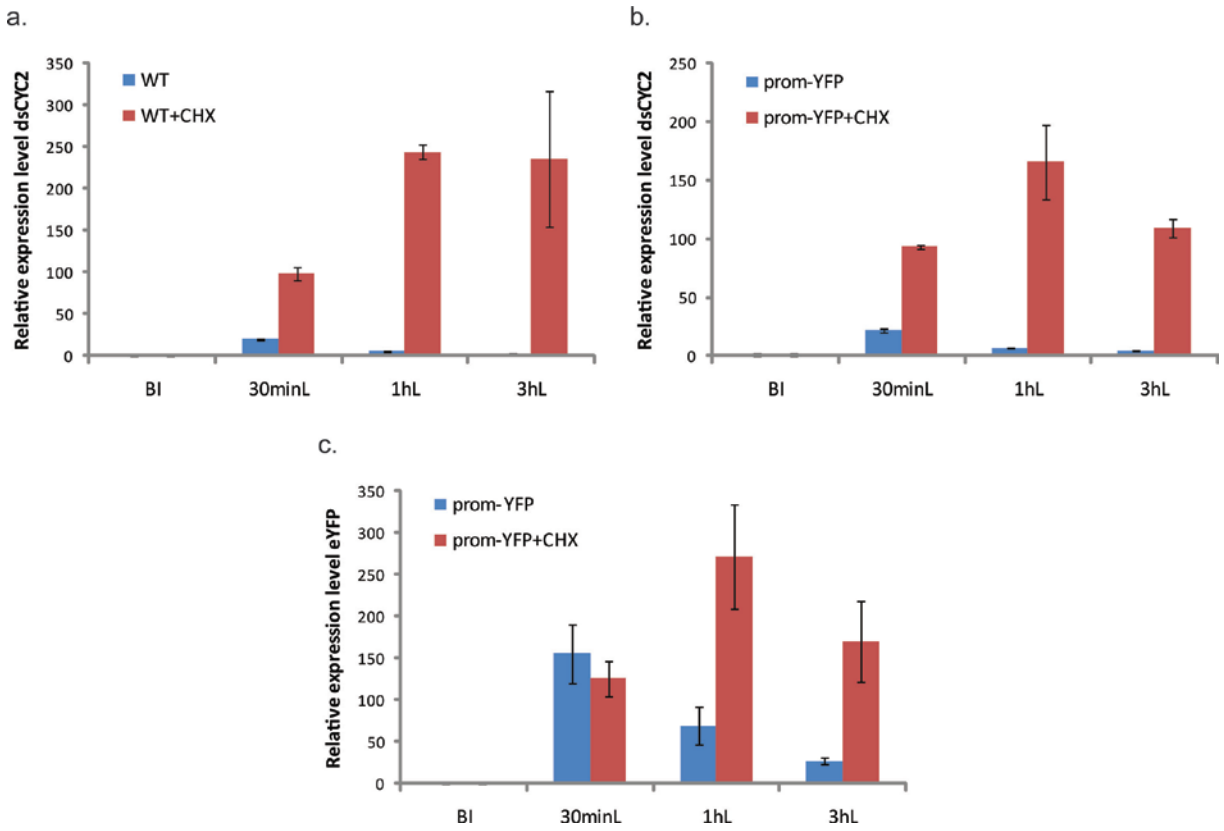


Figure 3.3: Protein synthesis is required for *dsCYC2* transcript downregulation during the light. Cultures were synchronized by 24 hour dark treatment (BI, before illumination)) and then exposed for 0.5 (30minL), 1 (1hL) and 3 hours (3hL) to light in the absence (blue) or the presence (red) of 2µg/ml cycloheximide (CHX). **a.** Relative expression levels of *dsCYC2* in WT cells. **b.** Relative expression levels of *dsCYC2* in prom-YFP cells. **c.** Relative expression levels of *eYFP* in prom-YFP cells. Values were normalized against the gene expression levels BI. Error bars represent standard errors of the mean of two independent experiments.

To determine whether this repressor remains active during the transfer from light to dark, *dsCYC2* transcript levels were monitored during normal (12L/12D) and constant light (12L/12L) conditions. To this end, WT cultures were grown exponentially in a 12L/12D

photoperiod and on the third day after medium refreshment, cultures were equally divided at the beginning of the normal dark period. One subculture was placed in darkness, while the other remained illuminated. When the dark period was replaced by a light period, cells show continuous low levels of *dsCYC2* transcript, in contrast to the cultures that were put in darkness, which showed slightly increasing levels of *dsCYC2* transcript (Figure 3.4). The latter is most likely due to an increased activity of the *dsCYC2* promoter during darkness, probably because of the absence of the repressor. By absence of the repressor in the dark, cells could be rendered sensitive for a new light signal by allowing the massive transcription and subsequent translation of *dsCYC2* shortly after light exposure.

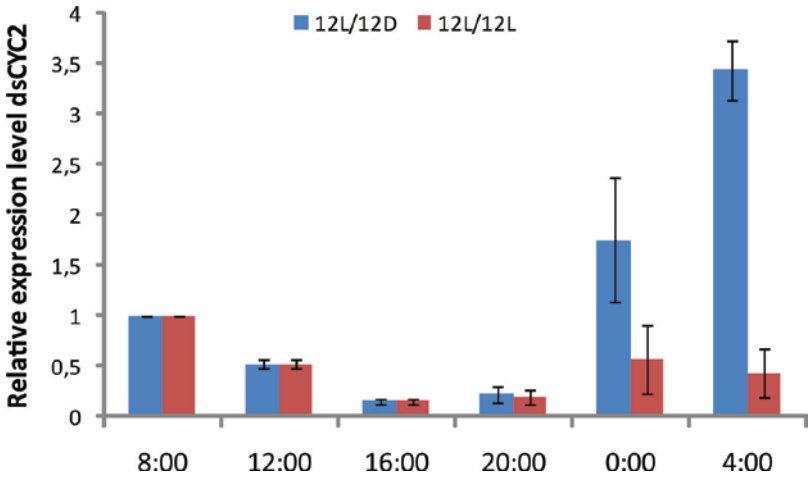


Figure 3.4: *dsCYC2* transcript levels are higher during darkness. Exponentially growing cultures entrained by a 12L/12D (lights on at 6:00/ lights off at 18:00) photoperiod were subdivided in two cultures at the end of the light period (at 18:00) three days after medium replenishment. Blue: Cells experiencing a normal dark period (lights off at 18:00). Red: Dark period replaced by light period (lights stay on at 18:00). Values were normalized against the gene expression levels at 8:00. Error bars represent standard errors of the mean of two biological replicates.

Changes in redox conditions affect dsCYC2 transcript and protein levels

As illustrated, *dsCYC2* transcript levels remain low during prolonged light periods, but slightly increase when cells are transferred to darkness. To investigate whether light-driven electron transport plays a role in the light-dependent regulation of *dsCYC2*, we tested the effect of 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU), during the light period (Figure 3.5). DCMU is a specific inhibitor of non-cyclic photosynthetic electron transport that blocks the plastoquinone (PQ) site of photosystem II and, as such, mimics dark conditions. In the

presence of DCMU, less reduction of *dsCYC2* transcipt levels was observed during the light period compared to the control cultures (Figure 3.5a). *dsCYC2* transcipt levels seem to be higher in DCMU-treated WT cells compared to HA marker cells, but this is probably due to experimental variation as both experiments were performed independently at a different time. Despite higher transcript levels, less dsCYC2 protein was produced in the DCMU-treated cells, especially apparent at 1 and 6 hour after illumination (Figure 3.5b).

Furthermore, preliminary results showed that cells treated during darkness with 1,4-dithiothreitol (DTT), a reducing agent, show a delayed accumulation of the transcript compared to non-treated cells (data not shown). Together, these data indicate that the redox state of the chloroplast can control the expression of the nuclear-encoded *dsCYC2* gene, probably through the regulation of the *dsCYC2* repressor protein.

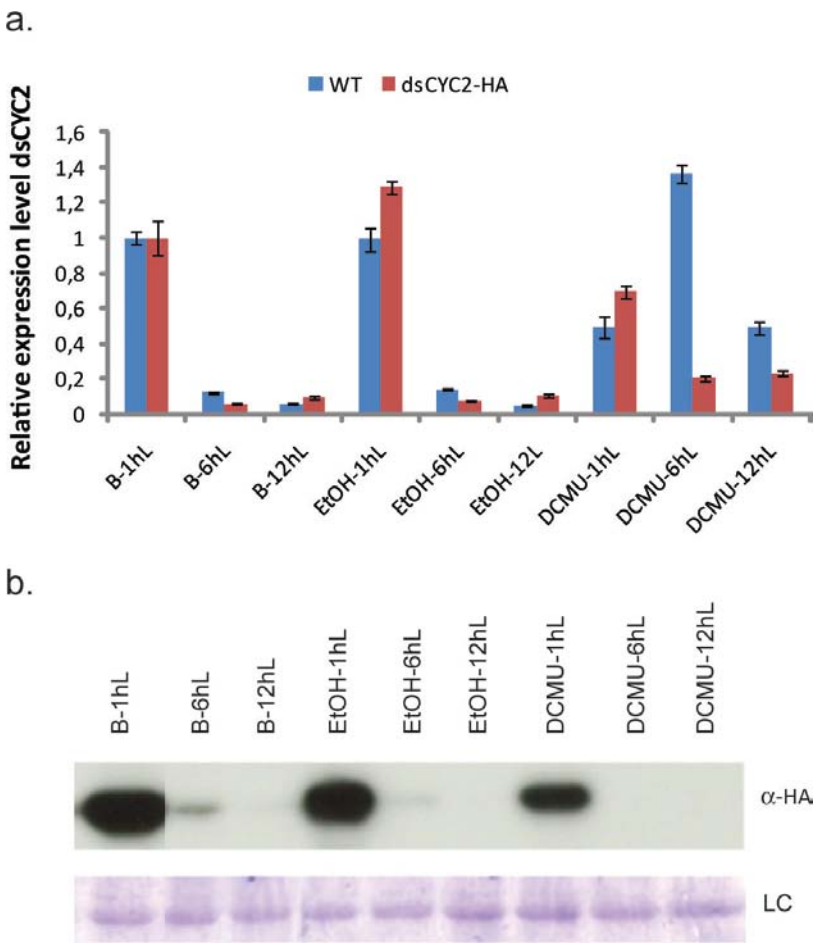


Figure 3.5. Effect of DCMU on *dsCYC2* transcript and protein levels. **a.** Transcript analysis of *dsCYC2* in WT and HA markerlines at 1h, 6h and 12h after illumination. Error bars represent standard errors. **b.** Western blot analysis of HA marker cells at 1h, 6h and 12h after illumination. B, Blanc control; EtOH, Solvent (ethanol) control; DCMU, Treatment with 20µM DCMU; LC, Loading control by Coomassie blue staining.

Interaction partners of dsCYC2

To identify candidate interactors of dsCYC2, we conducted a yeast-two-hybrid (Y2H) library screen. Y2H screening is a commonly used molecular technique to discover protein-protein interactions based on the reconstitution of the DNA binding and the activation domain of a transcription factor and hence the activation of a reporter gene when two proteins physically interact in the yeast system. (For more information see Fields and Song, 1989; Gietz et al., 1997 and http://en.wikipedia.org/wiki/Two-hybrid_screening).

A *P. tricornutum* Y2H prey library was constructed based on cDNA originating from mRNA pooled from different conditions. A screen using the full-length *dsCYC2* sequence as bait yielded no candidate interactors, probably because of low protein stability in the yeast. Therefore, a Y2H library screen using only the cyclin domain of *dsCYC2* as bait was performed. This screen yielded 20 candidate interactors, for which binding to dsCYC2 was confirmed through pairwise Y2H interaction assays. Nine proteins were identified to be strong (growth in the presence of ≥3mM 3AT) interactors in the yeast system, including transcription factors (two heat shock factors (HSFs), a MYB-related protein and a bZIP transcription factor), a phox-like protein, a helicase, a heavy metal binding/metal ion transport protein, an ariadne-like protein, and one unknown protein (Table 3.1).

Unlike expected, no CDK partner was found using the Y2H library screen. However, it is possible that such an interaction would be missed during the Y2H screen, because the CDK partner might be underrepresented in the cDNA library if it would be lower expressed compared to the candidate interactors identified during the screen. On the other hand, dsCYC2 might be working independently of CDK binding. To test if dsCYC2 can bind to the most classical and conserved type of CDKs, we performed a Y2H interaction assay in which *P. tricornutum* CDKA1, containing the PSTAIRE motif (Huysman et al., 2010), was used as bait and dsCYC2 as prey. Growth on His-lacking medium and a positive β-galactosidase test could be found for the combination of dsCYC2 with CDKA1, but not for any of the controls (Figure 3.6), indicating that both proteins might form a functional complex. This interaction of dsCYC2 with CDKA1 could be functionally relevant as *CDKA1* has previously been shown to be transcribed predominantly at the G1-S transition in synchronized cells (Huysman et al., 2010), coinciding with the peak of dsCYC2 protein levels (Figure 3.1e).

Table 3.1: Overview of the candidate dsCYC2 interactors identified by Y2H library screening.

Description	protein ID	#clones	interaction with dsCYC2FL
HSF (1)	**42824**	17	strong (10mM 3AT)
MYB	**37454**	5	strong (10mM 3AT)
phox-like	**45401**	4	strong (10mM 3AT)
Helicase	**43867**	2	strong (3mM 3AT)
heavy metal binding/metal ion transport	**43709**	1	strong (10mM 3AT)
HSF (2)	**49567**	1	strong (10mM 3AT)
chr10 (SAIP, similar to ankyrin repeat and IBR-domain-containing protein / ariadne)	**EST**	1	strong (3mM 3AT)
bZIP TF	**47686**	1	strong (10mM 3AT)
Unknown	**bd1360 (unmapped)**	1	strong (3mM 3AT)
HSP20A	**35158**	127	weak
HSP20	**54656**	20	weak
hypothetical protein	**48027**	11	weak
SWI/SNF	**45920**	2	weak
exonuclease	**19092**	1	weak
AP-1	**13511**	1	weak
chr23 (annexin-like?)	**EST**	1	weak
chr23 (annexin-like?)	**EST**	1	weak
alpha-tubulin	**54534**	1	weak
flavonol synthase?	**50451**	1	weak
no description	**1381**	1	weak

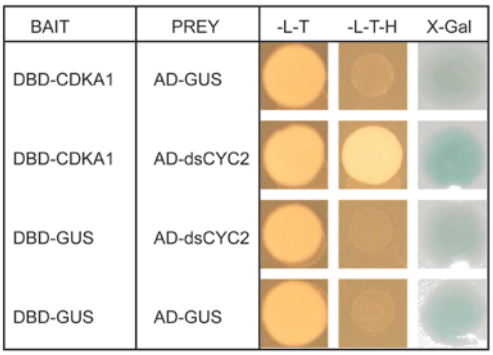


Figure 3.6: Interaction of dsCYC2 with CDKA1. Yeast PJ694-alpha cells were co-transformed with bait and prey plasmid as indicated. Co-transformation was analyzed on medium lacking leucine and tryptophan (-L-T). Co-transformants were tested for their ability to activate the histidine marker gene by assessing yeast growth on medium lacking leucin, tryptophan and histidin (-L-T-H) and for their ability to activate the *LacZ* reporter gene (X-Gal). Constructs containing GUS were used as negative controls. For each combination three independent colonies were screened, of which one is shown.

Overexpression of *dsCYC2* shortens G1 phase, but lengthens G2-M

To investigate the role of *dsCYC2* during the transition from dark to light, we generated overexpression lines by introducing a construct containing the full-length *dsCYC2* open reading frame fused N-terminally to *eYFP* under control of the *fucoxanthin chlorophyll binding protein B* (*fcpB*) promoter (Figure 3.7a). Overexpression of the fusion in different transgenic lines was evaluated by real-time qPCR, and two overexpressing lines (OE3 and OE14) were retained for further analysis (Figure 3.7b). We performed a growth analysis by absorbance measurements during eight consecutive days to examine whether *dsCYC2* overexpression could have an effect on cell cycle progression. To this end, cells were grown under continuous light conditions to ensure that overexpression of *dsCYC2* driven by the light-regulated *fcpB* promoter would be constant. The growth curve shows that both overexpression lines display no major differences in growth rate compared to WT cells, indicating that overexpression of *dsCYC2* would have no negative or positive effect on

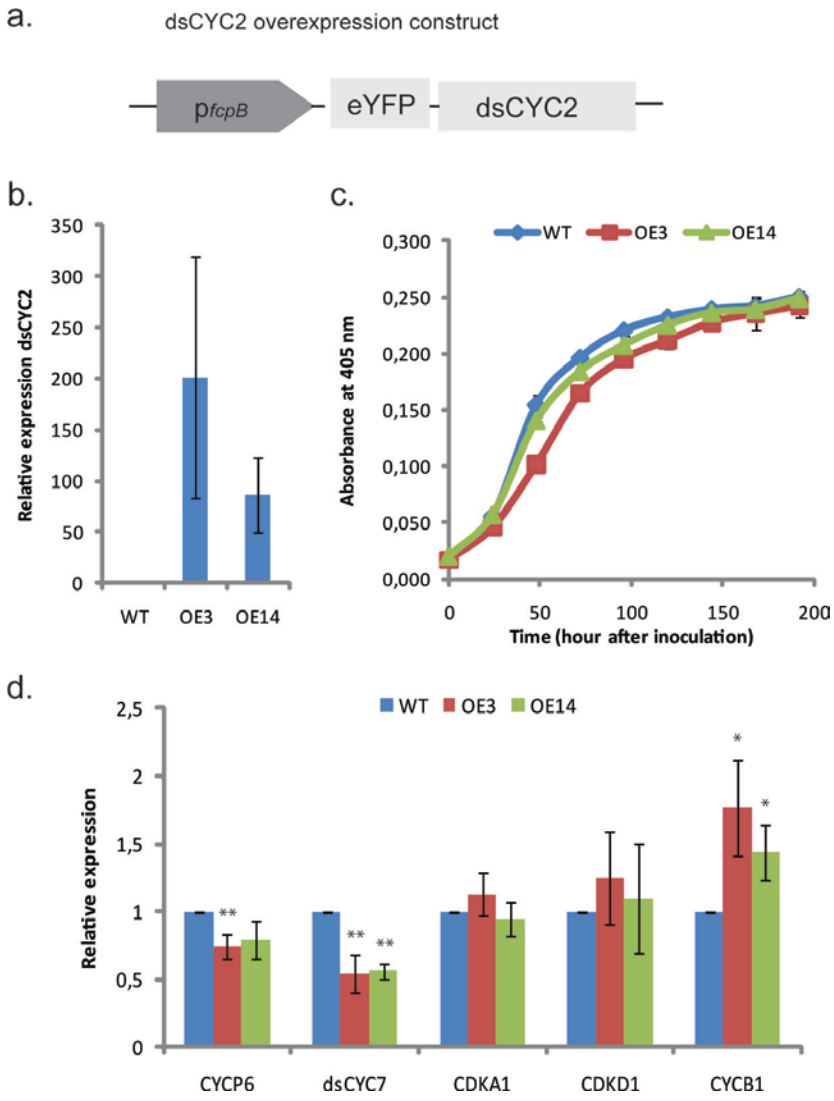
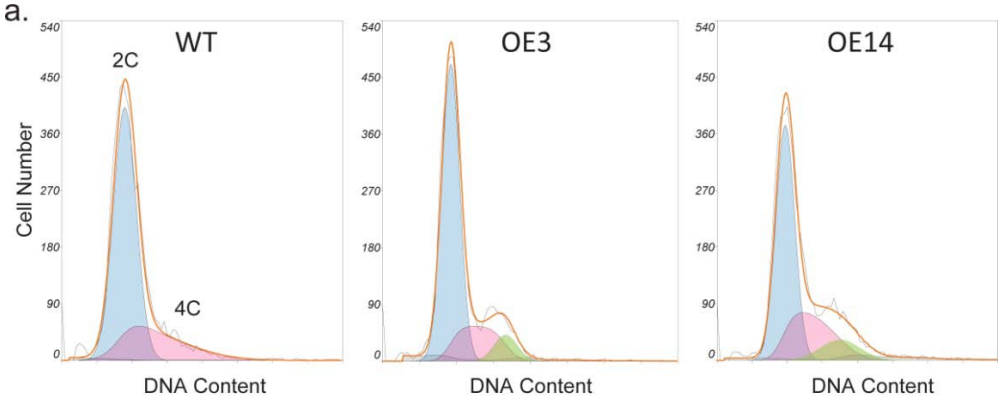


Figure 3.7: Effect of *dsCYC2* overexpression on cell cycle progression. **a.** Schematic representation of the dsCYC2 overexpression construct. **b.** Real-time qPCR analysis of *dsCYC2* transcript levels in WT, OE3 and OE14 lines. Expression values were normalized against the expression levels of WT cells. **c.** Growth curves of WT, OE3 and OE14 lines. **d.** Real-time qPCR analysis of different cell cycle marker genes in WT, OE3 and OE14 lines. * $p<0.1$; ** $p<0.05$ (One-tailed student's T-test). Error bars represent standard errors of the mean of three independent experiments.

overall cell cycle length (Figure 3.7c). However, determination of expression levels of different cell cycle marker genes (Huysman et al., 2010) during asynchronous growth, revealed slightly higher levels of the G2-M phase marker (*CYCB1*) and lower levels of both early G1 marker genes (*CYCP6* and *dsCYC7*) in the overexpression lines compared to the WT cells (Figure 3.7d). G1-S phase markers (*CDKA1* and *CDKD1*) were not differentially expressed in overexpression versus WT cells (Figure 3.7d). These data suggest that *dsCYC2* overexpressing cells proceed more rapidly through the G1 phase, and compensate this by spending a longer time in the G2-M phases. In agreement with this hypothesis, flow cytometric analysis of exponentially growing asynchronous cultures showed that *dsCYC2* overexpressing cells contain a small subpopulation of cells with a 4C DNA content (Figure 3.8a). Moreover, estimation of the cell cycle phase distribution in the histograms showed that overexpression cells contained a higher fraction of G2/M cells, but a reduced number of cells in G1 (Figure 3.8b). These data point to a role for *dsCYC2* as a positive regulator of G1 progression.

a.



b.

Strain	WT	OE3	OE14
%G1	**74**	**71**	**60**
%S	**26**	**21**	**30**
Intramodel 95% C.I. S phase	25-27	18-23	23-35
Intermodel 95% C.I. S phase	23-28	13-21	29-42
%G2-M	**1**	**8**	**11**
Intramodel 95% C.I. G2-M phase	0-1	7-9	8-14
Intermodel 95% C.I. G2-M phase	0-2	8-14	1-14

Figure 3.8: a. DNA distribution in WT cells versus OE lines (n=5000). Grey line indicates the raw histogram. Orange line shows the fitted curve obtained by applying the MultiCycle software for cell cycle analysis. G1, S and G2-M phase fittings are indicated in blue, pink and green, respectively. **b.** Cell cycle phase distribution estimates of WT and overexpression cells based on flow cytograms depcited in (a). 95 % confidence intervals (C.I.) for the S- and G2-M phase estimates are shown. Green and red numbers indicate high and low reliability, respectively, of the estimates by the intra- or intermodal error analyses (for more information see Materials and Methods section).

Silencing of *dsCYC2* slows down cell cycle progression by prolongation of the G1-phase

To study the effect of silencing of *dsCYC2* in *P. tricornutum*, we generated knock-down lines by introducing a hairpin construct under control of the *histone H4* promoter (De Riso et al., 2009) targeting the N-terminal part of *dsCYC2* (Figure 3.9a). Silencing was evaluated by comparing *dsCYC2* transcript levels at 15 minutes after illumination in WT cells and six independent mutant lines. Two lines, *dscyc2-2.4* and *dscyc2-2.8*, showed the same levels of *dsCYC2* transcript as the WT cells, while four other lines (*dscyc2-2.6*, *dscyc2-2.9*, *dscyc2-3.4* and *dscyc2-3.5*) showed 40 to 75% reduction of the transcript compared to WT cells (Figure 3.9b).

To test whether *dsCYC2* silencing had an effect on cell cycle progression, growth was monitored during 11 consecutive days by absorbance measurements on WT and mutant cells grown in a 12-hour light/ 12-hour dark (12L/12D) light regime. No effect could be observed for the non-silenced internal control lines (*dscyc2-2.4* and *dscyc2-2.8*) (Figure 3.9c). In contrast, all knock-down lines (*dscyc2-2.6*, *dscyc2-2.9*, *dscyc2-3.4* and *dscyc2-3.5*) showed a significant increase of their generation time compared to WT cells (Figure 3.9c), indicating that *dsCYC2* is crucial for proper cell cycle progression. In addition, we examined the effect of a 24-hour dark treatment on the strongest silenced line (*dscyc2-2.9*) and its cell cycle

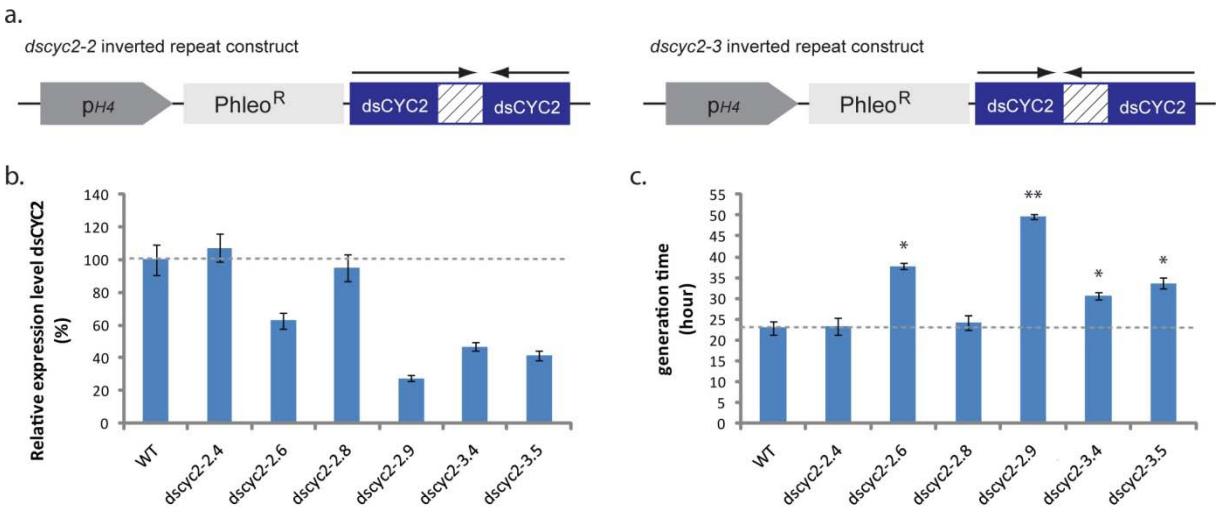
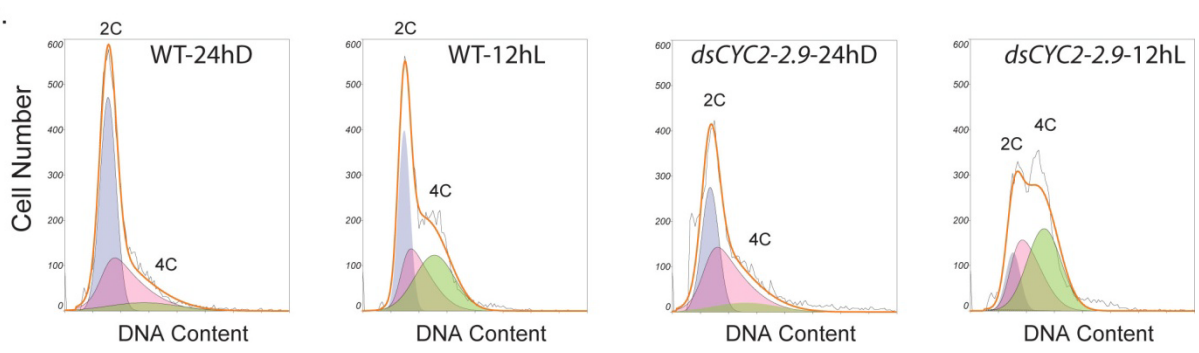


Figure 3.9: Effect of silencing of *dsCYC2* on cell cycle progression. **a.** Schematic representation of the *dsCYC2* inverted repeat constructs used for silencing analysis. In the *dscyc2-2* construct, the large fragment is positioned first and followed by the small fragment. In the *dscyc2-3* construct, the small fragment is followed by the large fragment (see arrows) **b.** Real-time qPCR analysis of *dsCYC2* transcript levels in WT and mutant lines. Cells were dark-adapted for 24 hours and transcript levels were measured 15 minutes after light exposure. Transcript levels of WT cells were set at 100% **c.** Generation times of WT and mutant lines. Error bars represent standard deviations of the mean of three independent experiments. * $p<0.005$; ** $p<0.001$ (two-tailed student's T-test).

progression after transfer to light by means of flow cytometry (Figure 3.10). In contrast to the WT cells, no predominant G1-arrest could be observed for the mutant cells after 24 hours of darkness. Instead, a broad shoulder was visible next to the 2C peak in the DNA histogram, indicative for S-phase cells (Figure 3.10a). Cell cycle analysis of the flow cytograms confirmed the S-phase identity of almost 50% of the dark-arrested silenced cells (Figure 3.10b). After 12 hours of illumination, WT and mutant cells showed both a 2C and 4C peak, but the 2C peak of WT cells was higher compared to mutant cells, probably because more WT cells already completed their first division cycle and entered the next one (Figure 3.10a). Cell cycle analysis of the flow cytograms showed that most of the silenced cells were situated in the S and G2-M phases, while the majority of the WT cells had a G2-M or G1-phase identity after 12 hour of light exposure (Figure 3.10b).

a.



b.

Strain	WT 24hD	WT 12hL	*dscyc2-2.9* 24hD	*dscyc2-2.9* 12hL
%G1	**56**	**35**	**40**	**14**
%S	**35**	**29**	**49**	**37**
Intramodel 95% C.I. S phase	34-37	26-33	44-55	33-40
Intermodel 95% C.I. S phase	35-45	19-53	47-52	37-44
%G2-M	**9**	**36**	**11**	**49**
Intramodel 95% C.I. G2-M phase	9-10	35-38	9-13	47-50
Intermodel 95% C.I. G2-M phase	1-9	4-36	8-14	42-49

Figure 3.10: DNA distribution in WT versus *dscyc2-2.9* cells (n=10000). Grey line indicates the raw histogram. Orange line shows the fitted curve obtained by applying the MultiCycle software for cell cycle analysis. G1, S and G2-M phase fittings are indicated in blue, pink and green, respectively. **b.** Cell cycle phase distribution estimates of WT and *dscyc2-2.9* cells based on flow cytograms depcited in (a). 95 % confidence intervals (C.I.) for the S- and G2-M phase estimates are shown. Green and red numbers indicate high and low reliability, respectively, of the estimates by the intra- or intermodal error analyses (for more information see Materials and Methods section). 24hD: cells incubated in darkness for 24 hours; 12hL: cells incubated in darkness for 24 hours and then illuminated for 12 hours.

Transcriptional analysis of several cell cycle marker genes during light-dependent cell cycle re-entry of 24-hour dark-arrested WT and *dscyc2-2.9* cells provided more evidence that silencing of *dsCYC2* results in the attenuation of the G1-S phase transition (Figure 3.11). Transcript levels of the G1-phase marker gene *dsCYC7* were reduced in mutant versus WT cells (Figure 3.11), in particular at the moment that *dsCYC2* shows normally its transcriptional peak, indicating that *dsCYC7* might be a downstream target of *dsCYC2*. Whereas in WT cells, the late G1 marker gene *CYCH1* shows an initial downregulation of its transcript after light exposure (Figure 3.11), to have a second and maximal expression peak at 2 hours after illumination in synchronized cells (Huysman et al., 2010), *CYCH1* transcript levels in mutant cells are only downregulated after 30 to 45 minutes of illumination, and a transcriptional peak is apparent only at 6h after reillumination. Moreover, the G1-S phase marker *CDKD1* was overall higher expressed in mutant versus WT cells, indicative for a higher proportion of cells at the G1-S transition in the mutant line. Surprisingly, *CYCB1* levels were also generally higher in mutant cells, although the flow cytometry data of dark-arrested cells did not show higher accumulation of 4C cells compared to WT cells (Figure 3.10). On the other hand, transcript levels of *hCDK1*, another M-phase marker showing the same cell cycle-dependent expression pattern as *CYCB1* (Huysman et al., 2010), were not markedly affected by *dsCYC2* silencing (Figure 3.11). Unscheduled *CYCB1* expression has been shown to be induced in G1-phase cancer cell lines and tissues (Shen et al., 2004), and it has been reported that *cyclin B1* possesses cryptic S-phase-promoting abilities (Moore et al., 2003). Moreover, in fission yeast *cyclin B* can regulate both the S- and M-phase progression in the absence of G1-cyclins (Fisher and Nurse, 1996). Together, these data point to an activating role for *dsCYC2* in cell cycle progression, probably through regulating cell cycle initiation after dark arrest.

DISCUSSION

In this study we showed that the expression of *dsCYC2* is regulated strictly by light. Upon illumination, transcript levels increase within 10 to 15 minutes, where after protein synthesis starts and transcript levels decline to stay low during the remaining light period. We have shown that the inhibition of protein synthesis at the dark-to-light transition delays the decrease of *dsCYC2* transcript levels in the light and results in the accumulation of higher

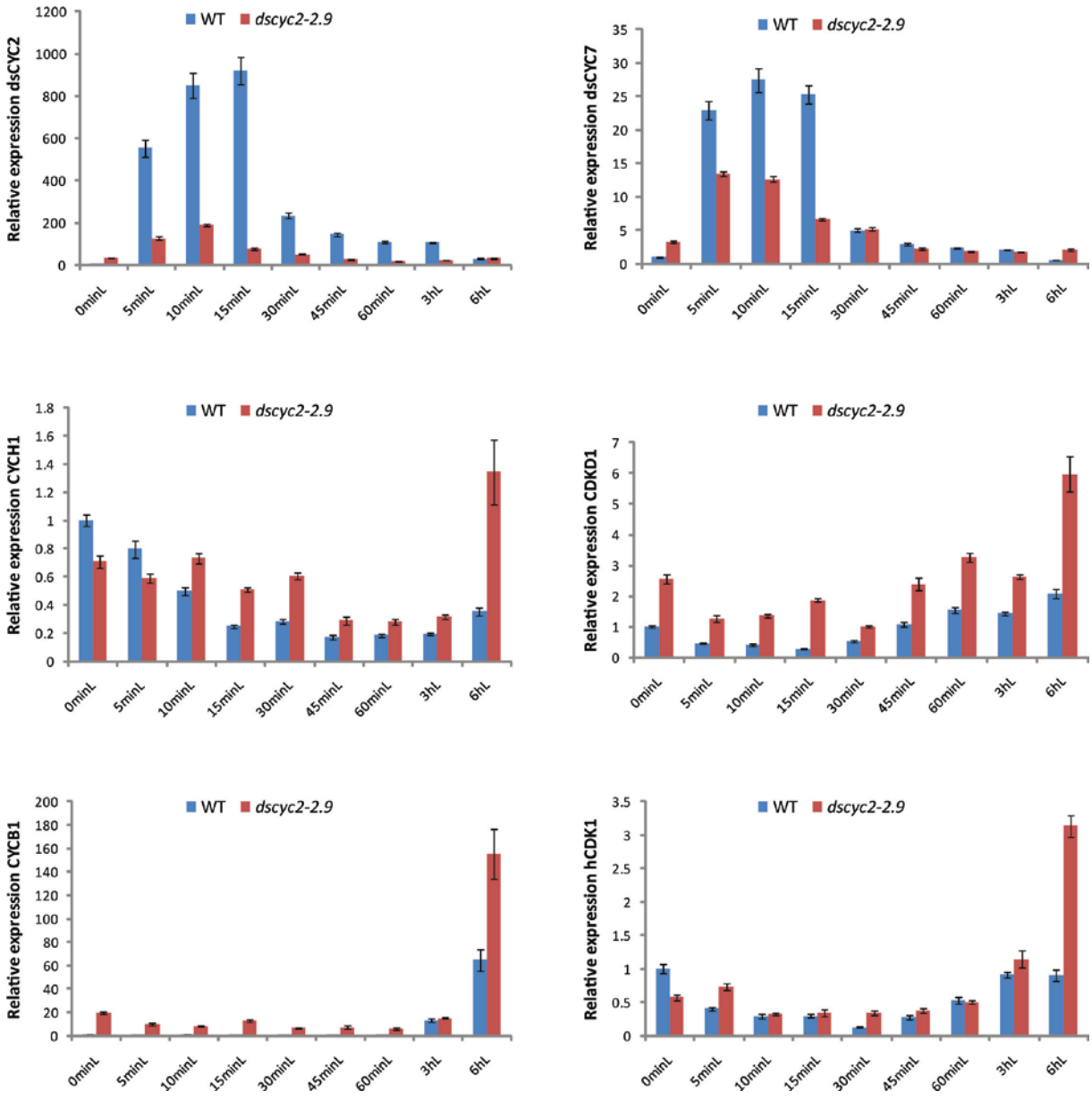


Figure 3.11: Real-time qPCR analysis of different cell cycle marker genes in WT and *dscyc2-2.9* cells. Cells were exponentially grown in a 12-hour light/12-hour dark photoperiod and dark-incubated for 24 hour at the third day after medium refreshment. Samples were taken at different timepoints after illumination (from 0 minutes (0minL) to 6 hour (6hL) after illumination). *dsCYC7* and *CYCH1* are G1-phase marker genes, *CDKD1* is an S-phase marker gene and *CYCB1* and *hCDK1* are G2-M phase marker genes (See chapter 2). Error bars represent standard errors.

transcript levels after illumination, suggesting that upon light a repressor of *dsCYC2* transcription is produced. Moreover, the CHX experiment on the prom-YFP reporter line indicated that this repressor would act on the promoter of *dsCYC2*. Examination of the

promoter region of *dsCYC2* for *cis*-regulatory elements using the signal scan database PLACE (Higo et al., 1999) (data not shown) revealed the presence of several light-regulated elements (LREs), including GATA- or I-boxes, GT-1 consensus sites (Terzaghi and Cashmore, 1995; Arguello-Astorga and Herrera-Estrella, 1998) and SORLIP sequences (Hudson and Quail, 2003). Next to these specific light-regulated elements, also several recognition sites for MYB transcription factors were detected that have been reported to be involved in transcriptional regulation of light-harvesting proteins in higher plants (Wang et al., 1997).

The mechanism(s) that controls the transduction of the light signal to the nucleus of the cell, the compartment where *dsCYC2* transcription should be initiated, still remains to be elucidated. However our results suggest that a plausible mechanism to relay the light signal might occur through the change of redox state of the chloroplast. Cultures treated during the light period with DCMU, an oxidative agent specifically targeting the PQ pool of photosystem II, show relatively higher levels of *dsCYC2* transcript, but lower protein levels during the light period compared to non-treated cells. Cells treated during darkness with DTT, a reducing agent, show a delayed accumulation of the transcript against non-treated cells. Similar redox control of transcript abundance has been reported for nuclear-encoded chloroplast genes in the green alga *Chlamydomonas reinhardtii* (Salvador and Klein, 1999) and for the *cab* gene regulation of *Dunaliella tertiolecta* (Escoubas et al., 1995).

Based on the results presented here, we propose a schematic model for *dsCYC2* regulation in *P. tricornutum* (Figure 3.12). Light triggers the activation of electron transport in the chloroplast compartment of the cell and this signal is transduced to the nucleus through an unknown signaling mechanism. It is possible that this signal-transduction occurs through a phosphorylation cascade, as already suggested for *cab* gene expression in *D. tertiolecta* (Escoubas et al., 1995), but to proof this in *P. tricornutum*, more experimental work will be required using specific phosphorylation inhibitors. This unknown signaling pathway results in the activation of a transcriptional activator (TA), which recognizes and activates a light-responsive element (LRE) in the promoter region of *dsCYC2*. Shortly after activation, a transcriptional repressor (R) is produced that either recognizes and inhibits the LRE or interferes with the activity of the TA. This repressor could be an early transcription factor. The opposite response of transcript and protein levels during the first hour after illumination, and the relatively lower *dsCYC2* protein, but higher transcript levels in DCMU-treated cultures, suggests that a negative autoregulatory feedback loop could be present, in which

dsCYC2 would act as its own transcriptional repressor or as an activator of a yet undefined repressor. The latter could possibly be one of the TFs identified in our interaction screen (Table 3.1). Such a feedback mechanism would ensure that *dsCYC2* is only expressed at a defined time window. The limited timing of *dsCYC2* expression is likely to be extremely important, as both overexpression and silencing of *dsCYC2* interfere with normal cell cycle progression. Most probably it contributes to resetting the cells to become sensitive for a new dark period.

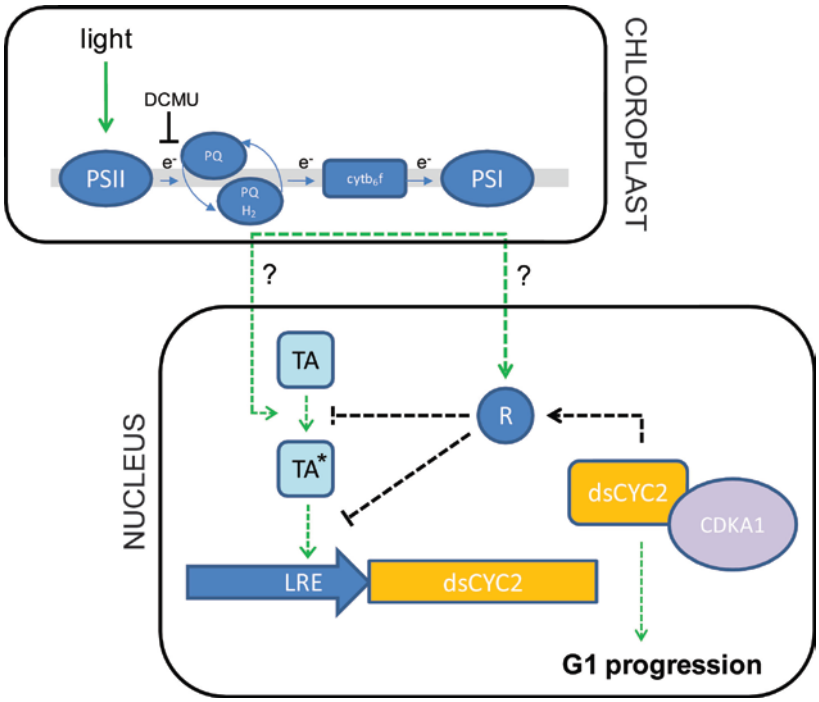


Figure 3.12: Schematic model of the light-dependent regulation of *dsCYC2* in *P. tricornutum*. For detailed description see text. Green and black lines indicate positive and negative regulation, respectively. LRE, light-responsive element; PSI, photosystem I; PSII, photosystem II; PQ, plastoquinone, R, repressor protein; TA, transcriptional activator in non-active state; TA*, transcriptional activator in active state.

The negative effect of *dsCYC2* silencing on the culture growth rate clearly indicates that *dsCYC2* is required for proper cell cycle progression after dark arrest. Our data suggest that lower *dsCYC2* levels at light exposure result in a prolongation of the G1-to-S-phase transition, as shown by an altered DNA distribution pattern in dark-arrested *dscyc2-2.9* cells, and by the delayed transcription of the late G1 marker *CYCH1* and the higher transcript levels of the S-phase marker *CDKD1* during illumination. In analogy with a putative role for *dsCYC2* to induce the cell cycle after dark arrest, cell cycle re-entry upon exogenous stimulatory signals in yeast, animals and plants is mediated by specific G1 cyclins. In yeast,

appropriate growth and nutrient status of the cell triggers G1 progression through activation of first Cln3 and later Cln1 and Cln2 in association with Cdc28 to finally activate the G1/S transcription factor SBF/MBF and transcription of S-phase genes (Tyers et al., 1993; Reviewed in Mendenhall and Hodge, 1998). In animals and plants, D-type cyclins are stimulated by serum growth factors and hormones or sucrose, respectively. D-type cyclins associate with CDKs and phosphorylate the Rb protein, leading to the release and activation of E2F transcription factors and G1-S phase transition (Reviewed in Oakenfull et al., 2002). Overexpression or silencing of these G1 cyclins has been reported to induce various effects on G1 phase duration and overall cell cycle length, dependent on the type of cyclin and the type of cells (Quelle et al., 1993; Resnitzky et al., 1994; Sherr, 1995; Menges et al., 2006). Both Cln1-3 and D-type cyclins are characterized by PEST sequences that render the proteins unstable and confer rapid turnover (Rechsteiner and Rogers, 1996; Renaudin et al., 1996; Mendenhall and Hodge, 1998). Furthermore, plant and animal D-type cyclins also possess an LxCxE motif at their N-terminal part that is responsible for the interaction with Rb (Dowdy et al., 1993; Renaudin et al., 1996). However, none of these motifs could be recognized in the dsCYC2 sequence (data not shown), suggesting that dsCYC2 turnover is regulated by alternative mechanisms and that the protein probably does not directly interacts with the Rb protein. Alternatively, it is possible that *dsCYC2* expression results in the transcription or activation of other G1 cyclins that regulate the Rb protein in *P. tricornutum*. Because diatom cell cycle progression depends not only on light, but also on other environmental factors, such as nutrient availability, it is to be expected that multiple cyclins are involved in G1 control, representing a complex integrative fine-tuning network of different signaling pathways.

To our knowledge this is the first report of a cyclin gene involved in cell cycle progression whose transcription and translation is strictly controlled by light. The presence of such a critical molecule that coordinates the activation of the cell cycle machinery upon changing light conditions is of major importance for diatoms living in highly variable environments and allows them to pace their cell division rate to the prevailing light conditions. Interestingly, *dsCYC2* transcription was recently shown to be higher induced in *P. tricornutum* cells overexpressing a blue light sensor CPF1 upon blue light treatment compared to WT cells (Coesel et al., 2009), suggesting that *dsCYC2* might be a downstream target of the CPF1 signaling cascade. At least 40 genes have been identified that belong to the light-harvesting complex (LHC) superfamily in *P. tricornutum* and some of them have been

characterized more in detail (Guglielmi et al., 2005; Siaut et al., 2007; Bowler et al., 2008; Coesel et al., 2009; Nymark et al., 2009; Bailleul et al., 2010). It would be interesting to see whether *dsCYC2* expression can be influenced by one of these antenna proteins. In addition, the effect of different combinations of light quality (white, blue and red) at different fluences on *dsCYC2* expression could be of great interest as well, as various spectral compositions have been reported to influence growth rates of several diatom species (Holdsworth, 1985; Mercado et al., 2004).

To understand more in detail the signaling cascades involved in the regulation of this gene, future experiments should aim to identify the light-responsive element in the *dsCYC2* promoter sequence and to functionally characterize the transcriptional activator and repressor acting on this promoter sequence. Although *dsCYC2* is clearly an important cell cycle regulator in *P. tricornutum*, the question remains to what extend this mechanism and its regulators are conserved among different diatom species.

MATERIALS AND METHODS

Diatom culture conditions

Phaeodactylum tricornutum (Pt1 8.6; accession numbers CCAP 1055/1 and CCMP2561) was grown in f/2 medium without silica (f/2-Si) (Guillard, 1975) made with filtered and autoclaved sea water collected from the North Sea (Belgium). Cultures were cultivated at 18°C-20°C in a 12-h light/12-h dark regime at 70-100 μmol photons m^{-2} s^{-1}. Liquid cultures were shaken at 100 rpm. For biolistic transformation, *P. tricornutum* cells were grown on solid f/2-Si medium containing 1% Select agar (Sigma).

Cloning constructs and biolistic transformation

The 1018 bp *dsCYC2* promoter sequence alone, the promoter sequence and the full length *dsCYC2* gene sequence, or the *dsCYC2* gene sequence of *P. tricornutum* alone was amplified with gene-specific primers (Table 3.2), cloned in the pDONR221 or pENTR-D-TOPO vector (Invitrogen) and subsequently recombined in a *P. tricornutum* destination vector (pDEST) by attL x attR recombination (Invitrogen) (Siaut et al., 2007). The *dsCYC2* promoter

sequence was recombined in the pDEST-C-EYFP for C-terminal fusion, to construct the prom-YFP reporter line. The promoter and gene construct was recombined in the pDEST-C-HA to construct the HA marker line. Both plasmids were subsequently digested with SacII and NotI (Promega) to remove the *fcpB* promoter sequence. The digested product was treated with T4 DNA polymerase in the presence of 10 mM dNTPs to produce blunt ends and then ligated using T4 DNA ligase according to the manufacturer's instructions (Promega). For the overexpression construct, the full sequence of *dsCYC2* was recombined in the pDEST-N-EYFP, resulting in an N-terminal fusion with the eYFP fluorescent marker (Siaut et al., 2007).

For the creation of *dsCYC2* inverted repeat constructs for silencing, a 167 bp fragment (corresponding to the *dsCYC2* gene sequence from 13 bp to 179 bp) and a 301 bp fragment (corresponding to the gene sequence from 13 to 313 bp) were amplified from the *dsCYC2* cDNA, respectively, with the primers *dsCYC2f1*_Fw (containing a *Eco*RI site) and *dsCYC2f1*_Rv (containing a *Xba*I site), and *dsCYC2f1*_Fw and *dsCYC2f2*_Rv (containing a *Xba*I site) (Table 3.2). The fragments were digested with *EcoR*I and *Xba*I (Promega) and ligated in sense and antisense orientations in the *Eco*RI site of the linearized hir-PtGUS vector (De Riso et al., 2009).

Table 3.2: Cloning primers.

Primer name	Sequence	Target
dsCYC2-CYCdomain_Fw	AAAAAAGCAGGCTTCCGACAGCAAATGTTCGACTGG	cyclin domain
dsCYC2-CYCdomain_Rv	GGGTCACCGCCTCCGGATCACAAGTGCCAGGAAAGAGCAGT	cyclin domain
dsCYC2_Fw	CACCATGATGAAACAGCAACATTCCC	gene
dsCYC2-N_Rv	TCAATTTCCATACAAAGCTTCCA	gene
dsCYC2-C_Rv	ATTTCCATACAAAGCTTCCAGTTG	gene
CDKA1_Fw	CACCATGGAGCGGTATCAAAGGAT	gene
CDKA1-N_Rv	TTACGCCGCCTGAAAGTAT	gene
pdsCYC2-attB1_Fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTTAGAATATGTGGGGTTTCGT	promoter (and gene)
pdsCYC2-attB2_Rv	GGGGACCACTTTGTACAAGAAAGCTGGGTCGACGTTATACTTAATCAGTATTTCAGG	promoter
dsCYC2-attB2_Rv	AGAAAGCTGGGTCATTTCCATACAAAGCTTCCAG	promoter and gene
*dsCYC2f1*_Fw	ACTGAATTCCAACATTCCCAGCAAGAA	hairpin construct
*dsCYC2f1*_Rv	ACTTCTAGATTGTTAGTGATACTGGCGC	hairpin construct
*dsCYC2f2*_Rv	ACTTCTAGAAAACTGGGGACGCTGTAG	hairpin construct

Constructs were introduced into *P. tricornutum* by microparticle bombardment as previously described (Falciatore et al., 1999). The pAF6 plasmid was used to confer the resistance to phleomycin (Falciatore et al., 1999). Individual phleomycin-resistant colonies were both restreaked on f/2-Si agar plates and grown in liquid f/2-Si medium without antibiotics for further analysis.

Inhibitor studies

To determine the effect of inhibition of protein translation on *dsCYC2* transcription, cells were treated with or without cycloheximide (CHX (Duchefa Biochemie), final concentration of 2µg/ml) 5 minutes before the onset of light.

To determine the effect of the photosynthetic electron transport inhibition, DCMU was dissolved in ethanol (EtOH) and delivered to the cells at a final concentration of 20µM 10 minutes before onset of light. Identical volumes of EtOH were added to the solvent controls, and had no effect on transcript expression.

Real-time quantitative PCR

For RNA extraction, 5 x 10^7 cells were collected either by centrifugation (long-term response) or by fast filtration (short-term response, CHX and DCMU experiments), and cell pellets were fast frozen in liquid nitrogen and stored at -70ºC. Cell lysis and RNA extraction was performed using TriReagent (Molecular Research Center, Inc., Cincinnati, OH, USA) according to the manufacturer's instructions. Contaminating genomic DNA was removed by DNaseI treatment (GE Healthcare, Little Chalfont, UK) and RNA was purified by amoniumacetate precipitation. To assess RNA concentration and purity, spectrophotometry was performed (NaNodrop ND-1000, Wilmington, DE). Total RNA was reverse transcribed using iScript reverse transcriptase (Roche) according to the manufacturer's instructions. Finally, an equivalent of 10 ng of reverse transcribed RNA (cDNA) was used as template in each qPCR reaction.

Samples in triplicate were amplified on the Lightcycler 480 platform with the Lightcycler 480 SYBR Green I Master mix (Roche Applied Science), in the presence of 0.5 µM gene-specific primers (dsCYC2_Fw: CTATCATCGCACTCGTCATCAAC and dsCYC2_Rv: TGTCCACCAAAGCCTCCAAAC, dsCYC2-HA_Fw: TCGCTCCTCTGGTGGAA and dsCYC2-HA_Rv: GTCGTAGGGGTAGGCGTAGT, other primer sequences see Huysman et al., 2010 and Siaut et al., 2007). The cycling conditions were 10 min polymerase activation at

95°C and 45 cycles at 95°C for 10 s, 58°C for 15 s and 72°C for 15 s. Amplicon dissociation curves were recorded after cycle 45 by heating from 65°C to 95°C. Data were analyzed using the ΔC_t relative quantification method using qBase (Hellemans et al., 2007), with the stably expressed *histone H4* (Figure 3.1, 3.2, 3.3, 3.4 and 3.5) or *TBP* and *UBI-4* (Figure 3.7, 3.9 and 3.11) as a normalization gene (Siaut et al., 2007).

Western blot analysis

Proteins were extracted by adding 200 µl Laemli buffer to 50 x 10^6 frozen cells and the cell lysates were incubated during 15 minutes on ice. Protein concentrations were determined by the Bradford assay (Bio-Rad Protein Assay, Bio-Rad). Equal amounts of protein extracts were resolved on 12% SDS-PAGE gels, and transferred to nitrocellulose membranes (Millipore) using the wet-blot method. The dsCYC2-HA fusion protein was detected by incubating proteins transferred to nitrocellulose membranes for 1h with a 1:500 dilution of anti-HA primary antibody at room temperature, followed by 1h incubation in a 1:10,000 dilution of horseradish peroxidase (HRP) anti-rat secondary antibody at room temperature. eYFP protein of the prom-YFP cells was detected by incubating proteins transferred to nitrocellulose membranes for 1h with a 1:3,000 dilution of anti-GFP primary monoclonal antibody (Rockland) at room temperature, followed by 1h incubation in a 1:10,000 dilution of horseradish peroxidase (HRP) anti-mouse secondary antibody at room temperature. eYFP proteins originating from the OE cells were detected by incubating proteins transferred to nitrocellulose membranes for 1h with a 1:1,000 dilution of anti-GFP primary polyclonal antibody (Rockland) at room temperature, followed by 1h incubation in a 1:10,000 dilution of horseradish peroxidase (HRP) anti-goat secondary antibody at room temperature. Signals were visualized using the Western Lightning™ detection kit (Pierce).

Yeast-two-hybrid analysis

Yeast-two-hybrid bait and prey plasmids were generated through recombinational GATEWAY cloning (Invitrogen). A 350 bp fragment encompassing the cyclin domain of *dsCYC2*, and the full length open reading frames of the *P. tricornutum dsCYC2* and *CDKA1* genes were amplified with gene-specific primers (Table 3.2), cloned in the pENTR-D-TOPO vector (Invitrogen) and subsequently recombined in the pDEST22 and pDEST32 vectors (Invitrogen) by attL x attR recombination, resulting in translational fusions between the proteins and the GAL4 transcriptional activators and GAL4 DNA-binding domains, respectively. For the

Y2H library screens, the yeast strain PJ694-alpha (MATa; trp1-901, leu2-3,112, ura3-52, his3-200, gal4D, gal80D, LYS2::GAL1-HIS3, GAL2-ADE2, met2GAL7-lacZ) was first transformed with the bait plasmid and transformed cells were selected on synthetic dextrose (SD) medium lacking leucin (Leu) by the LiAc method (Gietz et al., 1992). Subsequently the yeast containing the bait plasmid was transformed with 50µg of prey plasmids derived from a custom-made *P. tricornutum* Y2H library (Invitrogen) and yeast cells that hosted a successful interaction were selected on SD medium lacking Leu, tryptophan (Trp) and histidine (His). To detect strong interaction, positive colonies were restreaked on selective medium containing different concentrations of 3AT. For the co-transformation assay, plasmids encoding the bait and prey constructs were co-transformed in the yeast strain PJ694-alpha by the LiAc method (Gietz et al., 1992). Co-transformed yeast cells were selected on SD plates lacking Leu and Trp. Interaction between the introduced proteins was scored on SD plates lacking Leu, Trp and His, and by the LacZ test, as described by Boruc et al. (2010).

Growth analysis

To monitor growth, cells were grown at constant illumination (overexpression lines) or 12L/12D (silencing lines) in a 24-well plate (Falcon), in a total volume of 1 ml, over a time period of eleven days. Absorbances of the cultures were measured at 405 nm using the VICTOR^3 Multilabel Plate Reader (Perkin-Elmer) each day in the morning. Obtained growth curves of triplicate cultures were LN(2)-transformed and mean generation times were calculated by determination of the derivative of the values between the points of maximal slope (exponential growth phase).

Flow cytometric analysis

Flow cytometric analysis was performed on a Partec CyFlow ML platform (with data acquisition software Flomax; Partec GmbH, Münster, Germany) on cells fixed with 70% ethanol, washed three times with 1× phosphate buffered saline and stained with 4',6-diamidino-2-phenylindole (DAPI) (final concentration of 1 ng/ml). For each sample, 5,000 to 10,000 cells were processed. Flow cytograms were formatted and analyzed with Multicycle AV for Windows (Phoenix Flow Systems, San Diego, California, USA) software to determine relative representation estimates of the different cell cycle stages in the samples. Estimates were obtained by fitting the histograms (nonlinear least-square fitting) with the default parameters of background modeling and zero order S phase. In addition the software automatically fits five

more models. For each of these fitting models a 95% intramodel confidence range is provided based on successive analyses of the same histogram and this interval gives an confidence estimation of finding the same “best” fit. To address whether the model used is appropriate for the histogram, 95 % intermodel confidence intervals are provided, that indicate to wich extent the cell cycle parameters vary depending on the model chosen. In general if either confidence range is wide, the parameter is more likely to be approximate and subject to increased error. It should be noted that the wide intermodal confidence intervals observed in our analyses (marked in red in Figure 3.8b and 3.10b) could all be attributed to unrealistic fitting of only one of the models, while the other models gave similar cell cycle phase estimates.

REFERENCES

Arguello-Astorga G, Herrera-Estrella L (1998) Evolution of Light-Regulated Plant Promoters. Annual Review of Plant Physiology and Plant Molecular Biology **49:** 525-555

Bailleul B, Rogato A, de Martino A, Coesel S, Cardol P, Bowler C, Falciatore A, Finazzi G (2010) An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light. Proceedings of the National Academy of Sciences of the United States of America **42:** 18214-18219

Boruc J, Van den Daele H, Hollunder J, Rombauts S, Mylle E, Hilson P, Inzé D, De Veylder L, Russinova E (2010) Functional modules in the Arabidopsis core cell cycle binary protein-protein interaction network. Plant Cell **22:** 1264-1280

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kroger N, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin-Jezequel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Oudot-Le Secq MP, Napoli C, Obornik M, Parker MS, Petit JL, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature **456:** 239-244

Brzezinski MA, Olson RJ, Chisholm SW (1990) Silicon availability and cell-cycle progression in marine diatoms. Marine Ecology-Progress Series **67:** 83-96

Chen M, Chory J, Fankhauser C (2004) Light signal transduction in higher plants. Annual Review of Genetics **38:** 87-117

Coesel S, Mangogna M, Ishikawa T, Heijde M, Rogato A, Finazzi G, Todo T, Bowler C, Falciatore A (2009) Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. EMBO Rep **10:** 655-661

De Riso V, Raniello R, Maumus F, Rogato A, Bowler C, Falciatore A (2009) Gene silencing in the marine diatom Phaeodactylum tricornutum. Nucleic Acids Research **37:** e96

Dowdy SF, Hinds PW, Louie K, Reed SI, Arnold A, Weinberg RA (1993) Physical Interaction of the Retinoblastoma Protein with Human D-Cyclins. Cell **73:** 499-511

Escoubas JM, Lomas M, LaRoche J, Falkowski PG (1995) Light intensity regulation of cab gene transcription is signaled by the redox state of the plastoquinone pool. Proceedings of the National Academy of Sciences of the United States of America **92:** 10237-10241

Falciatore A, Casotti R, Leblanc C, Abrescia C, Bowler C (1999) Transformation of Nonselectable Reporter Genes in Marine Diatoms. Mar Biotechnol (NY) **1:** 239-251

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. Science **281:** 237-240

Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. Nature **340:** 245-246

Fisher DL, Nurse P (1996) A single fission yeast mitotic cyclin B p34cdc2 kinase promotes both S-phase and mitosis in the absence of G1 cyclins. The EMBO journal **15:** 850-860

Gietz D, St Jean A, Woods RA, Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res **20:** 1425

Gietz RD, Triggs-Raine B, Robbins A, Graham KC, Woods RA (1997) Identification of proteins that interact with a protein of interest: applications of the yeast two-hybrid system. Molecular and cellular biochemistry **172:** 67-79

Gillard J, Devos V, Huysman MJJ, De Veylder L, D'Hondt S, Martens C, Vanormelingen P, Vannerum K, Sabbe K, Chepurnov VA, Inzé D, Vuylsteke M, Vyverman W (2008) Physiological and transcriptomic evidence for a close coupling between chloroplast ontogeny and cell cycle progression in the pennate diatom Seminavis robusta. Plant Physiol **148:** 1394-1411

Guglielmi G, Lavaud J, Rousseau B, Etienne AL, Houmard J, Ruban AV (2005) The light-harvesting antenna of the diatom Phaeodactylum tricornutum. Evidence for a diadinoxanthin-binding subcomplex. The FEBS journal **272:** 4339-4348

Guillard RRL (1975) Culture of phytoplankton for feeding marine invertebrates. *In* WL Smith, MH Canley, eds, Culture of Marine Invertebrate animals. Plenum Press, New York, pp 29-60

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol **8:** R19

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Research **27:** 297-300

Holdsworth ES (1985) Effect of Growth-Factors and Light Quality on the Growth, Pigmentation and Photosynthesis of 2 Diatoms, Thalassiosira-Gravida and Phaeodactylum-Tricornutum. Marine Biology **86:** 253-262

Hudson ME, Quail PH (2003) Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. Plant Physiology **133:** 1605-1616

Huysman MJJ, Martens C, Vandepoele K, Gillard J, Rayko E, Heijde M, Bowler C, Inzé D, Van de Peer Y, De Veylder L, Vyverman W (2010) Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. Genome Biol **11:** R17

Inzé D, De Veylder L (2006) Cell cycle regulation in plant development. Annu Rev Genet **40:** 77-105

Kooistra WH, De Stefano M, Mann DG, Medlin LK (2003) The phylogeny of the diatoms. Prog Mol Subcell Biol **33:** 59-97

Mendenhall MD, Hodge AE (1998) Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast Saccharomyces cerevisiae. Microbiol Mol Biol Rev **62:** 1191-1243

Menges M, Samland AK, Planchais S, Murray JA (2006) The D-type cyclin CYCD3;1 is limiting for the G1-to-S-phase transition in Arabidopsis. The Plant cell **18:** 893-906

Mercado JM, Sanchez-Saavedra MD, Correa-Reyes G, Lubian L, Montero O, Figueroa FL (2004) Blue light effect on growth, light absorption characteristics and photosynthesis of five benthic diatom strains. Aquatic Botany **78:** 265-277

Moore JD, Kirk JA, Hunt T (2003) Unmasking the S-phase-promoting potential of cyclin B1. Science **300:** 987-990

Morgan DO (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu Rev Cell Dev Biol **13:** 261-291

Nymark M, Valle KC, Brembu T, Hancke K, Winge P, Andresen K, Johnsen G, Bones AM (2009) An integrated analysis of molecular acclimation to high light in the marine diatom Phaeodactylum tricornutum. PLoS One **4:** e7743

Oakenfull EA, Riou-Khamlichi C, Murray JA (2002) Plant D-type cyclins and the control of G1 progression. Philosophical transactions of the Royal Society of London. Series B, Biological sciences **357:** 749-760

Quelle DE, Ashmun RA, Shurtleff SA, Kato JY, Barsagi D, Roussel MF, Sherr CJ (1993) Overexpression of Mouse D-Type Cyclins Accelerates G(1) Phase in Rodent Fibroblasts. Genes & Development **7:** 1559-1571

Rechsteiner M, Rogers SW (1996) PEST sequences and regulation by proteolysis. Trends in biochemical sciences **21:** 267-271

Renaudin JP, Doonan JH, Freeman D, Hashimoto J, Hirt H, Inzé D, Jacobs T, Kouchi H, Rouze P, Sauter M, Savoure A, Sorrell DA, Sundaresan V, Murray JAH (1996) Plant cyclins: A unified nomenclature for plant A-, B- and D-type cyclins based on sequence organization. Plant Molecular Biology **32:** 1003-1018

Resnitzky D, Gossen M, Bujard H, Reed SI (1994) Acceleration of the G1/S phase transition by expression of cyclins D1 and E with an inducible system. Molecular and cellular biology **14:** 1669-1679

Salvador ML, Klein U (1999) The redox state regulates RNA degradation in the chloroplast of Chlamydomonas reinhardtii. Plant Physiology **121:** 1367-1374

Shen M, Feng Y, Gao C, Tao D, Hu J, Reed E, Li QQ, Gong J (2004) Detection of cyclin b1 expression in g(1)-phase cancer cell lines and cancer tissues by postsorting Western blot analysis. Cancer Research **64:** 1607-1610

Sherr CJ (1995) D-type cyclins. Trends in biochemical sciences **20:** 187-190

Siaut M, Heijde M, Mangogna M, Montsant A, Coesel S, Allen A, Manfredonia A, Falciatore A, Bowler C (2007) Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum. Gene **406:** 23-35

Sims PA, Mann DG, Medlin LK (2006) Evolution of the diatoms: insights from fossil, biological and molecular data. Phycologia **45:** 361-402

Smetacek V (1999) Diatoms and the ocean carbon cycle. Protist **150:** 25-32

Terzaghi WB, Cashmore AR (1995) Light-Regulated Transcription. Annual Review of Plant Physiology and Plant Molecular Biology **46:** 445-474

Tyers M, Tokiwa G, Futcher B (1993) Comparison of the Saccharomyces-Cerevisiae G1 Cyclins - Cln3 May Be an Upstream Activator of Cln1, Cln2 and Other Cyclins. Embo Journal **12:** 1955-1968

Van den Hoek C, Mann, D.G., and Jahns, H.M. (1995) Algae: An Introduction to Phycology. Cambridge University Press, Cambridge

Vaulot D, Olson RJ, Chisholm SW (1986) Light and Dark Control of the Cell-Cycle in 2 Marine-Phytoplankton Species. Experimental Cell Research **167:** 38-52

Wang ZY, Kenigsbuch D, Sun L, Harel E, Ong MS, Tobin EM (1997) A Myb-related transcription factor is involved in the phytochrome regulation of an Arabidopsis Lhcb gene. The Plant cell **9:** 491-507

Zhu SH, Green BR (2010) Photoprotection in the diatom Thalassiosira pseudonana: role of LI818-like proteins in response to high light stress. Biochimica Et Biophysica Acta **1797:** 1449-1457

4

Characterization of the Rb-mediated pathway for G1-S phase transition in the diatom *Phaeodactylum tricornutum*

Marie J.J. Huysman[1,2,3], Klaas Vandepoele[2,3], Cindy Martens[2,3], Joris Huylebroek[2,3], Dirk Inzé[2,3], Yves Van de Peer[2,3], Wim Vyverman[1] and Lieven De Veylder[2,3]

[1]Protistology and Aquatic Ecology, Department of Biology, Ghent University, 9000 Gent, Belgium.
[2]Department of Plant Systems Biology, Flanders Institute for Biotechnology (VIB), 9052 Gent, Belgium.
[3]Department of Plant Biotechnology and Genetics, Ghent University, 9052 Gent, Belgium.

Manuscript in preparation

Authors' contributions

MJJH, KV, CM, DI, YvdP, WV and LDV conceived and designed the study. KV and CM performed the annotation and phylogenetic analysis and the *in silico* motif prediction. MJJH performed the synchronization and expression experiments. MJJH and JH performed the interaction and transactivation assays. MJJH analyzed the data and wrote the chapter. WV and LDV revised the manuscript.

ABSTRACT

In plants and animals, the Retinoblastoma (Rb)-mediated pathway, including the Rb protein and the E2F family of transcription factors is crucial for control of the G1-to-S transition and the start of DNA replication. Here, we report the annotation, isolation and characterization of members of the Rb-pathway in the diatom *Phaeodactylum tricornutum*. One Rb-related gene, two classical E2F genes, two DP genes and one atypical E2F gene were identified. Transcript analysis during synchronized growth indicated that these genes are differentially expressed during the cell cycle. Using a yeast-two-hybrid interaction assay we found that E2F2 and both DP proteins associate, indicating that these heterodimers might form functional transcription factor complexes in the diatom. Furthermore, through *de novo* motif detection, we identified a diatom-specific E2F *cis*-regulatory element, deviating from the conserved E2F motif found in plants and animals. E2F2 together with DP1b could activate a reporter gene under control of promoter sequences containing this diatom-specific E2F element, indicating that the motif is functional.

INTRODUCTION

In plants and animals, but not in yeast, the Retinoblastoma (Rb)-mediated pathway is in charge of the G1-to-S transition and the start of DNA replication. Positive regulation of this pathway promotes the transcription of genes necessary for S-phase entry (e.g. DNA replication genes) by E2F transcriptional regulators that consist of a heterodimer of the related E2F and DP proteins. The transactivating potential of E2F transcription factors is inhibited through binding of the Rb protein and related pocket proteins in their hypophosphorylated state. Phosphorylation of the Rb and pocket proteins by cyclin-dependent kinase (CDK) activity inhibits their binding on the E2F transcription factor complex and hence triggers the G1-to-S-phase transition (Weinberg, 1995; de Jager and Murray, 1999; Claudio et al., 2002).

When forming heterodimers with DPs, E2F proteins can recognize and bind specific *cis*-regulatory elements, the E2F motif, in the promoter region of E2F target genes (Vandepoele et al., 2005). The domain responsible for this interaction is the N-terminally located DNA-binding domain (DBD), a domain structurally related to a winged helix motif that is highly conserved in animal and plant E2F and DP proteins (Ramirez-Parra et al., 1999; Zheng et al., 1999; Mariconti et al., 2002). This DBD is flanked towards the C-terminus by a domain containing a leucine zipper motif, called the dimerization domain (DD), enabling heterodimerization of E2F with a DP partner protein (Helin et al., 1993). The DD is followed by another conserved sequence called the marked box that has been shown to be involved in DNA bending (Cress and Nevins, 1996).

In human, six classical E2F (E2F1 to 6) and two DP (DP1 and DP2) proteins have been described, while *Arabidopsis thaliana* possesses three E2F (E2Fa, E2Fb and E2Fc) and two DP (DPa and DPb) proteins (Dyson, 1998; Vandepoele et al., 2002). All of these proteins contain the conserved DBD and DD and most of the E2F sequences, except animal E2F6, possess a Rb-binding region in their C-terminal part that in some cases overlaps with a transactivation domain (Mariconti et al., 2002). In addition, animals and plants also possess atypical E2F proteins, also called DP-E2F-like (DEL) proteins, that contain duplicated conserved DBDs, but lack any of the other conserved regions of animal and plant E2Fs (Reviewed in Lammens et al., 2009). These atypical E2Fs can recognize and interact with consensus E2F *cis*-acting elements in a DP-independent manner due to the presence of two DBDs (Kosugi and Ohashi, 2002; Di Stefano et al., 2003; Logan et al., 2004; Logan et al., 2005).

Diatoms are unicellular eukaryotic photoautotrophs that account for about one fifth of the global carbon fixation (Van den Hoek, 1995; Field et al., 1998). Besides their ecological importance, diatoms have a broad industrial potential as producers of various high value compounds including poly-unsaturated fatty acids and pigments. Aditionally, they are used as feed in aquaculture, and because of their highly structured mesoporous cell wall diatoms also gained interest for nanotechnological purposes (Lebeau and Robert, 2003b; Bozarth et al., 2009). In this respect, large-scale cultivation of diatoms needs optimization to be fully profitable, and therefore we need to understand the basic mechanisms controlling the diatom life cycle. Recently it was shown that the diatom core cell cycle machinery is conserved, but that some aspects appear to be diatom-specific, including the presence of an expanded cyclin family, of which a major part are diatom-specific (Huysman et al., 2010).

The main goal of this study was to examine whether the regulatory mechanism of G1-S transition governed by the Rb-mediated pathway is conserved in diatoms. As a first step, we performed a genome-wide annotation analysis in the diatom *Phaeodactylum tricornutum* to identify homologs of the Rb and E2F/DP family. Further subclassification of the putative homologs was examined by phylogenetic analysis and determination of domain organization by sequence analysis. Next, a functional analysis of the identified genes was performed including transcript profiling during synchronized cell cycle progression and identification of putative E2F/DP complexes by interaction assays. Finally, using a *de novo* motif search approach we identified a diatom-specific E2F *cis*-acting element that deviates from the motif found in animals and plants. To investigate whether this diatom-specific E2F motif is functional in diatoms we performed transient transactivation assays using promoter sequences that contain the E2F motif.

RESULTS

In silico annotation and sequence analysis of the members of the Rb-pathway in *Phaeodactylum tricornutum*

Using a family-based structural annotation approach we could identify several members of the Rb-pathway in the genome of *P. tricornutum* (Bowler et al., 2008). These include 1 Rb-related gene, 2 E2F genes, 2 DP genes and 1 atypical E2F gene (for an overview see Table 4.1). A phylogenetic analysis of the E2F/DP/DEL family members is shown in Figure 4.1.

Table 4.1: Overview of the annotated members of the Rb-pathway in *P. tricornutum*

Gene Name	Protein Name	Protein ID
Phatr;*RBR1*	RBR1	Phatr2_46245
Phatr;*DP1*	DP1	Phatr2_14805
Phatr;*DP2*	DP2	Phatr2_5060
Phatr;*E2F1*	E2F1	Phatr2_43065
Phatr;*E2F2*	E2F2	Phatr2_47264
Phatr;*DEL1*	DEL1	Phatr2_48930

Alignment of both *P. tricornutum* E2F proteins with representatives of the E2F family of animals and *Arabidopsis thaliana* shows that E2F1 has a shorter sequence than E2F2, mainly due to a shorter N-terminal part, while E2F2 displays a C-terminal extension that is not present in any of the other sequences (Figure 4.2). Both PtE2F proteins possess typical conserved domains, including the DNA binding domain (DBD) and dimerization domain (DD) (Figure 4.2 and Figure 4.3a). While in E2F2, the marked box can be recognized, E2F1 appears to lack this domain. Finally, the Rb-binding domain seems not to be conserved in either of the diatom sequences. When looking in more detail to the amino acid residues in the DNA-binding domain important for DNA recognition, it can be seen that in both sequences the typical RRXYD motif for recognition of the core DNA sequence is conserved (Figure 4.3b). E2F1 however, lacks a hydrophilic arginine (R) residue (substituted by a hydrophobic isoleucine (I) residue) that is invariant among E2F family members and important for base pair contact (Zheng et al., 1999). Furthermore E2F1 also lacks several other important residues, but these have been substituted

by amino acids with the same properties, unlikely to have an effect on DNA binding or heterodimerization (Figure 4.3b).

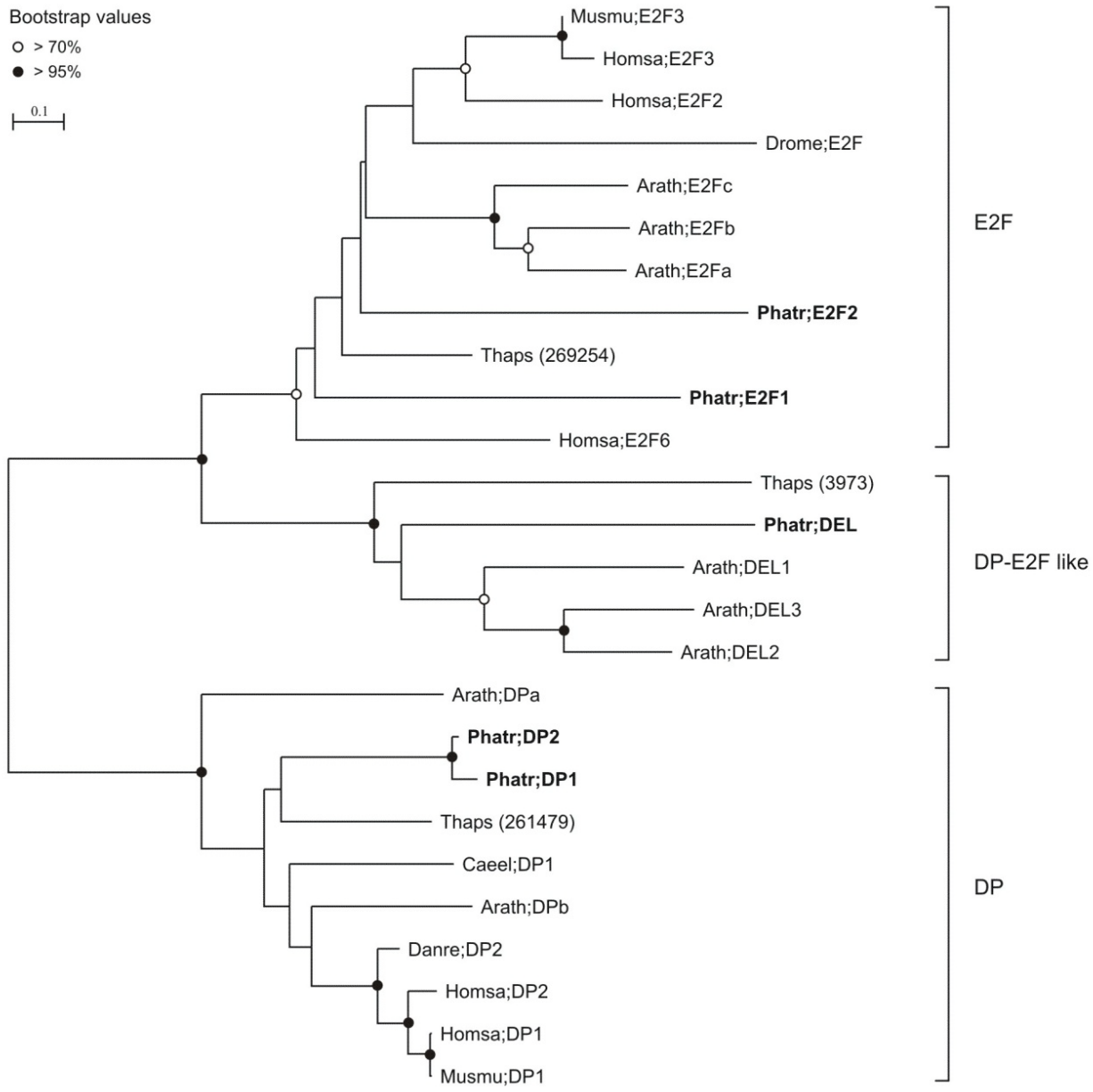


Figure 4.1: Phylogenetic analysis of the members of the E2F-family in *P. tricornutum*. Neighbor-joining tree (TREECON, based on multiple alignments of amino acids generated with MUSCLE, Poisson correction, 1000 replicates). The *P. tricornutum* sequences are shown in bold. Abbreviations: Arath, *Arabidopsis thaliana*; Caeel, *Caenorhabditis elegans*; Danre, *Danio rerio*; Drome, *Drosophila melegonaster*; Homsa, *Homo sapiens*; Musmu, *Mus musculus*; Phatr, *Phaeodactylum tricornutum*; Thaps, *Thalassiosira pseudonana*. The *P. tricornutum* sequences are show in bold.

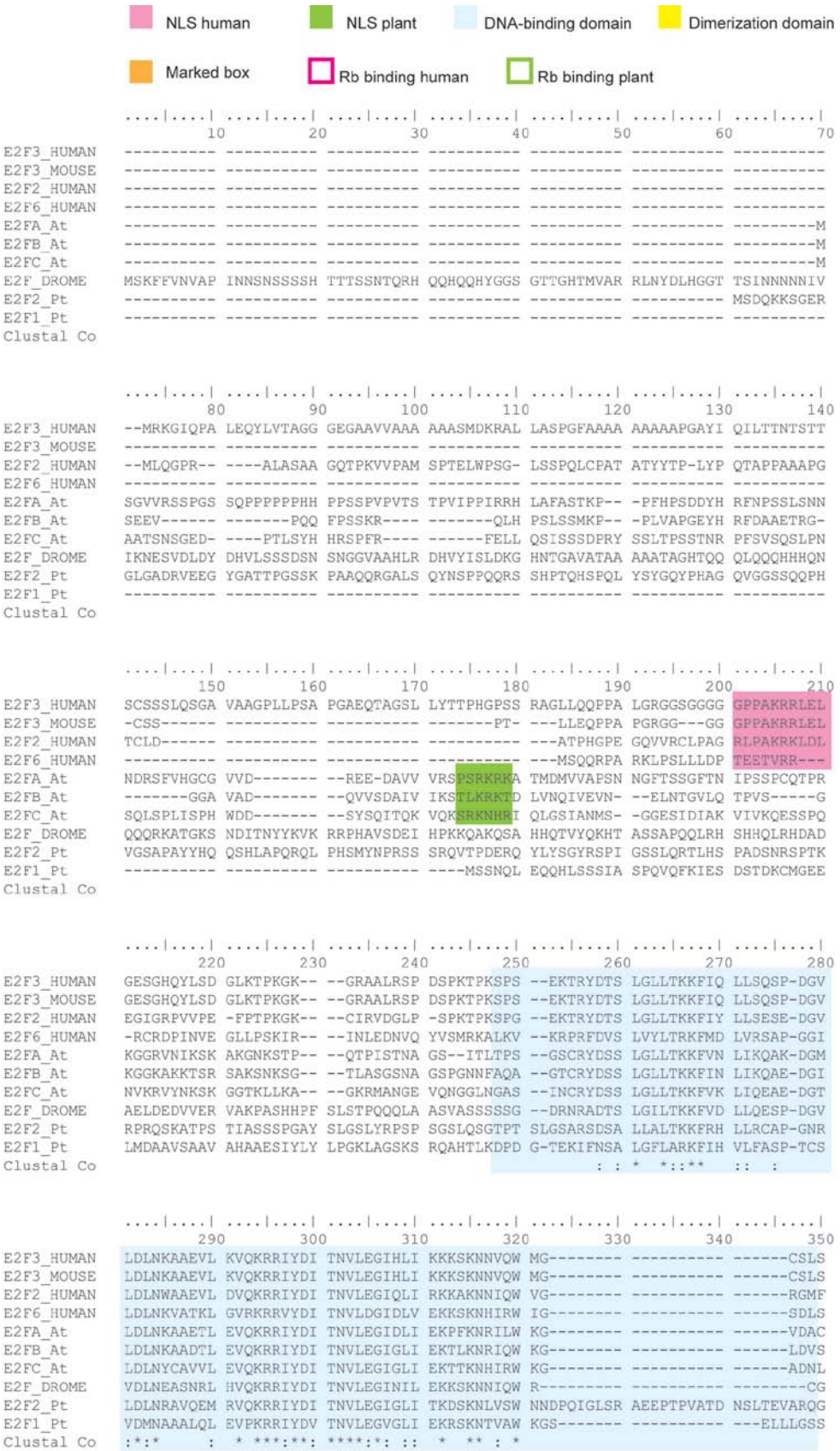


Figure 4.2: Alignment (generated with ClustalW) of the *P. tricornutum* E2F proteins (Pt) with E2F members of human, mouse, *Drosophila* (Drome) and *Arabidopsis* (At). Conserved domains are marked as indicated. Conserved residues are indicated by an asterix.

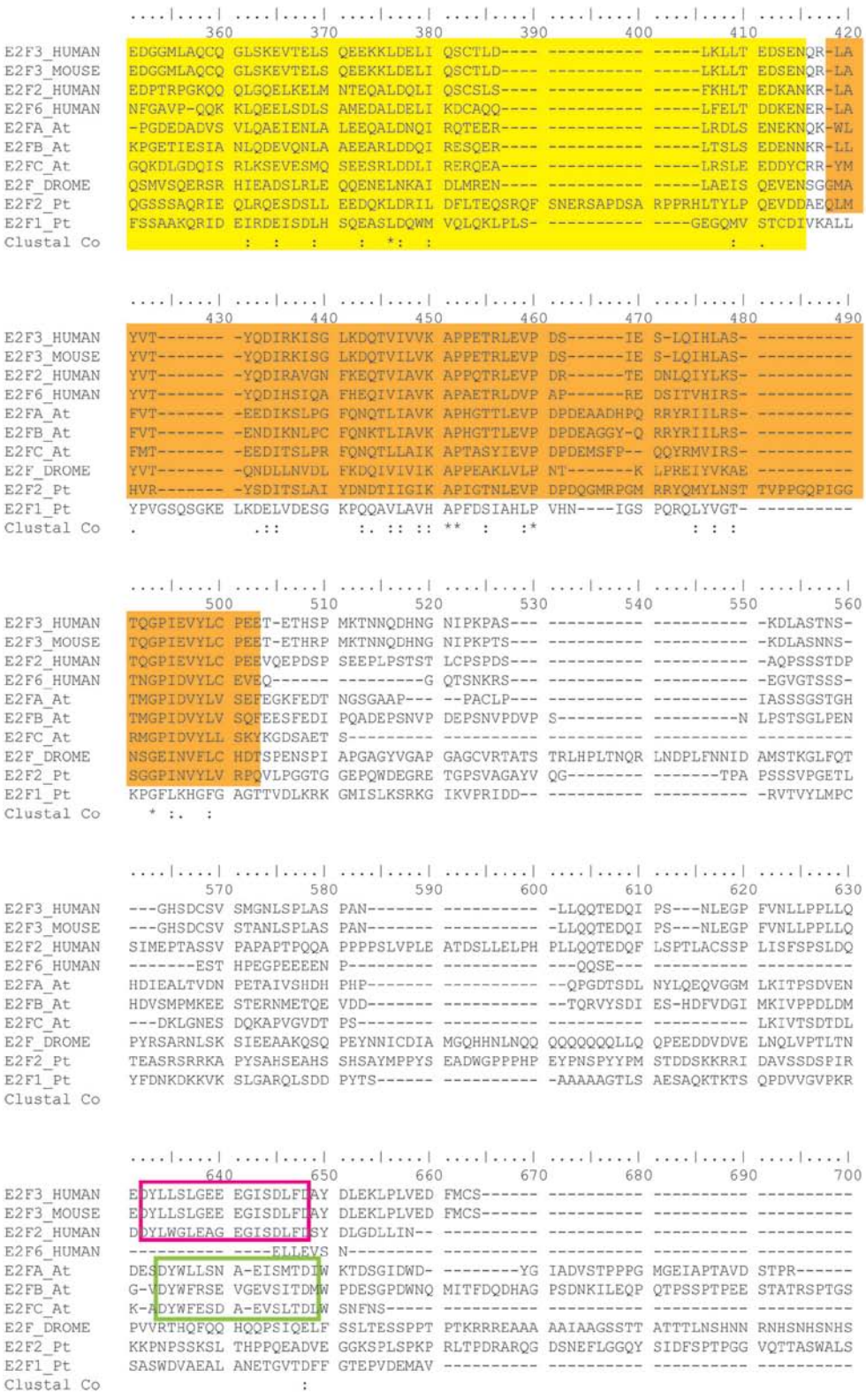


Figure 4.2: Continued. Alignment (generated with ClustalW) of the *P. tricornutum* E2F proteins (Pt) with E2F members of human, mouse, *Drosophila* (Drome) and *Arabidopsis* (At). Conserved domains are marked as indicated. Conserved residues are indicated by an asterix.

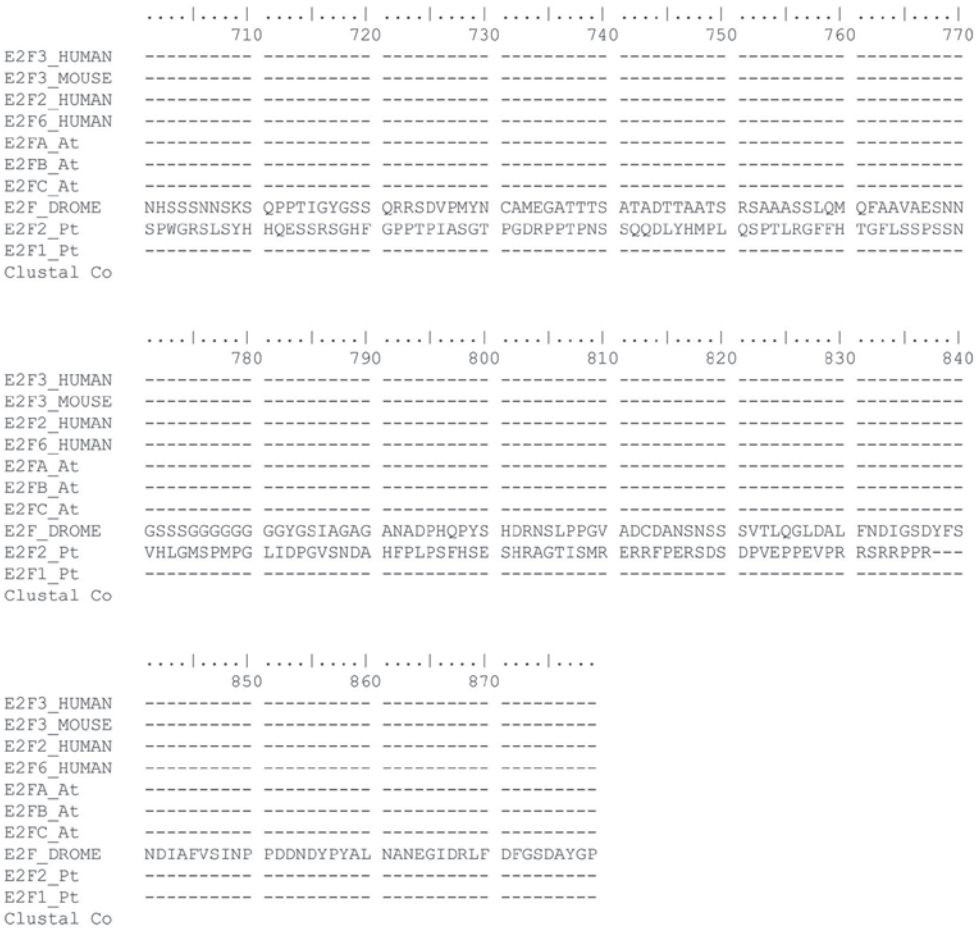


Figure 4.2: Continued. Alignment (generated with ClustalW) of the *P. tricornutum* E2F proteins (Pt) with E2F members of human, mouse, *Drosophila* (Drome) and *Arabidopsis* (At). Conserved domains are marked as indicated. Conserved residues are indicated by an asterix.

Similarly as for the E2F sequences, the alignment of the identified DP proteins with representatives of the DP family in human and *A. thaliana* shows that conserved domains, including the DBD and the DD, are present in both DP sequences (Figure 4.3a and Figure 4.4). Amino acid residues important for DNA recognition and heterodimerization contact are conserved in the diatom DP proteins (Figure 4.3c). Because of the apparent absence of NLSs in the diatom E2F sequences (Figure 4.2), and the possibility that their nuclear localization might be regulated in trans by the interacting DP partner, as reported for animal E2F4/5 proteins (de la Luna et al., 1996), signal motif prediction was performed for both DP sequences resulting in the detection of putative NLSs at their N-terminus (Figure 4.4).

The alignment of the single DEL protein with DEL homologs of *A. thaliana* shows the presence of two DBDs, typical for the DEL proteins (Figure 4.3a and Figure 4.5).

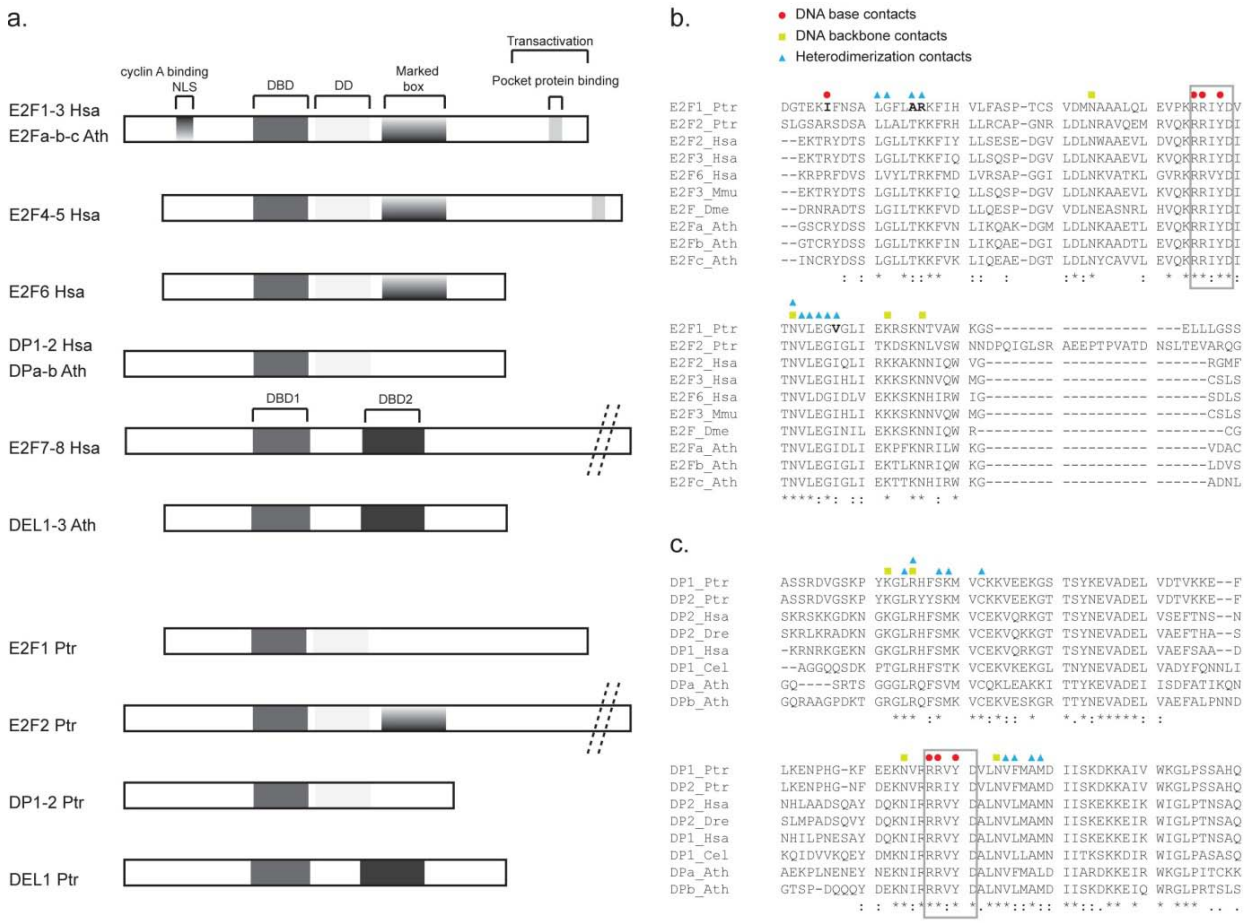


Figure 4.3: a. Schematic representation of the *P .tricornutum* E2F family members in comparison to human and plant E2F proteins. Conserved domains are indicated. NLS, nuclear localization signal; DBD, DNA-binding domain; DD, dimerization domain. **b and c.** Sequence alignment (generated with ClustalW) of the DNA-binding domains of known E2F and DP family members with the *P. tricornutum* E2F **(b)** and DP **(c)** proteins. Residues known to be important for DNA contact and heterodimerization contact according to Zheng et al. (1999) are marked as indicated by the legend. Residues that are not conserved in the E2F1 sequence are shown in bold. Conserved residues are indicated with an asterix. The RRXYD DNA recognition motif is indicated in the grey square.

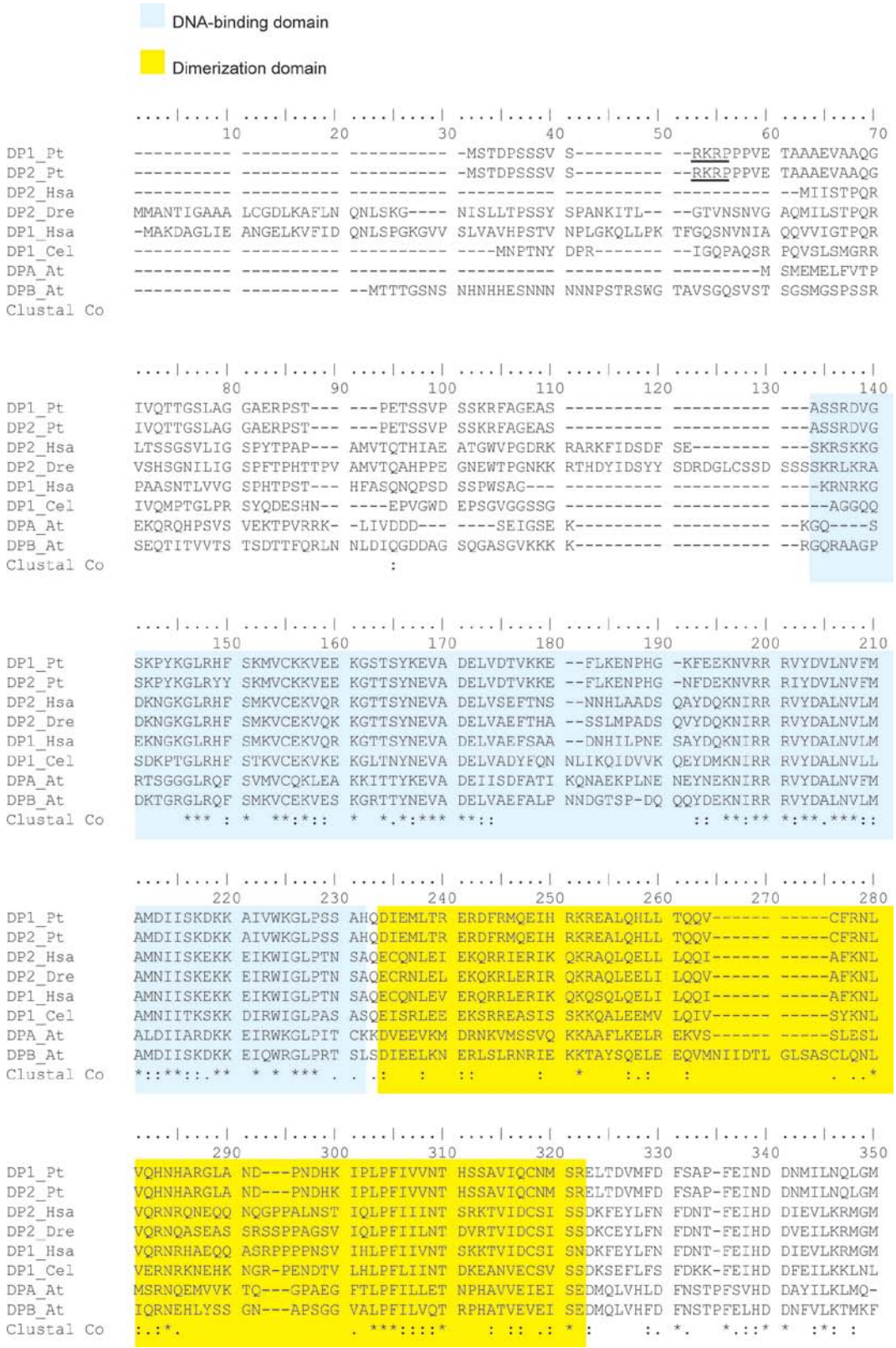


Figure 4.4: Alignment (generated with ClustalW) of the *P. tricornutum* DP proteins (Pt) with DP members of human (Hsa), *Danio rerio* (Dre), *Caenorhabditis elegans* (Cel) and *A. thaliana* (At). Conserved domains are marked as indicated. Conserved residues are indicated by an asterix. Predicted putative NLS signals are underlined.

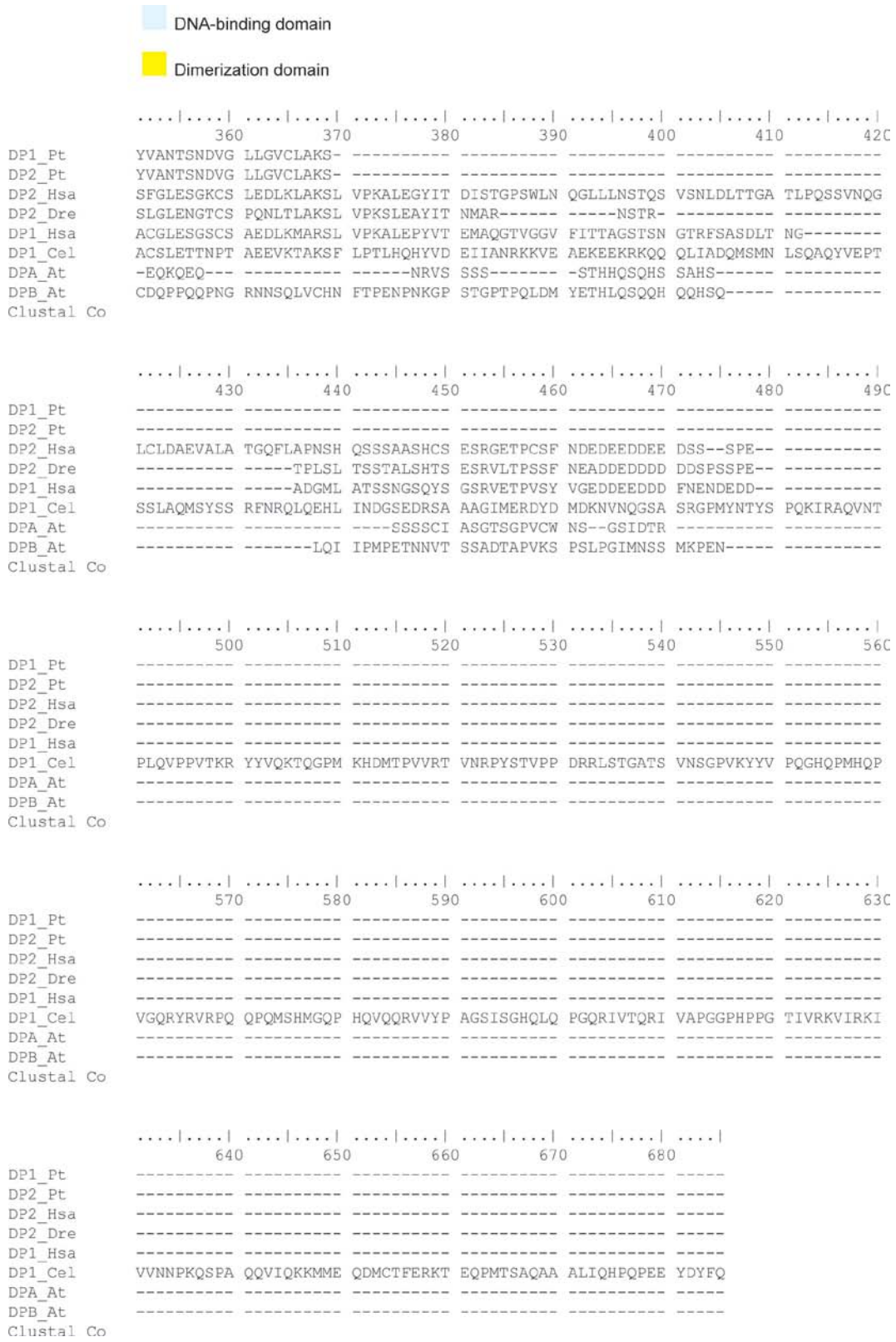


Figure 4.4: Continued. Alignment (generated with ClustalW) of the *P. tricornutum* DP proteins (Pt) with DP members of human (Hsa), *Danio rerio* (Dre), *Caenorhabditis elegans* (Cel) and *A. thaliana* (At). Conserved domains are marked as indicated. Conserved residues are indicated by an asterix. Predicted putative NLS signals are underlined.

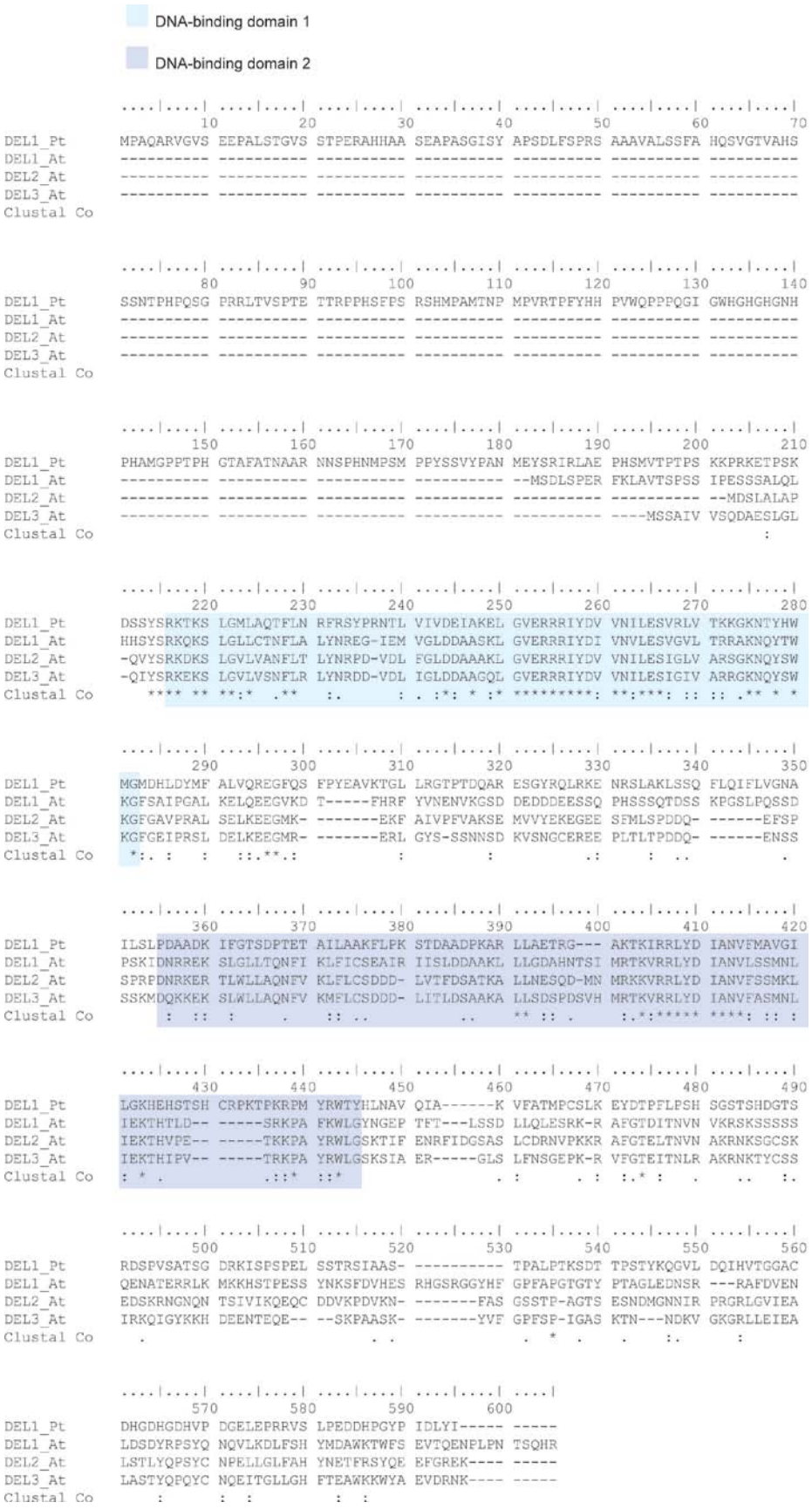


Figure 4.5: Alignment (generated with ClustalW) of the *P. tricornutum* DEL1 protein (Pt) with DEL proteins of *A. thaliana* (At). Conserved domains are marked as indicated. Conserved residues are indicated by an asterix.

The *P. tricornutum* members of the Rb-pathway are differentially expressed during the cell cycle

Transcript accumulation of all identified genes was monitored during the cell cycle in synchronized cultures using real-time quantitative PCR (real-time qPCR). Synchronization of cell division was performed by subjecting the cells to a prolonged dark period, which arrest the cells uniformly at the G1-phase, followed by their synchronous release by light exposure (Huysman et al., 2010, see Chapter 2). Transcript levels of *RBR1*, *E2F1*, *DP1* and *DP2* were all elevated at the late G1 phase and G1/S boundary (Figure 4.6), suggesting a similar control regulation for these genes as in other eukaryotes. *E2F2* shows a dual expression pattern during the cell cycle, displaying a peak of expression both at the G1/S and G2/M transition. *DEL1* showed elevated transcript levels during the G2- and M-phases.

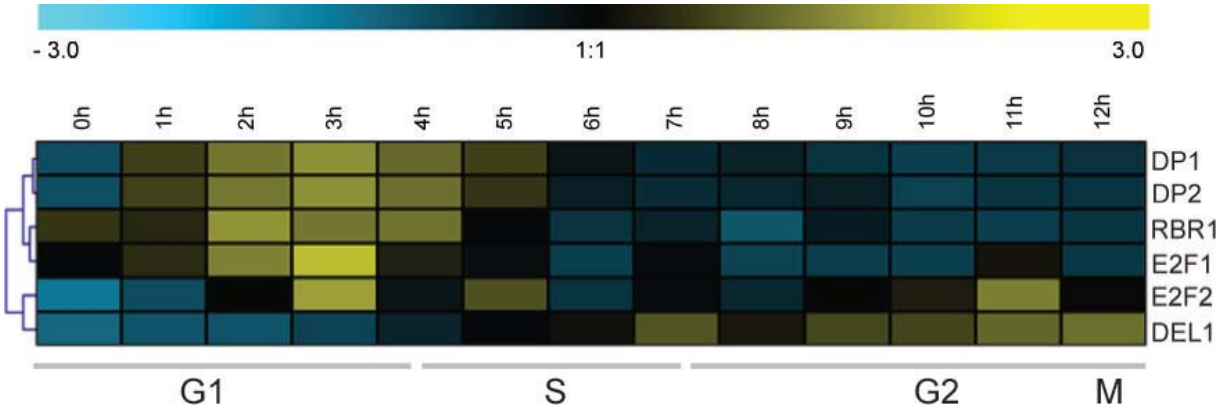


Figure 4.6: Hierarchical clustering of the transcript profiles of the diatom Rb- and E2F-family members during a synchronized cell cycle. Yellow: Relative upregulation. Blue: Relative downregulation. Approximate timing of the different cell cycle phases is indicated at the bottom. Transcript expression analysis was performed on the same set of samples as in Chapter 2. For cell cycle progression analysis by flow cytometry see Figure 2.1, Chapter 2. Values are the mean of three independent experiments normalized against the values of the consitutively expressed *histone H4* gene.

Isolation and cloning of the E2F and DP cDNA clones reveals the presence of different types of E2F1 and DP1

E2F and DP cDNA clones were isolated from cDNA prepared of an exponentially growing *P. tricornutum* culture. Plasmid DNA of the cloned fragments was isolated and the insert was sequenced. The obtained sequences were then compared to the respective sequence available from the *P. tricornutum* genome database (http://genome.jgi-

psf.org/Phatr2/Phatr2.home.html). Surprisingly, for the cDNA clones obtained for *E2F1* and *DP1* certain polymorphisms were detected in different clones originating from independent cloning procedures, indicating that multiple alleles are present of these genes in *P. tricornutum* cells. A multiple sequence alignment (ClustalW) was performed to check if the polymorphisms observed between the different forms resulted into important amino acid changes in the protein sequence, potentially influencing the structure or functional properties of the proteins (summarized in Table 4.2). For E2F1, all amino acid substitutions were positioned outside the conserved domains (Table 4.2), indicating that they probably do not interfere with the protein function. For DP1, a S>T substitution at residue 92 is positioned within the DNA binding domain (Table 4.2), but this concerns a amino acid residue which is only conserved between diatoms and human, but not in plants, indicating that it might not be crucial to structure and/or function of the protein.

Table 4.2: Different types of the annotated DP1 and E2F1 genes in *P. tricornutum.*

Sequence	Length(bp)	Nucleotide substitution*	Amino acid substitution*
PtE2F1a	1185	490G>A; 540T>C; 665G>A; 703G>T	164D>N; 221D>N; 234V>L
PtE2F1b	1185	120A>G;173C>A; 231; 836T>C; 836T>C; 918A>T; 1001C>T	55L>I; 278V>A; 334T>I
PtDP1a	839	785A>G; 794A>G	292Y>C; 295N>S
PtDP1b	839	169T>C; 181G>A; 192T>G; 274T>A; 411G>C; 609A>G; 615G>A; 732C>T; 742T>C; 785A>G; 794A>G	57F>L; 61A>T; 92S>T; 292Y>C; 295N>S
PtDP1c	839	No substitutions	No substitutions

*Substitutions compared to the available genome sequence (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html).

E2F2 can interact both with DP1 and DP2

To identify possible interactions between the different E2F and DP partners, a yeast-2-hybrid (Y2H) interaction screen was performed. Y2H screening is a commonly used technique to discover protein-protein interactions based on the reconstitution of the DNA binding and the activation domain of a transcription factor and hence the activation of a reporter gene when two proteins physically interact in the yeast system. (For more information see Fields and Song, 1989; Gietz et al., 1997 and http://en.wikipedia.org/wiki/Two-hybrid_screening). In this screen we tested all possible

interaction combinations between the *P. tricornutum* E2F and DP proteins. Yeast co-transformants carrying as bait either DP1 or DP2 and as prey E2F2 were able to grow on medium lacking histidine and resulted in a positive β-galactosidase test (figure 4.7), indicating that E2F2 can interact with both DP proteins. In the case of DP1, the interaction was confirmed for all tested alleles (DP1a, DP1b and DP1c), indicating that none of the substituted residues is necessary for heterodimer formation. No interaction was found for E2F1a (Figure 4.7), and we were unable to stably transform E2F1b in yeast (data not shown). A reciprocal test, in which E2F2 was used as bait and the DP proteins as prey, was unsuccessful due to self-activation of the E2F2 bait construct (data not shown).

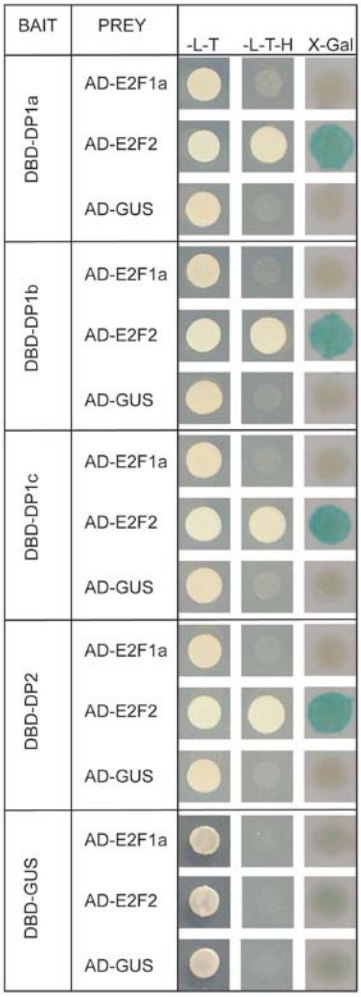


Figure 4.7: Y2H interaction assay of *P. tricornutum* E2F and DP proteins. Yeast PJ694-alpha cells were co-transformed with bait and prey plasmid as indicated. Co-transformation was analyzed on medium lacking leucine and tryptophan (-L-T). Co-transformants were tested for their ability to activate the histidine marker gene by assessing yeast growth on medium lacking leucine, tryptophan and histidine (-L-T-H) and for their ability to activate the *LacZ* reporter gene (X-Gal). For each combination three independent colonies were screened, of which one is shown.

A diatom-specific E2F binding site

Typically, *E2F* target genes possess E2F-binding sites in their promoter region that are recognized and bound by active E2F/DP complexes (Vandepoele et al., 2005). Through *de novo* motif detection with a set of 24 *P. tricornutum* genes homologous to known DNA replication genes from *Arabidopsis thaliana* (Table 4.3), we identified a candidate E2F consensus sequence (sGCsGGCs) (Figure 4.8a) which deviates from the most prominent motif found in animals and plants (wTTssCss) (Tao et al., 1997; Vandepoele et al., 2005). To investigate whether this E2F motif lacking a T-prefix is specific to the diatom-lineage, we repeated the *de novo* motif finding for *T. pseudonana* and *Phytophthora sojae* (using the same set of reference DNA replication genes). Whereas the former is another recently sequenced diatom (Armbrust et al., 2004), the latter is a member of the *Oomycetes* (Tyler et al., 2006), a group also belonging to the Stramenopiles. For *T. pseudonana* the detected motif was GC-rich without T prefix (CGGyGCCs) like in *P. tricornutum*, while the *P. sojae* motif showed a core GC-rich motif with an additional TTT-prefix (TTTTGGCG), as found in animals and plants (Figure 4.8b-c). These findings suggest that the diatom E2F consensus site has diverged from the sequence in other eukaryotes during evolution.

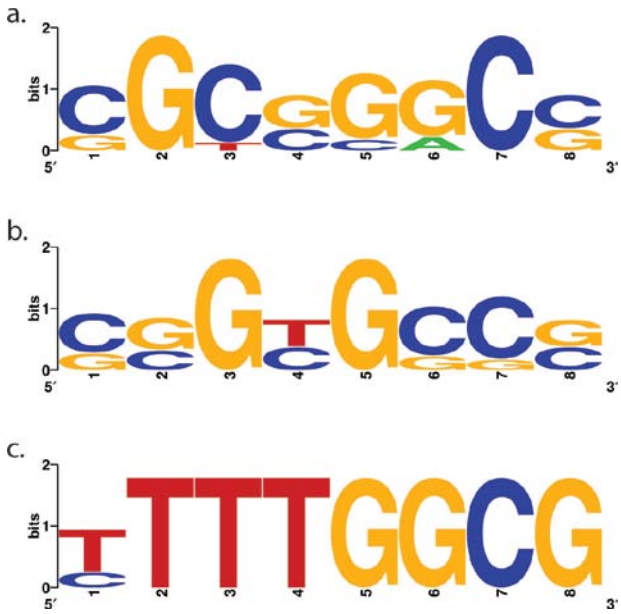


Figure 4.8: Sequence logos of the E2F consensus motifs detected by *de novo* motif finding in **a.** *P. tricornutum*, **b.** *T. pseudonana*, and **c.** *P. sojae*. The logos were created using WebLogo (Crooks et al., 2004) based on the motif instances present in the set of reference genes homologous to known DNA replication genes from *A. thaliana*. The overall height of each stack indicates the sequence conservation at that position (measured in bits), whereas the height of symbols within the stack reflects the relative frequency of the corresponding nucleic acid at that position.

Table 4.3. Identification of DNA replication genes in *P. tricornutum* by BlastP homolgy search starting from *A. thaliana* DNA replication genes.

A. thaliana	Description	*P. tricornutum*	% Identity	Length	E-value
AT1G07270	cell division control protein CDC6b, putative (CDC6b)	PT45115	24.40	332	2,00E-22
AT1G07370	Encodes putative proliferating cell nuclear antigen involved in cell cycle regulation;PCNA1	PT29196	54.62	260	5,00E-85
AT1G44900	ATP binding / DNA binding / DNA-dependent ATPase	PT18622	51.01	739	0.0
AT1G67320	DNA primase, large subunit family	PT46806	35.21	443	2,00E-68
AT2G06510	Encodes a homolog of Replication Protein A that is involved in meiosis I in pollen mother cells	PT14457	31.19	638	2,00E-77
AT2G07690	Member of the minichromosome maintenance complex, involved in DNA replication initiation	PT11490	49.09	658	3,00E-162
AT2G14050	MCM9	PT981	41.14	474	2,00E-77
AT2G21790	encodes large subunit of ribonucleotide reductase; RNR1	PT42726	66.24	776	0.0
AT2G24490	Encodes a component of Replication Protein A	PT32940	23.79	248	4,00E-14
AT2G27120	Encodes a protein with similarity to DNA polymerase epsilon catalytic subunit	PT52678	38.83	2166	0.0
AT2G29570	PCNA2 (PROLIFERATING CELL NUCLEAR ANTIGEN 2)	PT29196	53.46	260	3,00E-83
AT2G29680	Encodes cell division control protein 6 (CDC6).;CDC6 (CELL DIVISION CONTROL 6)	PT45115	20.70	401	5,00E-11
AT3G09660	MCM8	PT52561	36.24	665	3,00E-104
AT3G10690	DNA gyrase subunit A family protein	PT33633	42.09	841	1,00E-174
AT3G12530	PSF2	PT40942	26.97	178	4,00E-11
AT3G18524	Encodes a DNA mismatch repair homolog of human MutS gene, MSH6. MSH2 (MUTS HOMOLOG 2)	PT19604	46.33	354	1,00E-76
AT3G24320	Promotes re-arrangements of mitochondrial genome. MSH1 (MUTL PROTEIN HOMOLOG 1)	PT53933	30.74	680	3,00E-61
AT4G02060	Member of the minichromosome maintenance complex, involved in DNA replication initiation	PT13243	52.29	612	6,00E-154
AT4G02070	encodes a DNA mismatch repair homolog of human MutS gene, MSH6. MSH6 (MUTS HOMOLOG 6)	PT53969	31.69	1095	1,00E-123
AT4G25540	encodes a DNA mismatch repair homolog of human MutS gene, MSH6. MSH3 (Arabidopsis homolog of DNA mismatch repair protein MSH3)	PT53969	28.48	646	3,00E-60
AT4G35520	DNA mismatch repair protein similar to MutL. MLH3 (MUTL PROTEIN HOMOLOG 3)	PT54331	37.89	161	3,00E-11
AT5G08020	Encodes a homolog of Replication Protein A; RPA70B (RPA70-KDA SUBUNIT B)	PT14457	26.58	602	3,00E-53
AT5G27740	EMBRYO DEFECTIVE 2775 (EMB2775)	PT50732	50.14	363	2,00E-91
AT5G41880	POLA3, DNA primase	PT54992	43.27	349	7,00E-72
AT5G46280	DNA replication licensing factor, putative	PT51597	55.42	554	8,00E-168
AT5G63920	DNA topoisomerase III alpha, putative	PT47417	40.46	697	3,00E-132
AT5G63960	EMBRYO DEFECTIVE 2780 (EMB2780)	PT54677	50.64	1019	0.0
AT5G67100	Encodes the putative catalytic subunit of the DNA polymerase alpha. ICU2 (INCURVATA2)	PT12665	36.65	1135	3,00E-172

Mapping the candidate *P. tricornutum* consensus motif on all *P. tricornutum* promoter sequences yielded 572 genes containing the motif (Additional Table 4.1). Functional annotation based on the Gene Ontology (GO)-controlled vocabulary indicated that these genes were strongly enriched towards "DNA replication initiation" (9-fold enrichment; p-value < 0.01, hypergeometric test). Examples of DNA replication genes containing the E2F *cis*-acting element are *MCM7*, *MCM2* and *MCM3* (Phatr2_13243, Phatr2_18622 and Phatr2_51597, respectively); DNA mismatch repair protein *MSH2* (Phatr2_19604); DNA mismatch repair protein *MSH6-1* (Phatr2_53969); proliferating cell nuclear antigen *PCNA* (Phatr2_29196); *DNA gyrase subunit A* (Phatr2_33633) and a *DNA-directed DNA polymerase delta catalytic subunit* (Phatr2_54677). Interestingly, among the predicted E2F target genes, some cell cycle regulatory genes were identified, including *cyclin P1* (Phatr2_48210), *diatom-specific cyclin 1* (Phatr2_31942) and *diatom-specific cyclin 11* (Phatr2_41189), which were previously shown to be transcribed during the G1- and S-phases of the cell cycle (Huysman et al., 2010). Furthermore, also *E2F1* contains the predicted diatom consensus E2F motif in its promoter region, suggesting that its transcription might be under control of the Rb-mediated pathway itself.

E2F2-DP1b forms a functional transcription factor complex that activates promoters holding the diatom E2F motif

To investigate if the predicted diatom E2F *cis*-acting element can be recognized and regulated by the diatom E2F/DP heterodimers, we tested the effect of the identified E2F and DP proteins on promoters of candidate E2F target genes (*PCNA* and *MCM3*) using a transient reporter assay in tobacco 'Bright Yellow' (BY)-2 cells. Transient reporter assays are a valuable tool for the functional analysis of short regulatory sequences, because they can be performed rapidly and at a higher scale compared to reporter studies using stable transformation. In such reporter assays a reporter gene is placed under control of a target promoter (containing the regulatory sequence) and the activity of the reporter gene product is quantitatively measured upon co-transformation with putative transcription factors, in this particular case E2F/DP genes.

The promoter region from the *PCNA* and *MCM3* genes, both containing a diatom E2F motif (Figure 4.9a and 4.9b, respectively), was cloned in front of a reporter gene (*firefly luciferase* (*fluc*)) and the resulting construct was transiently expressed in the presence of different combinations of the E2F2 and DP proteins in BY-2 cells. Co-expression of *E2F2*

and *DP1b* was sufficient to significantly increase luciferase activity for both promoters (Figure 4.9c and 4.9d). Such an increase could not be observed when only one partner of the putative transcription factor complex was expressed in the BY-2 cells, indicating that both proteins interact and cooperate to form an active complex (Figure 4.9c and 4.9d). The combination of *E2F2* with the other *DP* types did not show any significant changes in luciferase activity (data not shown). It should be noted that both tested promoter sequences contain in addition to the predicted diatom E2F consensus also sequences that correspond to the plant/animal E2F consensus (Figure 4.9a and 4.9b), and therefore we cannot exclude that the observed activation of the *fluc* gene might be due to recognition and regulation of these sequences.

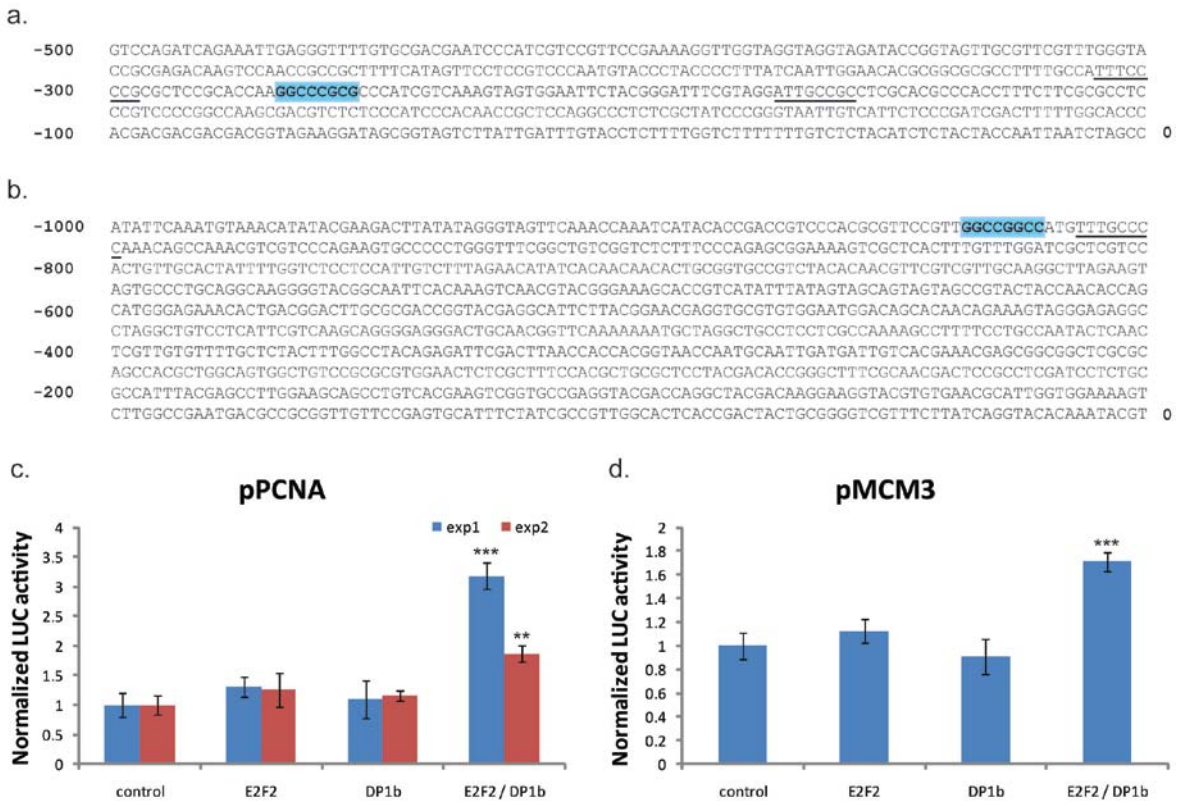


Figure 4.9: Promoter activity assay in BY-2 cells. **a.** Promoter sequence of *PCNA* (500 bp upstream of translational start). **b.** Promoter sequence of *MCM3* (1000 bp upstream of translational start). The putative diatom E2F motif is highlighted in blue and shown in bold. Sequences corresponding to the plant/animal E2F consensus (wTTssCss) are underlined. **c and d.** Tobacco BY-2 cells were transiently transformed with different combinations of expression vectors encoding E2F2 and DP1b, as indicated, with the pPCNA-fluc **(c)** or with the pMCM3-fluc **(d)** reporter construct. Regulatory properties of *E2F2* and *DP1b* were assessed by analyzing normalized fLUC activity (n>7). *** P<0.001; ** P<0.005 (Student's T-test). Error bars represent standard errors.

DISCUSSION

Structural domain organization of the diatom E2F family members

Here, we report the annotation of two classical E2F genes (*E2F1* and *E2F2*), two DP genes (*DP1* and *DP2*) and one atypical E2F gene (*DEL1*) in the diatom *P. tricornutum*. Overall, domain organization of the diatom E2F family members resembles mostly that of animal and plant E2F members. The DBD appears conserved in all diatom E2F members and the high conservation of the amino acid residues in this domain suggests that it may adopt a similar structure fold as already revealed for human heterodimer E2F4-DP2 (Zheng et al., 1999). Therefore, one should expect that diatom E2Fs would recognize and bind DNA sequences similar to the plant/animal E2F consensus. However, by applying *de novo* motif detection on promoters of diatom homologs of known plant E2F target genes, we could predict a consensus diatom E2F binding site (sGCsGGCs), that within its core resembles the animal/plant consensus site (wTTssCss), but that lacks the typical T-rich prefix. This T-rich portion renders the plant/animal E2F binding-site asymmetric, allowing orientation of the E2F-DP heterodimer on the promoter. An R residue conserved in the DBD of E2F, but not in DP proteins has been shown to be important for DNA binding near the T-rich portion (Zheng et al., 1999). Interestingly, whereas this R residue is present in diatom E2F2, it is lacking in E2F1. This observation, together with the finding of a symmetrical diatom E2F site, suggests a possible alternative mode of DNA binding for E2F1 in diatoms. On the other hand, E2F2 was found to be able to transactivate promoters that hold the symmetrical diatom E2F site, suggesting that orientation of the heterodimer might be less important for regulation of the diatom E2F site. The DD can be recognized in both classical diatom E2F and DP proteins, but is absent in the DEL1 protein, which contains two DBDs.

Most strikingly, we could not identify an Rb-binding motif in any of the diatom E2Fs. No sequence could be found that is even distantly related to the animal (DYX_7EX_3DLFD) (Cress et al., 1993; Helin et al., 1993) or plant (DYX_6DX_4DMWE) (Ramirez-Parra et al., 1999) Rb-binding motif, indicating that in diatoms this sequence has possibly diverged too much or, alternatively, that diatom E2Fs are not controlled by pocket proteins, but perhaps by other regulatory proteins. The latter is true for animal E2F6, which is regulated by polycomb group proteins through recognition of the marked box (Trimarchi et al., 2001). Furthermore, no NLS could be identified in the diatom E2F sequences, although present in the classical animal and plant E2Fs. However, nuclear localization of animal E2F4/5 proteins, that do not

possess NLSs, is driven by interaction with variants of DP2 that contain an NLS (de la Luna et al., 1996). Trans regulation of nuclear localization of E2Fs could be a more general mechanism in diatoms as signal motif prediction for both DP sequences identified putative NLSs in their N-terminal part.

During the cloning procedures it was found that *P. tricornutum* possesses different types of both *E2F1* and *DP1*, which was confirmed by examination of the available EST libraries (data not shown). Different polymorphisms seem to have accumulated in these genes, however without resulting in amino acid substitutions that would be expected to cause major structural or functional changes to the proteins. The presence of multiple single nucleotide polymorphisms does not appear to be unique for the *E2F1* and *DP1* genes, but seems to be more common for the gene repertoire of *P. tricornutum* (personal communication C. Bowler). This finding is rather astonishing, as the genome sequence available, the EST data obtained under different conditions, and the cultures used in the present study, were all derived from a monoclonal culture (Pt1 8.6) generated only ten years ago (De Martino et al., 2007; Bowler et al., 2008; Maheswari et al., 2010). The mechanisms responsible for this rapid mutation accumulation and its evolutionary and functional relevance remain to be elucidated.

G1-S transition regulation

Transcript analysis of the annotated diatom *E2F* family genes during synchronized cell cycle progression showed that all members, except *DEL1*, show a peak of transcript expression at the G1-to-S-phase transition, suggesting a functional role for them in the regulation of DNA replication. *RBR1* is somewhat earlier transcribed during the time course compared to both *E2F* genes, which suggests that the *E2F* genes could be target of RBR1 during the G1-phase. Release of RBR1 by its phosphorylation by G1-specific CDK/cyclin complexes could lead to the activation of the *E2F* genes, resulting finally in the activation of S phase E2F target genes. *E2F2* shows an additional transcriptional peak during the G2-phase, indicative of a dual role for this gene both during S- and G2-phase. Whether this concerns an activating or repressing role remains to be elucidated. The single *DEL1* gene identified showed elevated expression during the G2-phase. Analogously, *DEL* homologs of mice and *A. thaliana* preferentially accumulate during the S-G2 phase (Lammens et al., 2008; Li et al., 2008).

We found that E2F2 can associate with both DP proteins, whatever type of DP1 was tested, indicating that none of the amino acid changes present in the different DP1 types

interfere with its ability to bind E2F2. Furthermore, we showed that the transactivational potential of E2F2 to regulate promoters that hold the predicted diatom E2F motif depends on co-expression with DP1b in BY-2 cells. The functionality of this transcription factor complex in *P. tricornutum* and the necessity and sufficiency of the predicted diatom E2F motif to transactivate E2F target gene expression await further analysis.

Although domain organization of E2F1 is clearly conserved, we could not find any clear evidence of its functionality in *P. tricornutum*. Using the yeast-two-hybrid assay we were unable to detect interaction of either E2F1 alleles with a DP molecule. Possibly the method chosen here was not optimal, as we were unable to transform E2F1b in yeast, and therefore putative interactions could be revealed by applying other techniques, such as co-immunoprecipitation. However, its transcriptional peak at the G1-S transition and the presence of a diatom-specific E2F motif in its promoter region suggest that E2F1 might be involved in or at least under control of the G1-S regulatory mechanism.

To summarize, we showed that diatoms possess and express all members of the Rb-mediated pathway and, based on their expression profile during the cell cycle, they most likely control S-phase entry and DNA replication. Furthermore, we identified a diatom-specific E2F motif and characterized the E2F2/DP1b heterodimer as a transcriptional activator complex that is able to drive expression of potential E2F target genes. Together, these data form the basis to explore further the functions of the E2F family members in diatoms.

MATERIALS AND METHODS

Annotation and phylogenetic analysis of the Rb and E2F-family members

Animal and plant Rb and E2F family members were used to build a family specific HMMER profile (Eddy, 1998). With this profile, the predicted *P. tricornutum* proteome was screened to detect members of the Rb and E2F family. The putative *P. tricornutum* homologs were validated by comparing them with the reference family members in a multiple alignment. Multiple alignments based on amino acid sequences were generated with MUSCLE (Edgar, 2004) and manually improved using BioEdit (Hall, 1999). TREECON (Van de Peer and De Wachter, 1994) was used to generate the neighbor-joining tree based on Poisson-corrected

distances. To test the significance of the nodes, bootstrap analysis was done with 1000 replicates.

Sequence analysis

Multiple alignments were generated using the ClustalW application in BioEdit (Thompson et al., 1994; Hall, 1999). NLS prediction of the *P. tricornutum* DP sequences was performed with WoLF PSORT (Horton et al., 2007).

Real-time quantitative PCR

For RNA extraction, 5 x 10^7 cells were collected, fast frozen in liquid nitrogen and stored at -70ºC. Cell lysis and RNA extraction was performed using TriReagent (Molecular Research Center, Inc., Cincinnati, OH, USA) according to the manufacturer's instructions. Contaminating genomic DNA was removed by DNaseI treatment (GE Healthcare, Little Chalfont, UK). To assess RNA concentration and purity, spectrophotometry was used (NaNodrop ND-1000, Wilmington, DE). Total RNA was reverse transcribed using iScript reverse transcriptase (Roche). Finally, 1.25 ng (cell cycle expression) or 10 ng (overexpression lines) of cDNA was used as template for each qPCR reaction.

Table 4.4: gene specific primers used in qPCR analysis.

Protein Name	Protein ID	Forward gene specific primer	Reverse gene specific primer	Product length (bp)
RBR1	Phatr2_46245	CCCGCCACCGATGTCAAAC	TACGCCACGGATTCTACACG	165
DP1	Phatr2_14805	GGCTCAACGAGTTACAAGGA	GGCGCACATTCTTTTCTTC	106
DP2	Phatr2_5060	AAGGGCACAACGAGTTACAA	CGTTGAGGACGTCGTAGATT	133
E2F1	Phatr2_43065	CCCTAAGCGGCGGATTTACG	AAGGACGAGCCAAGAAGAAGC	117
E2F2	Phatr2_47264	CACTTTCCTCTGCCTTCGTTTC	CGTGGTACTTCTGGTGGTTCG	119
DEL1	Phatr2_48930	CACCGAAACAACCCGACCTC	CCGTGCCAGCCAATGCC	138

Samples in triplicate were amplified on the Lightcycler 480 platform with the Lightcycler 480 SYBR Green I Master mix (Roche Applied Science), in the presence of 0.5 μM gene-specific primers (Table 4.4). The cycling conditions were 10 min polymerase activation at 95°C and 45 cycles at 95°C for 10 s, 58°C for 15 s and 72°C for 15 s. Amplicon dissociation curves were recorded after cycle 45 by heating from 65°C to 95°C. In qBase (Hellemans et al., 2007), data were analyzed using the ΔC_t relative quantification method with the stably

expressed histone H4 as a normalization gene (Siaut et al., 2007). Expression profiles of the synchronized cell cycle series were mean relative expression from three independent sample series. After normalization, the mean profiles were clustered using hierarchical average linkage clustering (TMEV3D).

Yeast-two-hybrid analysis

Yeast-two-hybrid bait and prey plasmids were generated through recombinational gateway cloning (Invitrogen). Full length open reading frames of the *P. tricornutum* E2F and DP genes were isolated using gene-specific primers (Table 4.5), cloned in either the pDONR221 or pENTR-D-TOPO vector (Invitrogen) (Table 4.5), and subsequently recombined in the pDEST22 (prey) and pDEST32 (bait) vectors (Invitrogen) by attL x attR recombination, resulting in translational fusions between the proteins and the GAL4 transcriptional activators and GAL4 DNA-binding domains, respectively. Plasmids encoding the bait and prey constructs were co-transformed in the yeast strain PJ694-alpha (MATa; trp1-901, leu2-3,112, ura3-52, his3-200, gal4D, gal80D, LYS2::GAL1-HIS3, GAL2-ADE2, met2GAL7-lacZ) by the LiAc method (Gietz et al., 1992). Co-transformed yeast cells were selected on synthetic dextrose (SD) plates lacking Leu and Trp. Interaction between the introduced proteins was scored on SD plates lacking Leu, Trp and His and by the LacZ test, as described before (Boruc et al., 2010).

Promoter element detection

Promoter motifs were detected on 24 *P. tricornutum* genes homologous to E2F-containing *A. thaliana* genes upregulated during the S phase (Menges et al., 2002; Vandepoele et al., 2005) with MotifSampler 3.1 (Thijs et al., 2001) (settings: -w length of motif set to 8, -b background model of order 3 created using all 1-kb promoter sequences, -r number of MotifSampler repeats set to 25 runs). For all reported putative motifs enrichment analysis confirmed that these motifs were specifically overrepresented in the input set compared to the genome-wide occurrence. Genome-wide motif mapping was done using MotifLocater (settings: -t 0.95) (Thijs et al., 2001). A similar approach was applied for *T. pseudonana* and *P. sojae* DNA replication homologs. Genes were functionally annotated using Blast2GO (Conesa et al., 2005). Enrichment values were calculated as the ratio of the frequency in the selected set relative to the genome-wide frequency. The statistical significance of the functional GO enrichment within sets of genes was evaluated using the hypergeometric distribution adjusted

by the Bonferroni correction for multiple hypotheses testing. Corrected p-values<0.05 were considered as significant.

Transient reporter assay in BY2 cells

The 500- and 1000-bp promoter sequences upstream of the translational start of *PCNA* and *MCM3*, respectively, were amplified from genomic DNA using specific primers (Table 4.5), cloned in the pP4-P1 vector and subsequently cloned simultaneously with the fLUC sequence in the pm42GW7,3 pDEST vector (Karimi et al., 2007) by multisite gateway cloning (Invitrogen). The full-length open reading frames of the E2F and DP genes were recombined in the p2GW7 pDEST vector by gateway cloning, containing the 35S CaMV promoter. Both reporter and effector plasmids were used to transfect BY-2 protoplasts using the polyethylene glycol (PEG)/Ca^{2+} method, as described by De Sutter et al., 2005. Luciferase measurements were performed using the Dual-luciferase Reporter 1000 Assay System (Promega), according to the manufacturer's instructions and as described before (De Sutter et al., 2005).

Table 4.5: Cloning primers and cloning strategy.

Primer name	Sequence	Vector
DP-attB1_Fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTCGACTGATCCTTCAAG	pDONR221
DP-attB2_Rv	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAACTCTTTGCAAGACACACG	pDONR221
E2F1_Fw	CACCATGAGTAGCAATCAATTAGAGCAACA	pENTR-D-TOPO
E2F1_Rv	TCAAACGGCCATTTCATC	pENTR-D-TOPO
E2F2_Fw	CACCATGTCGGATCAGAAAAAATCG	pENTR-D-TOPO
E2F2_Rv	TCAGCGAGGTGGCCTACGAGA	pENTR-D-TOPO
pMCM3-attB4_Fw	ATAGAAAAGTTGATATTCAAATGTAAACATATACGAAGACTT	pP4-P1
pMCM3-attB1r_Rv	TACAAACTTGTACGTATTTGTGTACCTGATAAGAAAC	pP4-P1
pPCNA-attB4_Fw	ATAGAAAAGTTGGTCCAGATCAGAAATTGAGGGT	pP4-P1
pPCNA-attB1r_Rv	TACAAACTTGTGGCTAGATTAATTGGTAGTAGAGATGTAGA	pP4-P1

ADDITIONAL DATA

Additional Table 4.1 is available on the accompanying compact disc.

REFERENCES

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WW, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science **306:** 79-86

Boruc J, Van den Daele H, Hollunder J, Rombauts S, Mylle E, Hilson P, Inzé D, De Veylder L, Russinova E (2010) Functional modules in the Arabidopsis core cell cycle binary protein-protein interaction network. Plant Cell **22:** 1264-1280

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kroger N, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin-Jezequel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Oudot-Le Secq MP, Napoli C, Obornik M, Parker MS, Petit JL, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature **456:** 239-244

Bozarth A, Maier UG, Zauner S (2009) Diatoms in biotechnology: modern tools and applications. Appl Microbiol Biotechnol **82:** 195-201

Claudio PP, Tonini T, Giordano A (2002) The retinoblastoma family: twins or distant cousins? Genome Biol **3:** reviews3012

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21:** 3674-3676

Cress WD, Johnson DG, Nevins JR (1993) A genetic analysis of the E2F1 gene distinguishes regulation by Rb, p107, and adenovirus E4. Molecular and cellular biology **13:** 6314-6325

Cress WD, Nevins JR (1996) A role for a bent DNA structure in E2F-mediated transcription activation. Molecular and cellular biology **16:** 2119-2127

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. Genome research **14:** 1188-1190

de Jager SM, Murray JA (1999) Retinoblastoma proteins in plants. Plant Mol Biol **41:** 295-299

de la Luna S, Burden MJ, Lee CW, LaThangue NB (1996) Nuclear accumulation of the E2F heterodimer regulated by subunit composition and alternative splicing of a nuclear localization signal. Journal of Cell Science **109:** 2443-2452

De Martino A, Meichenin A, Shi J, Pan KH, Bowler C (2007) Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions. Journal of Phycology **43:** 992-1009

De Sutter V, Vanderhaeghen R, Tilleman S, Lammertyn F, Vanhoutte I, Karimi M, Inzé D, Goossens A, Hilson P (2005) Exploration of jasmonate signalling via automated and standardized transient expression assays in tobacco cells. Plant Journal **44:** 1065-1076

Di Stefano L, Jensen MR, Helin K (2003) E2F7, a novel E2F featuring DP-independent repression of a subset of E2F-regulated genes. The EMBO journal **22:** 6289-6298

Dyson N (1998) The regulation of E2F by pRB-family proteins. Genes & Development **12:** 2245-2262

Eddy SR (1998) Profile hidden Markov models. Bioinformatics **14:** 755-763

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:** 1792-1797

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. Science **281:** 237-240

Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. Nature **340:** 245-246

Gietz D, St Jean A, Woods RA, Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res **20:** 1425

Gietz RD, Triggs-Raine B, Robbins A, Graham KC, Woods RA (1997) Identification of proteins that interact with a protein of interest: applications of the yeast two-hybrid system. Molecular and cellular biochemistry **172:** 67-79

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser **41:** 95-98

Helin K, Harlow E, Fattaey A (1993) Inhibition of E2F-1 transactivation by direct binding of the retinoblastoma protein. Molecular and cellular biology **13:** 6501-6508

Helin K, Wu CL, Fattaey AR, Lees JA, Dynlacht BD, Ngwu C, Harlow E (1993) Heterodimerization of the transcription factors E2F-1 and DP-1 leads to cooperative trans-activation. Genes & Development **7:** 1850-1861

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol **8:** R19

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Research **35:** W585-W587

Huysman MJJ, Martens C, Vandepoele K, Gillard J, Rayko E, Heijde M, Bowler C, Inzé D, Van de Peer Y, De Veylder L, Vyverman W (2010) Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. Genome Biol **11:** R17

Karimi M, Depicker A, Hilson P (2007) Recombinational cloning with plant gateway vectors. Plant Physiology **145:** 1144-1154

Kosugi S, Ohashi Y (2002) E2Ls, E2F-like repressors of Arabidopsis that bind to E2F sites in a monomeric form. The Journal of biological chemistry **277:** 16553-16558

Lammens T, Boudolf V, Kheibarshekan L, Zalmas LP, Gaamouche T, Maes S, Vanstraelen M, Kondorosi E, La Thangue NB, Govaerts W, Inzé D, De Veylder L (2008) Atypical E2F activity restrains APC/CCCS52A2 function obligatory for endocycle onset. Proceedings of the National Academy of Sciences of the United States of America **105:** 14721-14726

Lammens T, Li J, Leone G, De Veylder L (2009) Atypical E2Fs: new players in the E2F transcription factor family. Trends in Cell Biology **19:** 111-118

Lebeau T, Robert JM (2003b) Diatom cultivation and biotechnologically relevant products. Part I: cultivation at various scales. Appl Microbiol Biotechnol **60:** 612-623

Li J, Ran C, Li E, Gordon F, Comstock G, Siddiqui H, Cleghorn W, Chen HZ, Kornacker K, Liu CG, Pandit SK, Khanizadeh M, Weinstein M, Leone G, de Bruin A (2008) Synergistic function of E2F7 and E2F8 is essential for cell survival and embryonic development. Developmental Cell **14:** 62-75

Logan N, Delavaine L, Graham A, Reilly C, Wilson J, Brummelkamp TR, Hijmans EM, Bernards R, La Thangue NB (2004) E2F-7: a distinctive E2F family member with an unusual organization of DNA-binding domains. Oncogene **23:** 5138-5150

Logan N, Graham A, Zhao X, Fisher R, Maiti B, Leone G, La Thangue NB (2005) E2F-8: an E2F family member with a similar organization of DNA-binding domains to E2F-7. Oncogene **24:** 5000-5004

Maheswari U, Jabbari K, Petit JL, Porcel BM, Allen AE, Cadoret JP, De Martino A, Heijde M, Kaas R, La Roche J, Lopez PJ, Martin-Jezequel V, Meichenin A, Mock T, Schnitzler Parker M, Vardi A, Armbrust EV, Weissenbach J, Katinka M, Bowler C (2010) Digital expression profiling of novel diatom transcripts provides insight into their biological functions. Genome biology **11:** R85

Mariconti L, Pellegrini B, Cantoni R, Stevens R, Bergounioux C, Cella R, Albani D (2002) The E2F family of transcription factors from Arabidopsis thaliana. Novel and conserved components of the retinoblastoma/E2F pathway in plants. The Journal of biological chemistry **277:** 9911-9919

Menges M, Hennig L, Gruissem W, Murray JA (2002) Cell cycle-regulated gene expression in Arabidopsis. J Biol Chem **277:** 41987-42002

Ramirez-Parra E, Xie Q, Boniotti MB, Gutierrez C (1999) The cloning of plant E2F, a retinoblastoma-binding protein, reveals unique and conserved features with animal G(1)/S regulators. Nucleic Acids Research **27:** 3527-3533

Siaut M, Heijde M, Mangogna M, Montsant A, Coesel S, Allen A, Manfredonia A, Falciatore A, Bowler C (2007) Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum. Gene **406:** 23-35

Tao Y, Kassatly RF, Cress WD, Horowitz JM (1997) Subunit composition determines E2F DNA-binding site specificity. Mol Cell Biol **17:** 6994-7007

Thijs G, Lescot M, Marchal K, Rombauts S, De Moor B, Rouze P, Moreau Y (2001) A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. Bioinformatics **17:** 1113-1122

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. Nucleic Acids Research **22:** 4673-4680

Trimarchi JM, Fairchild B, Wen J, Lees JA (2001) The E2F6 transcription factor is a component of the mammalian Bmi1-containing polycomb complex. Proceedings of the National Academy of Sciences of the United States of America **98:** 1519-1524

Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RH, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, Chapman J, Damasceno CM, Dorrance AE, Dou D, Dickerman AW, Dubchak IL, Garbelotto M, Gijzen M, Gordon SG, Govers F, Grunwald NJ, Huang W, Ivors KL, Jones RW, Kamoun S, Krampis K, Lamour KH, Lee MK, McDonald WH, Medina M, Meijer HJ, Nordberg EK, Maclean DJ, Ospina-Giraldo MD, Morris PF, Phuntumart V, Putnam NH, Rash S, Rose JK, Sakihama Y, Salamov AA, Savidor A, Scheuring CF, Smith BM, Sobral BW, Terry A, Torto-Alalibo

TA, Win J, Xu Z, Zhang H, Grigoriev IV, Rokhsar DS, Boore JL (2006) Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science **313:** 1261-1266

Van de Peer Y, De Wachter R (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Appl Biosci **10:** 569-570

Van den Hoek C, Mann, D.G., and Jahns, H.M. (1995) Algae: An Introduction to Phycology. Cambridge University Press, Cambridge

Vandepoele K, Raes J, De Veylder L, Rouze P, Rombauts S, Inzé D (2002) Genome-wide analysis of core cell cycle genes in Arabidopsis. Plant Cell **14:** 903-916

Vandepoele K, Vlieghe K, Florquin K, Hennig L, Beemster GTS, Gruissem W, Van De Peer Y, Inze D, De Veylder L (2005) Genome-wide identification of potential plant E2F target genes. Plant Physiology **139:** 316-328

Weinberg RA (1995) The retinoblastoma protein and cell cycle control. Cell **81:** 323-330

Zheng N, Fraenkel E, Pabo CO, Pavletich NP (1999) Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. Genes & Development **13:** 666-674

5

CDKA2 is involved in G2-M phase progression in the diatom *Phaeodactylum tricornutum*

Marie J.J. Huysman[1,2,3], Cindy Martens[2,3], Alessandra De Martino[4], Dirk Inzé[2,3], Chris Bowler[4], Wim Vyverman[1] and Lieven De Veylder[2,3]

[1]Protistology and Aquatic Ecology, Department of Biology, Ghent University, 9000 Gent, Belgium
[2]Department of Plant Systems Biology, Flanders Institute for Biotechnology (VIB), 9052 Gent, Belgium
[3]Department of Plant Biotechnology and Genetics, Ghent University, 9052 Gent, Belgium
[4]Département de Biologie, Ecole Normale Supérieure, CNRS, Unité Mixte de Recherche 8186, 75230 Paris Cedex 05, France

Manuscript in preparation

Authors' contributions

MJJH, DI, WV and LDV conceived and designed the study. MJJH performed the experiments, analyzed the data and wrote the chapter. CM helped with the phylogenetic analysis. WV and LDV revised the manuscript. ADM and CB provided overexpression tools, gave advise about genetic transformation and discussed the data with the other authors.

ABSTRACT

Cyclin-dependent kinases are crucial regulators of cell cycle progression in eukaryotes. The diatom *CDKA2* was originally assigned to the classical A-type CDKs, but its cell cycle-dependent transcription at the G2/M transition is typical for plant-specific B-type CDKs. Here, we report the functional characterization of *CDKA2* as a mitotic regulator during the diatom cell cycle. Using a yeast-two-hybrid library screen, CDKA2 was found to interact with both CKS1 and cyclinD1 (CYCD1), two cell cycle regulators transcribed during the G2-M phases. Localization of *CDKA2* was found to be nuclear in interphase cells, while in cells undergoing cytokinesis the signal extended to the cell division plane. Overexpression of *CDKA2* induced overall cell growth rate reduction, due to a prolongation of the mitotic phase. By contrast, overexpression of a kinase-null *CDKA2* allele did not result in any obvious growth defects, illustrating that the phenotype observed in CDKA2 overexpressing cells is primary due to increased kinase activity, rather than through competition with rate-limiting regulatory proteins.

INTRODUCTION

Control of cell cycle progression in eukaryotes is driven by an evolutionary conserved family of serine/threonine kinases, the cyclin-dependent kinases (CDKs). These CDKs form functional heterodimers with regulatory cyclin subunits (Morgan, 1997; Inzé and De Veylder, 2006). Together, CDK and cyclin partners form catalytic complexes that control cell cycle progression through phosphorylation of target proteins, involved in DNA replication and mitosis (Murray, 2004). The activity of CDK/cyclin complexes is regulated at multiple levels, including the interaction with inhibitors or scaffolding proteins, and phosphoregulation of the CDK subunit.

In contrast to fission yeast (*Schizosaccharomyces pombe*) and budding yeast (*Saccharomyces cerevisiae*), where only a single CDK, Cdc2/Cdc28, controls the cell cycle (Hartwell et al., 1974; Nurse and Thuriaux, 1980), animals and plants possess multiple CDKs (Morgan, 1997; Joubes et al., 2000). The most conserved cell cycle regulators possess a typical PSTAIRE cyclin-binding motif (Cdc2/Cdc28 in yeast, Cdk1/Cdk2 in animals and A-type CDKs in plants). In addition, a class of CDKs specific to plants, called the B-type CDKs, has been shown to be involved in cell cycle control as well (Porceddu et al., 2001; Lee et al., 2003; Boudolf et al., 2004; Boudolf et al., 2009). B-type CDKs in higher plants possess a variant of the PSTAIRE motif, either PPTALRE (CDKB1) or PPTTLRE (CDKB2) (Joubes et al., 2000) and, unlike A-type CDKs that are required for both G1/S and G2/M transition, they only play a role at G2/M (Mironov et al., 1999).

With the recent advances in sequencing techniques, several more genomes have become available, including genomes of different algal groups. This wealth of new data makes it possible to study CDKB evolution by comparative genomics. CDKB-like sequences have been identified in different algal species, including in the green algae *Ostreococcus tauri* (Robbens et al., 2005), *Chlamydomonas reinhardtii* (Bisova et al., 2005), *Micromonas species* and *Micromonas pusilla*, in the red alga *Cyanidioschyzon merolae* (Cizkova et al., 2008), and also recently in the brown algae *Ectocarpus siliculosus* (Bothwell et al., 2010). Remarkably, the CDKB-like sequences of *O. tauri* and *C. reinhardtii* have been reported to represent functional homologs of A-type CDKs, mainly by their ability to complement *cdc28* temperature-sensitive yeast mutants (Corellou et al., 2005; Cizkova et al., 2008), indicating that B-type CDKs might originate from a duplication of the A-type CDKs (Boudolf et al., 2005). However, to date, their functionality to complement higher plant B-type CDKs has not yet been investigated.

Phaeodactylum tricornutum is a unicellular marine diatom belonging to the heterokont lineage (De Martino et al., 2007). This diatom multiplies by binary division and, unlike most other diatoms, it lacks a sexual phase during its life cycle, rendering this diatom a perfect model species to study vegetative reproduction. Furthermore, due to the presence of a light-dependent phase during its cell cycle (Brzezinski et al., 1990; Huysman et al., 2010), the cell division process in *P. tricornutum* can easily be synchronized by implementation of alternating light/dark cycles (Huysman et al., 2010). Recently, we identified the core set of cell cycle genes in *P. tricornutum*, including CDKs (Huysman et al., 2010). By phylogenetic analysis, two A-type CDKs were identified, of which CDKA1 shows the classical PSTAIRE motif, while CDKA2 shows a PSTALRE cyclin-binding motif. The latter motif deviates only by one amino acid from the CDKA and CDKB hallmarks. Such a PSTALRE motif has also been found in the *Dictyostelium discoideum* CDC2 homolog (DdCDK1) (Michaelis and Weeks, 1992), the *C. merolae* CDKA protein (Cizkova et al., 2008) and in the *O. tauri* CDKB protein (Robbens et al., 2005). Moreover, unlike typical CDKAs, transcription of the *P. tricornutum CDKA2* gene is cell cycle regulated and shows a peak of transcription at the G2-to-M phase transition (Huysman et al., 2010). However, based on its current phylogenetic position and its transcription pattern it is impossible to define whether *CDKA2* represents a functional ortholog of A- or B-type CDKs.

The main objective of this study was to elucidate the role of CDKA2 during the cell cycle in *P. tricornutum*. First, we reanalyzed the phylogenetic position of CDKA2, this time including representatives of groups more closely related to the diatoms. Next, we determined its subcellular (re)localization during the cell cycle by microscopic analysis of transformant cells expressing a fluorescently-labeled version of CDKA2. In addition, a yeast-two-hybrid library screen was performed to fish out putative interactors of CDKA2, including cyclins. To discover the biological function of CDKA2 during cell division we monitored growth rate and cell cycle progression in CDKA2 overexpression lines. To determine whether the observed growth defects could be attributed to an increased kinase activity upon CDKA2 overexpression, we examined whether the same effects could be observed in transgenic lines expressing a kinase-null version of CDKA2.

RESULTS

Phylogenetic analysis

Considering the ambiguous classification of *CDKA2*, we performed a new phylogenetic analysis to determine the phylogenetic position of *CDKA2* amongst the CDK family members (Figure 5.1). The recent release of genome data from several algal groups more closely related to *P. tricornutum* allowed us to include CDK sequences of these species in the multiple sequence alignment. In the updated phylogenetic tree, CDKA2 forms together with the B-type CDK representatives of *E. siliculosus* a well-supported clade which may represent a stramenopile-specific cluster. However, based on this phylogenetic analysis it remains impossible to detemine whether CDKA2 is more related to the A-type or B-type CDKs (Figure 5.1). The tree indicates rapid radiation of CDK types near the base of the CDKA and CDKB lineage possibly representing a CDK gene diversification phase during the early stages of eukaryotic evolution.

CDKA2 relocalizes from the nucleus to the plane of division during cytokinesis

To determine the localization of CDKA2 during the cell cycle, we generated transgenic lines in which *CDKA2* was fused N-terminally to the fluorescent marker *eYFP* under control of the *fucoxanthin chlorophyll binding protein B* (*fcpB*) promoter (*eYFP-CDKA2*). In interphase cells that contain undivided or divided translocating chloroplasts, the eYFP signal could be observed in the nucleus (Figure 5.2a, 5.2b and 5.2c). This nuclear localization was confirmed by 4',6-diamidino-2-phenylindole (DAPI)-staining (Figure 5.2f). Remarkably, after karyokinesis, when daughter chloroplasts are fully translocated (De Martino et al., Manuscript in preparation), bright fluorescent dots appeared in both daughter nuclei (Figure 5.2d and 5.2f). During cytokinesis the signal extended to the division plane between the two daughter chloroplasts (Figure 5.2e). Following division, CDKA2 relocalized to the nucleus in both daughter cells.

CDKA2 can interact with CKS1 and CYCD1

To identify possible interactors of CDKA2, we conducted two independent yeast-two-hybrid (Y2H) library screens using as bait either the full-length wild-type CDKA2 or a kinase inactive CDKA2 protein (CDKA2.N148). The latter contains a point-mutation ($D^{148}>N^{148}$) in its predicted catalytic site, known to interfere with correct ATP binding, and hence kinase activity

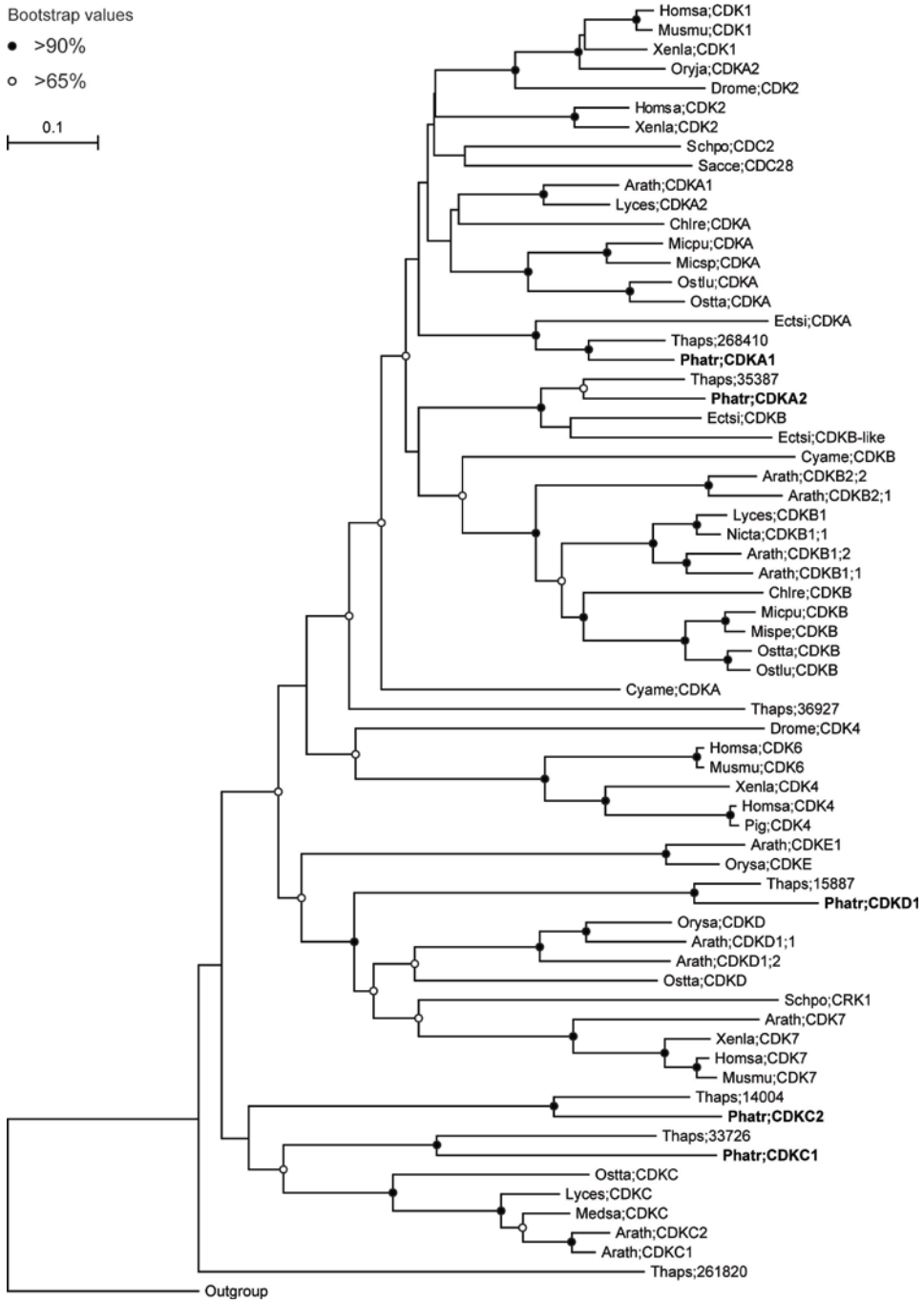


Figure 5.1: Updated phylogenetic analysis of the cyclin-dependent kinases of *P. tricornutum*. Neighbor-joining tree (TREECON, Poisson correction, 1000 replicates) of the CDK family. The *P. tricornutum* sequences are shown in bold. Abbreviations: Arath, *Arabidopsis thaliana*; Cyame, Cyanidioschyzon *merolae*; Drome, *Drosophila melanogaster*; Ectsi, *Ectocarpus siliculosus*; Homsa, *Homo sapiens*; Lyces, *Lycopersicon esculentum*; Medsa, *Medicago sativa*; Micpu, *Micromonas pusilus*; Mispe, *Micromonas species*; Musmu, *Mus musculus*; Nicta, *Nicotiana tabacum*; Oryja, *Oryza japonica*; Orysa, *Oryza sativa*; Ostlu, *Ostreococcus lucimarinus*; Ostta, *Ostreococcus tauri*; Phatr, *Phaeodactylum tricornutum*; Sacce, *Saccharomyces cerevisiae*; Schpo, *Schizosaccharomyces pombe*; Thaps, *Thalassiosira pseudonana*; and Xenla, *Xenopus laevis*. The utgroup is represented by human CDK10.

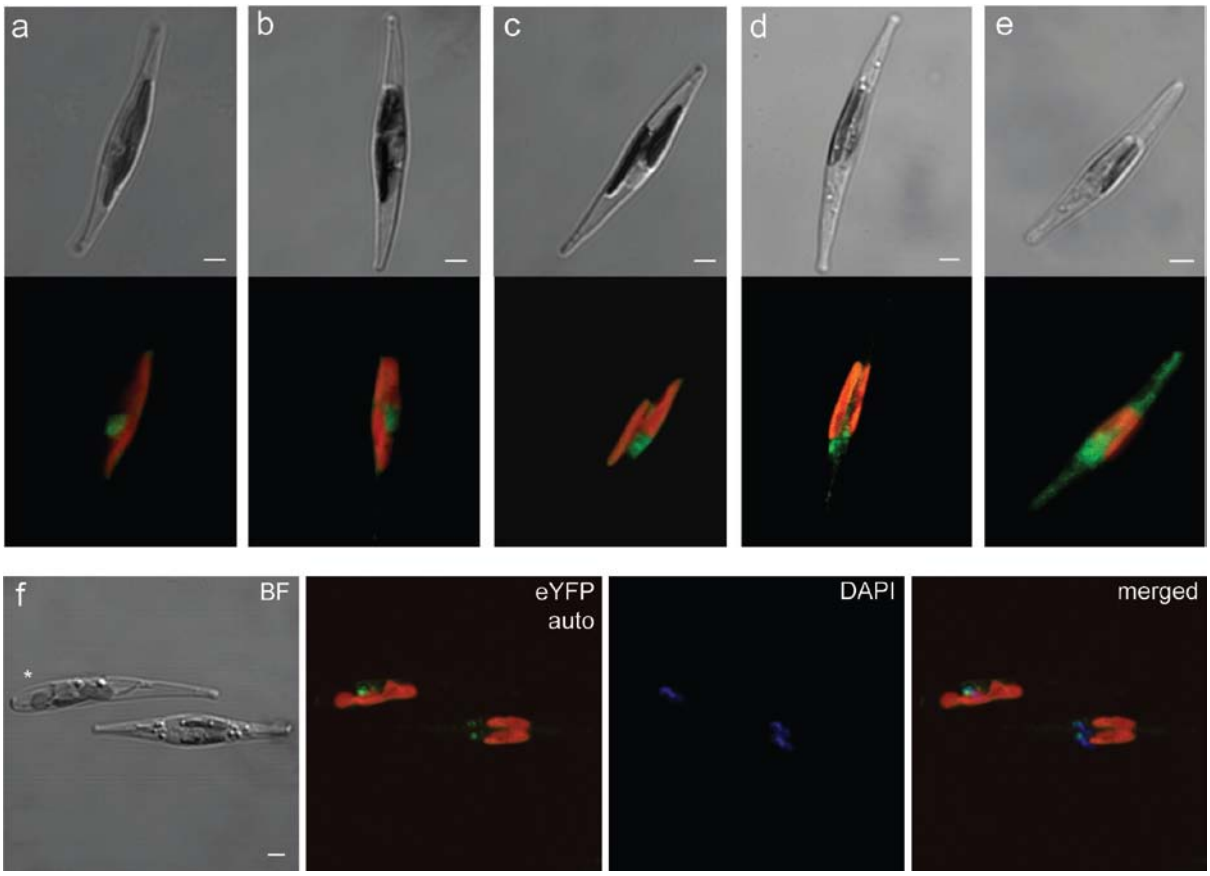


Figure 5.2: Localization of *YFP-CDKA2 in P. tricornutum*. Confocal images of normal fusiform *YFP-CDKA2* overexpression cells during different stages of the cell cycle. **(a)** Interphase cell with undivided chloroplast. **(b)** Interphase cell with twisted chloroplast. **(c)** Interphase cell with translocating chloroplast. **(d)** Cell with divided chloroplast and nucleus. Note the appearance of small bright fluorescent dots inside both daughter nuclei. **(e)** Cell during cytokinesis. At the time of cytokinesis, the signal is no longer prominent in the nucleus but it is also targeted to the cell plate. **(f)** Confirmation of the nuclear localization of the eYFP signal by colocalization of DAPI staining. BF, Bright field; Green, eYFP signal; Red, autofluorescence chloroplast (auto); Blue, DAPI signal. Baseball bat (BBB) shaped cell is indicated by an asterix. Scale bars represent 2µm.

without altering the protein's 3D structure (Taylor et al., 1993; Jeffrey et al., 1995). The screen using the wild-type allele of CDKA2 yielded four possible interactors, including the conservedCDK binding protein CKS1 (Table 5.1). The screen using CDKA2.N148 as a bait yielded seven possible interactors, including two cyclins (cyclin P6 and cyclin D1) (Table 5.1).

To confirm these interactions, pairwise Y2H assays were performed by co-transformation of CDKA2 or CDKA2.N148 fused to the DNA-binding domain of GAL-4 (DBD) and their respective candidate interactors fused to the GAL-4 activation domain (AD). Using CDKA2 as bait, we could only confirm the interaction with CKS1, whereas the assay using CDKA2.N148

Table 5.1: Overview of the results of the Y2H library screens using CDKA2 and CDKA2.N148 as bait.

BAIT	protein ID	Description
CDKA2	Phatr2_53229	CKS1
	EST	unknown1
	Phatr2_28797	PTD9, delta 9 desaturase
	Phatr2_47845	Sulfatase
CDKA2.N148	Phatr2_35894	protein phosphatase 2C-like, PP2C-like
	EST	unknown1
	Phatr2_6231	CYCP6
	Phatr2_50640	Unknown2 (cellulase E4)
	Phatr2_54511	COP beta
	Phatr2_29283	cyclin D1
	Phatr2_6231	CYCP6

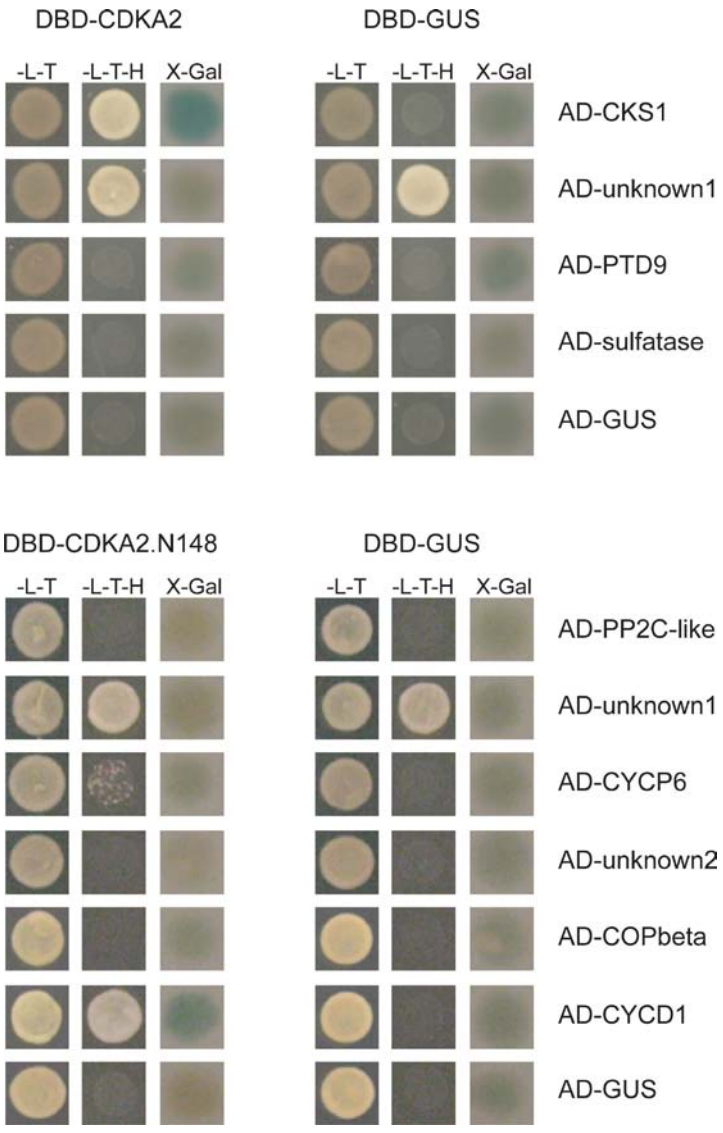


Figure 5.3: Confirmation of the candidate interactors by pairwise Y2H co-transformation assays. Yeast PJ694-alpha cells were co-transformed with bait (DBD) and prey (AD) plasmid as indicated. Co-transformation was analyzed on medium lacking leucine and tryptophan (-L-T). Co-transformants were tested for their ability to activate the histidine marker gene by assessing yeast growth on medium lacking leucine, tryptophan and histidine (-L-T-H) and for their ability to activate the *LacZ* reporter gene (X-Gal). For each combination three independent colonies were screened, of which one is shown.

as bait confirmed CYCD1 as an interactor (Figure 5.3). One candidate interactor (unknown1) was identified as a general false-positive as it was able to activate itself on medium lacking histidine (Figure 5.4). Interestingly, both confirmed interactors (CKS1 and CYCD1) were previously shown to be transcribed mainly during the G2-M phase, coinciding with the expression pattern of CDKA2 (Huysman et al., 2010).

Overexpression of *CDKA2* reduces cell growth rate by interfering with G2-M phase progression

To study the effect of *CDKA2* overexpression on cell cycle progression in *P. tricornutum*, we generated transgenic lines overexpressing *CDKA2* C-terminally fused to *eYFP* (*CDKA2-YFP*). Overexpression of *CDKA2-YFP* was evaluated by transcript analysis of *CDKA2* and *eYFP* using real-time quantitative PCR (real-time qPCR) in wild-type (WT) and transformed cells (Figure 5.4a and 5.4b, respectively). Two lines expressed a two- to three-fold higher level of *CDKA2* transcripts compared to WT cells (CDKA2-YFP A4 and CDKA2-YFP B1, respectively) (Figure 5.4a) due to the overexpression of the *CDKA2-YFP* fusion (Figure 5.4b), while another line (CDKA2-YFP B9) showed no increase in *CDKA2* (Figure 5.4a), rendering this a good internal transformation control. Western blot analysis using a GFP-specific antibody was performed to detect the expression of the fusion protein, and although a weak but clear band at the expected height could be detected for the strongest overexpression line CDKA2-YFP B1, no band could be detected for CDKA2-YFP A4 (Figure 5.4c). The latter might be related to the specificity of the GFP-antibody used or the amount of protein extract that was used. Increasing the protein amount and optimization of the probing conditions might be necessary in order to detect a clear signal on the blot.

To determine if *CDKA2* overexpression resulted in major growth changes, a growth analysis was performed on transgenic cells and control cells. Optical density determination, which can be used as a measurement of cell density, was performed in the morning during nine consecutive days, in lines grown under constant light. Cells expressing increased levels of *CDKA2-YFP* (CDKA2-YFP A4 and CDKA2-YFP B1) showed longer generation times, in a dose-dependent manner, compared to the control cells (WT and CDKA2-YFP B9) (Figure 5.4d). To determine the cell cycle phase that was affected during exponential growth, leading to a decreased growth rate, we measured transcript levels of *cyclinP6* (*CYCP6*), *E2F1* and *cyclinB1* (*CYCB1*) in exponentially growing asynchronous *CDKA2* overexpression and control cells. These genes represent diatom cell cycle marker genes specific for the G1-, S- and G2-M-

phases, respectively (Huysman et al., 2010) (see Chapter 2, Figure 2.6a; and Chapter 4, Figure 4.6). In both *CDKA2* overexpression lines, we detected higher transcript levels of the mitotic marker *CYCB1*, and lower *E2F1* levels compared to those of the control cells (figure 5.4e), indicating that *CDKA2* overexpressing cells spend a longer time in the G2-M phases.

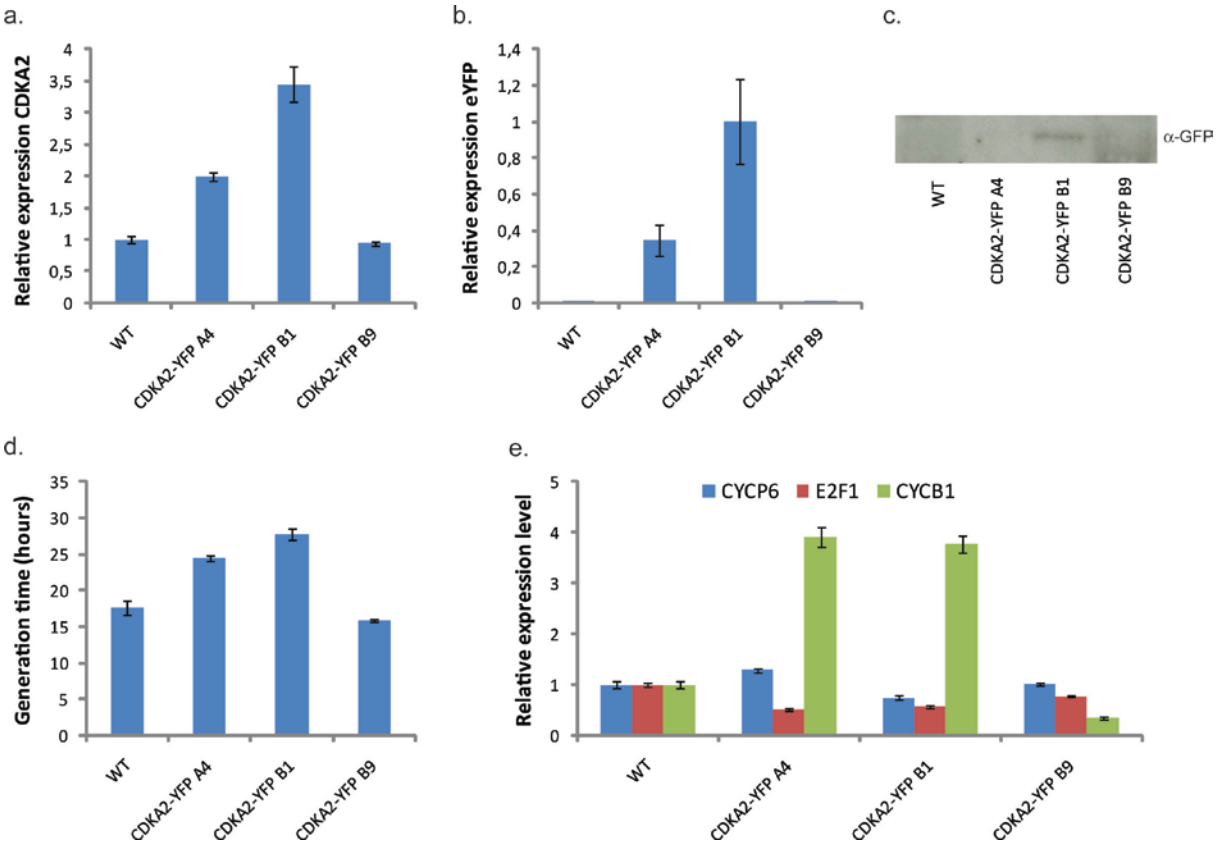


Figure 5.4: Effect of *CDKA2-YFP* overexpression on cell cycle progression. **a.** Real-time qPCR analysis of *CDKA2* transcript levels in WT and transgenic lines. **b.** Real-time qPCR analysis of *eYFP* transcript levels in WT and transgenic lines. **c.** Western blot analysis of CDKA2-YFP protein levels in WT and transgenic lines. **d.** Generation times of overexpression and control lines grown under constant light. **e.** Real-time qPCR analysis of different cell cycle marker genes in overexpression and control lines. **f.** DNA distribution in overexpression and control cells measured by flow cytometric analysis (n=5000). Error bars represent standard errors.

To test whether the observed growth phenotype of the transgenic cells was due to aberrant kinase activity of the CDKA2-YFP fusion protein or due to out titration of essential components, we generated transgenic lines harboring a dominant-negative version of *CDKA2* (CDKA2.N148) under the control of the overexpression *fcpB* promoter. Overexpression of the *CDKA2.N148-YFP* construct in *P. tricornutum* cells was evaluated by real-time qPCR and western blot analysis, and two lines were retained for further experiments (Figure 5.5a, 5.5b and

5.5c). Growth analysis was performed to detect growth effects and, although the selected *CDKA2.N148-YFP* lines expressed up to five-fold levels of the dominant-negative version versus the endogenous gene (Figure 5.5a), no reduction of cell growth rate could be observed for cells overexpressing the dominant-negative fusion compared to control cells (Figure 5.5d). Furthermore, transcript levels of the cell cycle marker genes were comparable to those of untransformed cells (Figure 5.5e). Together, these data indicate that the growth phenotype observed in *CDKA2* overexpressing cells is due to aberrant kinase activity levels in the cell, probably being responsible for misregulation of the G2-M phase progression.

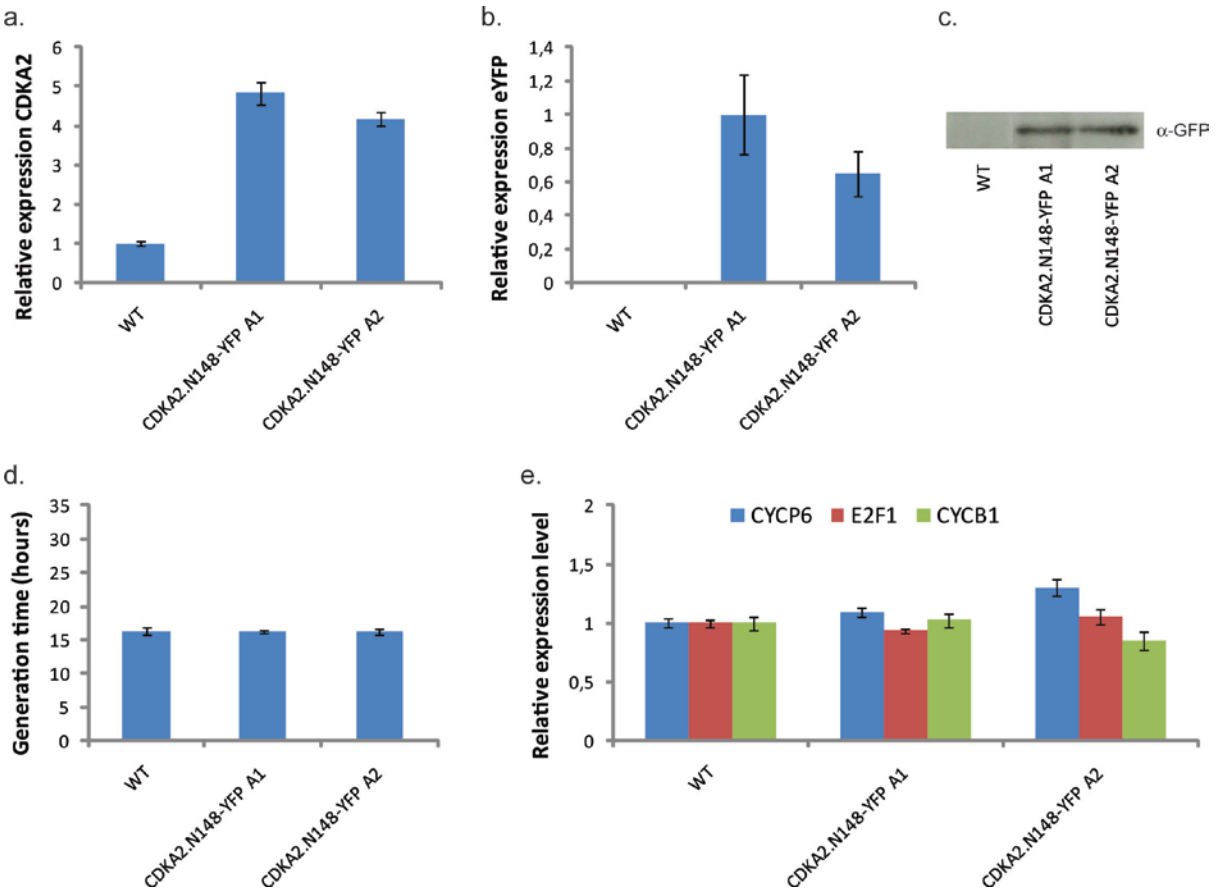


Figure 5.5: Effect of CDKA2.N148-YFP overexpression on cell cycle progression a. Real-time qPCR analysis of *CDKA2* transcript levels in WT and *CDKA2.N148-YFP* overexpressing lines. **b.** Real-time qPCR analysis of *eYFP* transcript levels. **c.** Western blot analysis of CDKA2.N148-YFP protein levels in WT and transgenic lines. **d.** Generation times of overexpression and WT cells grown in constant light. **e.** Real-time qPCR analysis of different cell cycle marker genes in overexpression and control lines. Error bars represent standard errors.

It is worthwhile to note that in several *CDKA2* overexpression lines, a subpopulation could be observed that showed aberrant cell morphology, called the baseball bat (BBB) shaped

cells (Figure 5.2f, indicated with an asterix). However, we were unable to find a clear correlation between the abundance of BBB-cells and the *CDKA2* overexpression level. We observed that within a single culture the frequency of BBB-shaped cells decreased over time, indicating that BBB-cells eventually cannot divide anymore and die. Therefore, the severity of this morphological phenotype might explain the lack of correlation between phenotype and transcript abundance levels, as higher levels of *CDKA2* might induce more severe phenotypes that are counter selected in the population, and hence can be missed during analysis.

DISCUSSION

In this study, we have attempted to elucidate the role of *CDKA2* in the diatom *P. tricornutum*. This CDK was previously assigned to the A-type family, but displays a PSTALRE motif, a motif in between the motifs of the conserved A-type and plant-specific B-type CDKs (Joubes et al., 2000; Huysman et al., 2010). Unlike for classical A-type CDKs, transcription of diatom *CDKA2* was reported to be fluctuating during the cell cycle in synchronized cells, starting to accumulate after S-phase and showing a peak at mitosis (Huysman et al., 2010), suggesting a role for *CDKA2* in the control of mitosis. Reexamination of the phylogenetic position of *CDKA2*, by including the sequences of representatives of more closely related groups in the phylogenetic analysis, assigned *CDKA2* to a cluster of CDKB-like sequences. Furthermore, the hypothesis that *CDKA2* might function as a mitotic regulator is supported by the observation of YFP-CDKA2 localization at the cell division plane in pre-cytokinetic cells. This localization pattern suggests that *CDKA2* might have an upstream regulatory function in the formation of the cell division plate and/or other mitotic structures, or in the recruitment of one or various components to the cell plate. In mammals the activity of PRC1, a microtubule (MT) binding and bundling protein responsible for MT stabilization during cytokinesis, is regulated through phosphorylation by CDKs (Jiang et al., 1998). Also in plant cells, the MT-dependent association of CDKs with mitotic structures and the functional involvement of CDKs in the organization of specific MT arrays has already been reported (Mineyuki et al., 1991; Colasanti et al., 1993; Stals et al., 1997; Ayaydin et al., 2000; Weingartner et al., 2001).

We identified the interaction of CDKA2 with two other cell cycle regulators, CKS1 and cyclinD1, which are both reported to be transcribed mainly at the G2 phase in synchronized

cells (Huysman et al., 2010). The interaction with CKS1 was not surprising as CKS1 is a member of a conserved family of small proteins that are believed to act as docking factors that mediate the interaction of CDKs with regulatory proteins and putative substrates (Pines, 1996). Therefore CKS1 might represent a good bait to identify putative substrates of CDKA2 in further studies. D-type cyclins in plants and animals have mainly been reported to function as integrators of extracellular signals to the cell cycle during the G1-phase (Sherr, 1995; Oakenfull et al., 2002). However, considering its transcription profile, the diatom CYCD1 protein might have acquired an alternative function at the G2-to-M transition. Similar G2-M transcript expression patterns have been reported for D3-type cyclins in tobacco and alfalfa cell cultures (Sorrell et al., 1999; Meszaros et al., 2000) and, moreover, a D-type cyclin has been shown to interact with a mitotic CDK in alfalfa cells (Meszaros et al., 2000). Furthermore, ectopic expression of CYCD3;1 in *A. thaliana* trichomes induces cell division (Schnittger et al., 2002).

Although in other organisms, overexpression of wild-type *CDKA* or *CDKB* sequences generally does not induce any phenotype (Hemerly et al., 1995; Porceddu et al., 2001; Boudolf et al., 2004), we observed major growth defects in *P. tricornutum* cells overexpressing *CDKA2*. Most obviously, the generation time almost doubled, despite on only a three-fold increase in *CDKA2* transcript level. By determination of cell cycle marker gene expression, we found that this delay probably results from a lengthening of the G2-M phases, as *CYCB1* levels were clearly elevated in *CDKA2* overexpressing cells. The study of kinase-null CDK mutants have been proven useful in yeast, animals and plants to assess the function of different CDKs during the cell cycle (Mendenhall et al., 1988; van den Heuvel and Harlow, 1993; Hemerly et al., 1995; Porceddu et al., 2001; Boudolf et al., 2004). Upon overexpression, these mutated alleles induce a dominant-negative effect, probably because of competition with the wild-type proteins for the association with rate-limiting interacting and regulatory proteins (van den Heuvel and Harlow, 1993; Labib et al., 1995). Surprisingly, overexpression of a dominant-negative form of *CDKA2*, *CDKA2.N148*, in *P. tricornutum* did not show any obvious cell cycle phenotype, suggesting that the mitotic arrest observed in *CDKA2* overexpressing cells might result from aberrant levels and timing of CDKA2 kinase activity, rather than from competition of the overexpressor construct with rate-limiting interacting proteins. Therefore, in analogy with other eukaryotes, exit from mitosis in *P. tricornutum* most likely depends on downregulation of CDK activity (Zachariae and Nasmyth, 1999; Weingartner et al., 2004).

Interestingly, all organisms in which B-type CDKs have been identified seem to lack a functional CDC25 phosphatase, except for *O. tauri* (Khadaroo et al., 2004; Robbens et al.,

2005). Despite the presence of almost all regulatory components of the eukaryotic cell cycle in *P. tricornutum*, diatoms also lack a clear *CDC25* phosphatase homolog (Huysman et al., 2010). In yeast and metazoans, CDC25 phosphatases are known to activate CDKs by opposing the activity of WEE1/MYT1/MIK1 family of inhibitory kinases (Perry and Kornbluth, 2007). Phosphorylation of Thr (T14) and/or Tyr (Y15) residues results in CDK inactivation upon activation of the cell cycle checkpoints caused by triggers that should stop the cell cycle, such as DNA damage or mitotic defects. Dephosphorylation of these residues by CDC25 renders the CDK/cyclin complex active and hence stimulates cell cycle progression (Lew and Kornbluth, 1996). Although Tyr phosphorylation in plants is important to arrest the cell cycle under stress conditions, it does not seem to be crucial for G2-M progression (Hemerly et al., 1995; Schuppler et al., 1998; Zhang et al., 2005; De Schutter et al., 2007). The detection of an intriguing number of parallels in transcriptional, biochemical and functional properties of mammalian *CDC25* and plant B-type CDKs, led to the suggestion that the CDC25-mediated regulatory mechanisms might have been replaced in plants by a mechanism governed by the plant-specific B-type CDKs (Boudolf et al., 2006). This would mean that B-type CDKs would probably have arisen from a duplication of the ancestral PSTAIRE-containing CDK in a CDC25-containing organism, followed by specification into A- and B-type CDKs and loss of the *CDC25* gene. The absence of a *CDC25* ortholog in *C. merolae*, *E. siliculosus* and *P. tricornutum* suggests that *CDC25* would have been lost before their divergence. However, the presence of a functional *CDC25* ortholog in *O. tauri* contradicts this possibility. Alternatively, it might be possible that along with the divergence of the green, red and heterokont lineage, there has been a co-evolution and specification of the regulatory mitotic phosphatase. The latter is supported by the high divergence of the *O. tauri CDC25* sequence, that only shows significant similarity to the C-terminal, but not the N-terminal part of the animal and yeast *CDC25* gene family (Khadaroo et al., 2004).

To find out more about the precise regulation of *CDKA2*, more biochemical experiments will be required to determine when its protein is active during the cell cycle by kinase assays and whether its activity is regulated by Tyr phosphorylation. Along with an in-depth cell biological study of the cellular structures in the *CDKA2* overexpressing cells, identification of CDKA2 substrates might help us to determine the precise function of *CDKA2* during diatom mitosis.

MATERIALS AND METHODS

Phylogenetic analysis

Multiple alignments based on CDK amino acid sequences were generated with MUSCLE (Edgar, 2004) and then manually improved with BioEdit (Hall, 1999). To define subclasses within the gene families, phylogenetic trees were built that included reference CDK sequences from animals, plants and algae. TREECON (Van de Peer and De Wachter, 1994) was used to construct the neighbor-joining trees based on Poisson-corrected distances. To test the significance of the nodes, bootstrap analysis was applied using 1000 replicates.

Diatom culture conditions

Phaeodactylum tricornutum (Pt1 8.6; accession numbers CCAP 1055/1 and CCMP2561) was grown in f/2 medium without silica (f/2-Si) (Guillard, 1975) made with filtered and autoclaved sea water collected from the North Sea (Belgium). Cultures were cultivated at 18°C-20°C in a 12-h light/12-h dark regime and 70-100 µmol photons m^{-2} s^{-1}. Liquid cultures were shaken at 100 rpm. For biolistic transformation, *P. tricornutum* cells were grown on solid f/2-Si medium containing 1% Select agar (Sigma).

Cloning constructs

All expression vectors were obtained via the Gateway system (Invitrogen) (Siaut et al., 2007). The full length sequence of *CDKA2* was amplified by PCR with gene-specific forward and reverse primers designed to allow either N- or C- terminal fusion (Table 5.2). The purified fragments were then cloned into the pENTR-D-TOPO vector (Invitrogen) and the obtained ENTRY clone was subsequently recombined into a diatom adapted destination vector for N-terminal or C-terrminal fusion with the eYFP fluorescent marker (Siaut et al., 2007) through attL×attR recombination reaction (Invitrogen). Point mutation of the CDKA2 sequence at the D148 residue (D>N) was performed by in vitro PCR-based mutagenesis using specific primers (Table 5.2).

Table 5.2: CDKA2 cloning primers.

Primer name	Sequence
CDKA2_Fw	CACCATGGAACGTTACCATAAGATAGAAAAG
CDKA2-N_Rv	TCAGATGTTTTCCTTATCCAAGTC
CDKA2-C_Fw	GATGTTTTCCTTATCCAAGTCATCA
CDKA2-DN_Fw	TCTTAAGATAGCGAACTTTGGTCTAGC
CDKA2-DN_Rv	GCTAGACCAAAGTTCGCTATCTTAAGA

Yeast-two-hybrid analysis

Yeast-two-hybrid bait plasmids were generated through recombinational GATEWAY cloning (Invitrogen). Full length open reading frames of the *P. tricornutum CDKA2* and *CDKA2.N148* genes (see above) were recombined in the pDEST32 (bait) vector (Invitrogen) by attL×attR recombination, resulting in translational fusions between the proteins and the GAL4 DNA-binding domains. Plasmids encoding the bait constructs were transformed in the yeast strain PJ694-alpha (MATa; trp1-901, leu2-3,112, ura3-52, his3-200, gal4D, gal80D, LYS2::GAL1-HIS3, GAL2-ADE2, met2GAL7-lacZ) by the LiAc method (Gietz et al., 1992). A library screen was performed using a custom-made Y2H cDNA library (Invitrogen). Yeast cells co-transformed with interaction combinations were selected on synthetic dextrose (SD) plates lacking Leu, Trp and His. Growing colonies were restreaked on SD medium lacking Leu, Trp and His, and plasmid was purified from patches using the Zymoprep I Yeast Plasmid Minipreparation Kit (Zymo Research) according to the manufacturers' instructions. Yeast plasmid was used as a template in a PCR reaction with primers flanking the gateway cloning site of pDEST22 (pDEST22_Fw: TATAACGCGTTTGGAATCACT and pDEST22_Rv: AGCCGACAACCTTGATTGGAGAC), and the obtained product was sequenced and blasted against the *P. tricornutum* genome database to identify the interactor (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html).

Biolistic transformation

Constructs were introduced into *P. tricornutum* by microparticle bombardment as previously described (Falciatore et al., 1999). The pAF6 plasmid was used to confer resistance to phleomycin (Falciatore et al., 1999). Transformants were initially selected based on their ability to grow on medium containing phleomycin (100µg/ml final concentration). Individual resistant colonies were both restreaked on f/2-Si agar plates and grown in liquid f/2-Si medium without antibiotics for further analysis.

Microscopic analysis

Images were obtained with a scanning confocal microscope 100 M (Zeiss) equipped with the software package LSM510 version 3.2 (Zeiss) and a C-Apochromat 63x (1.2 NA) water-corrected objective. eYFP fluorescence was excited with Argon illumination (488 nm) and chlorophyll autofluorescence with HeNe illumination (543 nm). DAPI (4',6-diamidino-2-phenylindole) staining was visualized by illumination at 351 nm.

Real-time quantitative PCR

For RNA extraction, 5 x 10^7 cells were collected, fast frozen in liquid nitrogen and stored at -70ºC. Cell lysis and RNA extraction was performed using TriReagent (Molecular Research Center, Inc., Cincinnati, OH, USA) according to the manufacturer's instructions. Contaminating genomic DNA was removed by DNaseI treatment (Promega). To assess RNA concentration and purity, spectrophotometry was used (NaNodrop ND-1000, Wilmington, DE). Total RNA was reverse transcribed using iScript reverse transcriptase (Roche). Finally, 10 ng of cDNA was used as template in each qPCR reaction.

Samples in triplicate were amplified on the Lightcycler 480 platform with the Lightcycler 480 SYBR Green I Master mix (Roche Applied Science), in the presence of 0.5 µM gene-specific primers (eYFP_Fw: TGCTTCGCCCGCTACCC and eYFP_Rv: ATGTTGCCGTCCTCCTTGAAG; other primers see Huysman et al., 2010 and Chapter 4). The cycling conditions were 10 min polymerase activation at 95°C and 45 cycles at 95°C for 10 s, 58°C for 15 s and 72°C for 15 s. Amplicon dissociation curves were recorded after cycle 45 by heating from 65°C to 95°C. In qBase (Hellemans et al., 2007), data were analyzed using the ΔC_t relative quantification method with the stably expressed *EF1a* and *TubA* as normalization genes (Siaut et al., 2007).

Western blot analysis

Proteins were extracted by adding 200 µl Laëmli buffer to frozen cells (50 * 10^6 cells) and incubation of cell lysates during 15 minutes on ice. Protein extracts (15 µg) were resolved on 12% SDS-PAGE gels, and transferred to nitrocellulose membranes (Millipore) using the wet-blot method. The YFP-fusion proteins were detected by incubating proteins transferred to nitrocellulose membranes for 1h with a 1:3,000 dilution of anti-GFP primary monoclonal antibody (Rockland) at room temperature, followed by 1h incubation in a 1:10,000 dilution of

horseradish peroxidase (HRP) anti-mouse secondary antibody at room temperature. Signals were detected using the Western Lightning^{TM} detection kit (Pierce).

Growth analysis

To monitor growth, cells were grown at constant illumination in a 24-well plate (Falcon), in a total volume of 1 ml, over a time period of nine days. Absorbances of the cultures were measured at 405 nm using the VICTOR^3 Multilabel Plate Reader (Perkin-Elmer) each day in the morning. Obtained growth curves of triplicate cultures were LN(2)-transformed and average generation times were calculated by determination of the derivative of the values between the points of maximal slope (exponential growth phase).

REFERENCES

Ayaydin F, Vissi E, Meszaros T, Miskolczi P, Kovacs I, Feher A, Dombradi V, Erdodi F, Gergely P, Dudits D (2000) Inhibition of serine/threonine-specific protein phosphatases causes premature activation of cdc2MsF kinase at G2/M transition and early mitotic microtubule organisation in alfalfa. Plant Journal **23:** 85-96

Bisova K, Krylov DM, Umen JG (2005) Genome-wide annotation and expression profiling of cell cycle regulatory genes in Chlamydomonas reinhardtii. Plant Physiol **137:** 475-491

Bothwell JH, Marie D, Peters AF, Cock JM, Coelho SM (2010) Role of endoreduplication and apomeiosis during parthenogenetic reproduction in the model brown alga Ectocarpus. The New phytologist **188:** 111-121

Boudolf V, Barroco R, Engler Jde A, Verkest A, Beeckman T, Naudts M, Inzé D, De Veylder L (2004) B1-type cyclin-dependent kinases are essential for the formation of stomatal complexes in Arabidopsis thaliana. The Plant cell **16:** 945-955

Boudolf V, Inzé D, De Veylder L (2006) What if higher plants lack a CDC25 phosphatase? Trends Plant Sci **11:** 474-479

Boudolf V, Lammens T, Boruc J, Van Leene J, Van Den Daele H, Maes S, Van Isterdael G, Russinova E, Kondorosi E, Witters E, De Jaeger G, Inzé D, De Veylder L (2009) CDKB1;1 forms a functional complex with CYCA2;3 to suppress endocycle onset. Plant Physiology **150:** 1482-1493

Brzezinski MA, Olson RJ, Chisholm SW (1990) Silicon availability and cell-cycle progression in marine diatoms. Marine Ecology-Progress Series **67:** 83-96

Cizkova M, Pichova A, Vitova M, Hlavova M, Hendrychova J, Umysova D, Galova E, Sevcovicova A, Zachleder V, Umen JG, Bisova K (2008) CDKA and CDKB kinases from Chlamydomonas reinhardtii are able to complement cdc28 temperature-sensitive mutants of Saccharomyces cerevisiae. Protoplasma **232:** 183-191

Colasanti J, Cho SO, Wick S, Sundaresan V (1993) Localization of the Functional p34cdc2 Homolog of Maize in Root Tip and Stomatal Complex Cells: Association with Predicted Division Sites. The Plant cell **5:** 1101-1111

Corellou F, Camasses A, Ligat L, Peaucellier G, Bouget FY (2005) Atypical regulation of a green lineage-specific B-type cyclin-dependent kinase. Plant Physiol **138:** 1627-1636

De Martino A, Amato A, Meichenin A, Montsant A, Mathieu B, Huysman MJJ, De Veylder L, Vyverman W, Bowler C (Manuscript in preparation) Dissection of mitotic features in a pennate diatom using molecular tools: *Phaeodactylum tricornutum* as a new model for cell division.

De Martino A, Meichenin A, Shi J, Pan KH, Bowler C (2007) Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions. Journal of Phycology **43:** 992-1009

De Schutter K, Joubes J, Cools T, Verkest A, Corellou F, Babiychuk E, Van Der Schueren E, Beeckman T, Kushnir S, Inzé D, De Veylder L (2007) Arabidopsis WEE1 kinase controls cell cycle arrest in response to activation of the DNA integrity checkpoint. Plant Cell **19:** 211-225

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:** 1792-1797

Falciatore A, Casotti R, Leblanc C, Abrescia C, Bowler C (1999) Transformation of Nonselectable Reporter Genes in Marine Diatoms. Mar Biotechnol (NY) **1:** 239-251

Gietz D, St Jean A, Woods RA, Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res **20:** 1425

Guillard RRL (1975) Culture of phytoplankton for feeding marine invertebrates. *In* WL Smith, MH Canley, eds, Culture of Marine Invertebrate animals. Plenum Press, New York, pp 29-60

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser **41:** 95-98

Hartwell LH, Culotti J, Pringle JR, Reid BJ (1974) Genetic control of the cell division cycle in yeast. Science **183:** 46-51

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol **8:** R19

Hemerly A, Engler Jde A, Bergounioux C, Van Montagu M, Engler G, Inzé D, Ferreira P (1995) Dominant negative mutants of the Cdc2 kinase uncouple cell division from iterative plant development. The EMBO journal **14:** 3925-3936

Huysman MJJ, Martens C, Vandepoele K, Gillard J, Rayko E, Heijde M, Bowler C, Inzé D, Van de Peer Y, De Veylder L, Vyverman W (2010) Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. Genome Biol **11:** R17

Inzé D, De Veylder L (2006) Cell cycle regulation in plant development. Annu Rev Genet **40:** 77-105

Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massague J, Pavletich NP (1995) Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature **376:** 313-320

Jiang W, Jimenez G, Wells NJ, Hope TJ, Wahl GM, Hunter T, Fukunaga R (1998) PRC1: a human mitotic spindle-associated CDK substrate protein required for cytokinesis. Molecular cell **2:** 877-885

Joubes J, Chevalier C, Dudits D, Heberle-Bors E, Inzé D, Umeda M, Renaudin JP (2000) CDK-related protein kinases in plants. Plant Mol Biol **43:** 607-620

Khadaroo B, Robbens S, Ferraz C, Derelle E, Eychenie S, Cooke R, Peaucellier G, Delseny M, Demaille J, Van de Peer Y, Picard A, Moreau H (2004) The first green lineage cdc25 dual-specificity phosphatase. Cell Cycle **3:** 513-518

Labib K, Moreno S, Nurse P (1995) Interaction of cdc2 and rum1 regulates Start and S-phase in fission yeast. Journal of Cell Science **108 (Pt 10):** 3285-3294

Lee J, Das A, Yamaguchi M, Hashimoto J, Tsutsumi N, Uchimiya H, Umeda M (2003) Cell cycle function of a rice B2-type cyclin interacting with a B-type cyclin-dependent kinase. The Plant journal : for cell and molecular biology **34:** 417-425

Lew DJ, Kornbluth S (1996) Regulatory roles of cyclin dependent kinase phosphorylation in cell cycle control. Current opinion in cell biology **8:** 795-804

Mendenhall MD, Richardson HE, Reed SI (1988) Dominant negative protein kinase mutations that confer a G1 arrest phenotype. Proceedings of the National Academy of Sciences of the United States of America **85:** 4426-4430

Meszaros T, Miskolczi P, Ayaydin F, Pettko-Szandtner A, Peres A, Magyar Z, Horvath GV, Bako L, Feher A, Dudits D (2000) Multiple cyclin-dependent kinase complexes and phosphatases control G2/M progression in alfalfa cells. Plant Molecular Biology **43:** 595-605

Michaelis C, Weeks G (1992) Isolation and characterization of a cdc 2 cDNA from Dictyostelium discoideum. Biochim Biophys Acta **1132:** 35-42

Mineyuki Y, Yamashita M, Nagahama Y (1991) P34cdc2 Kinase Homolog in the Preprophase Band. Protoplasma **162:** 182-186

Mironov VV, De Veylder L, Van Montagu M, Inzé D (1999) Cyclin-dependent kinases and cell division in plants- the nexus. The Plant cell **11:** 509-522

Morgan DO (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu Rev Cell Dev Biol **13:** 261-291

Nurse P, Thuriaux P (1980) Regulatory genes controlling mitosis in the fission yeast Schizosaccharomyces pombe. Genetics **96:** 627-637

Oakenfull EA, Riou-Khamlichi C, Murray JA (2002) Plant D-type cyclins and the control of G1 progression. Philosophical transactions of the Royal Society of London. Series B, Biological sciences **357:** 749-760

Perry JA, Kornbluth S (2007) Cdc25 and Wee1: analogous opposites? Cell Div **2:** 12

Pines J (1996) Cell cycle: reaching for a role for the Cks proteins. Curr Biol **6:** 1399-1402

Porceddu A, Stals H, Reichheld JP, Segers G, De Veylder L, Barroco RP, Casteels P, Van Montagu M, Inzé D, Mironov V (2001) A plant-specific cyclin-dependent kinase is involved in the control of G2/M progression in plants. The Journal of biological chemistry **276:** 36354-36360

Robbens S, Khadaroo B, Camasses A, Derelle E, Ferraz C, Inzé D, Van de Peer Y, Moreau H (2005) Genome-wide analysis of core cell cycle genes in the unicellular green alga Ostreococcus tauri. Mol Biol Evol **22:** 589-597

Schnittger A, Schobinger U, Bouyer D, Weinl C, Stierhof YD, Hulskamp M (2002) Ectopic D-type cyclin expression induces not only DNA replication but also cell division in Arabidopsis trichomes. Proc Natl Acad Sci U S A **99:** 6410-6415

Schuppler U, He PH, John PCL, Munns R (1998) Effect of water stress on cell division and cdc2-like cell cycle kinase activity in Wheat leaves (vol 117, pg 667, 1998). Plant Physiology **117:** 1528-1528

Sherr CJ (1995) D-type cyclins. Trends in biochemical sciences **20:** 187-190

Siaut M, Heijde M, Mangogna M, Montsant A, Coesel S, Allen A, Manfredonia A, Falciatore A, Bowler C (2007) Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum. Gene **406:** 23-35

Sorrell DA, Combettes B, Chaubet-Gigot N, Gigot C, Murray JA (1999) Distinct cyclin D genes show mitotic accumulation or constant levels of transcripts in tobacco bright yellow-2 cells. Plant Physiology **119:** 343-352

Stals H, Bauwens S, Traas J, Van Montagu M, Engler G, Inzé D (1997) Plant CDC2 is not only targeted to the pre-prophase band, but also co-localizes with the spindle, phragmoplast, and chromosomes. FEBS Lett **418:** 229-234

Taylor SS, Knighton DR, Zheng J, Sowadski JM, Gibbs CS, Zoller MJ (1993) A template for the protein kinase family. Trends in biochemical sciences **18:** 84-89

Van de Peer Y, De Wachter R (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Appl Biosci **10:** 569-570

van den Heuvel S, Harlow E (1993) Distinct roles for cyclin-dependent kinases in cell cycle control. Science **262:** 2050-2054

Weingartner M, Binarova P, Drykova D, Schweighofer A, David JP, Heberle-Bors E, Doonan J, Bogre L (2001) Dynamic recruitment of Cdc2 to specific microtubule structures during mitosis. The Plant cell **13:** 1929-1943

Weingartner M, Criqui MC, Meszaros T, Binarova P, Schmit AC, Helfer A, Derevier A, Erhardt M, Bogre L, Genschik P (2004) Expression of a nondegradable cyclin B1 affects plant development and leads to endomitosis by inhibiting the formation of a phragmoplast. The Plant cell **16:** 643-657

Zachariae W, Nasmyth K (1999) Whose end is destruction: cell division and the anaphase-promoting complex. Genes & Development **13:** 2039-2058

Zhang KR, Diederich L, John PCL (2005) The cytokinin requirement for cell division in cultured Nicotiana plumbaginifolia cells can be satisfied by yeast cdc25 protein tyrosine phosphatase. Implications for mechanisms of cytokinin response and plant development. Plant Physiology **137:** 308-316

6

General discussion and future perspectives

General discussion

In this thesis we aimed to unravel the physiological and molecular mechanisms that underlie the regulation of the diatom cell cycle. Diatoms can live and dominate in rapidly changing and sometimes extreme environments, suggesting that they possess the ability to adapt and survive under such conditions. Their complex evolutionary history, as illustrated by the recent genome sequencing of two diatom species (Armbrust et al., 2004; Bowler et al., 2008) most likely contributes to their high versatility towards changing conditions. Diatoms derived from a secondary endosymbiotic event, involving a photosynthetic red alga and a 'green' eukaryotic host cell (Li et al., 2006; Moustafa et al., 2009). Moreover, a substantial set of genes of bacterial origin was found in diatoms and phylogenomic analyses indicated that horizontal gene transfer is pervasive in diatoms (Bowler et al., 2008). The complex origin and combination of their genome repertoire has most likely provided the diatoms with novel metabolic capacities and possibilities for the perception and transduction of environmental signals.

In order to decipher the ecological success of diatoms and to be able to fully exploit their commercial value, a better understanding of diatom biology, and in particular of the mechanisms controlling diatom life cycle strategies, is needed. Our current understanding of the regulation of eukaryotic cell division relies mainly on studies in model systems, such as yeast, mammalian and plant cells (Morgan, 1997; Inzé and De Veylder, 2006; Lloyd and Chan, 2006; Doonan and Kitsios, 2009), representing only two of the eight major eukaryotic groups (Baldauf, 2003). Although diatoms have long been the subject of cell division studies since Lauterborn's remarkable microscopic observations of diatom mitosis at the end of the 19th century (Lauterborn, 1896), only recently significant progress has been made to uncover the molecular secrets of diatom cell cycle regulation (Hogan et al., 1992; Thamatrakoln and Hildebrand, 2007; Gillard et al., 2008; De Martino et al., 2009).

With the genome sequences of two diatom species (*Thalassiosira pseudonana* and *Phaeodactylum tricornutum*) available at the onset of this thesis work (Armbrust et al., 2004; Bowler et al., 2008), we were able to investigate the molecular conservation of the cell cycle machinery in diatoms (Figure 6.1). A common basis with other eukaryotic cell cycle regulatory systems was found, through the identification of conserved cell cycle regulators, including cyclin-dependent kinases (CDKs), cyclins and regulatory molecules (Chapter 2) (Doonan and Kitsios, 2009), and the presence of a functional Rb-pathway (Chapter 4). In addition, by genome-wide analysis we detected some interesting diatom-specific features of cell cycle regulation. We found that the major expansion of the cyclin gene family, as already reported by Bowler et al. (2008), was specific to diatoms and not shared by any closely related group of organisms, and, by phylogenetic analysis we identified a novel class of diatom-specific cyclins within the cyclin gene family (Chapter 2).

A diatom cell cycle study toolkit

To be able to study cell cycle related transcription of the annotated diatom cell cycle genes, a robust cell cycle synchronization method for our model organism *P. tricornutum* was established. Because in general both nutrient and light deprivation causes cell cycle arrest in diatoms (Olson et al., 1986; Vaulot et al., 1986; Vaulot et al., 1987; Brzezinski et al., 1990), both strategies could potentially be used to naturally synchronize the diatom cell cycle, without the need of chemical cell cycle inhibitors (Planchais et al., 2000), which first need optimization of concentration which might cause stress to the cells. The only nutrient that has been reported to induce a uniform arrest in diatoms is silicon (Vaulot et al., 1987), but as *P. tricornutum* is one of the few species without silicon requirement we did not consider this option. On the other hand, *P. tricornutum* displays a single light-dependent segment during its cell cycle situated at the G1-phase (Brzezinski et al., 1990). As expected, light limitation perturbed G1-phase progression in *P. tricornutum* and reillumination resulted in immediate and synchronous release of the cell cycle arrest (Chapter 2). Such a light-induced synchronization method has been proven useful to synchronize cell division in *S. robusta* as well (Gillard et al., 2008). Cytological observations of the chloroplast conformation during synchronized growth allowed us to determine the specific timing of chloroplast division and plastid movements during the cell cycle (Figure 6.1). Therefore, chloroplast cytology might be used as an easy and fast tool to identify the cell cycle phase of a diatom cell (Chapter 2) (De Martino et al., Manuscript in preparation). This method can complement other more

laborious techniques, such as flow cytometry or quantitative epifluorescence microscopy, to determine the cell cycle stage (von Dassow et al., 2008). In addition, applying this synchronization method for monitoring the transcription of the annotated diatom cell cycle genes during the cell cycle, provided us with a set of cell cycle marker genes that can be used both in the laboratory and in the field (Chapter 2). The development of molecular cell cycle probes might be valuable to assess and monitor diatom growth events, either species- or group-specific, in nature, and might for example help to detect early bloom stages of toxin producing species (Diercks et al., 2008). The type of probes that can be designed, either universal or specific, and their applicability will largely depend on the degree of sequence conservation of the genes of interest. In the case of the diatom-specific cyclins, development of universal probes might be difficult given their high sequence divergence, but in turn, they might be interesting candidates for the evaluation of species-specific responses to environmental stresses.

Diatom-specific cyclins link the environment to growth

During the genomic analysis a novel class of diatom-specific cyclins was identified (Chapter 2). Their principal transcription at the G1 phase, together with a remarkable lack of typical D-type cyclins in diatoms, suggested that these diatom-specific cyclins might be involved in coupling extracellular signals to the cell cycle machinery. This hypothesis was corroborated by the transcription of some of these cyclins upon phosphate and silicon availability (Chapter 2), and the strict light-regulation of *dsCYC2* (Chapter 3). The sequential timing of their transcription during a synchronized cell cycle, starting with the light-regulated *dsCYC2* and followed by the nutrient-dependent *dsCYCs*, combined with the persistent cell cycle arrest upon nutrient repletion of starved cells during darkness, suggests that the light control checkpoint precedes nutrient control in *P. tricornutum*, in contrast to what has been reported for *Thalassiosira weissflogii* (Olson and Chisholm, 1983). The hierarchy of these checkpoints might relate to the different spectral properties of their habitats. *P. tricornutum* is mostly found in dynamic and turbid estuarine waters where changes in light intensity occur much faster than in open-ocean waters, while *T. weissflogii* represents a coastal species experiencing more intermediately stable irradiances (MacIntyre et al., 2000; Lavaud et al., 2007).

Because light is a crucial factor for a photoautotrophic organism like *P. tricornutum*, we further functionally characterized *dsCYC2* to determine its precise role during the cell

cycle. *dsCYC2* was found to be highly transcribed within minutes after illumination of dark-arrested cells, readily followed by its translation. However, transcript levels decreased rapidly upon longer light exposure. By reporter line analysis, we showed that this strict regulation most likely occurs through regulation of *dsCYC2* promoter activity. In addition, a transcriptional repressor probably assures that *dsCYC2* is only transcribed during a defined time-window. Overexpression and silencing of *dsCYC2* both provoked subtle to serious effects on cell cycle progression, respectively, indicating that the controlled timing of *dsCYC2* expression is of major importance. The latter probably contributes to resetting the cells to become sensitive for a new light signal after a period of darkness. The differential expression levels of *dsCYC2* during constant light versus light-dark cycles suggest that in order to reset the light-sensitivity of the cells, the cells would require rapid down regulation of *dsCYC2* by the light-induced repressor during the light, as well as the inactivation of this repressor during darkness. Recently, *dsCYC2* was shown to be under control of a blue light sensor (CPF1) (Coesel et al., 2009). Therefore it would be interesting to test whether other recently discovered diatom light sensors or light receptors (Nymark et al., 2009; Bailleul et al., 2010) can regulate *dsCYC2* expression, and if different light quantities and qualities can alter its transcription. Determination of the sequence crucial to this regulation, presumably a light-responsive element located within the promoter of *dsCYC2*, might be of great importance to further unravel this light signaling pathway. Because co-expressed genes might be under control of similar promoter-based regulation, computational promoter analysis of a set of co-expressed genes, e.g. upon different light conditions or the dark-to-light transition, could lead to the identification of putative light-regulatory sequences. Such analyses have been proven useful in several organisms for the discovery of various regulatory elements (Reviewed in MacIsaac and Fraenkel, 2006) and even complete regulatory networks (Pilpel et al., 2001). Complementary to a co-expression analysis, phylogenetic footprinting, a technique that identifies conserved regions based on evolutionary relationships of species, could facilitate the search for specific regulatory motifs (Zhang and Gerstein, 2003). Although it is clear that *dsCYC2* plays a crucial role in the activation of cell cycle progression after dark arrest in *P. tricornutum*, it remains to be elucidated how well this mechanism is conserved among different diatom groups or even different species. We have shown in Chapter 2 that the diatom-specific cyclins form a large gene family with high sequence divergence and therefore it might be difficult to predict orthologs by sequence identity. However, inspection of the recently sequenced genome of the pennate diatom *Fragilariopsis cyclindrus* identified a good

candidate ortholog of *dsCYC2* showing an overall amino acid identity of 35% (data not shown). Further in-depth functional analysis will be required to assess the conservation of this light-dependent cell cycle regulator in other diatoms.

A conserved, yet specific G1/S control

In addition to the diatom-specific cyclins, we found another specific feature of diatom cell cycle regulation, related to the Rb-mediated pathway controlling the G1-to-S transition (Weinberg, 1995; de Jager and Murray, 1999) (Chapter 4). Using a *de novo* motif search approach, we predicted a diatom-specific E2F motif (sGCsGGCs) that resembles the plant/animal E2F consensus (wTTssCss) in its core, but lacks the typical T-rich prefix (Tao et al., 1997; Vandepoele et al., 2005). Cloning of the different *E2F* and *DP* genes for functional analysis, revealed that multiple alleles exist for *E2F1* and *DP1*. Interaction studies revealed that E2F2 could associate with both DP proteins, whatever type of DP1 was tested, providing evidence that none of the amino acid changes present in the different DP1 types interfere with its ability to bind E2F2. Furthermore, we showed that co-expression of *E2F2* with *DP1b* resulted in the transactivation of promoters of candidate E2F target genes (*PCNA* and *MCM3*) that hold the predicted diatom E2F motif. To further test the necessity and sufficiency of the predicted diatom E2F *cis*-regulatory element to transactivate E2F target gene expression, this reporter assay should be repeated by using an artificial minimal promoter sequence containing only the diatom E2F motif. In addition, it should be considered that at this moment we only demonstrated transcriptional activity of the E2F2/DP1b complex in a heterologous system, and therefore its function should be confirmed in diatom cells by e.g. overexpression or reporter assays.

***CDKA2*: Making the final cut**

Although *CDKA2* was originally assigned to the classical PSTAIRE-containing A-type of CDKs (Chapter 2), it showed some characteristics inconsistent with other members of this class. First, CDKA2 holds a PSTALRE cyclin-binding domain, a hallmark that deviates by only a single amino acid residue from those of the plant A-type (PSTAIRE) and B-type (PPTA/TLRE) CDKs. Furthermore, its cell cycle phase-dependent transcription at the G2-to-M transition and its localization at the cell division plane during cytokinesis, suggested a specific role for *CDKA2* during mitosis. This was supported by its association with CYCD1 and CKS1, two G2-M expressed cell cycle regulators (Chapter 2), and the observation that

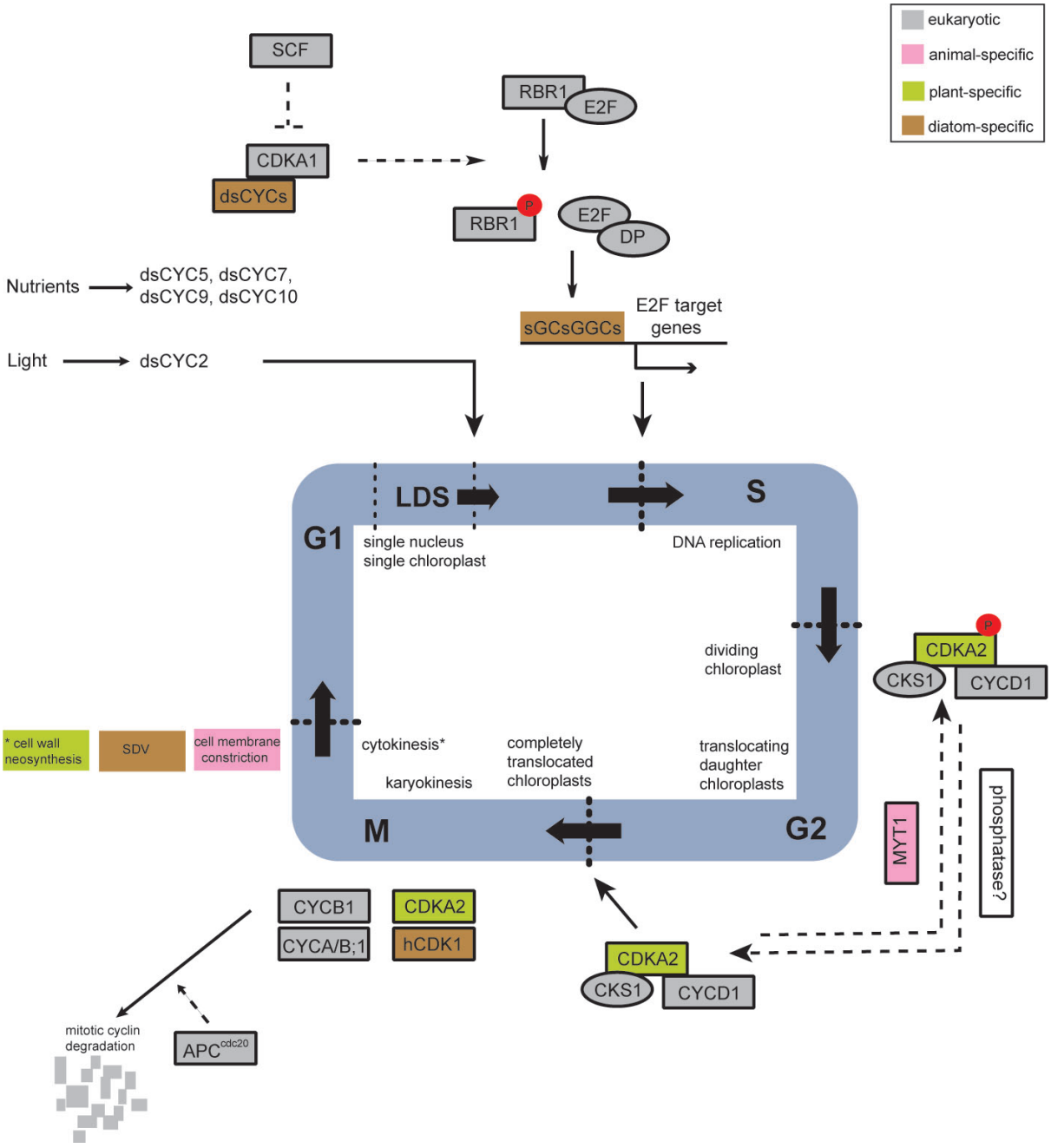


Figure 6.1: Overview of the cell cycle regulation and conservation in *P. tricornutum*. Start of the cell cycle at the G1-phase is initiated by the presence of non-limiting amounts of external stimulatory factors, including light and nutrients. Perception of these growth stimuli induces an intracellular signaling cascade that leads to the transcription of several diatom-specific cyclins (*dsCYCs*). These dsCYCs subsequently interact and activate G1-expressed CDKs, such as the conserved PSTAIRE-containing CDKA1. The active CDK/cyclin complex most likely phosphorylates, either directly or indirectly (through activation of other CDK/cyclin complexes), the RBR1 protein, resulting in the release and activation of the E2F transcription factor. E2F together with DP binds and transactivates a diatom-specific *cis*-regulatory E2F motif in the promoter of E2F target genes, leading to S-phase progression. After DNA synthesis has completed, a mitotic CDK, *CDKA2* is expressed, along with its interaction partners CYCD1 and CKS. In analogy with the DNA damage checkpoint known in other eukaryotes,

this complex is probably kept inactive during S-G2 through inhibitory phosphorylation by the MYT1 kinase. Possibly, this inhibitory phosphorylation is later removed by a yet to be specified phosphatase. Activation of CDKA2 leads to the progression of G2-M and finally results in nulear division (karyokinesis) and cell division (cytokinesis). Although not discussed in the thesis, note that the major cell cycle regulatory degradation complexes, the Skp1/Cdc53/F-box protein (SCF) complex and the anaphase promoting complex (APC), were also identified and have been found to be transcribed at the G1-S and M phases (Huysman et al., unpublished data), respectively, suggesting that they might control CDK/cylin activity at these phases. Cytokinesis involves the centripetal invagination of the plasma membrane, and the centrifugal neosynthesis of a new valve within the silica deposition vesicle (SDV). Chloroplast division and translocation both precede karyokinesis and cytokinesis. Conserved regulators or regulatory pathways in eukaryotes are indicated in grey, characteristics common with animals in pink, features common with plants in green, and diatom-specific properties in brown. Arrows and dashed lines indicate activation and hypothetical regulations, respectively. LDS, light-dependent segment.

overexpression of *CDKA2* resulted in a general increase in generation time, mainly due to a prolongation of the G2-M phases. The characterization of this putative B-type ortholog in an organism that lacks a functional CDC25 phophatase (see Chapter 2) suggests that B-type CDKs probably arose from a duplication of the ancestral PSTAIRE-containing CDK and that along with the diversification of the different lineages, the regulatory mitotic phosphatase has co-evolved. Next to the presence of this functional B-type homolog in diatoms, other CDKB-like sequences have been reported in *E. siliculosus*, another heterokont algae (Bothwell et al., 2010). This brings up the question whether the definition of B-type CDKs being strictly plant-specific still holds or whether it needs revision, and perhaps could be broadened to 'plant-alga'-specific.

Future directions

Although the findings discussed in this thesis along with the results of the cytological and physiological studies from the past, definitely allow us to fill in some major gaps in the understanding of the diatom cell and life cycle regulation, many questions remain unresolved. The results discussed above all resulted from so-called reverse genetics experiments, meaning that we started from the available genome data and tried to reveal the biological function of a specific gene by studying its effect on the phenotype. A major drawback of this approach lies in the dependence of what is already known, resulting in the ignorance of novel genes that show no homology to genes in other species. As diatoms possess several unique mitotic

structures (De Martino et al., 2009) and have evolved strategies to live in the most extremely varying conditions, it is to be expected that at least part of the unknown sequences, that account for almost half of the gene repertoire (Bowler et al., 2008), should play a role in the diatom life cycle regulation. As opposed to reverse genetic techniques that are now well established for diatoms, forward genetic screening in diatoms seems to be more difficult (Reviewed in Saade and Bowler, 2009). Forward genetics involves the mutagenization of the cells, followed by screening for alternated phenotypes of the mutants. However, due to the lack of a good experimental diatom model in which sexual reproduction can be controlled, and because of the time-consuming and laborious process of mutant mapping, reports on forward genetic screens in diatoms are scarce (Alonso et al., 1996; Huesemann et al., 2009; Saade and Bowler, 2009). Nevertheless, ongoing work may contribute to make this possible in near the future by investigating the potential of S*eminavis robusta*, a pennate diatom in which the sexual cycle can be easily manipulated, as a model system (see further) and by exploring alternative gene disruption strategies in *P. tricornutum*, such as the use of engineered zinc-finger nucleases (ZFN) (Santiago et al., 2008) to generate targeted gene knockouts or transposon-based mutagenesis like the Sleeping Beauty system that allows rapid mapping of the mutated gene (Dupuy et al., 2005).

In addition to the establishment of forward genetic techniques in diatoms, functional analysis of cell cycle regulators would greatly benefit from the use of inducible expression systems. When overexpressing or silencing regulators of a crucial process such as the cell cycle in any organism, one can expect that deleterious effects might occur, yielding transformants that have problems to reproduce. Indeed, from our functional analyses of *dsCYC2* and *CDKA2* we have witnessed that serious perturbations of cell cycle progression can be observed upon gene expression manipulation (see Chapter 3 and Chapter 5). Due to the relatively long selection time associated with the protocol of diatom transformation (lasting about 2-3 weeks), during which cells are selected based on their ability to confer resistance to antibiotics, transformants with the highest overexpression or silencing efficiencies of such vital genes would probably be overlooked due to perturbed reproduction. Inducible expression could overcome this problem and such systems have been developed for *Cylindrotheca fusiformis*, *T. pseudonana* and also recently for *P. tricornutum* based on promoter and terminator sequences of the nitrate reductase (NR) gene (Poulsen and Kroger, 2005; Poulsen et al., 2006; Hempel et al., 2009). The system represses the transcription of genes placed

under control of the NR regulatory sequences when ammonium is the only nitrogen source, and allows the rapid expression of genes upon transfer to nitrate-containing medium. Applying this system to our cell cycle studies in *P. tricornutum* would facilitate the analysis of phenotypic effects upon overexpression and silencing, especially in the case of key genes that are indispensable for growth. In addition, inducible expression might be useful to study the effect of a gene at a particular phase or checkpoint during the cell cycle. As an example, it would be interesting to induce the expression of *dsCYC2* specifically during darkness and examine whether cells can overcome the light-dependent G1 arrest in this way. In addition, inducible silencing of *dsCYC2* could provide a way to retain the mutants with the highest silencing levels and would allow to study its effect on cell cycle progression in a controlled manner. Also in the case of CDKA2, inducible overexpression could be valuable to trigger the baseball-bat phenotype and elucidate its origin by microscopical analysis of different stages during BBB formation.

Recently, the pennate diatom *S. robusta* was put forward as an new experimental model in which the sexual cycle can be easily manipulated (Chepurnov et al., 2008). As a consequence, *S. robusta* has been the subject of several cell cycle and life cycle studies in our lab during the last years. Gillard et al. described the first genome-wide transcript profiling, using a cDNA-AFLP (cDNA amplified fragment length polymorphism) approach, of *S. robusta* cells progressing synchronous through the mitotic cell cycle (Gillard et al., 2008) and ongoing work using the same approach aims to uncover the molecular mechanisms associated with the cell size reduction process and the switch from mitosis to meiosis (V. Devos and J. Gillard, personal communication). These studies have indicated that cell cycle modulated gene expression is abolished during the mating stage of mixed cultures and, in addition, point to the existence of a pheromone or cell pairing induced response pathway through the upregulation of e.g. hedgehog signaling molecules (J. Gillard, personal communication). Further functional charaterization of the latter genes could be done by overexpression analysis and monitoring of cell cycle responses in mixed cultures. However, a majority of the differentially expressed, relatively short transcript-derived fragments (TDFs) resulting from the cDNA-AFLP studies described above show no significant similarity to genes in other organisms, including diatoms. The current genome sequencing of *S. robusta* (A. Bones, personal communication) will facilitate functional analysis of these TDFs by obtaining full-length sequences. In addition, a deep transcriptomic sequencing (454 sequencing, performed

by the Joint Genome Institute JGI) of the mitotic and sexual cycle of four different species (including *S. robusta*) is ongoing. Without any doubt, these results will shed new light on a set of genes for which currently no expression or functional data exists and thanks to the broad phylogenetic coverage the dataset will provide more insights in life cycle gene evolution and conservation.

A crucial step towards the exploration of *S. robusta* as a model organism will be the establishment of a genetic transformation protocol and transformation tools to enable reverse-genetic functional analysis of the unknown genes identified in the expression studies. Transformation protocols of *S. robusta* by biolistic particle bombardment, the method routinely used to transform *P. tricornutum* (Falciatore et al., 1999) and also *T. pseudonana* (Poulsen et al., 2006), are currently being optimized in our lab. The development of the technique has been hampered by the high resistance of *S. robusta* for the antibiotics (e.g. phleomycin) commonly used for selection after transformation in diatoms (K. Vannerum, personal communication). However, testing other antibiotic substances, along with the generation of a transformation toolbox specific for *S. robusta*, including expression vectors containing endogenous promoter sequences for various purposes (e.g. overexpression, silencing, protein tagging, promoter analysis and inducible expression), will most likely make it possible to obtain stable *S. robusta* transformants in the near future.

An improved understanding of the diatom life and cell cycle will also require the confirmation of the function assigned to a gene by lab-based studies in the natural environment, e.g. by the development of molecular probes that can selectively detect the gene from environmental samples. In addition, metatranscriptomics approaches (consisting of the large-scale sequencing of transcripts retrieved from natural communities) combined with in-depth analysis of the prevailing physico-chemical parameters, hold enormous promise for understanding the regulation of mixed eukaryotic populations, including diatoms, in response to environmental variations and can facilitate functional assignment of specific encoded sequences (Rusch et al., 2007; Yooseph et al., 2007). In particular, this strategy would be extremely valuable for the functional characterization of the large set of unknown and diatom-specific genes identified in the recent genomic and transcriptomic analyses (Armbrust et al., 2004; Bowler et al., 2008; Gillard et al., 2008). However, the major bottleneck to apply metatranscriptomics approaches in diatoms comes from the availability of only a few diatom reference genome sequences to anchor the environmental sequences to. Nevertheless, the

advent of modern, so-called next-generation sequencing techniques that allow high-throughput parallel sequencing at a lower cost might tremendously expand our current diatom genome sequence resources in the coming years.

REFERENCES

Alonso DL, del Castillo CIS, Grima EM, Cohen Z (1996) First insights into improvement of eicosapentaenoic acid content in Phaeodactylum tricornutum (Bacillariophyceae) by induced mutagenesis. Journal of Phycology **32:** 339-345

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WW, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science **306:** 79-86

Bailleul B, Rogato A, de Martino A, Coesel S, Cardol P, Bowler C, Falciatore A, Finazzi G (2010) An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light. Proceedings of the National Academy of Sciences of the United States of America **107:** 18214-18219

Baldauf SL (2003) The deep roots of eukaryotes. Science **300:** 1703-1706

Bothwell JH, Marie D, Peters AF, Cock JM, Coelho SM (2010) Role of endoreduplication and apomeiosis during parthenogenetic reproduction in the model brown alga Ectocarpus. The New phytologist **188:** 111-121

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kroger N, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin-Jezequel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Oudot-Le Secq MP, Napoli C, Obornik M, Parker MS, Petit JL, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature **456:** 239-244

Brzezinski MA, Olson RJ, Chisholm SW (1990) Silicon availability and cell-cycle progression in marine diatoms. Marine Ecology-Progress Series **67:** 83-96

Chepurnov VA, Mann DG, von Dassow P, Vanormelingen P, Gillard J, Inzé D, Sabbe K, Vyverman W (2008) In search of new tractable diatoms for experimental biology. BioEssays : news and reviews in molecular, cellular and developmental biology **30:** 692-702

Coesel S, Mangogna M, Ishikawa T, Heijde M, Rogato A, Finazzi G, Todo T, Bowler C, Falciatore A (2009) Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. EMBO Rep **10:** 655-661

de Jager SM, Murray JA (1999) Retinoblastoma proteins in plants. Plant Mol Biol **41:** 295-299

De Martino A, Amato A, Bowler C (2009) Mitosis in diatoms: rediscovering an old model for cell division. Bioessays **31:** 874-884

De Martino A, Amato A, Meichenin A, Montsant A, Mathieu B, Huysman MJJ, De Veylder L, Vyverman W, Bowler C (Manuscript in preparation) Dissection of mitotic features in a

pennate diatom using molecular tools: *Phaeodactylum tricornutum* as a new model for cell division.

Diercks S, Metfies K, Medlin LK (2008) Molecular probe sets for the detection of toxic algae for use in sandwich hybridization formats. Journal of Plankton Research **30:** 439-448

Doonan JH, Kitsios G (2009) Functional Evolution of Cyclin-Dependent Kinases. Molecular Biotechnology **42:** 14-29

Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA (2005) Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. Nature **436:** 221-226

Falciatore A, Casotti R, Leblanc C, Abrescia C, Bowler C (1999) Transformation of Nonselectable Reporter Genes in Marine Diatoms. Mar Biotechnol (NY) **1:** 239-251

Gillard J, Devos V, Huysman MJJ, De Veylder L, D'Hondt S, Martens C, Vanormelingen P, Vannerum K, Sabbe K, Chepurnov VA, Inzé D, Vuylsteke M, Vyverman W (2008) Physiological and transcriptomic evidence for a close coupling between chloroplast ontogeny and cell cycle progression in the pennate diatom Seminavis robusta. Plant Physiol **148:** 1394-1411

Hempel F, Bullmann L, Lau J, Zauner S, Maier UG (2009) ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. Molecular Biology and Evolution **26:** 1781-1790

Hogan CJ, Stephens L, Shimizu T, Cande WZ (1992) Physiological Evidence for Involvement of a Kinesin-Related Protein during Anaphase Spindle Elongation in Diatom Central Spindles. Journal of Cell Biology **119:** 1277-1286

Huesemann MH, Hausmann TS, Bartha R, Aksoy M, Weissman JC, Benemann JR (2009) Biomass Productivities in Wild Type and Pigment Mutant of Cyclotella sp (Diatom). Applied Biochemistry and Biotechnology **157:** 507-526

Inzé D, De Veylder L (2006) Cell cycle regulation in plant development. Annu Rev Genet **40:** 77-105

Lauterborn R (1896) Untersuchugen über Bau, Kernteilung und Bewegung der Diatomeen. Verlag von Willelm Engelmann, Leipzig

Lavaud J, Strzepek RF, Kroth PG (2007) Photoprotection capacity differs among diatoms: Possible consequences on the spatial distribution of diatoms related to fluctuations in the underwater light climate. Limnology and Oceanography **52:** 1188-1194

Li S, Nosenko T, Hackett JD, Bhattacharya D (2006) Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. Mol Biol Evol **23:** 663-674

Lloyd C, Chan J (2006) Not so divided: the common basis of plant and animal cell division. Nature reviews. Molecular cell biology **7:** 147-152

MacIntyre HL, Kana TM, Geider RJ (2000) The effect of water motion on short-term rates of photosynthesis by marine phytoplankton. Trends in Plant Science **5:** 12-17

MacIsaac KD, Fraenkel E (2006) Practical strategies for discovering regulatory DNA sequence motifs. PLoS computational biology **2:** e36

Morgan DO (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu Rev Cell Dev Biol **13:** 261-291

Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D (2009) Genomic footprints of a cryptic plastid endosymbiosis in diatoms. Science **324:** 1724-1726

Nymark M, Valle KC, Brembu T, Hancke K, Winge P, Andresen K, Johnsen G, Bones AM (2009) An integrated analysis of molecular acclimation to high light in the marine diatom Phaeodactylum tricornutum. PLoS One **4:** e7743

Olson RJ, Chisholm SW (1983) Effects of Photocycles and Periodic Ammonium Supply on 3 Marine-Phytoplankton Species .1. Cell-Division Patterns. Journal of Phycology **19:** 522-528

Olson RJ, Vaulot D, Chisholm SW (1986) Effects of environmental stresses on the cell cycle of two marine phytoplankton species. Plant Physiol **80:** 918-925

Pilpel Y, Sudarsanam P, Church GM (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. Nature genetics **29:** 153-159

Planchais S, Glab N, Inzé D, Bergounioux C (2000) Chemical inhibitors: a tool for plant cell cycle studies. Febs Letters **476:** 78-83

Poulsen N, Chesley PM, Kroger N (2006) Molecular genetic manipulation of the diatom Thalassiosira pseudonana (Bacillariophyceae). Journal of Phycology **42:** 1059-1065

Poulsen N, Kroger N (2005) A new molecular tool for transgenic diatoms: control of mRNA and protein biosynthesis by an inducible promoter-terminator cassette. The FEBS journal **272:** 3413-3423

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers YH, Falcon LI, Souza V, Bonilla-Rosso G, Eguiarte LE, Karl DM, Sathyendranath S, Platt T, Bermingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Nealson K, Friedman R, Frazier M, Venter JC (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS biology **5:** e77

Saade A, Bowler C (2009) Molecular Tools for Discovering the Secrets of Diatoms. Bioscience **59:** 757-765

Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A, Collingwood TN (2008) Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. Proceedings of the National Academy of Sciences of the United States of America **105:** 5809-5814

Tao Y, Kassatly RF, Cress WD, Horowitz JM (1997) Subunit composition determines E2F DNA-binding site specificity. Mol Cell Biol **17:** 6994-7007

Thamatrakoln K, Hildebrand M (2007) Analysis of Thalassiosira pseudonana silicon transporters indicates distinct regulatory levels and transport activity through the cell cycle. Eukaryotic cell **6:** 271-279

Vandepoele K, Vlieghe K, Florquin K, Hennig L, Beemster GT, Gruissem W, Van de Peer Y, Inzé D, De Veylder L (2005) Genome-wide identification of potential plant E2F target genes. Plant Physiology **139:** 316-328

Vaulot D, Olson RJ, Chisholm SW (1986) Light and Dark Control of the Cell-Cycle in 2 Marine-Phytoplankton Species. Experimental Cell Research **167:** 38-52

Vaulot D, Olson RJ, Merkel S, Chisholm SW (1987) Cell-cycle response to nutrient starvation in 2 phytoplankton species, Thalassiosira weissflogii and Hymenomonas carterae. Marine Biology **95:** 625-630

von Dassow P, Petersen TW, Chepurnov VA, Armbrust EV (2008) Inter- and intraspecific relationships between nuclear DNA content and cell size in selected members of the centric diatom genus Thalassiosira (Bacillariophyceae). Journal of Phycology **44:** 335-349

Weinberg RA (1995) The retinoblastoma protein and cell cycle control. Cell **81:** 323-330

Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li W, Jaroszewski L, Cieplak P, Miller CS, Li H,

Mashiyama ST, Joachimiak MP, van Belle C, Chandonia JM, Soergel DA, Zhai Y, Natarajan K, Lee S, Raphael BJ, Bafna V, Friedman R, Brenner SE, Godzik A, Eisenberg D, Dixon JE, Taylor SS, Strausberg RL, Frazier M, Venter JC (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS biology **5:** e16

Zhang Z, Gerstein M (2003) Of mice and men: phylogenetic footprinting aids the discovery of regulatory elements. Journal of biology **2:** 11

7

Summary

Diatoms represent a highly diverse and ecologically important group of eukaryotic unicellular algae that are ubiquitous in marine and freshwater environments. Being one of the most species-rich classes among phytoplankton, diatoms account for about one fifth of the total oxygen production on our planet and they play a crucial role in the global biogeochemical cycling of carbon, nitrate, phosphate and silica. The complex evolutionary history of diatoms, including their secondary endosymbiotic origin and the pervasive lateral gene transfer from bacteria, resulted in a unique array of different metabolic processes and regulatory pathways. Combined with their diploid nature, obligate sexual reproduction and wide variation in mating systems, this probably contributed to their enormous diversity and ecological success. Furthermore, as producers of various lipids and pigments, and because of the mechanical properties of their finely ornamented silica cell wall, diatoms gained emerging interest for industrial applications, including biofuel production, nanotechnology and pharmaceutics. However, despite the major ecological importance of diatoms and their potential commercial value, many mysteries about the mechanisms that control their life cycle still remain. In particular, their ability to live and dominate in rapidly changing and sometimes harsh environments, suggests that diatoms have evolved specific strategies to adapt to and survive in such fluctuating conditions. Unraveling the regulatory mechanisms that underlie their unique life cycle strategies will be of crucial importance to understand diatom ecology and evolution and to further exploit their industrial potential.

During the last years, the development of molecular tools for diatoms, such as genetic transformation, silencing and gene expression analysis, along with the sequencing of several diatom genomes have resulted in the accumulation of functional and physiological studies to uncover different aspects of diatom biology. The major aim of this thesis was to gain insights into the genetic mechanisms and environmental factors that control the diatom cell cycle. At the onset of this work, our knowledge of diatom cell division largely depended on extensive

microscopy studies, describing the peculiar and unique mitotic structures of diatoms, and on physiological studies, reporting the influence of some physicochemical factors on cell cycle progression, while molecular and biochemical work was still at his infancy. Our quest for cell cycle regulatory mechanisms was started by a genome-wide analysis of the cell cycle in the pennate diatom *Phaeodactylum tricornutum*. In general, the diatom cell cycle machinery was found to share some important features with the eukaryotic regulatory system, including the presence of classical types of cyclin-dependent kinases (CDKs) and cyclins, and some regulatory proteins. To gain more insight in the biological function of the identified cell cycle regulators we developed a cell cycle synchronization method based on the release of dark-induced G1-arrested cells by illumination. This allowed us to monitor transcript expression during synchronized growth and assign the different annotated genes to specific cell cycle phases, yielding a set of cell cycle marker genes for diatoms.

From the first genome analyses, it became clear that the cyclin family in diatoms represents an expanded gene family in diatoms. By comparative analysis of cyclin abundance among several closely related Chromalveolate species, this expansion was found to be specific to diatoms. Phylogenetic analysis of the cyclin gene family lead to the discovery of a novel type of cyclins, the diatom-specific cyclins (dsCYCs). Their early expression, mainly at the G1 and G1-S phases of the cell cycle, and the induction of several members upon changing nutrient or light conditions, hints at a role for the dsCYCs in transduction of the environmental status of the cell to the cell cycle machinery. Related to this, a crucial role in regulating cell cycle progression upon changing light conditions was found for *dsCYC2*. This cyclin is probably one of the first cell cycle genes expressed upon release by illumination of dark-arrested cells, suggesting that it might play a role in the control of the G1-specific light-dependent checkpoint in *P. tricornutum*. Both silencing and overexpression of *dsCYC2* interfered with normal cell cycle progression, indicating that specific timing of *dsCYC2* expression is of major importance. Our results suggest that a light-dependent transcriptional repressor targeting the promoter of *dsCYC2* most likely plays a central role in the regulation of *dsCYC2* expression.

All members of the Rb-mediated pathway, a conserved pathway controlling G1-S regulation in animals and plants, were found in *P. tricornutum*. However, the discovery of a diatom-specifc E2F *cis*-regulatory element by applying *de novo* motif finding, suggests at least some specific regulation of the G1-S transition in diatoms. The possibility to

transactivate promoters holding this diatom-specific E2F motif by the active E2F2/DP1b transcription factor complex, indicates that the motif is functional.

Finally, we addressed the ambiguous nature of CDKA2, a CDK originally assigned to the A-type CDKs, but with some typical characteristics of the plant-specific B-type CDKs, including cell cycle-dependent transcription at the G2-M transition. Subcellular localization of CDKA2 at the cell division plane during cytokinesis and its interaction with other G2-M expressed cell cycle genes (CKS1 and CYCD1), point to a function for *CDKA2* during mitosis. In addition, *CDKA2* overexpression resulted in prolongation of the mitotic phase and an increase in cell cycle duration, demonstrating its role as mitotic regulator. This is the first functional characterization of a CDK with clear CDKB properties in a non-green lineage, and hence this report may question the current definition of B-type CDKs as being plant-specific.

In summary, our analyses demonstrated that diatoms use a common eukaryotic base of cell cycle regulatory components to control their cell division, complemented with some novel diatom-specific features, including an expanded set of diatom-specific cyclins, which most probably are part of a complex integrative network allowing them to pace the cell cycle with the surrounding conditions. Most likely, other interesting features that control diatom cell and life cycle are yet to be discovered. The wealth of information coming from more diatom genomes and the use of new functional approaches, such as forward genetic screens, will without any doubt help us to resolve some of the remaining mysteries.

7

Samenvatting

Diatomeeën zijn een uiterst diverse en ecologisch belangrijke groep van eukaryote eencellige algen welke wijdverspreid voorkomen in zowel mariene als zoetwater milieus. Als één van de meest soortenrijke fytoplankton klassen staan diatomeeën in voor ongeveer een vijfde van de totale zuurstofproductie op onze planeet en spelen ze een belangrijke rol in de biogeochemische cycli van koolstof, nitraat, fosfaat en silicium. Diatomeeën bezitten een complexe evolutionaire geschiedenis. Ze zijn ontstaan uit een secundaire endosymbiotische gebeurtenis en een deel van hun genen hebben ze verkregen via laterale gen overdracht van bacteriën, wat resulteerde in een unieke versmelting van verschillende metabolische processen en regulatorische biochemische routes. Dit samen met hun diploïde natuur, obligate seksuele voortplanting en verscheidenheid aan paringssystemen heeft waarschijnlijk bijgedragen tot de enorme diversiteit en het ecologisch succes van diatomeeën. Als producenten van verschillende soorten lipiden en pigmenten, en omwille van de mechanische eigenschappen van hun fijn gestructureerde silicium-houdende celwand, is er ook interesse voor diatomeeën vanuit verschillende industrietakken, o.a. voor biodiesel productie, nanotechnologie en farmaceutische toepassingen. Ondanks hun enorm ecologisch belang en hun economische waarde bestaan er nog vele mysteries omtrent de mechanismen die hun levenscyclus controleren. Meer bepaald, hun vermogen om te leven en te domineren in snel veranderende en soms extreme omgevingen suggereert dat diatomeeën specifieke strategieën hebben ontwikkeld om te adapteren aan en te overleven in dergelijke fluctuerende condities. Het blootleggen van de regulatorische mechanismen die aan de basis liggen van hun unieke levensstrategieën is van cruciaal belang om hun ecologie en evolutie te begrijpen en om hun industrieel potentieel verder te ontginnen.

Gedurende de laatste jaren heeft de ontwikkeling van moleculaire technieken voor diatomeeën, zoals genetische transformatie, silencing en gen expressie analyse, samen met de sequenering van verschillende genomen van diatomeeën, bijgedragen tot een toename van

functionele en fysiologische studies naar verschillende aspecten van de biologie van diatomeeën. Het hoofddoel van dit onderzoek was inzicht te verwerven in de genetische mechanismen en de omgevingsfactoren die de celcyclus in diatomeeën controleren. Bij het begin van dit werk berustte onze kennis over celdeling bij diatomeeën grotendeels op uitgebreide microscopische studies die de unieke mitotische structuren bij diatomeeën beschreven, en fysiologische studies die de invloed van een aantal fysisch-chemische factoren op de vooruitgang van de celcyclus rapporteerden, terwijl moleculair en biochemisch onderzoek nog maar net opkwam. We hebben onze zoektocht naar de mechanismen die de celcyclus reguleren gestart met een genoom-wijde analyse van de celcyclus in de pennate diatomee *Phaeodactylum tricornutum*. Over het algemeen werd vastgesteld dat het celcyclus controle systeem bij diatomeeën een aantal belangrijke eigenschappen deelt met het celcyclus systeem in andere eukaryoten, o.a. de aanwezigheid van de klassieke types van cycline-afhankelijke kinasen (CDKs) en cyclines, en een aantal regulatorische eiwitten. Om meer inzicht te verwerven in de biologische functie van de gevonden celcyclus regulatoren ontwikkelden we een methode om de celcyclus te synchroniseren, gebaseerd op alternerende periodes van licht en donker. Dit liet ons toe om gedurende synchrone groei de transcript expressie op te volgen en zo de verschillende geannoteerde genen aan specifieke celcyclus fasen toe te wijzen, resulterend in een set merkergenen voor de celcyclus in diatomeeën.

Na de eerste genoomanalyses werd het duidelijk dat de familie van de cyclines bij diatomeeën geëxpandeerd is. Door vergelijken van het aantal cyclines bij een aantal nauw verwante Chromalveolata soorten bleek deze expansie specifiek te zijn voor diatomeeën. Fylogenetische analyse van de cycline genfamilie resulteerde in de ontdekking van een nieuw type cyclines, de diatomee-specifieke cyclines (dsCYCs). Hun vroege expressie, voornamelijk tijdens de G1 en G1-S fasen van de celcyclus, en de inductie van verschillende leden bij variërende nutriënt- en lichtcondities, doet vermoeden dat deze dsCYCs instaan voor de link tussen de omgevingscondities en de regulatie van de celcyclus. Met betrekking hierop werd een cruciale rol gevonden voor *dsCYC2* bij het reguleren van de celcyclusvooruitgang bij wisselende lichtcondities. Dit cycline is mogelijk één van de eerste geëxpresseerde celcyclusgenen na belichting van donker-gearresteerde cellen, wat doet vermoeden dat *dsCYC2* een rol speelt in het G1-specifiek licht-afhankelijke controlepunt in *P. tricornutum*. Zowel het stilleggen als overexpressie van *dsCYC2* interfereerde met normale celcyclus progressie, wat aangeeft dat de specifieke timing van *dsCYC2* expressie van groot belang is. Onze resultaten suggereren dat een licht-afhankelijke transcriptionele repressor, welke de

promoter van *dsCYC2* beïnvloedt, mogelijks een centrale rol speelt in de regulatie van *dsCYC2* expressie.

Alle leden van de Rb-gemedieerde pathway, een geconserveede biochemische route voor G1-S regulatie in dieren en planten, werden gevonden in *P. tricornutum*. De ontdekking van een diatomee-specifiek *cis*-regulatorisch element via *de novo* motief detectie suggereert echter toch een specifieke regulatie van de G1-S overgang bij diatomeeën. De mogelijkheid om promoters welke dit diatomee-specifiek E2F motief bezitten te transactiveren via het actieve E2F2/DP1b transcriptie factor complex, geeft aan dat het motief functioneel is.

Tenslotte zijn we dieper ingegaan op het dubbelzinnige karakter van CDKA2, een CDK dat origineel werd ondergebracht bij de A-type CDKs, maar een aantal typische kenmerken vertoont van de plant-specifieke B-type CDKs, zoals de celcyclusafhankelijke transcriptie gedurende de G2-M overgang. Subcellulaire localisatie van CDKA2 ter hoogte van het celdelingsvlak tijdens cytokinese en de interactie met andere G2-M geëxpresseerde celcyclusgenen (CKS1 en CYCD1) duidden op een functie voor *CDKA2* tijdens de mitose. Daarbij resulteerde de overexpressie van *CDKA2* in de verlenging van de mitotische fase en een toename van de totale celcyclus duur, wat wijst op een rol als mitotisch regulator. Dit is de eerste functionele karakterisering van een CDK met duidelijke CDKB eigenschappen in een niet-groene lijn, en daarmee zou dit verslag dan ook de huidige definitie van B-type CDKs als zijnde plant-speciefiek in vraag kunnen stellen.

Onze analyses toonden aan dat diatomeeën een gemeenschappelijke eukaryote basis van celcyclus regulerende componenten gebruiken om hun celdeling te controleren, aangevuld met enkele nieuwe diatomee-specifieke kenmerken, waaronder een uitgebreide set van diatomee-specifieke cyclines, die waarschijnlijk deel uitmaken van een complex integratief controle netwerk waardoor ze de celcyclus kunnen aanpassen aan de omringende omstandigheden. Andere interessante kenmerken die de cel- en levenscyclus van diatomeeën reguleren moeten hoogst waarschijnlijk nog ontdekt worden. De rijkdom van informatie afkomstig van meerdere diatomee genomen en het gebruik van nieuwe functionele benaderingen, zoals “forward genetic screens”, zal ons zonder enige twijfel helpen om een deel van de resterende mysteries op te lossen.

Curriculum vitae

Personal information:

Name:	Huysman
First names:	Marie Jenny Jacqueline
Adress:	Robert Schumanstraat 11 9600 Ronse
Telephone:	0497 60 10 15
E-mail:	Marie.Huysman@gmail.com
Nationality:	Belgian
Date of birth:	August 16, 1982
Marital status:	Unmarried partners

Language skills:

Dutch:	native
English:	fluent
French:	good
German:	basic

Education:

2005-today	PhD student (sciences: Biotechnology) at Ghent University
2004-2005	Master in Molecular Biotechnology (distinction) at Ghent University option Molecular Medical Biotechnology *Research project: Determination by RACE technology of the ORF from a new bee poison protein similar to Major Royal Jelly Protein 8* *Promoter: Prof. Dr. F. Jacobs; Laboratory of Zoophysiology, Faculty of Sciences*
2002-2004	Master in Biotechnology (distinction) at Ghent University *Master dissertation: Molecular diagnostics of Burkholderia cepacia complex bacteria by means of RFLP-analysis of housekeeping genes.* *Promoter: Prof. Dr. P. Vandamme; Laboratory of Microbiology, Faculty of Sciences*
2000-2003	Bachelor in Biology at Ghent University

Research experience:

2005-today	Predoctoral fellow at the Protistology and Aquatic Ecology lab (Ghent University), the Department of Plant Biotechnology and Genetics (Ghent University) and the Department of Plant Systems Biology (VIB)

2007-2010 Aspirant at the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT)

Peer-reviewed publications:

- Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Vérret, F., Berges, J.A., Brownlee, C., Cadoret, J.-P., Chiovitti, A., Choi, C. J., Coesel, S., De Martino, A., Detter, J.C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., **Huysman, M.J.J**., Jenkins, B.D., Jiroutova, K., Jorgensen, R.E., Joubert, Y., Kaplan, A., Kroeger, N., Kroth, P.G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jézéquel, V., Lopez, P.J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L.K., Montsant, A., Oudot-Le Secq, M.-P., Napoli, C., Obornik, M., Petit, J.L., Porcel, B.M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T.A., Schmutz, J., Schnitzler Parker, M., Shapiro, H., Siaut, M., Stanley, M., Sussman, M.R., Taylor, A., Vardi, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L.S., Rokhsar, D.S., Weissenbach, J., Armbrust, E.V., Green, B.R., Van de Peer, Y., and Grigoriev, I.V. (2008). The Phaeodactylum genome reveals the dynamic nature and multi-lineage evolutionary history of diatom genomes. Nature 456(7219): 239-244.

- Gillard, J., Devos, V., **Huysman, M.J.J.**, De Veylder, L., D'Hondt, S., Martens, C., Vanormelingen, P., Vannerum, K., Sabbe, K., Chepurnov, V.A., Inzé, D., Vuylsteke, M. and W. Vyverman. (2008). Physiological and transcriptomic evidence for a close coupling between chloroplast ontogeny and cell cycle progression in the pennate diatom Seminavis robusta. Plant Physiology 148: 1394-1411.

- **Huysman M.J.J.**, Martens C., Vandepoele K., Gillard J., Rayko E., Heijde M., Bowler C., Inzé D., Van de Peer Y., De Veylder L. and Vyverman W. (2010) Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signalling. Genome Biology 11 (2):R17

Manuscripts in preparation:

- De Martino A., Amato A., Meichenin A., Montsant A., Mathieu B., **Huysman M.J.J.**, De Veylder L., Vyverman W., Bowler C. Dissection of mitotic features in a pennate diatom using molecular tools: Phaeodactylum tricornutum as a new model for cell division. At submission stage.

- Vannerum K., **Huysman M.J.J.**, De Rycke R., Vuylsteke M., Cocquyt E., Pollier J., Lütz-Meindl U., Gillard J., De Veylder L., Goossens A., Inzé D., Vyverman W. Transcriptional analysis of Micrasterias (Streptophyta) cell growth identifies a new expansin family. At submission stage.

Meetings and workshops with active participation:

- **DIATOMICS (European FP6) meeting**, 30 March 2007, Leibniz Institute, Kiel, Germany. Oral presentation.
- **Doctoraatsymposium Faculteit Wetenschappen 2007**, 24 April 2007, Universiteit Gent, Gent , Belgium. Poster presentation.
- **DIATOMICS (European FP6) meeting**, 23 November 2007, Interuniversity Institute of Eilat, Eilat, Israel. Oral presentation.
- **20th International Diatom Symposium**, 7-13 September 2008, Dubrovnik, Croatia. Oral presentation.
- **VLIZ Jongerencontactdag 2009**, 6 March 2009, Brugge, Belgium. Poster presentation.
- **EMBO Workshop, Evo-Devo meets Marine Ecology**, 9-11 October 2009, Sant'Angelo d'Ischia, Napels, Italy. Oral and poster presentation.
- **10th VLIZ young scientists' day**, 26 November 2009, Oostende, Belgium. Poster presentation.
- **Ocean Sciences Meeting 2010**, 22-26 February 2010, Portland, United States of America. Oral presentation (Invited speaker).
- **ESIL 2010: A decade of algal genomics**, 7-9 April 2010, Gent, Belgium. Oral presentation.

Other meetings and workshops:

- **Roche real-time quantitative PCR workshop**, 21-22 November, 2006, Leuven, Belgium
- **Effective Writing for Life Sciences Research** (by Dr. Jane Fraser), 21-22 January, 2009, Gent, Belgium
- **BITS Workshop: Understanding Protein Structures**, 10 June, 2009, Gent, Belgium
- **Illumina: Agriculture/Microbiology Seminar**, 12 November, 2009, Leuven, Belgium

Supervised students:

Joris Huylebroeck, Master in Biochemistry and Biotechnology (UGent) 2009-2010
Thesis title: *Functional analysis of the RB/E2F pathway in Phaeodactylum tricornutum*